8th Conference on the Theory of Quantum Computation, Communication and Cryptography

TQC 2013, May 21–23, 2013, Guelph, Ontario, Canada

Edited by
Simone Severini
Fernando Brandao

LIPIcs – Vol. 22 – TQC'13 www.dagstuhl.de/lipics

Editors

Simone Severini
Department of Computer Science
University College London
s.severini@ucl.ac.uk

Fernando Brandao
Department of Computer Science
University College London
f.brandao@ucl.ac.uk

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

ISBN 978-3-939897-55-2

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/978-3-939897-55-2.

Publication date
November, 2013

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.2013.i

ISBN 978-3-939897-55-2 **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

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Preface

The 8th Conference on the Theory of Quantum Computation, Communication and Cryptography was held at the University of Guelph, from the 21st to the 23rd May 2013.

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 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 Jop Briët (CWI, Amsterdam), Aram Harrow (MIT, Cambridge), Iordanis Kerenidis (CNRS – Université Paris Diderot-Paris 7, Paris), Thomas Vidick (MIT, Cambridge), and Stephanie Wehner (National University of Singapore, Singapore).

The conference was possible thanks to the financial support of the Institute for Quantum Computing (IQC) at the University of Waterloo, the Perimeter Institute for Theoretical Physics (PI), the Fields Institute for Research in Mathematical Sciences, and the University of Guelph.

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, and to Sarah Plosker, James Howard, and Tyler Jackson, for their help af the registration desk. 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!

October 2013 Fernando Brandao and Simone Severini

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Conference Organization

Local Organizing Commitee

Jianxin Chen	University of Guelph, Canada
Zhengfeng Ji	IQC and University of Waterloo, Canada
David Kribs *(Chair)*	University of Guelph, Canada
Bei Zeng *(Co-chair)*	University of Guelph, Canada

Program Commitee

Antonio Acin	ICFO Barcelona, Spain
Gorjan Alagic	Caltech, USA
Salman Beigi	Institute for Research in Fundamental Sciences, Iran
Michael Ben-Or	The Hebrew University of Jerusalem, Israel
Fernando Brandao *(Co-chair)*	ETH Zürich, Switzerland & UCL, UK
Sergey Bravyi	IBM, USA
Francesco Buscemi	University of Nagoya, Japan
Eric Chitambar	Southern Illinois University, USA
Runyao Duan	University of Technology Sydney, Australia
Michał Horodecki	University of Gdańsk, Poland
Kazuo Iwama	Kyoto University, Japan
Julia Kempe	University of Paris, France & Tel Aviv University, Israel
David Kribs	University of Guelph, Canada
Troy Lee	National University of Singapore, Singapore
Stefano Mancini	Università degli Studi di Camerino, Italy
Ashley Montanaro	University of Cambridge, UK
Ashwin Nayak	IQC and University of Waterloo, Canada
Harumichi Nishimura	Nagoya University, Japan
Stefano Pironio	Université Libre de Bruxelles, Belgium
Pranab Sen	Tata Institute of Fundamental Research, India
Simone Severini *(Chair)*	UCL, UK
Rolando Somma	Los Alamos National Laboratory, USA
Xiaoming Sun	China Academy of Science, P. R. China
Pawel Wocjan	University of Central Florida, USA
Bei Zeng	University of Guelph, Canada

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

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Ancilla Driven Quantum Computation with Arbitrary Entangling Strength

Kerem Halil Shah and Daniel K. L. Oi

SUPA Department Of Physics, University of Strathclyde
107 Rottenrow, Glasgow G4 0NG, UK
k.halil-shah@strath.ac.uk, daniel.oi@strath.ac.uk

Abstract

We extend the model of Ancilla Driven Quantum Computation (ADQC) by considering gates with arbitrary entangling power. By giving up stepwise determinism, universal QC can still be achieved through a variable length sequence of single qubit gates and probabilistic "repeat-until-succes" entangling operations. This opens up a new range of possible physical implementations as well as shedding light on the sets of resources sufficient for universal QC.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Ancilla, weak measurement, quantum computation, entanglement, random walks

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.1

1 Introduction

A fundamental question is "what set of resources are required for universal quantum computation?". Many models have been proposed ranging from the conventional gate-based [13], measurement-based [30], adiabatic [15], and topological [22] quantum computation. These models utilise different sets of resources and are suited to various physical implementations. By finding new models with different sets of sufficient resources, this may open up new ways of implementing quantum computation as well as enrich our understanding of quantum computation itself.

A hybrid model combining aspects of the gate-based and measurement-based approaches was introduced as Ancilla Driven Quantum Computation [3]. The key feature of ADQC is the restriction of resources to a single unitary interaction, and direct access (initialisation and measurement) only to an ancilla qubit that can be coupled sequentially to system (register) qubits. ADQC uses entanglement between the ancilla and register, and the kickback induced by measurement on the ancilla to drive unitary evolution of the register. By coupling the ancilla to various register qubits and choosing different measurements on the ancilla, universal QC can be achieved.

There is often a trade off between easy access and manipulation or long coherence times. ADQC lends itself to physical systems where register qubits with long decoherence times are difficult to manipulate while relatively short-lived ancillary systems are more easily controlled and can be quickly initialised and measured. There has been much work that looks at dealing with such properties with the use of ancillas for specific physical implementations. Work on optical clocks using aluminium ions have employed magnesium or beryllium ions as an ancillary system to account for a lack of optical accessibility with aluminium ions [9]. Ohshima [27] considered maintaining low decoherence of quantum dots by only activating access through an ancilla qubit in the same cell. It has been proposed that isolated, stable

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 1–19


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

NV centre nuclear spins be used as qubits manipulated by neighbouring electron spins [14, 5] and Bermudez *et al* developed an proposal for nuclear spin interactions mediated by electron spins that effects an Ising type interaction [5]. Ion trap+photon systems such as in [6] and solid state systems with ballistic electrons have been considered for a class of systems that involve generating quantum gates through scattering between a static and flying qubit [12]. In such cases we may have no interaction-time tuning [12] or restricted range of dynamic modulation which is particularly relevant for the work of this paper.

These proposals often look at one system or particular parameters with a focus on gate based QC. The proposals of Anders *et al* [3] and Kashefi *et al*[20] from which we will develop our proposal considers a scheme without reference to a specific physical system. A general description is provided, of the minimal resources require to achieve universality, expressed as a finite resource set that can be compared to Measurement Based Quantum Computation (MBQC) and the resources to form cluster states but aimed towards a Gate Based Quantum Computation (GBQC) style circuit.

In the original ADQC scheme, the computation proceeds step-wise deterministically up to Pauli corrections on the final output, very much in the spirit of MBQC [30, 8]. This requires that the entangling unitary between register and ancilla be of either of two specific gates, $H \otimes H.CZ$ or SWAP.CZ. No other entangling gate would allow the measurement-induced kickback to be unitary, and permit the different computation branches corresponding to the possible measurement results to be reunited via Pauli corrections [3].

Here, we show that by relaxing the requirement for stepwise determinism a much larger set of entangling gates can achieve universal QC which may open up a range of physical implementations with fixed arbitrary coupling strengths and interaction times. Investigation of to what extent the required entangling properties can be relaxed has been performed for MBQC [16, 17]. Unlike that, we do not use many-body physics techniques but expand the resource set using GBQC concepts.

Pauli corrections and the corresponding direct access of the register will not be required. This may also reduce the control requirements for characterisation as described in [28]. This is achieved by gate approximation, repeat-until-success strategies, and multi-step measurement. Single qubit unitary gates can still be implemented deterministically while two qubit gates and measurement and state initialisation will be probabilistic and require the development of probabilistic protocols of which we provide examples.

2 Overview of ADQC

We will review how ADQC implements single qubit unitary gates and which conditions on the available resources are necessary for the method to be step-wise deterministic. We highlight the resource requirements that our proposal will extend; a full description can be found in [3, 20, 2].

The evolution of a qubit in a quantum register can be driven by preparing an ancilla qubit system in a state $|a\rangle$, coupling that ancilla with an entangling interaction, E_{AR} and then measuring the ancilla in some basis $\{|m_+\rangle, |m_-\rangle\}$. After measurement, the evolution of the register qubit can be described by a Kraus operator $K_\pm = \langle m_\pm|E|a\rangle$ [26, 21, 32].

The interaction E_{AR} can generally be composed of a product of gates that act locally on the individual systems and a gate that produces some entanglement between the systems. For classification of two qubit gates, we turn to the canonical decomposition [24, 35, 31]: A unitary that acts on a system of a pair of qubits A and R, E_{AR}, can be expressed as

$$E_{AR} = (V_A \otimes V_R)\Delta(\alpha_x, \alpha_y, \alpha_z)(U_A \otimes U_R) \tag{1}$$

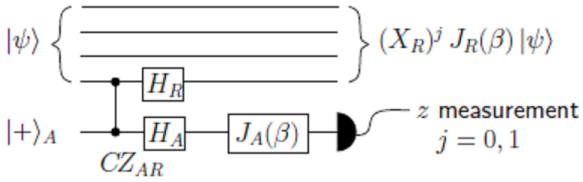


Figure 1 Anders *et al.* depiction of ancilla-driven implementation of a single qubit rotation [3]. The ancilla and register qubits are coupled with CZ and the local unitary gates are chosen such that the interaction remains symmetrical with respect to ancilla-register exchange. A rotation $X^j J(\beta)$ is enacted on the register a result $J(\beta)$ is enacted on the ancilla which is then measured in the z basis with a result j=0,1.

where U_A, U_R, V_A, V_R are unitary gates that act only on the local systems of qubits 1 and 2 and $\Delta(\alpha_x, \alpha_y, \alpha_z) = \exp(-i(\alpha_x \sigma_x \otimes \sigma_x + \alpha_y \sigma_y \otimes \sigma_y + \alpha_z \sigma_z \otimes \sigma_z))$.

If the final action on the register is to be unitary then the Kraus operator must be proportional to unitary such that $K_\pm K_\pm^\dagger = p_\pm \mathbb{I}$ where $p_\pm$ is the probability ot]f the measurement. This does not depend on the local unitary gates acting on the register. The $(\alpha_x, \alpha_y, \alpha_z)$ parameters are the only parameters unique to E_{AR} that determine whether the measurement back-action will be unitary. The preparation of the state of the ancilla system can be adapted for any unitary that immediate follows it; U_A is effectively under full control and can be represented entirely by ancilla state preparation. Similarly choice of measurement basis equates to freedom of choice of the unitary after interaction, V_A. Only the $\Delta(\alpha_x, \alpha_y, \alpha_z)$ component is of interest when considering the entangling capabilities of the interaction [24, 35].

$K_\pm$ are generated probabilistically and are not equivalent. In order to be deterministic, a key idea from MBQC is utilised: if $PK_- = K_+$ where P is a Pauli operator correction, this correction can then be commuted through several applications of the Kraus operator and local unitary gates. Different computation branches associated with the measurement outcomes can be reunited by a Pauli correction.

$$(V_R P K_- U_R).(V_R P K_- U_R).(V_R P K_- U_R) = P' P'' P'''(V_R K_- U_R).(V_R K_- U_R).(V_R K_- U_R)$$

To have the capability to enact any arbitrary unitary gate on the register, a sequence of $V_R K_- U_R$ should be universal for single qubit unitary gates.

2.1 Control-Z Hadamard example

An example can be performed with $E_{AR} = (H_A \otimes H_R).CZ_{AR}$. The ancilla is prepared in the $|+\rangle$ state, couples with the register qubit with E_{AR} and then undergoes a unitary $J(\beta) = HR_{\hat{z}}(\beta)$ before being measured in the computational basis. The action on the register qubit for a measurement result $|j\rangle$ is $X_R^j J_R(\beta)$. The difference between the two resulting unitary operations can be corrected by X_R. Any arbitrary single qubit unitary can be decomposed into four rotations $e^{i\alpha} J(0)J(\beta)J(\gamma)J(\delta)$ so to implement any arbitrary single qubit unitary up to a global phase, four ancilla interaction-measurements are performed changing the parameter of the local unitary on the ancilla to be the Euler angles δ, γ, β and then 0,

$$X^i J(0) X^j J(\beta) X^k J(\gamma) X^l J(\delta). \tag{2}$$

As in MBQC, the Pauli corrections can be commuted through each application of $J(\beta)$ if we make adaptations to the local unitary applied to the ancilla measurement basis [30, 8, 20].

$$X^i J(0) X^j J(\beta) X^k J(\gamma) X^l J(\delta) = X^i Z^j X^k Z^l J(0) J((-1)^k \beta) J((-1)^l \gamma) J(\delta) \tag{3}$$

To extend to larger computations all corrections on register qubits are required to interchange with future entangling operations $\Delta(\alpha_x, \alpha_y, \alpha_z)$ so that they remain local corrections on the register and on the ancilla preparation or measurement basis choice. This allows all the Pauli corrections to accumulate at the final step. This provides the condition that the entangling operation $\Delta(\alpha_x, \alpha_y, \alpha_z)$ tensor commutes with the corrections [3]; as a result only two classes of coupling are universal.

2.2 Conditions on the class of interactions

In order to allow achronical Pauli operator corrections ensuring general stepwise deterministic single qubit gates exactly, we must have interactions that belong to the classes of interactions that are locally equivalent to CZ or $CZ + SWAP$ gates. Expressed in terms of $\Delta(\alpha_x, \alpha_y, \alpha_z)$, these classes are $\Delta(\frac{\pi}{4}, 0, 0)$ and $\Delta(\frac{\pi}{4}, \frac{\pi}{4}, 0)$. There exists broader classes of interactions which fulfil the weaker condition that the Kraus operators acting on the register qubit are proportional to unitary. These are the classes for which there is at least one α_i parameter equal to zero- $\Delta(\alpha, 0, 0)$ and $\Delta(\alpha_1, \alpha_2, 0)$ up to symmetries. It is symmetrical up to local unitary gate corrections with respect to $\pm$ exchanges, $\frac{\pi}{2}$ shifts and reflections in $\frac{\pi}{4}$ acting on the parameters $(\alpha_x, \alpha_y, \alpha_z)$ [24, 31]. The values are also symmetrical with respect to permutations [31]. This allows us to consider only cases where $|\alpha_j| < \frac{\pi}{4}$ and to classify an interaction by the number of non-zero parameters, e.g. $e^{-i(\alpha \sigma_z \otimes \sigma_z)}$ is equivalent, up to local unitary corrections to any case where $(\alpha_i = 0, \alpha_j = 0, \alpha_k \neq 0)$. We will demonstrate a way in which we can eschew a stepwise construction of unitary gates, thus not requiring Pauli corrections and allowing the broader class $\Delta(\alpha, 0, 0)$, locally equivalent to the Control-unitary set of gates, for ancilla driven quantum computation with arbitrary interactions. As a cost, we will not be performing exact unitary gates but efficiently generated approximation.

Some characteristics are general for any member of the class so we will use $e^{-i\alpha\sigma_z \otimes \sigma_z}$ to represent them. Other effects are dependent on specific interactions with descriptions of the local unitary gates. In the next section we will explain our choice of interaction in those cases.

3 Single qubit gates using $(H \otimes H).C\text{-}T$

An interaction $\Delta(\alpha, 0, 0)$ where $0 < \alpha < \frac{\pi}{4}$ has a restricted set of conditions on the choice of ancilla preparation state $|a\rangle$ and measurement basis $\{|m_j\rangle\}$ that allow for a measurement induced back action represented by the Kraus operator $\langle m_j|\Delta|a\rangle$ to be unitary. The resulting set of unitary gates that can be implemented deterministically is smaller than for $\Delta(\frac{\pi}{4}, 0, 0)$. Using $\Delta(\alpha, 0, 0)$, $\alpha < \frac{\pi}{4}$, only two gates can be implemented deterministically. If we consider just using the non-local part of the Cartan decomposition $e^{-i\alpha\sigma_z \otimes \sigma_z}$, the only two possible unitary gates that can be enacted independently of the random measurement result are achieve by preparing the ancilla in the computational basis. A $|0\rangle, |1\rangle$ ancilla input always corresponds to a $R_{\hat{z}}(2\alpha), R_{\hat{z}}(-2\alpha)$ unitary gate on the register qubit respectively.

However, with local unitary gate corrections any member of the class $\Delta(\alpha, 0, 0)$ is equivalent to a Control-Unitary gate. Consider this in conjunction with a fixed local unitary post-action, U_b, on the register. $U_{b,R}.C\text{-}U_a$ can implement a two gate finite set $\{U_0, U_1\}$

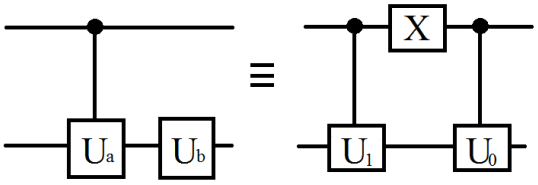


Figure 2 A generalised description of the action by which members of the interaction class $\Delta(0 < \alpha < \frac{\pi}{4}, 0, 0)$ can generate two unitary gates whose commutation rules do not form a closed set. $U_0 = U_b$ and $U_1 = U_b U_a$

where $U_0 = U_b$, $U_1 = U_a U_b$. If the Lie algebra closure of U_0 and U_1 covers SU(2) then these two gates form a universal set for single qubit unitary gates [7].

To demonstrate our proposal we choose the fundamental unitary for ancilla driven quantum computation $E = (H \otimes H).C\text{-}R_{\hat{z}}\left(\frac{\pi}{4}\right)$ (which corresponds to $\alpha = \frac{\pi}{16}$). This specific gate is chosen for two reasons; 1) It is directly comparable with the gate $E_{AR} = (H \otimes H).CZ$ from [3] but with a smaller rotation angle parameter of the Control-unitary, 2) It will generate $\{T, HT\}$ which generates the same group as the finite set $\{H, T\}$ which is well studied and proven to be universal [7] and of interest for its applications in fault tolerant quantum computation [7, 1]. The method can then be generalised to arbitrary coupling strengths and local unitary gate products. With this method, the circuit is programmed by the ancilla state preparation and requires no further manipulation nor measurement of the ancilla. This may have benefits for particular physical systems depending on the lifetime and robustness of the ancilla.

With this choice of interaction for producing single qubit gates, it is necessary to demonstrate the ability to perform measurements, initialise the register qubit into a specific state and enact a two qubit entangling gate in order to achieve universal quantum computation. We will address these issues in later sections.

3.1 Two parameter interactions and stochastic ADQC

The above method can only be employed with "one-parameter" interactions – the class $\Delta(\alpha, 0, 0)$. There does not exist an ancilla preparation basis for which the resulting action is deterministic in the case of the $\Delta(\alpha_x, \alpha_y, 0)$-"two-parameter" class. Instead there must be a measurement of the ancilla which results in an action that is dependent on the measurement result.

A two-parameter class interaction can be seen as a succession of one-parameter class interactions with local unitary corrections to account for permutation of the parameters e.g. $e^{-i(\alpha_x \sigma_x \otimes \sigma_x + \alpha_z \sigma_z \otimes \sigma_z)} \equiv e^{-i\alpha_z \sigma_z \otimes \sigma_z}(H \otimes H) e^{-i\alpha_x \sigma_x \otimes \sigma_x}(H \otimes H) \equiv C\text{-}R_{\hat{z}}(4\alpha_z)(H \otimes H) C\text{-}R_{\hat{z}}(4\alpha_x)$ (up to local unitary corrections). Viewing it this way demonstrates why it should not have a strictly deterministic set of parameters but also why the form of the unitary actions brought about is always $R_{\hat{z}}(\beta) R_{\hat{x}}(\gamma) = HJ(\beta)J(\gamma)H$. Considering this the issue for using the two parameter class is just the inability to make a deterministic choice between two pairs of $\{\beta, \gamma\}$ to produce a universal single qubit gate set. It is similar in effect to flipping the resource requirements for the one-parameter case in time so that rather than preparing in the $\{|j\rangle\}$ basis, we only measure the ancilla in that basis. The resulting gates on the register would still be the same $\{U_0, U_1\}$ as described before but randomly generated with probabilities p_0, p_1 dependent on the initial ancilla state.

In fact, we can generate an approximation to any arbitrary single qubit unitary gate in this way. The stochastically generated sequence $\prod_k U_{i(k)}$ will perform a random walk on the compact set of unitaries and be guaranteed by Poincaré recurrence to reach an approximation of any unitary eventually. The generation of the strings $\prod_k U_{i(k)}$, though probabilistic, are still consistent with the conditions of the Solovay-Kitaev theorem [26] and so the efficiency of these approximations might be improved.

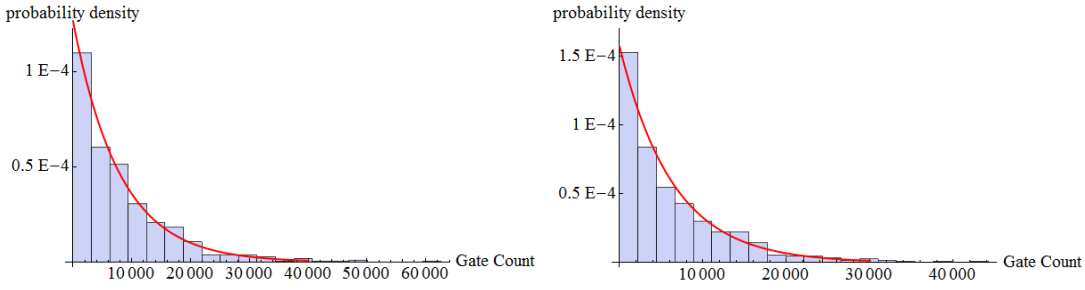
Future research will look at finding the gate count required to hit a target averaged over all unitaries for a given error and the scaling with the error. In addition to the relevance to our work, it may serve as a benchmark for any circuit compilation sequence: to see how much better inputting it through a string of ancilla is compared with than "measuring them all and letting God sort it out".

As a preliminary investigation, we simulated the random generation of gates using the interactions $(H \otimes H)\Delta(0, 0, \frac{\pi}{16})$ and $\Delta(\frac{\pi}{16}, 0, \frac{\pi}{16})$, ancilla preparation state $|+\rangle$ and measurement in the computational state basis. For the one-parameter class interaction the resulting unitary gates are $U_0 = HR_{\hat{z}}(\frac{\pi}{8})$,$U_1 = HR_{\hat{z}}(-\frac{\pi}{8})$, $p_0 = p_1 = \frac{1}{2}$; for the two-parameter class,$U_0 = R_{\hat{z}}(\frac{\pi}{8})R_{\hat{x}}(\frac{\pi}{8})$,$U_1 = R_{\hat{z}}(-\frac{\pi}{8})R_{\hat{x}}(\frac{\pi}{8})$, $p_0 = p_1 = \frac{1}{2}$. The target unitary was $U_T = R_{\hat{x}}(\frac{\pi}{2})$. At each step a gate corresponding to the $\{U_0, U_1\}$ of each interaction was multiplied to the product of the previous step starting with the identity operator. The normalised trace distance of the product,$V = \prod_k U_{i(k)}$, and the target unitary was evaluated at each step until within a chosen error size $\epsilon < 0.05$.

$$\|V - U_T\| = \sqrt{\frac{2 - |\mathrm{Tr}[V^\dagger U_T]|}{2}} \leq \epsilon \tag{4}$$

The number of gates required for this to occur was collected 1000 times and used to create a probability distribution for the gate count required to reach the target unitary (see Figure 3).

This small investigation reveals some behaviours to be considered in further research. While the target is not achievable in a single step and there is no finite probability for success per step, the aggregate behaviour over many steps, taken over a large number of simulations, can be modelled as a geometric or discretised exponential distribution. This is true for both one-parameter and two-parameter interactions. Particular target unitaries may cause anomalous effects; the two-parameter case is able to produce an exact solution of the target unitary in 4 steps which causes a large peak in the distribution and then suppresses the probability of a result for several steps after (see Figure 13d in Appendix B) but with a



(a) $H_R\Delta(0, 0, \frac{\pi}{16})$ **(b)** $\Delta(\frac{\pi}{16}, 0, \frac{\pi}{16})$

Figure 3 Probability distribution of required gate count to achieve target $U_T = R_{\hat{x}}(\frac{\pi}{2})$. a) Use of a single parameter interaction in a 20 bin histogram, b) Use of a two parameter interaction in a 20 bin histogram. The probability distribution corresponding to the exponential distribution parametrised by the mean of the results is displayed by the solid red curve (colour online).

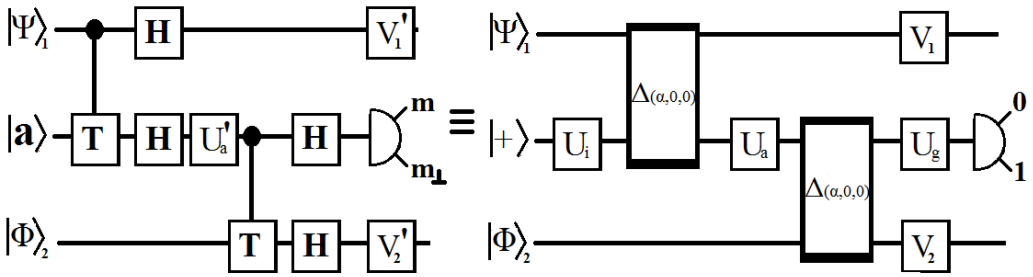


Figure 4 The circuit for implementing a two qubit gate and its generalisation. $V_1 \otimes V_2$ allows us to make any local unitary corrections to the register qubits. The preparation of the ancilla state is represented by U_i while the choice of measurement basis is represented by U_g.

large enough bin size the exponential model dominates. This provides some features to test when extending the average over a large number of target unitaries. An average over many random target unitaries may smooth out such effects and verify the general applicability of the exponential distribution model. Under this model, we can calculate the probability of reaching any target unitary within a fixed number of steps and increase the number of steps until it provides a desired fidelity.

4 Entangling gates

Given the ability to implement any single qubit unitary gate, universal quantum computation requires that we are also able to implement an entangling two-qubit unitary gate between register qubits. Note that direct interaction between register qubits not allowed, the interaction must be mediated by an ancilla qubit in the ADQC model. ADQC is capable of implementing an entangling gate by the use of one ancilla and two implementations of the same fundamental interaction E_{AR} as used in the implementation of single qubit gates. The measurement of the ancilla results in two outcomes but they are equivalent up to local unitary corrections.

In our proposal, we attempt the same use of a single ancilla and two interactions but with weaker coupling strength $\Delta(0 < \alpha < \frac{\pi}{4})$. We have the freedom to prepare and measure the ancilla in any state/basis, as well as perform local unitary operations and post corrections on it and the register qubits. Due to this, we will focus on the $e^{-i\alpha\sigma_z\otimes\sigma_z}$ form of the interaction, with any local unitary corrections being incorporated into a single gate; the local H post-interaction unitary on the ancilla can be removed by the appropriate choice of U_a – see Figure 4 .

4.1 Interpreting the unitary and entangling conditions

The operation on the register qubits after measuring the ancilla must be unitary and entangling. This will restrict the unitary operations, prepared states and measurement bases we can use on the ancilla. We need to be able to write the unitary and entangling conditions as some parameter restraints on the circuit. Because the interaction is of the class $\Delta(\alpha, 0, 0)$, we can make some simplifications. If a register qubit is in a computational basis state $|i\rangle$, $i = 0, 1$, then it will, through the interaction $\Delta(\alpha, 0, 0)$ cause a unitary $R_{\hat{z}}((-1)^i 2\alpha)$ action on the ancilla. So for two register qubits there are four potential final ancilla states

corresponding to the computational basis $|a_{00}\rangle$ for $|00\rangle_{12}$ etc.. Therefore the transformation of a general register qubit and ancilla state, $|a\rangle|\Phi\rangle_{12}$ will be transformed by the circuit thus:

$$|a\rangle|\Phi\rangle_{12} = |a\rangle \sum_{ij} C_{ij}|ij\rangle_{12} \rightarrow \sum_{ij} C_{ij}|a_{ij}\rangle|ij\rangle_{12} \tag{5}$$

Each final ancilla state is a unitary evolution of the original determined by the parameters of the circuit:

$$|a_{ij}\rangle = U_g R_{\hat{z}}((-1)^j 2\alpha) U_a R_{\hat{z}}((-1)^i 2\alpha)|a\rangle \tag{6}$$

After a measurement of the ancilla in a state $|m\rangle$ the (unnormalised) state of the register will be

$$\sum_{ij} \langle m|a_{ij}\rangle C_{ij}|ij\rangle_{12} \tag{7}$$

The evolution of the register can be represented by a Kraus operator in a matrix representation:

$$\mathbf{K}_m = \mathrm{diag}(\langle m|a_{00}\rangle, \langle m|a_{01}\rangle, \langle m|a_{10}\rangle, \langle m|a_{11}\rangle) \tag{8}$$

For this operator to be proportional to unitary $|\langle m|_{ij}\rangle|$ must be the same $\forall i, j$. This encapsulates an expression of the equivalence between the unitary condition and the requirement that the measurement of the ancilla extracts no information about the register system. $\langle m|a_{ij}\rangle = |\langle m|a_{ij}\rangle| e^{i\phi_{ij}}$; the probability of the measurement does not distinguish between the register states and the term $|\langle m|a_{ij}\rangle|$ drops out and the effective unitary on the register pair of qubits is:

$$\mathbf{U} = \mathrm{diag}(e^{i\phi_{00}}, e^{i\phi_{01}}, e^{i\phi_{10}}, e^{i\phi_{11}}) \tag{9}$$

The result is also of the class $\Delta(\alpha, 0, 0)$ and is thus also equivalent to a Control-unitary gate. Being diagonal in the computation basis, we can apply local unitary corrections to convert it into a gate of the form $C - R_{\hat{z}}(\Phi)$. Given a two qubit unitary of the form $\mathrm{diag}(e^{i\phi_{00}}, e^{i\phi_{01}}, e^{i\phi_{10}}, e^{i\phi_{11}})$ we can multiply it by local unitary gates

$$\begin{pmatrix} e^{-ia_1} & \\ & e^{-ia_2} \end{pmatrix} \otimes \begin{pmatrix} e^{-ib_1} & \\ & e^{-ib_2} \end{pmatrix} = \mathrm{diag}(e^{-i(a_1+b_1)}, e^{-i(a_1+b_2)}, e^{-i(a_2+b_1)}, e^{-i(a_2+b_2)}) \tag{10}$$

to give $\mathrm{diag}(e^{i\phi_{00}-(a_1+b_1)}, e^{i\phi_{01}-(a_1+b_2)}, e^{i\phi_{10}-(a_2+b_1)}, e^{i\phi_{11}-(a_2+b_2)})$

We can choose a_1, b_1, a_2, b_2 such that

$$\phi_{00} - (a_1 + b_1) = 0 \tag{11}$$
$$\phi_{01} - (a_1 + b_2) = 0 \tag{12}$$
$$\phi_{10} - (a_2 + b_1) = 0 \tag{13}$$
$$\rightarrow a_2 + b_2 = a_2 + b_1 - (a_1 + b_1) + (a_1 + b_2) = \phi_{10} - \phi_{00} + \phi_{01} \tag{14}$$

the resulting gate (9) must therefore be equivalent to the form

$$\tilde{\mathbf{U}} = \mathrm{diag}(1, 1, 1, e^{i((\phi_{11}-\phi_{10})-(\phi_{01}-\phi_{00})}) \tag{15}$$

We therefore will use $\Phi = \delta\phi_1 - \delta\phi_0 = (\phi_{11} - \phi_{10}) - (\phi_{01} - \phi_{00})$ to characterise the entangling power of each gate.

4.2 A geometric picture of the unitary condition

We have four final ancilla states $|a_{ij}\rangle$ corresponding to the basis states of the register. To allow a measurement basis $\{|m\rangle, |m_\perp\rangle\}$ for which $|\langle m|a_{ij}\rangle|$ is constant $\forall i, j$, these four points must all lie in the same plane and form a ring around a cap with $|m\rangle$ at the midpoint. For four $|a_{ij}\rangle$ given by the preparation of the ancilla state, the two ancilla-register couplings and the intermediate local unitary gate on the ancilla, $|m\rangle$ will be fixed and unique. The relative phases of $\langle m|a_{ij}\rangle = |\langle m|a_{ij}\rangle|e^{i\phi_{ij}}$ corresponds to the angles between the points on the ring (see Figure 5).

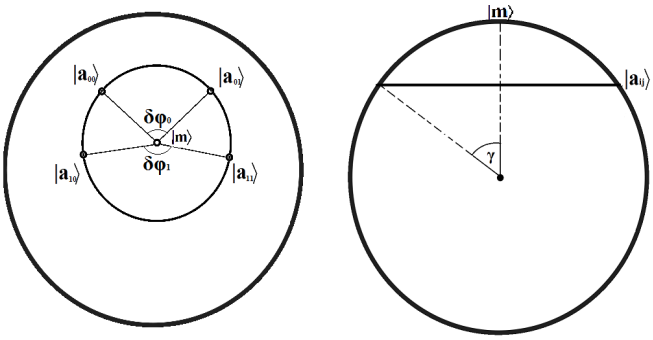


Figure 5 For four states that have the same value $|\langle m|a_{ij}\rangle|$, there are four points on the Bloch sphere that define a ring that encircle and thus define the state $|m\rangle$. $|\langle m|a_{ij}\rangle|^2 = \cos^2(\frac{\gamma}{2})$.

4.2.1 A geometric picture of the entanglement condition

After the first interaction but before the second the ancilla will be in one of two states $|a_i\rangle$ corresponding to $|\Psi\rangle_1 = |i\rangle_1$, $i = 0, 1$. The second interaction induces a unitary on the second register qubit that is given by $\langle m|\Delta|a_i\rangle$. For entanglement between the first and second register qubit, this unitary must be distinguishable by i. Therefore, each $|a_i\rangle$ must have a distinct value of $\langle\sigma_z\rangle_i$. In the geometric picture, two points given by a_i that are on the same horizontal plane, before the second interaction, will produce four a_{ij} on the same horizontal plane afterwards. However, the angle between states marked by different j in the pairs $|a_{0j}\rangle$ and $|a_{1j}\rangle$ would be the same for each i. To be entangling $\delta\phi_1 - \delta\phi_0$ must be non-zero.

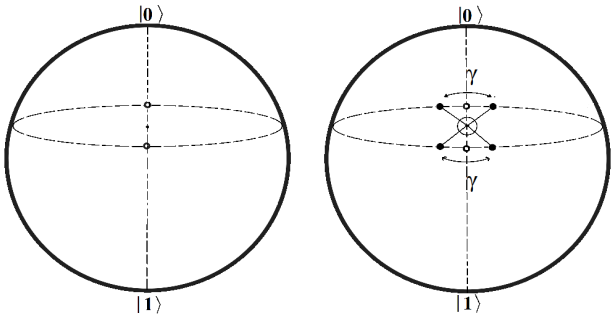


Figure 6 All horizontal planes have a cap with a midpoint at the poles thus $|m\rangle = |0\rangle$. An ancilla state in the computational basis corresponds to the same constant unitary enacted with each interaction and extracts no information from the first interaction to transmit in the intermediate stage.

From this geometric picture, we can show that the unitary and entangling conditions are fulfilled when the intermediate ancilla states $|a_i\rangle$ are restricted to the same vertical plane (see Appendix A and Figure 12).

4.3 Construction of the ancilla states

With the intermediate states $|a_i\rangle$ being of the same vertical plane but with some angle between them that we will label as 2β, the final states $|a_{ij}\rangle$ will be

$$\cos\left(\frac{\theta-(-1)^i 2\beta}{2}\right)|0\rangle + e^{i(-1)^j 2\alpha}\sin\left(\frac{\theta-(-1)^i 2\beta}{2}\right) \tag{16}$$

We can construct the vertically split states $|a_i\rangle$ and manipulate the angle 2β by the choice of preparation state and intermediate unitary U_a. Rotations around the z axis will form a gauge transformation since $|m\rangle$ is specified by $\{|a_{ij}\rangle\}$ so we can, without a loss of generality, assume that the ancilla preparation state is on the x–z plane: $|a\rangle = \cos(\frac{\theta'}{2})|0\rangle + \sin(\frac{\theta'}{2})|1\rangle$. After the first interaction this becomes $|a\rangle = \cos(\frac{\theta'}{2})|0\rangle + e^{i(-1)^i 2\alpha}\sin(\frac{\theta'}{2})|1\rangle$. Now the solid angle between the two resulting points is not given by 2α but by 2β where $\sin(2\beta) = \sin(\theta')\sin(2\alpha)$. There will always exist a unitary that can rotate the two points such that they lie in the x–z plane: $|a\rangle = \cos(\frac{\theta'}{2})|0\rangle + e^{i(-1)^i 2\alpha}\sin(\frac{\theta'}{2})|1\rangle \rightarrow \cos(\frac{\theta-(-1)^i 2\beta}{2})|0\rangle + \sin(\frac{\theta-(-1)^i 2\beta}{2})$. In the geometric picture, this is a rotation around the point where the great circle that connects the two points and the x–z plane cross, followed by any rotation around $\hat{y}$ of our choice so that θ is a parameter under control (see Figure 7). The next interaction produces the points $\vec{a}_{ij}$ which are, under this order of construction, dependent on the intermediate unitary and the ancilla preparation state.

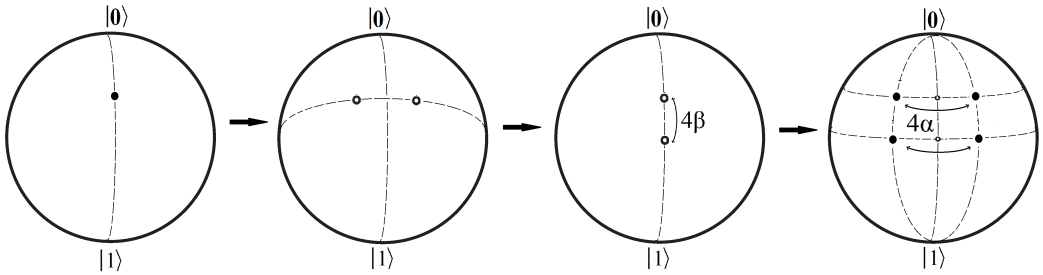


Figure 7 A geometric representation of the method of constructing an entangling two qubit unitary. The first step corresponds to the ancilla preparation, the second and fourth to the interactions with register qubits and the third to unitary actions on the ancilla in between the interactions.

4.4 An example of the relative entanglement powers of the Kraus operators

As an example, take the intermediate state to have $\theta = \frac{\pi}{2}$ so that the $|a_i\rangle$ are vertically split about the $|+\rangle$ state. It is helpful to think of the intermediate state as $R_{\hat{x}}(\frac{\pi}{2})R_{\hat{z}}(\pm 2\beta)|+\rangle$ while the ancilla was prepared in $|+\rangle$. The effect of the preparation choice and the reduced solid angle are treated like an effective reduction in the interaction strength of the first interaction while the intermediate $U_a = R_{\hat{x}}(\frac{\pi}{4})$. The four states of $|a_{ij}\rangle = R_{\hat{z}}((-1)^j 2\alpha)R_{\hat{x}}(\frac{\pi}{2})R_{\hat{z}}((-1)^i 2\beta)|+\rangle$ will be all symmetrically placed around $|+\rangle$ so we can also say $|m\rangle = |+\rangle$ and measure in the $\{|+\rangle, |-\rangle\}$ basis.

$$\langle m|a_{ij}\rangle = \langle +|R_{\hat{z}}((-1)^j 2\alpha)R_{\hat{x}}\left(\frac{\pi}{2}\right)R_{\hat{z}}((-1)^i 2\beta)|+\rangle \tag{17}$$

$$\mathbf{R}_{\hat{z}}(2\alpha)\mathbf{R}_{\hat{x}}\left(\frac{\pi}{2}\right)\mathbf{R}_{\hat{z}}(2\beta) = \frac{1}{\sqrt{2}}\begin{pmatrix} e^{-iA} & -ie^{-iB} \\ -ie^{iB} & e^{iA} \end{pmatrix} \tag{18}$$

$$\rightarrow \langle +|a_{00}\rangle = \frac{1}{2\sqrt{2}}(e^{-iA} - ie^{-iB} - ie^{iB} + e^{iA}) = \frac{1}{\sqrt{2}}(\cos(A) - i\cos(B)) \tag{19}$$

$$\langle -|a_{00}\rangle = \frac{1}{2\sqrt{2}}(e^{-iA} - ie^{-iB} + ie^{iB} - e^{iA}) = \frac{1}{\sqrt{2}}(-i\sin(A) - \sin(B)) \tag{20}$$

where we define $A = \alpha + \beta$, $B = \alpha - \beta$. Each other element of the two qubit evolution operator will just be a transformation of A and B (which thus fulfils the unitary condition) and because of the $\pm$ symmetries of sine and cosine we will be able to express the final Kraus operator and Φ in terms of just ϕ_{00}.

$$\Phi_+ = \delta\phi_1 - \delta\phi_0 = 4\phi_{00}^+ + \pi \tag{21}$$

$$\Phi_- = 4\phi_{00}^- + \pi \tag{22}$$

Table 1 Table of transformations of A and B for different computational states of the register.

i	j	A→	B→	ϕ_{ij}^+	ϕ_{ij}^-
0	0	A	B	ϕ_{00}^+	ϕ_{00}^-
0	1	-B	-A	$-\phi_{00}^+ - \frac{\pi}{2}$	$-\phi_{00}^- + \frac{\pi}{2}$
1	0	B	A	$-\phi_{00}^+ - \frac{\pi}{2}$	$-\phi_{00}^- - \frac{\pi}{2}$
1	1	-A	-B	ϕ_{00}^+	$\phi_{00}^- + \pi$

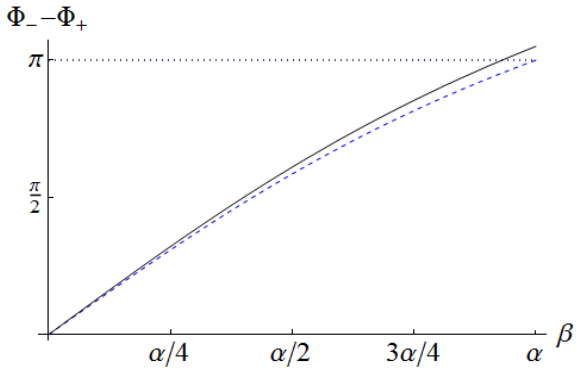


Figure 8 By manipulating the effective coupling strength, β, with the ancilla state preparation, the difference in Φ for the two possible operator outputs can be adjusted.For $\alpha = \frac{\pi}{16}$, the difference between outputs is plotted with the solid (black) curve while the value of Φ_- is plotted with the dashed (blue) curve (colour online). At $\beta \approx 0.183$, $\delta\Phi = \pi$

When the ancilla is prepared and measured in the Pauli-X eigenstates, it is possible to perform this procedure with a two parameter interaction and induce only additional local unitaries on the register qubits. A $e^{-i\alpha_x\sigma_x\otimes\sigma_x}$ contribution would not perturb the ancilla in that case and leave ancilla and register separable notwithstanding other parameter contributions.

4.5 Repeat-until-success entangling gate generation (EGG)

Without any optimisation of the ancilla initial state and measurement, the generation of entangling gates may, at the very least, be able to perform a random walk (a classical random

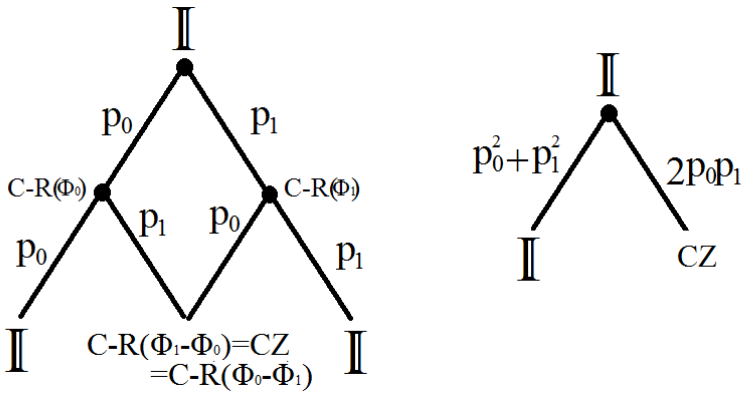


Figure 9 Since all enacted gates are of the same group, Control-$R_{\hat{z}}(\Phi)$ gates, multiple gates can be easily combined. Random gate production is a Markovian process. By manipulating the relative values, the output can be limited to a finite probability tree, including a case equivalent to a single outcome "success/fail" gate suitable for a repeat-until-success method.

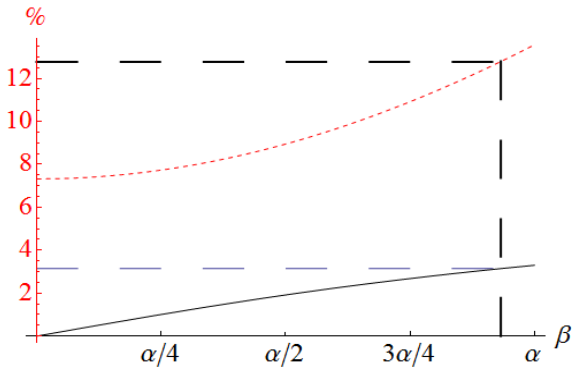


Figure 10 The probability of achieving two opposing ancilla measurements, $2p_0p_1$, against β, displayed against the resulting Φ. At the case where $\delta\Phi = \pi$, $p \approx 0.128$ (colour online).

walk, not to be confused with quantum walks) through the group of Control-$R_{\hat{z}}(\gamma)$. With optimisation, we can apply a protocol to convert the output into a "success/fail" scenario with failure corresponding to generation of identity.

In this protocol EGG is enacted twice: once as described before and once again with local unitary changes that convert $\Phi \to -\Phi$. The latter can be done either by local Pauli-X gate pre- and post-corrections on the register or, if there is access to the ancilla between interactions, by making a correction to the intermediate ancilla unitary to exchange the resulting intermediate states $|a_0\rangle$—$|a_1\rangle$. The change in the second implementation enacts a change of the resulting gate $C\text{-}R_{\hat{z}}(\gamma) \to C\text{-}R_{\hat{z}}(-\gamma)$. If these two output gates are (under local operator corrections) $C\text{-}R_{\hat{z}}(\Phi_0)$ and $C\text{-}R_{\hat{z}}(\Phi_1)$ with probabilities p_0 and p_1 then after two two EGG the Control-Unitary is one of $C\text{-}R_{\hat{z}}(\Phi_0 - \Phi_1)$, $C\text{-}R_{\hat{z}}(\Phi_1 - \Phi_0)$ or $\mathbb{I}$. If $\Phi_0 - \Phi_1 = \pi$ then we have performed probabilistic CZ generation with a success probability of $2p_0p_1$.

For example, in Figure 8, we see that we can lower x to match such conditions. In Figure 10, the value of $2p_0p_1$ is calculated, giving us a value of $p \approx 0.128$.

We provide this one adaptation as a single example of the general principle of using

a multi-step protocol to adapt the behaviour of a Markov chain. Though an arbitrary interaction strength leads to the probabilistic generation of gates where the outputs are not local unitary correctable, they are of the same interaction class and thus can be adapted to the group of Control-Phase Rotations and generate group members. Specific physical parameter constraints and choice of target gates may lead to methods that involve different schemes and different groups of gates such as $C\text{-}R_{\hat{z}}(\frac{m\pi}{n})$, $m, n \in \mathbb{Z}$ for a specific n or a continuous parameter group $C\text{-}R_{\hat{z}}(\gamma)$. We provide a method that allows one to enact a repeat-until success method similar to other probabilistic gate proposals for linear optical systems [25, 18, 23].

5 Initialisation-Measurement

In ADQC, the ability to perform a projective measurement with an ancilla qubit was straightforward and generalized measurements could be performed by introducing a second ancilla system. With a weak coupling, enacting a projective measurement is a less straightforward proposition requiring many steps and possible fidelity loss. We will provide a protocol for doing so in a naive adaptation of the Control-Z+Hadamard example, followed by discussion of potential future investigation and improvement.

With a general register qubit $|\Psi\rangle_R = \alpha|0\rangle + \beta|1\rangle$ and ancilla in initialised state $|+\rangle_A$, the interaction E acting on $|\Psi\rangle_R|+\rangle_A$ would produce $\alpha|+\rangle_R|0\rangle_A + \beta|-\rangle_R|1\rangle_A$. A z basis measurement on the ancilla would replicate a z basis measurement on the register state before the interaction; this can also be used to initialise the register qubit with a result $|j\rangle$ producing a register qubit in the state $Z^j|+\rangle$. Local unitary gate preparations on the register could be used to enable measurements or state initialisation in other abses.

In our proposal, the interaction between register and ancilla can be weaker, enabling us to perform only non-projective measurements. Fortunately Oreskhov & Brun [29] show that weak measurements are universal and provides an explicit construction for how to use weak measurements to achieve projective and general measurements in a random walk, for two outcome measurements. This can be extended to higher dimensions [29] and it has been shown that one can build up many result general measurements from two output measurements [4]. Thus we expect that it should be possible to achieve at least an approximation of projective measurements using multiple successive non-projective measurements from a single ancilla. We have a specific interest in being able to do so using our interaction of interest $(H_A \otimes H_R).C\text{-}T_{AR}$.

5.1 Initialisation and Measurement in logarithmic time with a fixed interaction

At the heart of our proposal is the capability to perform a two qubit unitary that is equivalent up to local unitary gates to a Control-$R_z(\theta)$ gate. Consider $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$ and an ancilla in the $|+\rangle$ state:

$$(H_A \otimes H_R).C\text{-}R_z(\theta)|\Psi\rangle_R|+\rangle_A = H(\alpha|0\rangle + \beta\cos\left(\frac{\theta}{2}\right)|1\rangle)_R|0\rangle_A + \beta\sin\left(\frac{\theta}{2}\right)H|1\rangle_R|0\rangle_A \quad (23)$$

Now if we measure the ancilla qubit in the $\{|0\rangle, |1\rangle\}$ eigenstate basis and make a local Hadamard correction on the register, the register qubit will be projected into the respective states $\frac{1}{\sqrt{N}}(\alpha|0\rangle + \beta\cos(\frac{\theta}{2})|1\rangle)$ (with probability $N = |\alpha|^2 + |\beta|^2\cos^2(\frac{\theta}{2})$) and $|1\rangle$ (with probability $|\beta|^2\sin^2(\frac{\theta}{2})$).

Consider the effect of many such measurements in the case where every successive result on the ancilla is $|0\rangle$. The iterated effect over n steps is for the register qubit to be projected into the state $\frac{1}{\sqrt{N_n}}(\alpha|0\rangle + \beta\cos^n(\frac{\theta}{2})|1\rangle)$ (where N_n is the appropriate normalization factor). As n increases the state tends to $|0\rangle$ exponentially. The probability of achieving such a chain is $N_n = |\alpha|^2 + |\beta|^2\cos^{2n}(\frac{\theta}{2})$ which will tend to $|\alpha|^2$. Thus we can replicate a projective measurement in n steps up to a state fidelity error of $\beta\cos^n(\frac{\theta}{2}) \leq \cos^n(\frac{\theta}{2})$ with the following procedure:

Set a desired state fidelity error ϵ. Calculate $n \in \mathbb{Z}$ such that $n \geq ln(\epsilon)/ln(\cos(\frac{\theta}{2}))$. Prepare ancilla qubit in the $|+\rangle$ state, couple to register qubit with the interaction E_{AR}, measurement ancilla qubit in computational basis, apply H gate correction to register qubit. Repeat until measurement has been performed n times or until ancilla is measured as $|1\rangle$. Label a result of $|0_1 0_2 0_3 ... 0_n\rangle_A$ as $|0\rangle_R$, any other result where the process was terminated by a $|1\rangle_A$ as $|1\rangle_R$.

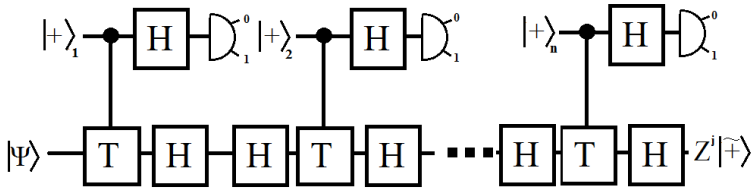


Figure 11 The circuit description of the iterative measurement protocol using the fundamental interaction $E_{AR} = (H_a \otimes H_R)C\text{-}T$. The local unitary gate components of the interaction require a Hadamard gate correction; though not depicted, this is implemented using ancilla qubits as described in the previous section, leading to a total number of ancilla qubits and interactions of $2n$.

In this way, the measurement does not follow a random walk but occurs within a finite bound on the number of steps that is logarithmic with respect to the state fidelity error.

6 Avenues of further investigation

Both the classes $\Delta(\alpha, 0, 0)$ and $\Delta(\alpha_x, \alpha_y, 0)$ can be used to enact proportional-to-unitary Kraus operators however our work has mainly dealt with the former. Stochastic ADQC motivates extending the method of enacting two qubit gates to the two parameter interaction class by providing means in which the $\Delta(\alpha_x, \alpha_y, 0)$ may be viable for single qubit gates. It also may provide a useful tool for examining the trade off between control or time. The random gate generation method we have described incoporates little control over the initial ancilla state or measurement basis. Potentially future work could consider employing feedback to optimise the choice of ancilla state preparation and measurement basis in each step, resulting in a variable step size and guided random walk. The measurement we perform in each step is biased and we incorporate no feedback in our protocol save for the decision to continue or halt based on the previous measurement result. We have also only considered measurement on one qubit while the use of an ancilla provides us with a natural system for making higher dimensional measurements and generalised measurements. Combes & Jacobs [10] considered using feedback to speed up the purification of states via continuous measurements. A case where the measurement basis is kept unbiased to the state density operator through feedback is faster than measurement alone [11]. Much work has been done considering the case where we have a sequence of weak measurement with feedback enacted through unitary operations

between steps in time for various goals and measures [33, 34]. Jacobs [19] raised the question of whether there is a trade off between speed of projection and loss of initial information. We propose the work of coming up with an unbiased basis feedback protocol for our specific interaction and non-continuous discrete protocol for future investigation.

7 Conclusion and summary

We have show how, by relaxing the requirement for stepwise determinism up to Pauli corrections, a broader class of interactions can be used to implement a form of ancilla driven quantum computation. This is achieved by gate approximation, repeat-until-success strategies, and generalised measurement. Single qubit unitary gates can still be implemented deterministically while two qubit gates and measurement and state initialisation will be probabilistic and require the development of probabilistic protocols which we provide examples of. We provide a measurement process that is logarithmic to the bound on state fidelity error and reduces to a deterministic approximation of projective measurements, with the potential for further speed up to be examined in future research. The conditions for the generation of a two qubit entangling unitary gate are described in geometric terms. We demonstrate how choice of ancilla state preparation and local unitaries enable us to manipulate the entangling strength of the output gates. This allows us to convert the result into a "repeat-until-success" gate. Further research may reveal other protocols based around adaptations of the entangling gate power. Future research into probabilistic protocols for enacting specific unitary gates may also aid in expanding possible interactions to include the class $\Delta(\alpha_x, \alpha_y, 0)$.

We would like to acknowledge discussions with Dan Browne and Erika Andersson, and comments on the draft manuscript by Janet Anders and Elham Kashefi. DKLO and KHS are supported by the Quantum Information Scotland (QUISCO) network.

References

1 Panos Aliferis, Daniel Gottesman, and John Preskill. Quantum accuracy threshold for concatenated distance-3 codes. *eprint arXiv:quant-ph/0504218*, April 2005.

2 J. Anders, E. Andersson, D.E. Browne, E. Kashefi, and D.K.L. Oi. Ancilla-driven quantum computation with twisted graph states. *Theoretical Computer Science*, 430(0):51–72, 2012. Mathematical Foundations of Programming Semantics (MFPS XXV).

3 Janet Anders, Daniel K. L. Oi, Elham Kashefi, Dan E. Browne, and Erika Andersson. Ancilla-driven universal quantum computation. *Phys. Rev. A*, 82(2):020301, Aug 2010.

4 Erika Andersson and Daniel K. L. Oi. Binary search trees for generalized measurements. *Phys. Rev. A*, 77:052104, May 2008.

5 A. Bermudez, F. Jelezko, M. B. Plenio, and A. Retzker. Electron-mediated nuclear-spin interactions between distant nitrogen-vacancy centers. *Phys. Rev. Lett.*, 107:150503, Oct 2011.

6 B. B. Blinov, D. L. Moehring, L.-M. Duan, and C. Monroe. Observation of entanglement between a single trapped atom and a single photon. *Nature*, 428:153–157, Mar 2004.

7 P. Oscar Boykin, Tal Mor, Matthew Pulver, Vwani Roychowdhury, and Farrokh Vatan. On universal and fault-tolerant quantum computing. *eprint arXiv:quant-ph/9906054*, June 1999.

8 Dan Browne and Hans J. Briegel. One-way quantum computation. In *Lectures on Quantum Information*, pages 359–358. Wiley-VCH, Weinheim,Germany, 2007.

9 C. W. Chou, D. B. Hume, J. C. J. Koelemeij, D. J. Wineland, and T. Rosenband. Frequency comparison of two high-accuracy al$^+$ optical clocks. *Phys. Rev. Lett.*, 104:070802, Feb 2010.

10 Joshua Combes and Kurt Jacobs. Rapid state reduction of quantum systems using feedback control. *Phys. Rev. Lett.*, 96:010504, Jan 2006.

11 Joshua Combes, Howard M. Wiseman, Kurt Jacobs, and Anthony J. O'Connor. Rapid purification of quantum systems by measuring in a feedback-controlled unbiased basis. *Phys. Rev. A*, 82:022307, Aug 2010.

12 G. Cordourier-Maruri, F. Ciccarello, Y. Omar, M. Zarcone, R. de Coss, and S. Bose. Implementing quantum gates through scattering between a static and a flying qubit. *Phys. Rev. A*, 82:052313, Nov 2010.

13 D. Deutsch. Quantum computational networks. *Proceedings of the Royal Society of London. A. Mathematical and Physical Sciences*, 425(1868):73–90, 1989.

14 M. V. Gurudev Dutt, L. Childress, L. Jiang, E. Togan, J. Maze, F. Jelezko, A. S. Zibrov, P. R. Hemmer, and M. D. Lukin. Quantum register based on individual electronic and nuclear spin qubits in diamond. *Science*, 316(5829):1312–1316, 2007.

15 Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. Quantum Computation By Adiabatic Evolution. *eprint arXiv:quant-ph/0001106*, January 2000.

16 D. Gross and J. Eisert. Novel schemes for measurement-based quantum computation. *Phys. Rev. Lett.*, 98:220503, May 2007.

17 D. Gross, J. Eisert, N. Schuch, and D. Perez-Garcia. Measurement-based quantum computation beyond the one-way model. *Phys. Rev. A*, 76:052315, Nov 2007.

18 D. Gross, K. Kieling, and J. Eisert. Potential and limits to cluster-state quantum computing using probabilistic gates. *Phys. Rev. A*, 74:042343, Oct 2006.

19 Kurt Jacobs. How to project qubits faster using quantum feedback. *Phys. Rev. A*, 67:030301, Mar 2003.

20 E. Kashefi, D.K.L. Oi, D. Browne, J. Anders, and E. Andersson. Twisted graph states for ancilla-driven universal quantum computation. *Electronic Notes in Theoretical Computer Science*, 249(0):307–331, 2009. Proceedings of the 25th Conference on Mathematical Foundations of Programming Semantics (MFPS 2009).

21 M. Keyl and R.F. Werner. Channels and maps. In *Lectures on Quantum Information*, pages 73–86. Wiley-VCH, Weinheim,Germany, 2007.

22 A.Yu. Kitaev. Fault-tolerant quantum computation by anyons. *Annals of Physics*, 303(1):2–30, 2003.

23 Pieter Kok, W. J. Munro, Kae Nemoto, T. C. Ralph, Jonathan P. Dowling, and G. J. Milburn. Linear optical quantum computing with photonic qubits. *Rev. Mod. Phys.*, 79:135–174, Jan 2007.

24 B. Kraus and J. I. Cirac. Optimal creation of entanglement using a two-qubit gate. *Phys. Rev. A*, 63:062309, May 2001.

25 Yuan Liang Lim, Almut Beige, and Leong Chuan Kwek. Repeat-until-success linear optics distributed quantum computing. *Phys. Rev. Lett.*, 95:030505, Jul 2005.

26 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, Cambridge, United Kingdom, 2000.

27 Toshio Ohshima. All-optical electron spin quantum computer with ancilla bits for operations in each coupled-dot cell. *Phys. Rev. A*, 62:062316, Nov 2000.

28 D. K. L. Oi and S. G. Schirmer. Quantum system characterization with limited resources. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 370(1979):5386–5395, 2012.

29 Ognyan Oreshkov and Todd A. Brun. Weak measurements are universal. *Phys. Rev. Lett.*, 95:110409, Sep 2005.

30 Robert Raussendorf and Hans J. Briegel. A one-way quantum computer. *Phys. Rev. Lett.*, 86:5188–5191, May 2001.

31 A. T. Rezakhani. Characterization of two-qubit perfect entanglers. *Phys. Rev. A*, 70:052313, Nov 2004.

32 W. Forrest Stinespring. Positive functions on c*-algebras. *Proceedings of the American Mathematical Society*, 6(2):pp. 211–216, 1955.

33 H M Wiseman and J F Ralph. Reconsidering rapid qubit purification by feedback. *New Journal of Physics*, 8(6):90, 2006.

34 Howard M. Wiseman and Luc Bouten. Optimality of feedback control strategies for qubit purification. *Quantum Information Processing*, 7:71–83, 2008.

35 Jun Zhang, Jiri Vala, Shankar Sastry, and K. Birgitta Whaley. Geometric theory of nonlocal two-qubit operations. *Phys. Rev. A*, 67:042313, Apr 2003.

A Restrictions of the parameters of the ancilla found in a geometric proof

Each point can be represented by the spherical coordinates that describe the state $\vec{a}_k = (\theta_k, \phi_k)$, $k = 1, 2, 3, 4$. For out notation the values of k will correspond to the values expressed by ij in binary.

Each point can be expressed in Cartesian coordinates by the relationships:

$$|\vec{a}_k.\vec{x}| = \sin(\theta_k)\cos(\phi_k) \tag{24}$$

$$|\vec{a}_k.\vec{y}| = \sin(\theta_k)\sin(\phi_k) \tag{25}$$

$$|\vec{a}_k.\vec{z}| = \cos(\theta_k) \tag{26}$$

Three points alone can always be found to be on the same plane. We will define a plane from three points and then find the expression for the distance from the fourth point to that plane. Thus we will find the conditions for the fourth point to be in the same plane as the other three. The equation for a plane defined by three points is

$$a.x + b.y + c.z + d = 0 \tag{27}$$

$$a = \frac{-d}{D}\begin{vmatrix} 1 & y_1 & z_1 \\ 1 & y_2 & z_2 \\ 1 & y_3 & z_3 \end{vmatrix}, b = \frac{-d}{D}\begin{vmatrix} x_1 & 1 & z_1 \\ x_2 & 1 & z_2 \\ x_3 & 1 & z_3 \end{vmatrix}, c = \frac{-d}{D}\begin{vmatrix} x_1 & y_1 & 1 \\ x_2 & y_2 & 1 \\ x_3 & y_3 & 1 \end{vmatrix}, D = \begin{vmatrix} x_1 & y_1 & z_1 \\ x_2 & y_2 & z_2 \\ x_3 & y_3 & z_3 \end{vmatrix} \tag{28}$$

There is a freedom of choice of d so it can be set $d = D$. The distance of point 4 to the plane is given by

$$\text{distance} = \frac{|a.x_4 + b.y_4 + c.z_4 + d|}{\sqrt{a^2 + b^2 + c^2}} \tag{29}$$

As we are only interested in the case where the distance is zero, we can ignore the normalisation factor and simply examine

$$\text{distance}' = |a.x_4 + b.y_4 + c.z_4 + d| \tag{30}$$

We now make use of some restrictions on the formation of these four points. Before the second interaction, there are intermediate ancilla states corresponding to the computational basis of the first register qubit. The final four points are constructed from these two points

by the second interaction which imparts a rotation around the $\vec{z}$ axis by $\pm 2\alpha$ with $\pm$ corresponding to the second qubit being in the computational basis state $|j\rangle$, $j = 0, 1$. This means that we can set

$$\theta_1 = \theta_2, \theta_3 = \theta_4 \tag{31}$$

and

$$\phi_2 - \phi_1 = \phi_4 - \phi_3 = 4\alpha \tag{32}$$

Using (31), the distance between the fourth point and the plane can be simplified to

$$2(\cos(\theta_2) - \cos(\theta_4)) \left[\cos\left(\phi_2 - \frac{\phi_3 + \phi_4}{2}\right) - \cos\left(\phi_1 - \frac{\phi_3 + \phi_4}{2}\right)\right] \sin(\theta_2)\sin(\theta_4)\sin(\frac{\phi_3 - \phi_4}{2}) \tag{33}$$

This generates several possible conditions for the fourth point to lie in the plane, some more trivial than others. If $\cos(\theta_2) = \cos(\theta_4)$ then all four points must lie on the same horizontal plane which means that there is no entangling power. $\sin(\theta_2) = 0$ and $\sin(\theta_4) = 0$ would mean that there are only three distinct points with one being at the pole i.e. the $|0\rangle$ state. Due to the construction of these four points it is not possible for $\phi_3 - \phi_4 = 0$ to be true. The final condition is that

$$\cos\left(\phi_2 - \frac{\phi_3 + \phi_4}{2}\right) = \cos\left(\phi_1 - \frac{\phi_3 + \phi_4}{2}\right) \tag{34}$$

If (32) is then substituted in, this becomes

$$\cos\left(\phi_1 + 4\alpha - \frac{2\phi_3 + 4\alpha}{2}\right) = \cos\left(\phi_1 - \frac{2\phi_3 + 4\alpha}{2}\right) \tag{35}$$

$$\cos\left(\phi_1 - \phi_3 + 2\alpha\right) = \cos\left(\phi_1 - \phi_3 - 2\alpha\right) \tag{36}$$

Since α is non-zero, this requires $\phi_1 = \phi_3 + n\pi$ for $n \in \mathbb{Z}$ i.e. the two points are in the same vertical plane.

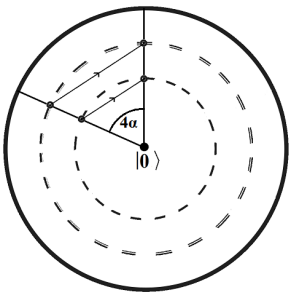
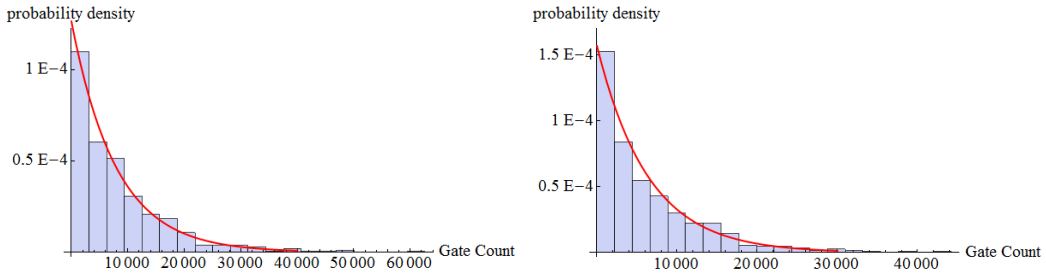
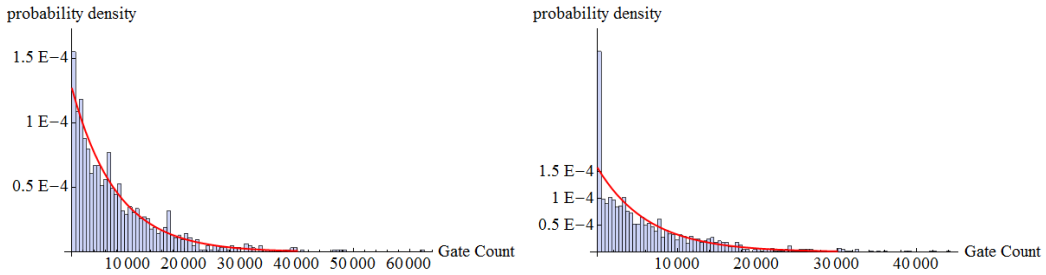


Figure 12 A 2d projection of the construction of four point on the Bloch sphere. By restricting the points to only one of two polar and one of two azimuthal angles, the vectors that connect two points of the same polar angle will be parallel. This guarantees that all four points lie on the same plane.

B Graphs of stochastic ADQC gate count and the exponential distribution

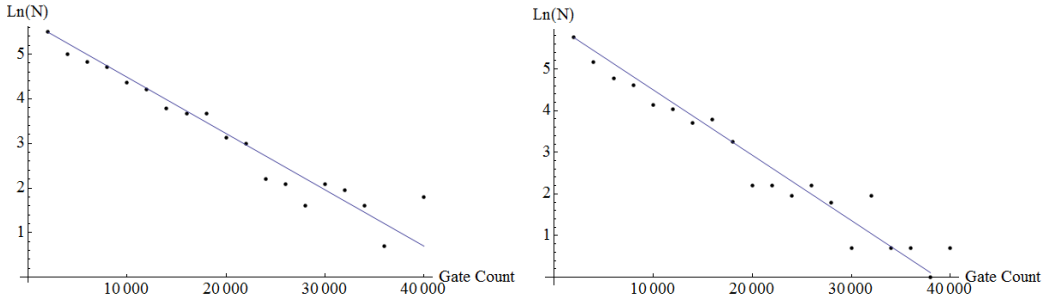


(a) $H_R\Delta(0,0,\frac{\pi}{16})$ **(b)** $\Delta(\frac{\pi}{16},0,\frac{\pi}{16})$



(c) $H_R\Delta(0,0,\frac{\pi}{16})$ **(d)** $\Delta(\frac{\pi}{16},0,\frac{\pi}{16})$

Figure 13 Probability distribution of required gate count to achieve target $U_T = R_{\hat{x}}(\frac{\pi}{2})$. a) Use of a single parameter interaction in a 20 bin histogram, b) Use of a two parameter interaction in a 20 bin histogram,c) Use of a single parameter interaction in a 100 bin histogram, d) Use of a two parameter interaction in a 100 bin histogram; note the large aberration in the first division. The probability distribution corresponding to the exponential distribution parametrised by the mean of the results is displayed by the solid red curve (colour online).



(a) $H_R\Delta(0,0,\frac{\pi}{16})$ **(b)** $\Delta(\frac{\pi}{16},0,\frac{\pi}{16})$

Figure 14 The natural logarithm of the bin counts of the number of gates required to achieve target $U_T = R_{\hat{x}}(\frac{\pi}{2})$. By taking the natural logarithm the exponential distribution model forms a linear curve which fits the points generated by the simulation. Noise distrupts the linear behaviour for very low counts of high gate number. Figure (a) matches the one parameter interaction, Figure (b) matches the two parameter interaction.

Another Subexponential-time Quantum Algorithm for the Dihedral Hidden Subgroup Problem*

Greg Kuperberg

University of California, Davis
greg@math.ucdavis.edu

Abstract

We give an algorithm for the hidden subgroup problem for the dihedral group D_N, or equivalently the cyclic hidden shift problem, that supersedes our first algorithm and is suggested by Regev's algorithm. It runs in $\exp(O(\sqrt{\log N}))$ quantum time and uses $\exp(O(\sqrt{\log N}))$ classical space, but only $O(\log N)$ quantum space. The algorithm also runs faster with quantumly addressable classical space than with fully classical space. In the hidden shift form, which is more natural for this algorithm regardless, it can also make use of multiple hidden shifts. It can also be extended with two parameters that trade classical space and classical time for quantum time. At the extreme space-saving end, the algorithm becomes Regev's algorithm. At the other end, if the algorithm is allowed classical memory with quantum random access, then many trade-offs between classical and quantum time are possible.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases quantum algorithm, hidden subgroup problem, sieve, subexponential time

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.20

1 Introduction

In a previous article [7], we established a subexponential-time algorithm for the dihedral hidden subgroup problem, which is equivalent to the abelian hidden shift problem. That algorithm requires $\exp(O(\sqrt{\log N}))$ time, queries, and quantum space to find the hidden shift s in the equation $g(x) = f(x+s)$, where f and g are two injective functions on $\mathbb{Z}/N$. In this article we present an improved algorithm, Algorithm 7, which is much less expensive in space, as well as faster in a heuristic model. Our algorithm was inspired by and generalizes Regev's algorithm [10]. It uses $\exp(O(\sqrt{\log N}))$ *classical* space, but only $O(\log N)$ *quantum* space. We heuristically estimate a total computation time of $\widetilde{O}(2^{\sqrt{2\log_2 N}})$ for the new algorithm; the old algorithm takes time $\widetilde{O}(3^{\sqrt{2\log_3 N}})$.

The algorithm also has two principal adjustable parameters. One parameter allows the algorithm to use less space and more quantum time. A second parameter allows the algorithm to use more classical space and classical time and less quantum time, if the classical space has quantum access [5]. (See also Section 2.) Finally, the new algorithm can take some advantage of multiple hidden shifts; somewhat anomalously, our old algorithm could not.

The new algorithm can be called a *collimation sieve*. As in the original algorithm and Regev's algorithm, the weak Fourier measurement applied to a quantum query of the hiding function yields a qubit whose phases depend on the hidden shift s. The sieve makes larger qudits from the qubits which we call *phase vectors*. It then collimates the phases of the

* Partly supported by NSF grant DMS CCF-1013079

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 20–34

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

qudits with partial measurements, until a qubit is produced whose measurement reveals the parity of s. We also use a key idea from Regev's algorithm to save quantum space. The sieve is organized as a tree with $O(\sqrt{\log N})$ stages, and we can traverse the tree depth first rather than breadth first. The algorithm still uses a lot of classical space to describe the coefficients of each phase vector when it lies in a large qudit. If the qudit has dimension ℓ, then this is only $O(\log \ell)$ quantum space, but the classical description of its phases requires $\widetilde{O}(\ell)$ space.

The main discussion of the dihedral hidden subgroup problem has been as an algorithm with a black-box hiding function. Recently Childs, Jao, and Soukharev [4] found a classical, white-box instance of the dihedral hidden subgroup problem, or the abelian hidden shift problem. The instance is that an isogeny between isogenous, ordinary elliptic curves can be interpreted as a hidden shift on a certain abelian group. Thus, just as Shor's algorithm allows quantum computers to factor large numbers, an abelian hidden shift algorithm allows quantum computers to find isogenies between large elliptic curves. This is a new impetus to study algorithms for the dihedral hidden shift problem.

Before describing the algorithm, we review certain points of quantum complexity theory in general, and quantum algorithms for hidden structure problems. We adopt the general convention that if X is a finite set of orthonormal vectors in a Hilbert space $\mathcal{H}$ (but not necessarily a basis), then

$$|X\rangle \stackrel{\text{def}}{=} \sqrt{|X|} \sum_{x \in X} |x\rangle$$

is the constant pure state on X. Also if X is an abstract finite set, then $\mathbb{C}[X]$ is the Hilbert space in which X is an orthonormal basis. Also we use the notation

$$[n] = \{0, 1, \ldots, n-1\},$$

so that $\mathbb{C}[[n]]$ becomes another way to write the vector space $\mathbb{C}^n$.

2 Quantum time and space

As with classical algorithms, the computation "time" of a quantum algorithm can mean more than one thing. One model of quantum computation is a quantum circuit that consists of unitary operators and measurements, or even general quantum operations, and is generated by a classical computer. (It could be adaptively generated using quantum measurements.) Then the circuit depth is one kind of quantum time, a type of parallel time. The circuit gate complexity is another kind of quantum time, a type of serial time. We can justify serial quantum time with the following equivalence with a RAM-type machine.

▶ Proposition 1. The gate complexity of a classically uniform family of quantum circuits is equivalent, up to a constant factor, to the computation time of a RAM-type machine with a classical address register, a quantum data register, a classical tape, and a quantum tape.

We will discuss Proposition 1 more rigorously in Section 2.1. From either the circuit viewpoint or the RAM machine viewpoint, serial computation time is a reasonable cost model: in practice, gate operations are more expensive than simple memory multiplied by clock time.

An interesting and potentially important variation of the random-access model is quantum random access memory, or QRAM [5]. In this model, there is an address register composed of qubits and a memory can be accessed in quantum superposition, whether or not the cells of the memory tape are classical. Of course, if the memory is classical, only read operations can be made in quantum superposition. A RAM quantum computer thus has four

possible types of memory tapes: classical access classical memory (CRACM), quantum access classical memory (QRACM), classical access quantum memory (CRAQM), and quantum access quantum memory (QRAQM).

Hypothetically, one could cost quantum access classical memory (QRACM) simply as quantum memory. But for all we know, quantum access classical memory (QRACM) and classical-access quantum memory (CRAQM) are non-comparable resources. We agree with the suggestion [3] that quantum-access classical memory could be cheaper than quantum memory with either classical or quantum access. After all, such memory does not need to be preserved in quantum superposition. Our own suggestion for a QRACM architecture is to express classical data with a 2-dimensional grid of pixels that rotate the polarization of light. (A liquid crystal display has a layer that does exactly that.) When a photon passes through such a grid, its polarization qubit reads the pixel grid in superposition. Such an architecture seems easier to construct than an array of full qubits.

A good example of an algorithm that uses QRACM is the Brassard-Høyer-Tapp algorithm for the 2-to-1 collision problem [3], as the authors themselves point out. Given a function $f : X \to Y$ where X has N elements, the algorithm generates $N^{1/3}$ values of f at random and then uses a Grover search over $N^{2/3}$ values to find a collision; thus the time complexity is $\widetilde{O}(N^{1/3})$. This is a large-memory algorithm, but the bulk of the memory only needs to be quantumly addressable classical memory. By contrast, Ambainis' algorithm [2] for the single collision problem uses true quantum memory.

▶ Proposition 2. In the RAM model, a quantum access memory with N quantum or classical cells can be simulated with a classical linear access memory, with the same cells, with $\widetilde{O}(N)$ time overhead.

2.1 Some rigor

Here we give more precise definitions of quantum RAM machine models, and we argue Propositions 1 and 2. We would like models that have no extraneous polynomial overhead, although they might have polylogarithmic overhead. On the other hand, it seems very difficult to regularize polylogarithmic overhead. In our opinion, different models of computation that differ in polylogarithmic overhead could be equally good. Actually, at some level a physical computer has at most the computational strength of a 3-dimensional cellular automaton, where again, the total number of operations is as important as the total clock time. (Or even a 2-dimensional cellular automaton; a modern computer is approximately a 2-dimensional computer chip.) Procedural programming languages typically create a RAM machine environment, but usually with polylogarithmic overhead that depends on various implementation details.

A classical Turing machine M is a tuple (S, Γ, δ), where S is a finite set of states, Γ is a finite alphabet, and δ is a transition map. The Turing machine has a tape which is linear in one direction with a sequence of symbols in Γ, which initially are all the blank symbol $b \in \Gamma$ except for an input written in the alphabet $\Sigma = \Gamma \setminus \{b\}$. The state set S includes an initial state, a "yes" final state, and a "no" final state. Finally the transition map δ instructs the Turing machine to change state, write to the tape, and move along the tape by one unit.

In one model of a RAM machine, it is a Turing machine M with two tapes, an address tape T_A with the same rules as a usual linear tape; and a main work tape T_W. The machine M (as instructed by δ) can now also read from or write to $T_W(T_A)$, meaning the cell of the tape T_W at the address expressed in binary (or some other radix) on the tape T_A. It is known [8, 9] that a RAM machine in this form is polylog equivalent to a *tree Turing machine*, meaning a standard Turing machine whose tape is an infinite rooted binary tree.

It is useful to consider an intermediate model in which the transition map δ is probabilistic, *i.e.*, a stochastic matrix rather than a function. (Or a substochastic matrix rather than a partial function.) Then the machine M arrives at either answer, or fails to halt, with a well-defined probability. This is a non-deterministic Turing machine, but it can still be called classical computation, since it is based on classical probability.

One workable model of a RAM quantum computer is all of the above, except with two work tapes T_C and T_Q, and a register (a single ancillary cell) R_Q. In this model, each cell of T_Q has the Hilbert space $\mathbb{C}[\Gamma]$, and the cell R_Q does as well. The machine M can apply a joint unitary operator (or a TPCP) to the state of R_Q and the state of the cell of T_Q at the classical address in T_A. Or it can decide its next state in S by measuring the state in R_Q. Or it can do some classical computation using the classical tape T_C to decide what to do next. All of this can be arranged so that δ is a classical stochastic map (which might depend on quantum measurements), T_A and T_C are classical but randomized, and all of the quantum nondeterminism is only in the tape T_Q and the register R_Q. In some ways this model is more complicated than necessary, but it makes it easy to keep separate track of quantum and classical resources. T_C is a CRACM and T_Q is a CRAQM.

Proposition 1 is routine in this more precise model. The machine can create a quantum circuit drawn from a uniform family using T_A and T_C. Either afterwards or as it creates the circuit, it can implement it with unitary operations or quantum operations on T_Q and R_Q. Finally it can measure R_Q to decide or help decide whether to accept or reject the input. At linear time or above, it doesn't matter whether the input is first written onto T_C or T_Q.

The basic definition of quantum addressability is to assume that the address tape T_A is instead a quantum tape. For simplicity, we assume some abelian group structure on the alphabet Γ. Then adding the value of $T_C(T_A)$ to R_Q is a well-defined unitary operator on the joint Hilbert space of T_A and R_Q; in fact it is a permutation operator. This is our model of QRACM. Analogously, suppose that we choose a unitary operator U_{QR} that would act on the joint state of $T_Q(T_A)$ and R_Q if T_A were classical. Then it yields a unitary operator U_{QAR} on the joint state of T_Q, T_A, and R_Q that, in superposition, applies U_{QR} to $T_Q(T_A)$ and R_Q. This is a valid model of QRAQM.

To prove Proposition 2, we assume that T_C can no longer be addressed with T_A, and that instead the Turing machine has a position n on the tape T_C that can be incremented or decremented. Then to emulate a quantum read of $T_C(T_A)$, the machine can step through the tape T_C and add $T_C(n)$ to R_Q on the quantum condition that n matches T_A. This is easiest to do if the machine has an auxiliary classical tape that stores n itself. Even otherwise, the machine could space the data on T_C so that it only uses the even cells, and with logarithmic overhead drag the value of n itself on the odd cells.

3 Hide and seek

3.1 Hidden subgroups

This section is strictly a review of ideas discussed in our earlier article [7].

In the usual hidden subgroup problem, G is a group, X is an unstructured set, and $f : G \to X$ is a function that hides a subgroup H. This means that f factors through the coset space G/H (either left or right cosets), and the factor $f : G/H \to X$ is injective. In a quantum algorithm to find the subgroup H, f is implemented by a unitary oracle U_f that adds the output to an ancilla register. More precisely, the Hilbert space of the input register is the group algebra $\mathbb{C}[G]$ when G is finite (or some finite-dimensional approximation to it

when G is infinite), the output register is $\mathbb{C}[X]$, and the formula for U_f is

$$U_f|g, x_0\rangle = |g, f(g) + x_0\rangle.$$

All known subexponential algorithms for the hidden subgroup problems make no use of the output when the target set X is unstructured. (We do not know whether it is even possible to make good use of the output with only subexponentially many queries.) The best description of what happens is that the algorithm discards the output and leave the input register in a mixed state ρ. However, it is commonly said that the algorithm measures the output. This is a strange description if the algorithm then makes no use of the measurement; its sole virtue is that it leaves the quantum state of the input register in a pure state $|\psi\rangle$. The state $|\psi\rangle$ is randomly chosen from a distribution, which is the same as saying that the register is in a mixed state ρ.

If the output of f is always discarded, then the algorithm works just as well if the output of f is a state $|\psi(g)\rangle$ in a Hilbert space $\mathcal{H}$. The injectivity condition is replaced by the orthogonality condition $\langle\psi(g)|\psi(h)\rangle = 0$ when g and h lie in distinct cosets of H. In this case f would be implemented by a unitary

$$U_f|g, x_0\rangle = |g\rangle \otimes U_g|x_0\rangle,$$

with the condition that if $x_0 = 0$, then

$$U_g|0\rangle = |\psi(g)\rangle.$$

Or we can have the oracle, rather than the algorithm, discard the output. In this case, the oracle is a quantum operation (or quantum map) $\mathcal{E}_{G/H}$ that measures the name of the coset gH of H, and only returns the input conditioned on this measurement.

Suppose that the group G is finite. Then it is standard to supply the constant pure state $|G\rangle$ to the oracle U_f, and then discard the output. The resulting mixed state,

$$\rho_{G/H} = \mathcal{E}_{G/H}(|G\rangle\langle G|),$$

is the uniform mixture of $|gH\rangle$ over all (say) left cosets gH of H. This step can also be relegated to the oracle, so that we can say that the oracle simply broadcasts copies of $\rho_{G/H}$ with no input.

Like our old algorithm, our new algorithm mainly makes use of the state $\rho_{G/H}$, in the special case of the dihedral group $G = D_N$. When $N = 2^n$, it is convenient to work by induction on n, so that technically we use the state $\rho_{D_{2^k}/H_k}$ for $1 \le k \le n$. However, this is not essential. The algorithm can work in various ways with identical copies of $\rho_{D_N/H}$.

An important point is that the state $\rho_{G/H}$ is block diagonal with respect to the weak Fourier measurement on $\mathbb{C}[G]$. More precisely, the group algebra $\mathbb{C}[G]$ has a Burnside decomposition

$$\mathbb{C}[G] \cong \bigoplus_V V^* \otimes V,$$

where the direct sum is over irreducible representations of G and also the direct sum is orthogonal. The weak Fourier measurement is the measurement the name of V in this decomposition. Since $\rho_{G/H}$ is block diagonal, if we have an efficient algorithm for the quantum Fourier transform on $\mathbb{C}[G]$, then we might as well measure the name of V and condition the state $\rho_{G/H}$ to a state on $V^* \otimes V$, because the environment already knows[1] the

[1] In other words, Schrödinger's cat is out of the bag (or box).

name of V. Moreover, the state on the "row space" V^* is known to be independent of the state on V and carry no information about H [6]. So the algorithm is left with the name of V, and the conditional state $\rho_{V/H}$ on V. The difference in treatment between the value $f(g)$, and the name of the representation V, both of which are classical data that have been revealed to the environment, is that the name of V is materially useful to existing quantum algorithms in this situation. So it is better to say that the name of V is measured while the value $f(g)$ is discarded. (In fact, the two measurements don't commute, so in a sense, they discredit each other.)

3.2 Hidden shifts

In our earlier work [7], we pointed out that if A is an abelian group, then the hidden subgroup problem on the generalized dihedral group $G = (\mathbb{Z}/2) \ltimes A$ is equivalent to the abelian hidden shift problem. The hard case of a hidden subgroup on G consists of the identity and a hidden reflection. (By definition, a reflection is an element in $G \setminus A$, which is necessarily an element of order 2.) In this case, a single hiding function f on G is equivalent to two injective functions f and g on A that differ by a shift:

$$f(a) = g(a + s).$$

(Note that we allow an algorithm to evaluate them jointly in superposition.) Finding the hidden shift s is equivalent to finding the hidden reflection.

In this article, we will consider multiple hidden shifts. By this we mean that we have a set of endomorphisms

$$\phi_{j \in J} : A \to A$$

and a set of injective functions

$$f_{j \in J} : A \to X$$

such that

$$f_j(a) = f_0(a + \phi_j(s)).$$

Here J is an abstract finite indexing set with an element $0 \in J$. We assume that we know each ϕ_j explicitly (with $\phi_0 = 0$) and that we would like to find the hidden shift s. In the cyclic case $A = \mathbb{Z}/N$, we can write these relations as

$$f_j(a) = f_0(a + r_j s)$$

for some elements $r_j \in \mathbb{Z}/N$. Note that, for s to be unique, the maps ϕ_j or the factors r_j must satisfy a non-degeneracy condition. Since we will only address multiple hidden shifts in the initial input heuristically, we will not say too much about non-degeneracy when $|J| > 2$. If $|J| = 2$ then r_1 or ϕ_1 must be invertible to make s unique, in which case we might as well assume that they are the identity.

As a special case, we can look at the hidden subgroup problem in a semidirect product $G = K \ltimes A$, where K is a finite group, not necessarily abelian. Our original algorithm was a sieve that combined irreducible representations of such a group G to make improved irreducible representations. Anomalously, the sieve did not work better when $|K| > 2$ than in the dihedral case. The new algorithm can make some use of multiple hidden shifts, although the acceleration from this is not dramatic.

The principles of Section 3.1 apply to the hidden shift or multiple hidden shift problem. For the following, assume that A is a finite group. We write

$$f(j, a) = f_j(a),$$

and we can again make a unitary oracle U_f that evaluates f as follows:

$$U_f|j, a, x_0\rangle = |j, a, f(j, a) + x_0\rangle.$$

Suppose also that we can't make any sense of the value of $f(j, a)$, so we discard it. As in Section 3.1, the unitary oracle U_f is thus converted to a quantum map $\mathcal{E}$ that makes a hidden measurement of the value of f and returns only the input registers, *i.e.*, a state in $\mathbb{C}[J] \otimes \mathbb{C}[A]$. Suppose that we provide the map $\mathcal{E}$ with a state of the form

$$\rho = \sigma \otimes (|A\rangle\langle A|) \tag{1}$$

where σ is some possibly mixed state on $\mathbb{C}[J]$. As in Section 3.1, we claim that we might as well measure the Fourier mode $\widehat{b} \in \widehat{A}$ of the state $\mathcal{E}(\rho)$, because the environment already knows what it is. To review, the dual abelian group $\widehat{A}$ is by definition the set of group homomorphisms

$$\widehat{b} : A \to S^1 \subset \mathbb{C},$$

and the Fourier dual state $|\widehat{b}\rangle$ is defined as

$$|\widehat{b}\rangle = \frac{1}{\sqrt{|A|}} \sum_{a \in A} \overline{\widehat{b}(a)}|a\rangle.$$

We state the measurement claim more formally.

▶ Proposition 3. Let $\mathcal{E}$ be the partial trace of U_f given by discarding the output, and let the state ρ be as in (1). Then the state $\mathcal{E}(\rho)$ is block diagonal with respect to the eigenspaces of the measurement of $|\widehat{b}\rangle$. Also, the measurement has a uniformly random distribution.

Proof. The key point is that ρ is an A-invariant state and $\mathcal{E}$ is an A-invariant map, where A acts by translation on the $\mathbb{C}[A]$ register. The state $|A\rangle$ is A-invariant by construction, while A has no action on the $\mathbb{C}[J]$ register. Meanwhile $\mathcal{E}$ is A-invariant because it discards the output of f, and translation by A can be reproduced by permuting the values of f. Since ρ is an A-invariant state, and since the elements of A are unitary, this says exactly that ρ as an operator commutes with A. The eigenspaces of the action of A on $\mathbb{C}[J] \otimes \mathbb{C}[A]$ are all of the form $\mathbb{C}[J] \otimes |\widehat{b}\rangle$, so the fact that ρ commutes with A is equivalent to the conclusion that ρ is block diagonal with respect to the eigenspaces of the measurement $|\widehat{b}\rangle$.

To prove the second part, imagine that we also measure $|j\rangle$ on the register $\mathbb{C}[J]$. This measurement commutes with both measuring the Fourier mode $|\widehat{b}\rangle$ and measuring or discarding the output register $\mathbb{C}[X]$, so it changes nothing if we measure $|j\rangle$ first. So we know j, and since $f_j : A \to X$ is injective, measuring its value is the complete measurement of $|a\rangle$ starting with the constant pure state $|A\rangle$. This yields the uniform state ρ_{unif} on $\mathbb{C}[A]$, so the value of $|\widehat{b}\rangle$ is also uniformly distributed. ◀

Suppose further that in making the state ρ, the state σ on the $\mathbb{C}[J]$ register is the constant pure state $|J\rangle$. If the measured Fourier mode is $\widehat{b} \in \widehat{A}$, then the state of the j register after measuring this mode is:

$$|\psi\rangle \propto \sum_{j \in J} \widehat{b}(\phi_j(s))|j\rangle. \tag{2}$$

This can be written more explicitly in the cyclic case $A = \mathbb{Z}/N$. In this case there is an isomorphism $A \cong \widehat{A}$, and we can write any element $\widehat{b} \in \widehat{A}$ as

$$\widehat{b}(a) = \exp(2\pi i a b/N),$$

and we can also write

$$\phi_j(a) = r_j a$$

for some elements $r_j \in \mathbb{Z}/N$. So we can then write

$$|\psi\rangle \propto \sum_{j \in J} \exp(2\pi i b r_j s)|j\rangle. \tag{3}$$

At this point we know both b and each r_j, although for different reasons: r_j is prespecified by the question, while b was measured and is uniformly random. Nonetheless, we may combine these known values as $b_j = r_j b$ and write:

$$|\psi\rangle \propto \sum_{j \in J} \exp(2\pi i b_j s)|j\rangle. \tag{4}$$

To conclude, the standard approach of supplying the oracle U_f with the constant pure state and discarding the output leads us to the state (2), or equivalently (3) or (4). (Because measuring the Fourier mode does not sacrifice any quantum information.) In the rest of this article, we will assume a supply of states of this type.

4 The algorithm

4.1 The initial and final stages

For simplicity, we describe the hidden shift algorithm when $A = \mathbb{Z}/N$ and $N = 2^n$. The input to the algorithm is a supply of states (4). As explained in our previous work [7], the problem for any A, even A infinite as long as it is finitely generated, can be reduced to the cyclic case with overhead $\exp(O(\sqrt{d}))$. Also for simplicity, we will just find the parity of the hidden shift s. Also as explained in our previous work [7], if we know the parity of s, then we can reduce to a hidden shift problem on $\mathbb{Z}/2^{n-1}$ and work by induction. Finally, just as in our previous algorithm, we seek a wishful special case of (4), namely the qubit state

$$|\psi\rangle \propto |0\rangle + \exp(2\pi i (2^{n-1}) s/2^n)|1\rangle = |0\rangle + (-1)^s|1\rangle. \tag{5}$$

If we measure whether $|\psi\rangle$ is $|+\rangle$ or $|-\rangle$, that tells us the parity of s.

Actually, although we will give all of the details in base 2, we could just as well work in any fixed base, or let N be any product of small numbers. This generalization seems important for precise optimization for all values of N, which is an issue that we will only address briefly in the conclusion section.

4.2 Combining phase vectors

Like the old algorithm, the new algorithm combines unfavorable qubits states $|\psi\rangle$ to make more favorable ones in stages, but we change what happens in each stage. The old algorithm was called a sieve, because it created favorable qubits from a large supply of unfavorable qubits, just as many classical sieve algorithms create favorable objects from a large supply of candidates [1]. The new algorithm could also be called a sieve, but all selection is achieved with quantum measurement instead of a combination of measurement and matching. The process can be called collimation, by analogy with its meaning in optics: Making rays parallel.

Consider a state of the form (4), where we write the coefficient b_j instead as a function $b(j)$, except that we make no assumption that $b_j = r_j b$ for a constant b. We also assume that the index set is explicitly the integers from 0 to $\ell - 1$ for some ℓ, the *length* of $|\psi\rangle$:

$$J = [\ell] = \{0, 1, \ldots, \ell - 1\}.$$

We obtain:

$$|\psi\rangle \propto \sum_{0 \le j < \ell} \exp(2\pi i b(j) s / 2^n) |j\rangle.$$

Call a vector of this type a *phase vector*. We view a phase vector as favorable if every difference $b(j_1) - b(j_2)$ is divisible by many powers of 2, and we will produce new phase vectors from old ones that are more favorable. In other words, we will *collimate* the phases. The algorithm collimates phase vectors until finally it produces a state of the form (5). Note that the state $|\psi\rangle$ only changes by a global phase if we add a constant to the function b. (Or we can say that as a quantum state, it does not change at all.) If $2^m | b(j_1) - b(j_2)$ for some $m \le n$, then we can both subtract a constant from b and divide the numerator and denominator of $b(j)/2^n$ by 2^m. So we can $|\psi\rangle$ as

$$|\psi\rangle \propto \sum_{0 \le j < \ell} \exp(2\pi i b(j) s / 2^h) |j\rangle,$$

where $h = m - n$ is the *height* of $|\psi\rangle$. (We do not necessarily assign the smallest height h to a given $|\psi\rangle$.) We would like to collimate phase vectors to produce one with length 2 and height 1 (but not height 0).

Given two phase vectors of height h,

$$|\psi_1\rangle \propto \sum_{0 \le j_1 < \ell_1} \exp(2\pi i b_1(j_1) s / 2^h) |j_1\rangle$$
$$|\psi_2\rangle \propto \sum_{0 \le j_2 < \ell_2} \exp(2\pi i b_2(j_2) s / 2^h) |j_2\rangle,$$

their joint state is a double-indexed phase vector that also has height h:

$$\begin{aligned} |\psi_1, \psi_2\rangle &= |\psi_1\rangle \otimes |\psi_2\rangle \\ &\propto \sum_{\substack{0 \le j_1 < \ell_1 \\ 0 \le j_2 < \ell_2}} \exp(2\pi i (b_1(j_1) + b_2(j_2)) s / 2^h) |j_1, j_2\rangle. \end{aligned}$$

We can now collimate this phase vector by measuring

$$c \equiv b_1(j_1) + b_2(j_2) \pmod{2^m}$$

for some $m < h$. Let P_c be the corresponding measurement projection. The result is another phase vector

$$|\psi\rangle = P_c |\psi_1, \psi_2\rangle,$$

but one with a messy indexing set:

$$J = \{(j_1, j_2) | b_1(j_1) + b_2(j_2) \equiv c \pmod{2^m}\}.$$

We can compute the index set J, in fact entirely classically, because we know c. We can compute the phase multiplier function b as the sum of b_1 and b_2. Finally, we would like to reindex $|\psi\rangle$ using some bijection $\pi : J \to [\ell_{\text{new}}]$, where $\ell_{\text{new}} = |J|$. As we renumber J, we also permute the phase vector $P_c |\psi_1, \psi_2\rangle$. Then there is a subunitary operator

$$U_\pi : \mathbb{C}^{\ell_1} \otimes \mathbb{C}^{\ell_2} \to \mathbb{C}^{\ell_{\text{new}}}$$

that annihilates vectors orthogonal to $\mathbb{C}[J]$ and that is unitary on $\mathbb{C}[J]$. Then

$$|\psi_{\text{new}}\rangle = U_\pi |\psi\rangle.$$

The vector $|\psi_{\text{new}}\rangle$ has height $h - m$.

Actually, collimation generalizes to more than two input vectors. Given a list of phase vectors

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_r\rangle,$$

and given a collimation parameter m, we can produce a collimate state $|\psi_{\text{new}}\rangle$ from them. We summarize the process in algorithm form:

▶ **Algorithm 4** (Collimation)**.** Input: A list of phase vectors

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_r\rangle$$

of length $\ell_1, \ldots, \ell_r$, and a collimation parameter m.

1. Notionally form the phase vector

$$|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle \otimes \ldots \otimes |\psi_r\rangle$$

with indexing set

$$[\ell_1] \times [\ell_2] \times \cdots \times [\ell_r]$$

and phase multiplier function

$$b(\vec{j}) = b(j_1, j_2, \ldots, j_r) = b_1(j_1) + b_2(j_2) + \cdots + b_r(j_r).$$

2. Measure $|\psi\rangle$ according to the value of

$$c = b(\vec{j}) \bmod 2^m \tag{6}$$

to obtain $P_c|\psi\rangle$.

3. Find the set J of tuples $\vec{j}$ that satisfy (6). Set $\ell_{\text{new}} = |J|$ and pick a bijection

$$\pi : J \to [\ell_{\text{new}}].$$

4. Apply π to the value of b on J and apply U_π to $|\psi\rangle$ to make $|\psi_{\text{new}}\rangle$ and return it.

Algorithm 4 is our basic method to collimate phase vectors. We can heuristically estimate the length ℓ by assuming that $b(\vec{j})$ is uniformly distributed mod 2^m. In this case,

$$\ell_{\text{new}} \approx 2^{-m}\ell_1\ell_2 \ldots \ell_r. \tag{7}$$

So ℓ stays roughly constant when $\ell \approx 2^{m/(r-1)}$.

4.3 The complexity of collimation

▶ Proposition 5. Let $|\psi_1\rangle$ and $|\psi_2\rangle$ be two phase vectors of length ℓ_1 and ℓ_2 and height h, and suppose that they are collimated mod 2^m to produce a phase vector $|\psi_{\text{new}}\rangle$ of length ℓ_{new}. Suppose also that the quantum computer is allowed QRACM. Then taking $\ell_{\max} = \max(\ell_1, \ell_2, \ell_{\text{new}})$ and $r = 2$, Algorithm 4 needs

- $\widetilde{O}(\ell_{\max})$ classical time (where "$\widetilde{O}$" allows factors of both $\log \ell_{\max}$ and $h \leq n = \log N$).
- $O(\ell_{\max} h)$ classical space,
- $O(\ell_{\max} \max(m, \log \ell_{\max}))$ classical space with quantum access,
- $\text{poly}(\log \ell_{\max})$ quantum time, and
- $O(\log \ell_{\max})$ quantum space.

Proof. First, we more carefully explain the data structure of a phase vector $|\psi\rangle$. The vector $|\psi\rangle$ itself can be stored in $\lceil \log_2 \ell_{\max} \rceil$ qubits. The table b of phase multipliers is a table of length $O(\ell_{\max})$ whose entries have h bits, so this is $O(\ell_{\max} h)$ bits of classical space. Algorithm 4 needs the low m bits of each entry in the table, so $O(\ell_{\max} m)$ bits are kept in quantum access memory. We also assume that the table b is sorted on low bits.

We follow through the steps of Algorithm 4, taking care to manage resources at each step. First, measuring

$$c \equiv (b_1(j_1) + b_2(j_2)) \pmod{2^m}$$

can be done in quantum time $\mathrm{poly}(\log \ell, m)$ by looking up the values and adding them. As usual, when performing a partial quantum measurement, the output must be copied to an ancilla and the scratch work (in this case the specific values of b_1 and b_2) must be uncomputed.

The other step of collimation is the renumbering. To review, the measurement of c identifies a set of double indices

$$J \subseteq [\ell_1] \times [\ell_2].$$

These indices must be renumbered with a bijection

$$\pi : J \to [\ell_{\mathrm{new}}],$$

indeed the specific bijection that sorts the new phase multiplier table $b = b_1 + b_2$. The function π can be computed in classical time $\widetilde{O}(\ell)$ using standard algorithms, using the fact that b_1 and b_2 are already sorted. More explicitly, we make an outer loop over decompositions

$$c = c_1 + c_2 \in \mathbb{Z}/2^m.$$

In an inner loop, we write all solutions to the equations

$$b_1(j_1) \equiv c_1 \pmod{2^m} \qquad b_2(j_2) \equiv c_2 \pmod{2^m}$$

using sorted lookup. This creates a list of elements of J in some order. We can write the values of

$$b(j_1, j_2) = b_1(j_1) + b_2(j_3)$$

along with the pairs $(j_1, j_2) \in J$ themselves. Then b can be sorted and J can be sorted along with it.

This creates a stored form of the *inverse* bijection π^{-1}, which is an ordinary 1-dimensional array. We will want this, and we will also want quantum access to the *forward* bijection π stored as an associative array. Since we will need quantum access to π, we would like to limit the total use of this expensive type of space. We can make a special associative array to make sure that the total extra space is $O(\ell_{\max}(\log \ell_{\max}))$ bits. For instance, we can make a list of elements of J sorted by (j_1, j_2), a table of π sorted in the same order, and an index of pointers from $[\ell_1]$ to the first element of J with any given value of j_1.

The final and most delicate step is to apply the bijection π to $|\psi\rangle$ in quantum polynomial time in $\log \ell$. Imagine more abstractly that $|\psi\rangle$ is a state in a Hilbert space $\mathbb{C}^s$ supported on a subset $X \subseteq [s]$, and that we would like to transform it to a state in a Hilbert space $\mathbb{C}^t$ supported on a subset $Y \subset [t]$ of the same size, using a bijection $\pi : X \to Y$. We use the group structures $[s] = \mathbb{Z}/s$ and $[t] = \mathbb{Z}/t$, and we assume quantum access to both π and π^{-1}. Then we will use these two permutation operators acting jointly on a $\mathbb{C}^s$ register and a $\mathbb{C}^t$ register:

$$U_1|x, y\rangle = |x, y + \pi(x)\rangle \qquad U_2|x, y\rangle = |x - \pi^{-1}(y), y\rangle.$$

A priori, $\pi(x)$ is only defined for $x \in X$ and $\pi^{-1}(y)$ is only defined for $y \in Y$; we extend them by 0 (or extend them arbitrarily) to other values of x and y. Then clearly

$$U_2 U_1 |x, 0\rangle = |0, \pi(x)\rangle.$$

Thus

$$|\psi_{\text{new}}\rangle = U_2 U_1 |\phi, 0\rangle$$

is what we want. Following the rule of resetting the height to 0, we can also let

$$b_{\text{new}}(j) = b(j)/2^m.$$

◀

▶ **Corollary 6.** *Taking the hypotheses of Proposition 5, if the quantum computer has no quantum access memory, then Algorithm 4 can be executed with $r = 2$ with*

- $\widetilde{O}(\ell_{\max})$ *quantum time (and classical time),*
- $\widetilde{O}(\ell_{\max})$ *classical space, and*
- $O(\log \ell_{\max})$ *quantum space.*

Corollary 6 follows immediately from Proposition 5 and Proposition 2. The point is that, even though there is a performance penalty in the absence of quantum access memory, the same algorithm still seems competitive.

4.4 The outer algorithm

In this section we combine the ideas of Sections 3.2, 4.1, 4.2, and 4.3 to make a complete algorithm. We present the algorithm with several free parameters. We will heuristically analyze these parameters in Section 4.5. Then in Section 2.1 we will simply make convenient choices for the parameter to prove that the algorithm has quantum time and classical space complexity $\exp(O(\sqrt{n}))$.

The algorithm has a recursive subroutine to produce a phase vector of height 1. The subroutine uses a collimation parameter $0 < m(h) \leq n - h$ and a starting minimum length ℓ_0.

▶ **Algorithm 7** (Collimation sieve)**.** Input: A height h, a collimation parameter $m = m(h)$, a branching parameter $r = r(h)$, a starting minimum length ℓ_0, and access to the oracle U_f. Goal: To produce a phase vector of height h.

1. If $h = n$, extract phase vectors

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_s\rangle$$

of height n from the oracle as described in Section 3 until the length of

$$|\psi_{\text{new}}\rangle = |\psi_1, \psi_2, \ldots, \psi_s\rangle$$

is at least ℓ_0. Return $|\psi_{\text{new}}\rangle$.

2. Otherwise, recursively and *sequentially* obtain a sequence of phase vectors

$$|\psi_1\rangle, |\psi_2\rangle, \ldots, |\psi_r\rangle$$

of height $h + m$.

4. Collimate the vectors mod 2^m using Algorithm 4 to produce a phase vector $|\psi_{\text{new}}\rangle$ of height h. Return it.

When called with $h = 1$, Algorithm 7 produces a phase vector

$$|\psi\rangle \propto \sum_{0 \le j < \ell} (-1)^{b(j)s} |j\rangle.$$

Otherwise, we pick a maximal subset $X \subseteq [\ell]$ on which b is equally often 0 and 1. (Note that this takes almost no work, because the collimation step sorts b.) If X is empty, then we must run Algorithm 7 again. Otherwise, we measure whether $|\psi\rangle$ is in $\mathbb{C}[X]$. If the measurement fails, then again we must run Subroutine A again. Otherwise the measured form of $|\psi\rangle$ has a qubit factor of the form

$$|0\rangle + (-1)^s |1\rangle,$$

and this can be measured to obtain the parity of s.

Algorithm 7 recursively makes a tree of phase vectors that are more and more collimated, starting with phase vectors obtained from the hiding function $f(j, a)$ by the weak Fourier measurement. An essential idea, which is due to Regev and is used in his algorithm, is that with the collimation method, the tree can be explored depth-first and does not need to be stored in its entirety. Only one path to a leaf needs to be stored. No matter how the collimation parameter is set, the total quantum space used is $O(n^2)$, while the total classical space used is $O(n \max(\ell))$. (But the algorithm is faster with quantum access to the classical space.)

An interesting feature of the algorithm is that its middle part, the collimation sieve, is entirely *pseudoclassical*. The algorithm begins by applying QFTs to oracle calls, as in Shor's algorithm. It ends with the same parity measurement as Simon's algorithm. These parts of the algorithm are fully quantum in the sense that they use unitary operators that are not permutation matrices. However, collimation consists entirely of permutations of the computational basis and measurements in the computational basis.

4.5 Heuristic analysis

Heuristically the algorithm is the fastest when $r = 2$.

Suppose that the typical running time of the algorithm is $f(n)$, with some initial choice of $m = m(1)$. First, creating a phase vector of height h is similar to running the whole algorithm with $n' = n - h$. So the total computation time (both classical and quantum) can be estimated as

$$f(n) \approx \min_m \left(2^m + 2f(n-m)\right).$$

Here the first term is dominated by the classical work of collimation, while the second term is the recursive work. The two terms of the minimand are very disparate outside of a narrow range of values of m. So we can let $g(n) = \log_2 f(n)$, and convert multiplication to addition and approximate addition by max. (This type of asymptotic approximation is lately known in mathematics as *tropicalization*.) We thus obtain

$$g(n) \approx \min_m \left(\max(m, g(n-m) + 1\right).$$

The solutions to this equation are of the form

$$g(\frac{m(m+1)}{2} + c) = m,$$

where c is a constant. We obtain the heuristic estimate

$$f(n) \stackrel{?}{=} \widetilde{O}(2^{\sqrt{2n}}) \tag{8}$$

for both the quantum plus classical time complexity and the classical space complexity of the algorithm. We put a question mark because we have not proven this estimate. In particular, our heuristic calculation does not address random fluctuations in the length estimate (7).

If the quantum computer does not have QRACM or if it is no cheaper than quantum memory, then the heuristic (8) is the best that we know how to do. If the algorithm is implemented with QRACM, then the purely quantum cost is proportional to the number of queries. In this case, if there is extra classical space, we can make m larger and larger to fill the available space and save quantum time. This is the "second parameter" mentioned in Section 1. However, this adjustment only makes sense when classical time is much cheaper than quantum time. In particular, (8) is our best heuristic if classical and quantum time are simply counted equally.

If classical space is limited, then equation (7) tells us that we can compensate by increasing r. To save as much space as possible, we can maintain $\ell = 2$ and adjust in each stage of the sieve r to optimize the algorithm. In this case the algorithm reduces to Regev's algorithm.

5 Conclusions

At first glance, the running time of our new algorithm for DHSP or hidden shift is "the same" as our first algorithm, since both algorithms run in time $2^{O(\sqrt{\log N})}$. Meanwhile Regev's algorithm runs in time $2^{O(\sqrt{(\log N)(\log\log N)})}$, which may appear to be almost as fast. Of course, these expressions hide the real differences in performance between these algorithms, simply because asymptotic notation has been placed in the exponent. All polynomial-time algorithms with input of length n run in time

$$n^{O(1)} = 2^{O(\log n)}.$$

Nonetheless, polynomial accelerations are taken seriously in complexity theory, whether they are classical or quantum accelerations.

For many settings of the parameters, Algorithm 7 is superpolynomially faster than Regev's algorithm. It is Regev's algorithm if we have exponentially more quantum time than classical space. However, in real life, classical computation time has only scaled polynomially faster than available classical computer memory. So it is reasonable to consider a future regime in which quantum computers exist, but classical memory is cheaper than quantum time, or is only polynomially more expensive.

Regev [11] established a reduction from certain lattice problems (promise versions of the short vector and close vector problems) to the version of DHSP or hidden shift in which $f(a)$ and $g(a+s)$ are overlapping quantum states. At first glance, our algorithms apply to this type of question. However, we have not found quantum accelerations for these instances. The fundamental reason is that we have trouble competing with classical sieve algorithms for these lattice problems [1]. The classical sieve algorithms work in position space, while our algorithms work in Fourier space, but otherwise the algorithms are similar. Instead, DHSP seems potentially even more difficult than related lattice problems (since that is the direction of Regev's reduction) and the main function of our algorithms is to make DHSP roughly comparable to lattice problems on a quantum computer.

One significant aspect of Algorithm 7, and also in a way Regev's algorithm, is that it solves the hidden subgroup problem for a group $G = D_N$ without staying within the representation theory of G in any meaningful way. It could be interesting to further explore non-representation methods for other hidden structure problems.

Acknowledgments. The author would like to thank Scott Aaronson and Oded Regev for useful discussions.

References

1 Miklós Ajtai, Ravi Kumar, and Dandapani Sivakumar, *A sieve algorithm for the shortest lattice vector problem*, Proceedings of the thirty-third annual ACM symposium on Theory of computing, 2001, pp. 601–610.

2 Andris Ambainis, *Quantum walk algorithm for element distinctness*, SIAM J. Comput. **37** (2007), no. 1, 210–239, `arXiv:quant-ph/0311001`.

3 Gilles Brassard, Peter Høyer, and Alain Tapp, *Quantum algorithm for the collision problem*, `arXiv:quant-ph/9705002`.

4 Andrew M. Childs, David Jao, and Vladimir Soukharev, *Constructing elliptic curve isogenies in quantum subexponential time*, `arXiv:1012.4019`.

5 Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone, *Architectures for a quantum random access memory*, Phys. Rev. A **78** (2008), no. 5, 052310, `arXiv:0807.4994`.

6 Michelangelo Grigni, Leonard J. Schulman, Monica Vazirani, and Umesh V. Vazirani, *Quantum mechanical algorithms for the nonabelian hidden subgroup problem*, ACM Symposium on Theory of Computing, 2001, pp. 68–74.

7 Greg Kuperberg, *A subexponential-time quantum algorithm for the dihedral hidden subgroup problem*, SIAM J. Comput. **35** (2005), no. 1, 170–188, `arXiv:quant-ph/0302112`.

8 Michael C. Loui and David R. Luginbuhl, *Optimal on-line simulations of tree machines by random access machines*, SIAM J. Comput. **21** (1992), no. 5, 959–971.

9 W. Paul and R. Reischuk, *On time versus space. II*, J. Comput. System Sci. **22** (1981), no. 3, 312–327.

10 Oded Regev, *A subexponential time algorithm for the dihedral hidden subgroup problem with polynomial space*, `arXiv:quant-ph/0406151`.

11 Oded Regev, *Quantum computation and lattice problems*, SIAM J. Comput. **33** (2004), no. 3, 738–760, `arXiv:cs.DS/0304005`.

Universal Entanglers for Bosonic and Fermionic Systems*

Joel Klassen[1,2], Jianxin Chen[3,2,4], and Bei Zeng[3,2]

1 **Department of Physics, University of Guelph**
50 Stone Road East, Guelph, Ontario, Canada
joeldavidklassen@gmail.com

2 **Institute for Quantum Computing**
200 University Avenue West, Waterloo, Ontario, Canada

3 **Department of Mathematics & Statistics, University of Guelph**
50 Stone Road East, Guelph, Ontario, Canada
{jianxinc,zengb}@uoguelph.ca

4 **UTS-AMSS Joint Research Laboratory for Quantum Computation and Quantum Information Processing**
Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing, China

Abstract

A universal entangler (UE) is a unitary operation which maps all pure product states to entangled states. It is known that for a bipartite system of particles $1, 2$ with a Hilbert space $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$, a UE exists when $\min(d_1, d_2) \geq 3$ and $(d_1, d_2) \neq (3, 3)$. It is also known that whenever a UE exists, almost all unitaries are UEs; however to verify whether a given unitary is a UE is very difficult since solving a quadratic system of equations is NP-hard in general. This work examines the existence and construction of UEs of bipartite bosonic/fermionic systems whose wave functions sit in the symmetric/antisymmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$. The development of a theory of UEs for these types of systems needs considerably different approaches from that used for UEs of distinguishable systems. This is because the general entanglement of identical particle systems cannot be discussed in the usual way due to the effect of (anti)-symmetrization which introduces "pseudo entanglement" that is inaccessible in practice. We show that, unlike the distinguishable particle case, UEs exist for bosonic/fermionic systems with Hilbert spaces which are symmetric (resp. antisymmetric) subspaces of $\mathbb{C}^d \otimes \mathbb{C}^d$ if and only if $d \geq 3$ (resp. $d \geq 8$). To prove this we employ algebraic geometry to reason about the different algebraic structures of the bosonic/fermionic systems. Additionally, due to the relatively simple coherent state form of unentangled bosonic states, we are able to give the explicit constructions of two bosonic UEs. Our investigation provides insight into the entanglement properties of systems of indistinguishable particles, and in particular underscores the difference between the entanglement structures of bosonic, fermionic and distinguishable particle systems.

1998 ACM Subject Classification J.2 Physics

Keywords and phrases Universal Entangler, Bosonic States, Fermionic States

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.35

* This work was partially supported by NSERC and CIFAR

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 35–49

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Entanglement sits at the core of the counterintuitive and useful properties of quantum mechanics. At its inception Schrödinger labeled entanglement "the characteristic trait of quantum mechanics, the one that enforces its entire departure from classical lines of thought." [19] This observation remains true today, and with the advent of quantum computing, its practical consequences have never before been more real. However after decades of effort, entanglement remains poorly understood [13, 1, 11]. A promising avenue for furthering our understanding of entanglement is cataloguing and analyzing the various means of generating it. There is a sense that those mechanisms which generate maximal amounts of entanglement, or most consistently generate entanglement, are especially enlightening because they serve as bounds on what can and can not be done, thus restricting our domain of inquiry.

One outcome of this line of thought is the concept of a universal entangler (UE). A UE is a unitary operator which maps any non-entangled state to an entangled state [3]. A UE can act as a useful tool, both theoretically and experimentally, due to its generality. This generality is derived from the fact that a UE admits any non-entangled quantum states. However this generality also makes demonstrating the properties of UEs very difficult. For instance, while it has been shown that UEs do exist for a system with Hilbert space $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ when $\min(d_1, d_2) \geq 3$ and $(d_1, d_2) \neq (3, 3)$, proving this fact has been nontrivial, requiring techniques from algebraic geometry [3]. To date no elementary method is known which can achieve the same results. Additionally, although it has been shown that whenever UEs exist almost all unitaries are UEs [4], explicit constructions of UEs remain ellusive. This is due to the fact that the problem of verifying whether a given unitary is a UE is in general intractable since the verification is equivalent to solving a quadratic system of equations which is hard in general [6]. So far the only explicitly known UE is an example for the $(d_1, d_2) = (3, 4)$, from an order 12 Hadamard matrix [4]. In general more advanced methods may be needed in order to construct UEs, as well as to verify their universality.

The theory of entanglement of systems of indistinguishable particles has garnered much attention during the past decade [17, 16, 12, 14, 5, 1]. The entanglement of systems of indistinguishable particles cannot necessarily be approached in the same way as the distinguishable particle case because the symmetry requirement of the wave functions (i.e. symmetrization for bosonic system and antisymmetrization for fermionic system) may introduce 'pseudo entanglement' which is not accessible in practice [5, 16, 17, 12, 14]. It is now widely agreed that non-entangled states correspond to the coherent states $|v\rangle^{\otimes N}$ [15] for indistinguishable bosonic systems and to Slater determinants for indistinguishable fermionic systems [5, 1]. A natural line of inquiry is to identify the existence and construction of UEs for systems of indistinguishable particles. Indistinguishable bipartite bosonic/fermionic states are symmetric/antisymmetric states of the Hilbert space $\mathbb{C}^d \otimes \mathbb{C}^d$. This does not necessarily mean that the theory of UEs of distinguishable particles is readily generalizable to UEs for indistinguishable particles. Some obvious reasons for this are: 1. although almost all unitaries are UEs when $d > 3$, the lack of understanding of explicit constructions prevents us from directly verifying whether there exist any UEs which are symmetric under particle permutation; 2. the definition of a non-entangled state for systems of indistinguishable fermions is dramatically different from that of systems of distinguishable particles (in fact, a single Slater determinant, when viewed as an antisymmetric distinguishable particle state, is indeed entangled).

This paper discusses the existence and construction of UEs for both indistinguishable bipartite bosonic (BUE) and fermionic (FUE) systems. Employing techniques in algebraic

geometry, considering the different algebraic structures of the bosonic and fermionic systems, we show that, in contrast to the distinguishable particle case, BUEs exist for bosonic systems if and only if the single particle Hilbert space has dimension $d \geq 3$, and FUEs exist for fermionic systems if and only if the single particle Hilbert space has dimension $d \geq 8$. We also show, similarly to the distinguishable particle case, that for dimensions where BUEs/FUEs exist, almost all unitaries are BUEs/FUEs. Finally, because the unentangled states of indistinguishable bosonic systems are of a relatively simple coherent state form $|v\rangle \otimes |v\rangle$, which implies a hidden linear structure for the product states (i.e. the set of all single particle states $|v\rangle$ form a vector space), the construction of BUEs becomes significantly simpler. We have found a simple explicit construction of a BUE based on permutation matrices which holds for all $d \geq 3$, and another one based on Householder-type gates [10] which holds for all $d \geq 5$. Unfortunately the explicit construction and verification of FUEs, like distinguishable particle UEs, remains a significantly more intractable problem.

We believe that our investigation provides insight into the entanglement properties of identical particle systems, and in particular the different entanglement structures between bosonic, fermionic and distinguishable particle systems.

We organize our paper as follows. In section 2 we review some previously established results about UEs and provide some preliminaries about bosonic and fermionic systems to help establish our main results. In section 3 we give a proof for the existence and prevalence of BUEs, and give two explicit examples of their construction. In Section 4 we give a proof for the existence and prevalence of FUEs. Finally, in section 5, we provide a brief summary of our results and a discussion of future directions.

2 Preliminaries

This section provides preliminaries to help establish our main results for BUEs and FUEs. We first briefly review UEs for distinguishable particle systems established in [3]. We then further briefly review basic entanglement theory for bosonic and fermionic systems.

2.1 Universal entanglers

For the case of distinguishable particles, it is known that any given quantum system is identified with some finite (or infinite) Hilbert space $\mathcal{H}$. Moreover, two unit vectors are indistinguishable if they differ only by a global phase factor. Hence, distinct pure states can be put in correspondence with "rays" in $\mathcal{H}$, or equivalently, points in the projective Hilbert space $\mathbb{P}(\mathcal{H})$.

We consider pure states for bipartite systems, whose Hilbert space is $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$. A bipartite quantum state is a product state if $|\psi\rangle = |\psi_1\rangle \otimes |\psi_2\rangle$ for some $|\psi_1\rangle \in \mathbb{C}^{d_1}$ and $|\psi_2\rangle \in \mathbb{C}^{d_2}$. Otherwise, it is an entangled state. It is straightforward to see that the set of all the product states do not form a linear vector space, so one does not expect that the UE problem can be examined using basic tools from linear algebra.

Instead, it is observed that the set of normalized product states in a composite system associated with $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ is isomorphic to a projective variety in $\mathbb{P}^{d_1 d_2 - 1}$, a well studied object in algebraic geometry. Before continuing, we need some basic notations and necessary background materials from algebraic geometry [8].

For any positive integer n, the set of all n-tuples from $\mathbb{C}$ is called an n-dimensional *affine space* over $\mathbb{C}$. An element of $\mathbb{C}^n$ is called a point, and if point $P = (a_1, a_2, \cdots, a_n)$ with $a_i \in \mathbb{C}$, then the a_i's are called the coordinates of P. Informally, an affine space is what is left of a vector space after forgetting its origin.

We define *projective n-space*, denoted by $\mathbb{P}^n$, to be the set of equivalence classes of $(n+1)-$tuples $(a_0, \cdots, a_n)$ from $\mathbb{C}$, not all zero, under the equivalence relation given by $(a_0, \cdots, a_n) \sim (\lambda a_0, \cdots, \lambda a_n)$ for all $\lambda \in \mathbb{C}$, $\lambda \neq 0$. We use $[a_0 : \cdots : a_n]$ to denote the projective coordinates of this point.

The *polynomial ring* in n variables, denoted by $\mathbb{C}[x_1, x_2, \cdots, x_n]$, is the set of polynomials in n variables with coefficients in field $\mathbb{C}$.

A subset Y of $\mathbb{C}^n$ is an *algebraic set* if it is the common zeros of a finite set of polynomials $f_1, f_2, \cdots, f_r$ with $f_i \in \mathbb{C}[x_1, x_2, \cdots, x_n]$ for $1 \leq i \leq r$, which is also denoted by $Z(f_1, f_2, \cdots, f_r)$.

One may observe that the union of a finite number of algebraic sets is an algebraic set, and the intersection of any family of algebraic sets is again an algebraic set. Therefore, by taking the open subsets to be the complements of algebraic sets, we can define a topology, called the *Zariski topology* on $\mathbb{C}^n$.

A nonempty subset Y of a topological space X is called *irreducible* if it cannot be expressed as the union of two proper closed subsets. The empty set is not considered to be irreducible.

An *affine algebraic variety* is an irreducible closed subset of $\mathbb{C}^n$, with respect to the induced topology.

A notion of algebraic variety may also be introduced in projective spaces, called projective algebraic variety: a subset Y of $\mathbb{P}^n$ is an *algebraic set* if it is the common zeros of a finite set of homogeneous polynomials $f_1, f_2, \cdots, f_r$ with $f_i \in \mathbb{C}[x_0, x_1, \cdots, x_n]$ for $1 \leq i \leq r$. We call open subsets of irreducible projective varieties quasi-projective varieties.

Observe that a product state in $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ can be written as the Kronecker product of a vector $v_1 \in \mathbb{C}^{d_1}$ and another vector $v_2 \in \mathbb{C}^{d_2}$. Let's further write these vectors in the computational basis, say $v_1 = \begin{pmatrix} x_1, & x_2, & \cdots & , x_{d_1} \end{pmatrix}$ and $v_2 = \begin{pmatrix} y_1, & y_2, & \cdots & , y_{d_2} \end{pmatrix}$. Their product state is a $d_1 d_2$-dimensional vector

$$\begin{pmatrix} z_1, & z_2, & \cdots & z_{d_2}, & z_{d_2+1}, & \cdots & , z_{d_1 d_2} \end{pmatrix}$$
$$= \begin{pmatrix} x_1 y_1, & x_1 y_2, & \cdots & x_1 y_{d_2}, & x_2 y_1, & \cdots & , x_{d_1} y_{d_2} \end{pmatrix}$$

Hence $z_{(i-1)d_2+j} = x_i y_j$ for any $1 \leq i \leq d_1, 1 \leq j \leq d_2$. It follows that

$$z_{(i_1-1)d_2+j_1} z_{(i_2-1)d_2+j_2} = z_{(i_1-1)d_2+j_2} z_{(i_2-1)d_2+j_1}$$

for any $1 \leq i_1, i_2 \leq d_1, 1 \leq j_1, j_2 \leq d_2$. On the other hand, any $d_1 d_2$-dimensional vector $(z_k)_{k=1}^{d_1 d_2}$ satisfying the above polynomials can be written as the tensor product of $v_1 \in \mathbb{C}^{d_1}$ and $v_2 \in \mathbb{C}^{d_2}$ [8]. This implies that the set of normalized product states in $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ is isomorphic to a projective variety in $\mathbb{P}^{d_1 d_2 - 1}$ which is called a "Segre variety" and denoted as Σ_{d_1, d_2}. This simple observation provides an algebraic geometric description of product states and entangled states.

Therefore, a unitary operator U acting on $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ is a UE if and only

$$U(\Sigma_{d_1, d_2}) \bigcap \Sigma_{d_1, d_2} = \emptyset.$$

From the geometric point of view, a UE will rotate the set of product states to another set which is completely void of product states.

In [3], it is proved that UEs exist if and only if $\min\{d_1, d_2\} \geq 3$ and $(d_1, d_2) \neq (3, 3)$. Surprisingly, it is further illustrated that a random unitary operator acting on such a bipartite system will even rotate the set of product states to another set which contains nothing but nearly maximally entangled states [4].

Although it has been shown that a random unitary gate will almost surely be a UE of a bipartite quantum system $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$ if $\min\{d_1, d_2\} \geq 3$ and $(d_1, d_2) \neq (3, 3)$, constructing an explicit UE for any bipartite quantum system is not that easy. One simple strategy is to randomly pick a unitary gate acting on $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$, and then verify whether it is a UE by solving a family of polynomial equations. Unfortunately, there is no known efficient way to solve quadratic polynomial systems [6]. So far, explicit UEs are only known for $(d_1, d_2) = (3, 4)$ [4].

2.2 Bosonic systems

It is known that bosonic states lie in the 2nd symmetric tensor power of $\mathbb{C}^d$, denoted by $\vee^2\mathbb{C}^d$. A state in $\vee^2\mathbb{C}^d$ is a product state if it can be written as some $|\alpha\rangle \otimes |\alpha\rangle$, i.e. it is a coherent state [14, 5]. Any state which cannot be written as such a symmetric product form does demonstrate correlation which can be potentially used in quantum information processing [14], and hence is considered entangled.

Any bipartite bosonic pure state is local unitarily equivalent to $\sum_\alpha \lambda_\alpha |\alpha\rangle \otimes |\alpha\rangle$ [12, 14]. This then indicates a hidden linear structure for bipartite bosonic pure states because the single particles states $|\alpha\rangle$ form a vector space.

From the algebraic geometric point of view, any bosonic product state $|\alpha\rangle \otimes |\alpha\rangle$ can be written as a vector with projective coordinates

$$[a_1a_1 : a_1a_2 : \cdots : a_1a_d : a_2a_1 : a_2a_2 : \cdots : a_2a_d : a_3a_1 : \cdots : a_da_d]$$

where $[a_1 : \cdots : a_d]$ are the projective coordinates of $|\alpha\rangle$.

Such points can be characterized by a family of polynomials again. In fact, the set of projective points with coordinates

$$[a_1a_1 : a_1a_2 : \cdots : a_1a_d : a_2a_1 : a_2a_2 : \cdots : a_2a_d : a_3a_1 : \cdots : a_da_d]$$

is obviously isomorphic to the set of the following points

$$[a_1^2 : a_2^2 : \cdots : a_d^2 : a_1a_2 : a_1a_3 : \cdots : a_1a_d : a_2a_3 : \cdots : a_{d-1}a_d]$$

which is known as the Veronese variety in algebraic geometry [7].

Hence the set of bosonic product states corresponds to a special case of Veronese variety whose dimension is $d - 1$. This fact will be used in our further investigation.

2.3 Fermionic systems

Consider the pure states of a bipartite fermionic system whose Hilbert space is the antisymmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$. The Pauli exclusion principle requires that $d \geq 2$. We denote the 2nd exterior power of $\mathbb{C}^d$, i.e. the antisymmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$ by $\wedge^2\mathbb{C}^d$. For any $|\alpha\rangle, |\beta\rangle \in \mathbb{C}^d$, we use the notation

$$|\alpha\rangle \wedge |\beta\rangle = \frac{1}{\sqrt{2}}(|\alpha\rangle \otimes |\beta\rangle - |\beta\rangle \otimes |\alpha\rangle), \tag{1}$$

to denote a single Slater determinant.

A quantum state $|\psi\rangle$ in $\wedge^2\mathbb{C}^d$ is said to be decomposable if it can be written as an exterior product of individual vectors from $\mathbb{C}^d$, i.e. there exists $|\alpha\rangle, |\beta\rangle \in \mathbb{C}^d$ such that $|\psi\rangle = |\alpha\rangle \wedge |\beta\rangle$. Decomposable states are considered unentangled, as any correlation results purely from the fermionic statistics, and so is not useful for quantum information processing [17, 16]. Any

state which cannot be written in such a decomposable form does demonstrate correlation which can be potentially used in quantum information processing [17, 16] , and hence is considered to be entangled.

Any bipartite fermionic pure state is local unitarily equivalent to $\sum_\alpha \lambda_i|\alpha_i\rangle \wedge |\beta_i\rangle$ [17, 16], where $|\alpha_i\rangle, |\beta_i\rangle \in \mathbb{C}^d$, $\langle\alpha_i|\beta_j\rangle = 0$, $\langle\alpha_i|\alpha_j\rangle = \delta_{ij}$, and $\langle\beta_i|\beta_j\rangle = \delta_{ij}$. This is an analogue of the Schmidt decomposition of a distinguishable particle system and hence is called the Slater decomposition. Similarly to the distinguishable particle case, the set of all decomposable states do not form a linear vector space, so one does not expect that the FUE problem can be examined using basic tools from linear algebra.

Again, let's look over the decomposable (or fermionic product) states from the algebraic geometric point of view. As we showed before, a decomposable state can be written as $|\psi\rangle = |\alpha\rangle \wedge |\beta\rangle$ where $|\alpha\rangle$ and $|\beta\rangle$ are two vectors in $\mathbb{C}^d$. Let S_ψ be the 2-dimensional subspace spanned by $|\alpha\rangle$ and $|\beta\rangle$. A different basis for S_ψ will give a different exterior product, but the two exterior products will differ only by a nonzero scale. Ignoring the nonzero scale, any decomposable state corresponds to a 2-dimensional subspace in $\mathbb{C}^d$ and vice versa. Hence the set of decomposable states is isomorphic to the set of 2-dimensional subspaces which is known as a Grassmannian $G(2, d)$ [7]. It is not that obvious that $G(2, d)$ can be characterized by a set of polynomials, but it can be. The correspondence we have just shown is known as the Plücker embedding of a Grassmannian into a projective space:

$$\tau : G(2, d) \to \mathbb{P}(\wedge^2\mathbb{C}^d).$$

This embedding satisfies certain simple quadratic polynomials and is called the Grassmann-Plücker relations (see e.g. p. A III.172 Eq. (84-(J,H)) in [2], Prop 11-32 in [9], and [5]). This implies the Grassmannian embeds as an algebraic variety of $\mathbb{P}(\wedge^2\mathbb{C}^d)$.

3 Bosonic Universal Entanglers

3.1 Existence and Prevalence

Recall that a bosonic state in $\vee^2\mathbb{C}^d$ is a product state if it can be written as $|\alpha\rangle \otimes |\alpha\rangle$ for some $|\alpha\rangle$. A quantum gate acting on $\vee^2\mathbb{C}^d$ is said to be a bosonic universal entangler (BUE) if it will map every product state to some entangled state.

Note that the set of product states of a bosonic system can also be characterized by a set of polynomials. Indeed, let $\Lambda = \{|\alpha\rangle \otimes |\alpha\rangle : |\alpha\rangle \in \mathbb{C}^d\}$, this is a precisely the Veronese variety [7]. Furthermore, Λ is isomorphic to $\mathbb{C}^d$. For any $|\psi\rangle \in \vee^2\mathbb{C}^2$, let us denote $\text{rank}|\psi\rangle \equiv \min\{r : |\psi\rangle = \sum_{i=1}^{r} |a_i\rangle|a_i\rangle\}$.

▶ **Theorem 1.** *There is a BUE acting on $\vee^2\mathbb{C}^d$ if and only if $d \geq 3$. Furthermore, when $d \geq 3$, almost every quantum gate acting on $\vee^2\mathbb{C}^d$ is a BUE.*

Proof. For $d \leq 2$, we have

$$\dim U(\Lambda) + \dim \Lambda = 2 \dim \Lambda = 2(d-1) \geq \binom{d+1}{2} - 1 = \dim \mathbb{P}(\vee^2\mathbb{C}^d). \tag{2}$$

This implies there is no BUE for $\vee^2\mathcal{C}^d$. This assertion follows from the dimension counting theorem which states that the intersection of any two projective varieties $\mathcal{A}$ and $\mathcal{B} \subseteq \mathbb{P}^m$ is nonempty if $\dim \mathcal{A} + \dim \mathcal{B} \geq m$. More specifically, we have $U(\Lambda) \bigcap \Lambda \neq \emptyset$.

On the other hand, consider the set of quantum gates acting on a system of two indistinguishable bosons. Any quantum gate acting on this system should be a symmetric gate, i.e., $SUS = U$, where S is the swap operator. Equivalently, U is a quantum gate acting on $\vee^2\mathbb{C}^d$.

Let $\mathcal{X} = \{\Phi | \Phi \in \mathcal{U}(\vee^2\mathbb{C}^d), \Phi(\Lambda) \cap \Lambda \neq \emptyset\}$. Our aim is to show that $\mathcal{X}$ is a proper subset of $\mathcal{U}(\vee^2\mathbb{C}^d)$. If this is so, then the existence of BUEs will be automatically guaranteed.

Let's consider the Zariski topology on the projective space. In this setting, the unitary group $\mathcal{U}(\vee^2\mathbb{C}^d)$ is Zariski dense in the general linear group $GL(\vee^2\mathbb{C}^d)$ [18]. We further define $\mathcal{X}' = \{\Phi | \Phi \in GL(\vee^2\mathbb{C}^d), \Phi(\Lambda) \cap \Lambda \neq \emptyset\}$. It is easy to see that $\mathcal{X} \subseteq \mathcal{X}'$.

The dimension of its Zariski closure $\dim \overline{\mathcal{X}'}$ is bounded by $\binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + 2(d-1)$. See Lemma 3 in Appendix C for details.

Now we prove the existence of a BUE as follows. If U is not a BUE, $\mathcal{U}(\vee^2\mathbb{C}^d) \subset \mathcal{X}'$, then $GL(\vee^2\mathbb{C}^d) = \overline{\mathcal{U}(\vee^2\mathbb{C}^d)} \subset \overline{\mathcal{X}'}$. However, $\dim(\overline{\mathcal{X}'}) \leq \binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + 2(d-1) < \binom{d+1}{2}^2 = \dim GL(\vee^2\mathbb{C}^d)$. This is a contradiction. So $\mathcal{U}(\vee^2\mathbb{C}^d) \not\subset \mathcal{X}'$, i.e. a unitary operator $\Phi \in \mathcal{U}(\vee^2\mathbb{C}^d)$ with universal entangling power exists.

We will now show that $\mathcal{X}$ is not only a proper subset, but also a negligible subset of $\mathcal{U}(\vee^2\mathbb{C}^d)$.

$\mathcal{U}(\vee^2\mathbb{C}^d)$ is a locally compact Lie group of dimension $\binom{d+1}{2}^2$. Recall that $\dim(\overline{\mathcal{X}'})$ is at most $\binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + 2(d-1) < \binom{d+1}{2}^2 = \dim(\mathcal{U}(\vee^2\mathbb{C}^d))$.

We have shown $\dim(\overline{\mathcal{X}'}) < \binom{d+1}{2}^2 = \dim(\mathcal{U}(\vee^2\mathbb{C}^d))$. $\overline{\mathcal{X}'}$ is Noetherian (i.e. any descending sequence of its closed subvarieties is stationary), then $\overline{\mathcal{X}'}$ is a union of finitely many smooth subvarieties of $GL(\vee^2\mathbb{C}^d)$ with lower dimensions. Hence $\overline{\mathcal{X}'} \cap \mathcal{U}(\vee^2\mathbb{C}^d)$ (which contains $\mathcal{X}' \cap \mathcal{U}(\vee^2\mathbb{C}^d)$, the set of our main interest) is a union of finite many submanifolds of $\mathcal{U}(\vee^2\mathbb{C}^d)$ with lower dimensions. Therefore, $\mathcal{X}' \cap \mathcal{U}(\vee^2\mathbb{C}^d)$ is measure zero in $\mathcal{U}(\vee^2\mathbb{C}^d)$ which implies that a random unitary operator U is almost surely a BUE.

◀

3.2 Explicit Construction

As we have shown in Theorem 1, a random unitary acting on $\vee^2\mathbb{C}^d$ will almost surely be a BUE. Hence we can pick an arbitrary unitary acting on $\vee^2\mathbb{C}^d$ and verify whether it will map some product state to another product state. Recall that the set of product states in a bosonic system is isomorphic to $\mathbb{C}^d$. This will make it easier to verify whether a unitary is a BUE. Here we provide verifications of two different classes of BUEs.

3.2.1 Householder-type Bosonic Universal Entanglers

For $d \geq 5$ and any subspace $S \subset \vee^2\mathbb{C}^d$, let's consider the following gate $U = \mathbb{I}_{\vee^2\mathbb{C}^d} - 2P_S$ where P_S is a projection to some subspace S. These gates are known as Householder matrices in linear algebra [10] and they are widely used to perform QR decomposition.

A gate U constructed in this way will be a BUE if the subspace S is chosen properly to satisfy the following two constraints:

1. There is no product state in $S^\perp$.
2. $\text{rank}|\psi\rangle \geq 3$ for any $|\psi\rangle \in S$.

This claim can be proved by contradiction. Assume there are two product states $|\psi\rangle|\psi\rangle$ and $|\phi\rangle|\phi\rangle$ such that $(\mathbb{I}_{\vee^2\mathbb{C}^d} - 2P_S)|\psi\rangle|\psi\rangle = |\phi\rangle|\phi\rangle$, we have $2P_S|\psi\rangle|\psi\rangle = |\psi\rangle|\psi\rangle - |\phi\rangle|\phi\rangle$. $P_S|\psi\rangle|\psi\rangle \neq 0$ since there is no product state in $S^\perp$. On the other hand, $P_S|\psi\rangle|\psi\rangle$ is a vector in S which is a subspace completely void of states with rank no more than 2. This contradicts our assumption.

In this subsection, we will construct a subspace S to satisfy the above two constraints for any $d \geq 5$. A family of BUEs will follow immediately.

Let S be the span of the following vectors.

$$
\begin{aligned}
&|11\rangle + |23\rangle + |32\rangle,\\
&|22\rangle + |34\rangle + |43\rangle,\\
&\qquad \cdots,\\
&|d-2, d-2\rangle + |d-1, d\rangle + |d, d-1\rangle,\\
&|d-1, d-1\rangle + |d, 1\rangle + |1, d\rangle,\\
&|d, d\rangle + |12\rangle + |21\rangle.
\end{aligned}
$$

We first show there is no product state in $S^{\perp}$. Assume $|\psi\rangle|\psi\rangle \perp S$ where $|\psi\rangle = \sum_{i=1}^{d} a_i|i\rangle$. The orthogonality implies the following equations.

$$
(E1)\begin{cases}
a_1^2 + 2a_2a_3 &= 0,\\
a_2^2 + 2a_3a_4 &= 0,\\
&\vdots\\
a_d^2 + 2a_1a_2 &= 0.
\end{cases}
$$

The only common solution to the above equations is $(a_1, a_2, \cdots, a_d) = (0, 0, \cdots, 0)$ when $d \geq 3$. See Appendix A for details.

Hence, there is no product state in $S^{\perp}$.

Next, we will verify that rank$|\psi\rangle \geq 3$ for any $|\psi\rangle \in S$.

Assume there is some state $|\psi\rangle \in S$ with rank no more than 2. Let's say

$$
\begin{aligned}
& c_1(|11\rangle + |23\rangle + |32\rangle), && (3)\\
+\ & c_2(|22\rangle + |34\rangle + |43\rangle), && (4)\\
+\ & \qquad \cdots, && (5)\\
+\ & c_d(|d, d\rangle + |12\rangle + |21\rangle), && (6)\\
=\ & (x_1|1\rangle + x_2|2\rangle + \cdots + x_d|d\rangle)(x_1|1\rangle + x_2|2\rangle + \cdots + x_d|d\rangle), && (7)\\
+\ & (y_1|1\rangle + y_2|2\rangle + \cdots + y_d|d\rangle)(y_1|1\rangle + y_2|2\rangle + \cdots + y_d|d\rangle). && (8)
\end{aligned}
$$

Then we have the following equations.

$$
(E2)\begin{cases}
x_1^2 + y_1^2 &= c_1,\\
x_2^2 + y_2^2 &= c_2,\\
&\vdots\\
x_d^2 + y_d^2 &= c_d,\\
x_1x_2 + y_1y_2 &= c_d,\\
x_2x_3 + y_2y_3 &= c_1,\\
&\vdots\\
x_dx_1 + y_dy_1 &= c_{d-1},\\
x_ix_j + y_iy_j &= 0 \forall |i-j| \geq 2.
\end{cases} \qquad (9)
$$

There is no nonzero $(c_0, c_2, \cdots, c_d)$ satisfying the above equations when $d \geq 5$. See Appendix B for details. Hence rank$|\psi\rangle \geq 3$ for any $|\psi\rangle \in S$.

This implies that $U = I - 2P_S$ is a bosonic universal entangler for any $d \geq 5$.

3.2.2 Permutation Universal Entanglers

Any product state can be written as the following.

$$|\phi\rangle|\phi\rangle = (\sum_{i=1}^{d} a_i|i\rangle)(\sum_{j=1}^{d} a_j|j\rangle) \tag{10}$$

$$= \sum_{i,j=1}^{d} a_i a_j |ij\rangle \tag{11}$$

$$= \sum_{i=1}^{d} a_i^2|ii\rangle + \sum_{1\le i<j\le d} \sqrt{2}a_i a_j(\frac{|ij\rangle+|ji\rangle}{\sqrt{2}}). \tag{12}$$

Any bosonic state $|\psi\rangle \in \vee^2\mathbb{C}^d$ can be denoted as a $\binom{d+1}{2}$-dimensional vector

$$(x_{11}, x_{22}, \cdots, x_{dd}, x_{12}, \cdots, x_{1d}, x_{21}, \cdots, x_{d-1d})$$

since we can always write $|\psi\rangle$ as a linear combination of bosonic basis states

$$x_{11}|11\rangle + x_{22}|22\rangle + \cdots + x_{dd}|dd\rangle + x_{12}\frac{|12\rangle+|21\rangle}{\sqrt{2}} + x_{13}\frac{|13\rangle+|31\rangle}{\sqrt{2}} + \cdots + x_{d-1,d}\frac{|d-1,d\rangle+|d,d-1\rangle}{\sqrt{2}}.$$

$|\psi\rangle$ is a product state if and only if there exists some nonzero vector $(a_1, a_2, \cdots, a_d)$ such that

$$(x_{11}, \cdots, x_{dd}, x_{12}, x_{13}, \cdots, x_{1d}, x_{23}, \cdots, x_{d-1d}) \tag{13}$$

$$= (a_1^2, \cdots, a_d^2, \sqrt{2}a_1a_2, \sqrt{2}a_1a_3, \cdots, \sqrt{2}a_1a_d, \sqrt{2}a_2a_3, \cdots, \sqrt{2}a_{d-1}a_d). \tag{14}$$

A permutation matrix U acting on the $\binom{d+1}{2}$-dimensional vector space is certainly a bosonic quantum gate.

For any $d \ge 3$, let's define a permutation matrix U as the following:

$$U = \sum_{i=1}^{d}(\frac{|i,i+1\rangle+|i+1,i\rangle}{\sqrt{2}})\langle ii| + \sum_{i=1}^{d}|ii\rangle(\frac{\langle i,i+1|+\langle i+1,i|}{\sqrt{2}}) + \sum_{1\le i<i+1<j\le d}\frac{(|ij\rangle+|ji\rangle)(\langle ij|+\langle ji|)}{2}.$$

Here the addition and subtraction are all modulo d, but the results range from 1 to d.

U is a unitary matrix since it is simply a rotation of the $\binom{d+1}{2}$-dimensional vector space. Let's assume U will map some (bosonic) product state to another (bosonic) product state. Without loss of generality, let's assume

$$U(\sum_{i=1}^{d} a_i^2|ii\rangle + \sum_{1\le i<j\le d}\sqrt{2}a_ia_j(\frac{|ij\rangle+|ji\rangle}{\sqrt{2}})) = \sum_{i=1}^{d} b_i^2|ii\rangle + \sum_{1\le i<j\le d}\sqrt{2}b_ib_j(\frac{|ij\rangle+|ji\rangle}{\sqrt{2}}).$$

It follows that

$$\begin{aligned} a_1^2 &= \sqrt{2}b_1b_2, \\ a_2^2 &= \sqrt{2}b_2b_3, \\ &\vdots\ , \\ a_d^2 &= \sqrt{2}b_db_1, \\ \sqrt{2}a_1a_2 &= b_1^2, \\ \sqrt{2}a_2a_3 &= b_2^2, \\ &\vdots\ , \\ \sqrt{2}a_da_1 &= b_d^2. \end{aligned}$$

Hence we have $\prod_{i=1}^{d} a_i^2 = (\sqrt{2})^d \prod_{i=1}^{d} b_ib_{i+1} = \sqrt{2}^d \prod_{i=1}^{d} b_i^2$. Similarly, $\prod_{i=1}^{d} b_i^2 = \sqrt{2}^d \prod_{i=1}^{d} a_i^2$. The above two equations imply that there exists some $1 \leq t \leq d$ such that $a_t = 0$.

The equation $b_t^2 = \sqrt{2}a_ta_{t+1}$ implies $b_t = 0$. Then $a_{t-1}^2 = \sqrt{2}b_{t-1}b_t = 0$ will implies $a_{t-1} = 0$. By repeating the above procedure, we will eventually have $a_i = 0$ for any $1 \leq i \leq d$. This contradicts our assumption that U will map some (bosonic) product state to another (bosonic) product state. Hence U is a bosonic universal entangler.

4 Fermionic Universal Entanglers

Given a bipartite system of indisitinguishable fermions $\wedge^2\mathbb{C}^d$, a 2-vector in $\wedge^2\mathbb{C}^d$ is said to be decomposable if it can be written as an exterior product of individual vectors from $\mathbb{C}^d$. Decomposable 2-vectors are also considered to be unentangled states in this fermionic system.

We say a quantum gate U is a fermionic universal entangler (FUE) if U will transform every product state to some entangled state.

▶ **Theorem 2.** *There is some FUE acting on a bipartite system of indisitinguishable fermions $\wedge^2\mathbb{C}^d$ if and only if $d \geq 8$. Furthermore, almost every quantum gate acting on $\wedge^2\mathbb{C}^d$ is an FUE when $d \geq 8$.*

Proof. Let $\Gamma_d = \{|\phi\rangle \in \wedge^2\mathbb{C}^d : |\phi\rangle = |\psi_1\rangle \wedge |\psi_2\rangle \text{ for some } |\psi_1\rangle, |\psi_2\rangle \in \mathbb{C}^d\}$. A quantum gate U is an FUE if and only if

$$U(\Gamma_d) \bigcap \Gamma_d = \emptyset. \tag{15}$$

Observe that decomposable 2-vectors in $\wedge^2\mathbb{C}^d$ correspond to weighted 2-dimensional linear subspaces of $\mathbb{C}^d$. If we ignore the phase factor, decomposable 2-vectors can be characterized by the Grassmannian of 2-dimensional subspaces of $\mathbb{C}^d$, an algebraic subvariety of the projective space $\mathbb{P}(\wedge^2\mathbb{C}^d)$[8]. We will denote the Grassmannian of r-dimensional subspaces of $\mathbb{C}^d$ as $G(r, d)$.

First, we examine the necessary condition.

According to the intersection theorem, if $\dim U(\Gamma_d) + \dim \Gamma_d \geq \dim \mathbb{P}(\wedge^2\mathbb{C}^d)$, or equivalently, $2 \times 2(d-2) = 2 \dim G(2, d) \geq \binom{d}{2} - 1$, then for any U, $U(\Gamma_d) \bigcap \Gamma_d \neq \emptyset$. This inequality holds only for $2 \leq d \leq 7$ which implies the fermionic universal entangling device does not exist for $d \leq 7$.

Now, let's look into the sufficient condition.

The set of quantum gates acting on a bipartite system of indisitinguishable fermions $\wedge^2\mathbb{C}^d$ is the unitary group acting on $\wedge^2\mathbb{C}^d$, denoted as $\mathcal{U}(\wedge^2\mathbb{C}^d)$.

Similarly, let $\mathcal{Y} = \{\Phi | \Phi \in \mathcal{U}(\wedge^2\mathbb{C}^d), \Phi(\Gamma_d) \cap \Gamma_d \neq \emptyset\}$. We will show that $\mathcal{Y}$ is a proper subset in $\mathcal{U}(\wedge^2\mathbb{C}^d)$.

Again, let's consider the Zariski topology on the projective space. In this setting, the unitary group $\mathcal{U}(\wedge^2\mathbb{C}^d)$ is Zariski dense in the general linear group $GL(\wedge^2\mathbb{C}^d)$[18]. We further define $\mathcal{Y}' = \{\Phi | \Phi \in GL(\wedge^2\mathbb{C}^d), \Phi(\Gamma_d) \cap \Gamma_d \neq \emptyset\}$. It is easy to see $\mathcal{X} \subseteq \mathcal{Y}'$.

Similar to the proof of Lemma 3 in Appendix C, the dimension of $\mathcal{Y}'$'s Zariski closure $\dim \overline{\mathcal{Y}'}$ is bounded by $\binom{d}{2}^2 - (\binom{d}{2} - 1) + 2 \times 2(d-2)$.

Now we prove the existence of an FUE U as follows. If it does not exist, $\mathcal{U}(\wedge^2\mathbb{C}^d) \subset \mathcal{Y}'$, then $GL(\wedge^2\mathbb{C}^d) = \overline{\mathcal{U}(\wedge^2\mathbb{C}^d)} \subset \overline{\mathcal{Y}'}$. However, $\dim(\overline{\mathcal{Y}'}) \leq \binom{d}{2}^2 - (\binom{d}{2} - 1) + 4(d-2) < \binom{d}{2}^2 = \dim GL(\wedge^2\mathbb{C}^d)$. This is a contradiction. So $\mathcal{U}(\wedge^2\mathbb{C}^d) \not\subset \mathcal{Y}'$, i.e. an FUE $\Phi \in \mathcal{U}(\wedge^2\mathbb{C}^d)$ exists.

Following the lines of the proof of Theorem 1, we can prove that $\mathcal{Y}$ is not only a proper subset, but also a neglectable subset in $\mathcal{U}(\wedge^2\mathbb{C}^d)$. ◀

5 Summary and Discussion

Employing properties of algebraic geometry, we have shown that for bipartite systems of indistinguishable bosons with a Hilbert space that is the symmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$, bosonic universal entanglers (BUEs) exist if and only if $d \geq 3$. Similarly, we have shown that for bipartite systems of indistinguishable fermions with a Hilbert space that is the antisymmetric subspace of $\mathbb{C}^d \otimes \mathbb{C}^d$, fermionic universal entanglers (FUEs) exist if and only if $d \geq 8$. These two results are in contrast to previous results regarding bipartite systems of distinguishable particles with a Hilbert space $\mathbb{C}^{d_1} \otimes \mathbb{C}^{d_2}$, for which universal entanglers exist if and only if $\min(d_1, d_2) \geq 3$ and $(d_1, d_2) \neq (3, 3)$. This illustrates some of the important differences between the entanglement of systems of indistinguishable particles to the entanglement of systems of distinguishable particles.

In contrast, we have illustrated one feature which holds for both systems of distinguishable and indistinguishable particles. Previous work has shown that, for systems of distinguishable particles, if a universal entangler exists for some Hilbert space, then almost all unitaries operating on that space are universal entanglers. We have shown that this result also holds for systems of indistinguishable bosons and fermions. However to verify whether or not a given bipartite unitary is a universal entangler is in general an intractable problem for both distinguishable particle systems and fermionic systems. This intractability arises from the fact that solving a system of quadratic equations is, in general, NP-hard.

Bosonic systems turn out to be special though. Because the set of all product states is isomorphic to a linear vector space, it is possible to use elementary methods to verify bosonic universal entanglers. We have given explicit constructions of two types of BUE, one is of the Householder type which is valid for $d \geq 5$ and the other is of a permutation type which is valid for $d \geq 3$. Both are very simple constructions.

It is our hope that our success in finding explicit constructions of BUEs will help inform the search for explicit constructions of both FUEs and UEs, problems which remain intractable in general. We can not rule out the possibiliy that there might be some other structure, beyond just the corresponding general algebraic varieties, which would provide some special family of explicitly verifiable UEs or FUEs. In fact, the explicit construction for the $(3, 4)$ system from an order 12 Hadamard matrix demonstrated in [4] provides a hint of the possibility of such families.

Another natural direction of inquiry is to explore the entangling power of these BUEs

and FUEs. As demonstrated in [4], a random unitary is not only almost surely a UE, but it also almost surely maps the set of product states to another set which contains nothing but nearly maximally entangled states, with respect to almost any kind of entanglement measure. One would expect similar properties for BUEs and FUEs. However to go further in that direction one would need to first establish reasonable entanglement measures for bosonic and fermionic systems (see, e.g. entanglement measures discussed in [5]).

Finally, it would be useful to generalize these results to multipartite bosonic and fermionc systems. Our guess is that the bosonic systems might remain easy to solve since they retain the nice property that the set of all product states is isomorphic to a linear vector space. The fermionic case is expected to be much more complicated given that in the multipartite case even the Grassmann-Plücker relations themselves are harder to describe [2, 9, 5]. We would leave these cases for future investigation.

Acknowledgements. JK is supported by NSERC. JC is supported by NSERC, UTS-AMSS Joint Research Laboratory for Quantum Computation and Quantum Information Processing and NSF of China (Grant No. 61179030). BZ is supported by NSERC and CIFAR.

References

1 L. Amico, R. Fazio, A. Osterloh, and V. Vedral. Entanglement in many-body systems. *Reviews of Modern Physics*, 80:517–576, April 2008.

2 N. Bourbaki. *Èlèments de Mathèmatique, Algébre*. Hermann, 1970.

3 Jianxin Chen, Runyao Duan, Zhengfeng Ji, Mingsheng Ying, and Jun Yu. Existence of universal entangler. *Journal of Mathematical Physics*, 49(1):012103, 2008.

4 Jianxin Chen, Zhengfeng Ji, David W. Kribs, and Bei Zeng. Minimum Entangling Power is Close to Its Maximum. *arXiv:1210.1296*, 2012.

5 K. Eckert, J. Schliemann, D. Bruß, and M. Lewenstein. Quantum Correlations in Systems of Indistinguishable Particles. *Annals of Physics*, 299:88–127, July 2002.

6 Aviezri S. Fraenkel and Yaacov Yesha. Complexity of solving algebraic equations. *Information Processing Letters*, pages 178,179, 1980.

7 J. Harris. *Algebraic geometry: a first course*, volume 133. Springer, 1992.

8 Robin Hartshorne. Algebraic geometry. Springer-Verlag, 1983.

9 B. Hasset. *Introduction to algebraic geometry*. Cambridge University Press, Cambridge, 2007.

10 R.A. Horn and C.R. Johnson. *Matrix analysis*. Cambridge university press, 1990.

11 R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki. Quantum entanglement. *Reviews of Modern Physics*, 81:865–942, April 2009.

12 Y. S. Li, B. Zeng, X. S. Liu, and G. L. Long. Entanglement in a two-identical-particle system. *Physical Review A*, 64(5):054302, November 2001.

13 M. Nielsen and I. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, England, 2000.

14 R. Paškauskas and L. You. Quantum correlations in two-boson wave functions. *Physical Review A*, 64(4):042310, October 2001.

15 Ravider R. Puri. *Mathematical Methods of Quantum Optics (Springer Series in Optical Sciences*. Springer, 2001.

16 J. Schliemann, J. I. Cirac, M. Kuś, M. Lewenstein, and D. Loss. Quantum correlations in two-fermion systems. *Physical Review A*, 64(2):022303, August 2001.

17 J. Schliemann, D. Loss, and A. H. MacDonald. Double-occupancy errors, adiabaticity, and entanglement of spin qubits in quantum dots. *Physical Review B*, 63(8):085311, February 2001.

18 Alexander H W Schmitt. *Geometric invariant theory and decorated principal bundles*. Amer Mathematical Society, July 2008.

19 E. Schrödinger. Discussion of probability relations between separated systems. *Mathematical Proceedings of the Cambridge Philosophical Society*, 31(04):555–563, October 1935.

20 Patrice Tauvel and Rupert W T Yu. *Lie algebras and algebraic groups*. Springer Verlag, 2005.

A There Is No Nonzero Solution For Polynomial System (E1)

Here we will show there is no nonzero solution $(a_1, \cdots, a_d)$ satisfying the following equations.

$$(E1)\begin{cases} a_1^2 + 2a_2a_3 &= 0, \\ a_2^2 + 2a_3a_4 &= 0, \\ &\vdots \\ a_d^2 + 2a_1a_2 &= 0. \end{cases}$$

Assume $a_i \neq 0$, then a_{i+1}, a_{i+2} are nonzero. This follows all a_i's are nonzero.

$$\prod_{i=1}^{d} a_i^2 = \prod_{i=1}^{d} (-2a_{i+1}a_{i+2}) = (-2)^d \prod_{i=1}^{d} a_i^2. \tag{16}$$

This implies $\prod_{i=1}^{d} a_i^2 = 0$. Hence it is a contradiction.

Therefore, the only solution to this polynomial system is $(a_1, \cdots, a_d) = (0, \cdots, 0)$.

B There Is No Nonzero Solution For Polynomial System (E2)

Here we will show there is no nonzero solution $(c_0, c_1, \cdots, c_d)$ satisfying the following equations.

$$(E2)\begin{cases} x_1^2 + y_1^2 &= c_1, \\ x_2^2 + y_2^2 &= c_2, \\ &\vdots \\ x_d^2 + y_d^2 &= c_d, \\ x_1x_2 + y_1y_2 &= c_d, \\ x_2x_3 + y_2y_3 &= c_1, \\ &\vdots \\ x_dx_1 + y_dy_1 &= c_{d-1}, \\ x_ix_j + y_iy_j &= 0 \forall |i-j| \geq 2. \end{cases} \tag{17}$$

For $d \geq 5$, let's assume there is some $1 \leq i \leq d$ such that $x_i = 0$ and $y_i \neq 0$.

It follows from $x_ix_j + y_iy_j = 0$ for any $|j-i| \geq 2$ that $y_j = 0$ for any $|j-i| \geq 2$. So, $y_{i+2} = y_{i+3} = 0$.

Then, $0 \neq x_i^2 + y_i^2 = x_{i+1}x_{i+2} + y_{i+1}y_{i+2} = x_{i+1}x_{i+2}$. This implies $x_{i+1}, x_{i+2} \neq 0$.

From $x_{i+1}x_{i+3} + y_{i+1}y_{i+3} = 0$, we have $x_{i+3} = 0$. So, $0 \neq x_{i+2}^2 + y_{i+2}^2 = x_{i+3}x_{i+4} + y_{i+3}y_{i+4} = 0$. This is a contradiction.

So, for any $1 \leq i \leq d$, we have $x_i = y_i = 0$ or $x_iy_i \neq 0$.

Let's look into the various situations.

1. There is some i such that $x_i = y_i = 0$. Then we have $x_{i-1}^2 + y_{i-1}^2 = 0$. If $x_{i-1} = y_{i-1} = 0$, we consider $x_{i-2}^2 + y_{i-2}^2 = 0$. By repeating this procedure, if all x_j's,y_j's are not all zero, we will find some i' such that $y_{i'} = ix_{i'}$ or $y_{i'} = -ix_{i'}$ and $x_{i'+1} = y_{i'+1} = 0$. We will further have $y_{i'+k} = \pm x_{i'+k}$ for any $k = 2, \cdots, d-1$. This implies that $(c_1, \cdots, c_d) = 0$.
2. All x_i's, y_i's are nonzero. For any fixed i, $\frac{y_j}{x_j} = -\frac{x_i}{y_i}$ for any $j = i+2, \cdots, i+d-2$. This implies $\frac{y_k}{x_k}$ is a constant i or $-i$. This also implies $(c_1, \cdots, c_d) = 0$.

C Proof of Lemma 3

▶ **Lemma 3.** $\dim(\overline{\mathcal{X}'}) \leq \binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + 2(d-1)$, *where* $\overline{\mathcal{X}'}$ *is the Zariski closure of* $\mathcal{X}'$.

The following technical lemmas will be needed.

▶ **Lemma 4** ([20]). *If* Z_1 *and* Z_2 *are both irreducible varieties over* $\mathbb{C}$, *and* $\phi : Z_1 \to Z_2$ *is a dominant morphism, then* $\dim(Z_2) \leq \dim(Z_1)$. *Here, dominant means* $\Phi(Z_1)$ *is dense in* Z_2.

▶ **Lemma 5** ([20]). *If* Z_1 *and* Z_2 *are both varieties over* $\mathbb{C}$, *and* $\phi : Z_1 \to Z_2$ *is a morphism, then* $\dim(Z_1) \leq \dim(Z_2) + \max_{z \in Z_2} \dim(\phi^{-1}(z))$.

Lemma 4 and Lemma 5 establish a connection between the dimensions of domain and codomain of a variety morphism.

Proof. We have a morphism $F : GL(\vee^2\mathbb{C}^d) \times \mathbb{P}^{\binom{d+1}{2}-1} \to \mathbb{P}^{\binom{d+1}{2}-1}$ which is just the left action of $GL(\vee^2\mathbb{C}^d)$ on $\mathbb{P}^{\binom{d+1}{2}-1}$, defined by $F(g, [w]) = [g \cdot w]$.

We let $y_0 = (1, 0, \cdots, 0)$ be a row vector with $\binom{d+1}{2}$ entries, and for any given y_1, $y_2 \in \mathbb{P}^{\binom{d+1}{2}-1}$, we choose proper g_1 and $g_2 \in GL(\vee^2\mathbb{C}^d)$, such that $[g_1 \cdot y_0] = [y_1]$ and $[g_2 \cdot y_0] = [y_2]$. Then we have

$$[g \cdot y_2] = [y_1] \iff [gg_2 \cdot y_0] = [gg_1 \cdot y_0] \iff [g_1^{-1} g g_2 \cdot y_0] = [y_0]. \tag{18}$$

From the above observations, F has the following property: for any y_1, $y_2 \in \mathbb{P}^{\binom{d+1}{2}-1}$, $F^{-1}(y_2) \cap \{GL(\vee^2\mathbb{C}^d) \times \{y_1\}\} \cong \{\begin{pmatrix} z_1 & \alpha \\ 0 & g' \end{pmatrix} : z_1 \in \mathbb{C}\backslash\{0\}, g' \in GL(\binom{d+1}{2} - 1), \alpha \in \mathbb{C}^{\binom{d+1}{2}-1}$ *is a row vector.*$\}$. Hence $\dim(F^{-1}(y_2) \cap GL(\vee^2\mathbb{C}^d) \times \{y_1\}) = \binom{d+1}{2}^2 - (\binom{d+1}{2} - 1)$.

Let P_1, P_2 be projections of $GL(\vee^2\mathbb{C}^d) \times \mathbb{P}^{\binom{d+1}{2}-1}$ to $GL(\vee^2\mathbb{C}^d)$, $\mathbb{P}^{\binom{d+1}{2}-1}$ respectively. Now we only look at $GL(\vee^2\mathbb{C}^d) \times \Lambda \subseteq GL(\vee^2\mathbb{C}^d) \times \mathbb{P}^{\binom{d+1}{2}-1}$, to get $F : GL(\vee^2\mathbb{C}^d) \times \Lambda \to \mathbb{P}^{\binom{d+1}{2}-1}$. Then we have a characterization of $\mathcal{X}'$: $\mathcal{X}' = P_1 F^{-1}(\Lambda)$. In fact

$$\begin{aligned}
& g \in \mathcal{X}' \\
\iff & g(\Lambda) \cap \Lambda \neq \emptyset \\
\iff & \exists z_1, z_2 \in \Lambda, s.t. g(z_1) = z_2 \\
\iff & \exists z_1, z_2 \in \Lambda, s.t. (g, z_1) \in F^{-1}(z_2) \\
\iff & \exists z_2 \in \Lambda, s.t. g \in P_1 F^{-1}(z_2) \\
\iff & g \in P_1 F^{-1}(\Lambda).
\end{aligned}$$

So $\overline{\mathcal{X}'} \subseteq GL(\vee^2\mathbb{C}^d)$ is the Zariski closure of $\mathcal{X}'$, which is also an algebraic variety.

Next, we assert that $P_1 : F^{-1}(\Lambda) \to \overline{\mathcal{X}'}$ is a dominant morphism.

Furthermore, consider $\Psi : F^{-1}(\Lambda) \to \Lambda \times \Lambda$ given by $\Psi(g, [z]) = ([z], [g \cdot z])$.

For $\forall z_1 \in \Lambda$, $z_2 \in \Lambda$, we have $\Psi^{-1}(z_1, z_2) = (g_2 T g_1^{-1}, z_1)$, where $T = \{\begin{pmatrix} z_0 & \alpha \\ 0 & g' \end{pmatrix}$: $z_0 \in \mathbb{C}\backslash\{0\}, g' \in GL(\binom{d+1}{2} - 1), \alpha \in \mathbb{C}^{\binom{d+1}{2}-1}$ *is a row vector*$\}$, and $g_1, g_2 \in GL(\vee^2\mathbb{C}^d)$, s.t. $g_1(y_0) = z_1$, $g_2(y_0) = z_2$. So this is a dominant morphism. Then we obtain

$$\begin{aligned}\dim(F^{-1}(\Lambda)) \leq & \dim(T) + \dim(\Lambda \times \Lambda) \\ = & \binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + \dim(\Lambda) + \dim(\Lambda).\end{aligned}$$

It is required in Lemma 4 that varieties Z_1 and Z_2 be irreducible. Actually, this condition can be weakened. Lemma 4 is still true for the more general case that Z_1 and Z_2 are closed subsets of irreducible varieties[8]. Through this approach, we can fill out the gap and apply this lemma without danger of confusion. Indeed, the irreduciblity of Z_1 and Z_2 really holds, but verification of this is not easy.

Then from Lemma 4 and Lemma 5, we will have

$$\begin{aligned}\dim(\overline{\mathcal{X}'}) \leq & \dim(F^{-1}(\Lambda)) \\ \leq & \binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + \dim(\Lambda) + \dim(\Lambda) \\ = & \binom{d+1}{2}^2 - (\binom{d+1}{2} - 1) + 2(d-1).\end{aligned}$$

◀

Easy and Hard Functions for the Boolean Hidden Shift Problem

Andrew M. Childs[1], Robin Kothari[2], Maris Ozols[3(1,4)], and Martin Roetteler[4]

1 **Department of Combinatorics & Optimization and Institute for Quantum Computing, University of Waterloo**
200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
amchilds@uwaterloo.ca

2 **David R. Cheriton School of Computer Science and Institute for Quantum Computing, University of Waterloo**
200 University Avenue West, Waterloo, ON, N2L 3G1, Canada
rkothari@uwaterloo.ca

3 **IBM TJ Watson Research Center**
1101 Kitchawan Road, Yorktown Heights, NY 10598, USA
marozols@yahoo.com

4 **NEC Laboratories America**
4 Independence Way, Suite 200, Princeton, NJ 08540, USA
mroetteler@nec-labs.com

Abstract

We study the quantum query complexity of the Boolean hidden shift problem. Given oracle access to $f(x+s)$ for a known Boolean function f, the task is to determine the n-bit string s. The quantum query complexity of this problem depends strongly on f. We demonstrate that the easiest instances of this problem correspond to bent functions, in the sense that an exact one-query algorithm exists if and only if the function is bent. We partially characterize the hardest instances, which include delta functions. Moreover, we show that the problem is easy for random functions, since two queries suffice. Our algorithm for random functions is based on performing the pretty good measurement on several copies of a certain state; its analysis relies on the Fourier transform. We also use this approach to improve the quantum rejection sampling approach to the Boolean hidden shift problem.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Boolean hidden shift problem, quantum algorithms, query complexity, Fourier transform, bent functions

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.50

1 Introduction

Many computational problems for which quantum algorithms can achieve superpolynomial speedup over the best known classical algorithms are related to the *hidden subgroup problem* (see for example [1]).

▶ **Problem 1** (Hidden subgroup problem)**.** For any finite group G, say that a function $f\colon G \to X$ *hides* a subgroup H of G if it is constant on cosets of H in G and distinct on different cosets. Given oracle access to such an f, find a generating set for H.

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 50–79

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Two early examples of algorithms for hidden subgroup problems are the Deutsch–Jozsa algorithm [2] and Simon's algorithm [3]. Inspired by the latter, Shor discovered efficient quantum algorithms for factoring integers and computing discrete logarithms [4]. Kitaev subsequently introduced the Abelian stabilizer problem and derived an efficient quantum algorithm for it that includes Shor's factoring and discrete logarithm algorithms as special cases [5]. Eventually it was observed that all of the above algorithms solve special instances of the hidden subgroup problem [6, 7, 8].

This early success created significant interest in studying various instances of the hidden subgroup problem and led to discovery of many other quantum algorithms. For example, period finding over the reals was used by Hallgren to construct an efficient quantum algorithm for solving Pell's equation [9]. Moreover, the hidden subgroup problem over symmetric and dihedral groups are related to the graph isomorphism problem [10, 11, 12, 13, 14] and certain lattice problems [15], respectively. The possibility of efficient quantum algorithms for these problems remains a major open question. Kuperberg has provided a subexponential-time quantum algorithm for the dihedral subgroup problem [16, 17, 18], which has been used to construct elliptic curve isogenies in quantum subexponential time [19].

The *hidden shift problem* (also known as the *hidden translation problem*) is a natural variant of the hidden subgroup problem.

▶ **Problem 2** (Hidden shift problem)**.** Let G be a finite group. Given oracle access to functions $f_0, f_1 \colon G \to X$ with the promise that $f_0(x) = f_1(x \cdot s)$ for some $s \in G$, determine s.

If G is Abelian and f_0 is injective, this problem is equivalent to the hidden subgroup problem in the semidirect product group $G \rtimes \mathbb{Z}_2$, where the group operation is defined by $(x_1, b_1) \cdot (x_2, b_2) := \left(x_1 \cdot x_2^{(-1)^{b_1}}, b_1 + b_2\right)$ and the hiding function $f \colon G \rtimes \mathbb{Z}_2 \to X$ is defined as $f[(x, b)] := f_b(x)$. One can check that f is constant on cosets of $H := \langle (s, 1) \rangle$ and that injectivity of f_0 implies that f is distinct on different cosets. Thus, f hides the subgroup H in $G \rtimes \mathbb{Z}_2$.

Notice that if $G = \mathbb{Z}_d$ then $G \rtimes \mathbb{Z}_2$ is the dihedral group. Ettinger and Høyer [20] showed that the dihedral hidden subgroup problem reduces to the special case of a subgroup $\langle (s, 1) \rangle$. Thus the hidden shift problem in $\mathbb{Z}_d$ (with f_0 injective) is equivalent to the dihedral hidden subgroup problem, motivating further study of the hidden shift problem for various groups [21, 22, 23, 24, 25, 26].

While the case where f_0 is injective is simply related to the hidden subgroup problem, one can also consider the hidden shift problem without this promise. For example, van Dam, Hallgren, and Ip [21] gave an efficient quantum algorithm to solve the shifted Legendre symbol problem, a non-injective hidden shift problem. Their result breaks a proposed pseudorandom function [27], showing the potential for cryptographic applications of hidden shift problems. Work on hidden shift problems can also inspire new algorithmic techniques, such as quantum rejection sampling [28]. Moreover, negative results could have applications to designing classical cryptosystems that are secure against quantum attacks [15].

For the rest of the paper we restrict our attention to the *Boolean hidden shift problem*, in which the hiding function has the form $f_0 \colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ for some integer $n \geq 1$. For this problem (with $n > 1$), f_0 is necessarily non-injective. This problem has previously been studied in [29, 30, 31, 28, 32].

Notice that to determine the hidden shift of an injective function f_0, it suffices to find x_0 and x_1 such that $f_0(x_0) = f_1(x_1)$. However, this does not hold in the non-injective case, so it is nontrivial to verify a candidate hidden shift (see [28, Appendix B]). In fact, sometimes the hidden shift cannot be uniquely determined in principle (see Sect. D.1). On the other

hand, by considering functions with codomain $\mathbb{Z}_2$, we have more structure than in the hidden subgroup problem or the injective hidden shift problem, where the codomain is arbitrary. We exploit this structure by encoding the values of the function as phases and using the Fourier transform.

More precisely, the main problem studied in this paper, sometimes denoted BHSP_f, is as follows.

▶ **Problem 3** (Boolean hidden shift problem)**.** Given a complete description of a function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ and access to an oracle for the *shifted function* $f_s(x) := f(x+s)$, determine the *hidden shift* $s \in \mathbb{Z}_2^n$.

Note that in degenerate cases, when the oracle does not contain enough information to completely recover the hidden shift, no algorithm can succeed with certainty.

Let us highlight the main differences between the above problem and other types of hidden shift problem. In the Boolean hidden shift problem,

- the function f is *not* injective, and
- we are given a *complete description* of the unshifted function f instead of having only oracle access to f.

Moreover, we are interested only in the *query complexity* of the problem and do not consider its time complexity. This means that we can pre-process the description of f (which may be exponentially large) at no cost before we start querying the oracle.

This problem has been considered previously, e.g., by [28]. Note that some prior work does not give complete description of f but only oracle access to it [29, 30, 31, 32] (and in some cases [30] also gives oracle access also to $\tilde{f}$, the dual bent function of f).

To address this problem on a quantum computer, we use an oracle that computes the shifted function in the phase. Such an oracle can be implemented using only one query to an oracle that computes the function in a register.

▶ **Definition 1.** The quantum *phase oracle* is $O_{f_s}\colon |x\rangle \mapsto (-1)^{f(x+s)}|x\rangle$.

More generally, one can use a controlled phase oracle $\bar{O}_{f_s}\colon |b,x\rangle \mapsto (-1)^{bf(x+s)}|b,x\rangle$ for $b \in \{0,1\}$, which is equivalent to an oracle that computes the function in the first register up to a Hadamard transform. Some of our algorithms do not make use of this freedom, although our lower bounds always take it into account.

Ultimately, we would like to characterize the classical and quantum query complexities of the hidden shift problem for any Boolean function (or more generally, for any function $f\colon \mathbb{Z}_d^n \to \mathbb{Z}_d$). While we do not resolve this question completely, we make progress by providing a new quantum query algorithm (see Sect. 4) and improving an existing one (see Sect. 5). However, it remains an open problem to better understand both the classical and quantum query complexities of the BHSP for general functions.

While general functions are difficult to handle, the quantum query complexity of the hidden shift problem is known for two extreme classes of Boolean functions:

- If f is a *bent function*, i.e., it has a "flat" Fourier spectrum (see Sect. 3.1), then one quantum query suffices to solve the problem exactly [30].
- If f is a *delta function*, i.e., $f(x) := \delta_{x,x_0}$ for some $x_0 \in \mathbb{Z}_2^n$, then the hidden shift problem for f is equivalent to unstructured search—finding $x_0 + s$ among the 2^n elements of $\mathbb{Z}_2^n$—so the quantum query complexity is $\Theta(\sqrt{2^n})$ [33, 34].

Intuitively, other Boolean functions should lie somewhere between these two extreme cases. In this paper, we give formal evidence for this: we show that the problem can be solved exactly with one query only if f is bent, and we show that it can be solved for any function

with $O(\sqrt{2^n})$ queries, with a lower bound of $\Omega(\sqrt{2^n})$ only if the truth table of f has Hamming weight $\Theta(1)$ or $\Theta(2^n)$. This is similar to the weighing matrix problem considered by van Dam [35], which also interpolates between two extreme cases: the Bernstein-Vazirani problem [36] and Grover search [33].

Aside from delta and bent functions, the Boolean hidden shift problem has previously been considered for several other families of functions. Boolean functions that are quadratic forms or are close to being quadratic are studied in [29]. Random Boolean functions have been considered in [31, 32]. Finally, [28] uses quantum rejection sampling to solve the BHSP for any function, although its performance in general is not well understood.

Apart from algorithms designed specifically for the BHSP, there are generic classical and quantum algorithms for the BHSP derived from learning theory. In particular, the BHSP can be viewed as an instantiation of the problem of exact learning through membership queries. The resulting algorithms are optimal for classical and quantum query complexity up to polynomial factors in n. More precisely, for any learning problem, Servedio and Gortler define a combinatorial parameter γ [37]. For the problem BHSP_f, we denote the parameter as γ_f. From their results it follows that the classical query complexity of BHSP_f is lower bounded by $\Omega(n)$ and $\Omega(1/\gamma_f)$ and upper bounded by $O(n/\gamma_f)$. For quantum algorithms, they show a lower bound of $\Omega(1/\sqrt{\gamma_f})$. Atıcı and Servedio [38] later showed an upper bound of $O(n \log n/\sqrt{\gamma_f})$ queries.

The rest of this paper is organized as follows. In Sect. 2 we briefly review some basic Fourier analysis to establish notation. Next, in Sect. 3 we explore the extreme cases of the BHSP. In Sect. 4 we introduce a new approach to the BHSP based on the pretty good measurement. We analyze its performance for delta, bent, and random Boolean functions in Sect. 4.3. In Sect. 5 we propose an alternative method for boosting the success probability of the quantum rejection sampling algorithm from [28]. Finally, Sect. 6 presents conclusions and open questions.

This paper has several appendices. In Appendix A we show that the easy instances of the BHSP correspond to bent functions. In Appendix B, we show that with one quantum query we can succeed on a constant fraction of all functions, whereas in Appendix C we prove that two quantum queries suffice to solve the BHSP for random functions. Finally, in Appendix D we analyze the structure of zero Fourier coefficients of Boolean functions.

2 Fourier analysis

Our main tool is Fourier analysis of Boolean functions [39]. Here we state the basic definitions and properties of the Fourier transform and convolution. Readers who are familiar with the topic might skip this section, except for Definition 6.

▶ **Definition 2.** The *Hadamard gate* is $H := \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$.

▶ **Definition 3.** The *Fourier transform* of a function $F\colon \mathbb{Z}_2^n \to \mathbb{R}$ is a function $\hat{F}\colon \mathbb{Z}_2^n \to \mathbb{R}$ defined as $\hat{F}(w) := \langle w|H^{\otimes n}|F\rangle$ where $|F\rangle := \sum_{x\in\mathbb{Z}_2^n} F(x)|x\rangle$. Here $\hat{F}(w)$ is called the *Fourier coefficient* of F at $w \in \mathbb{Z}_2^n$. Explicitly, $\hat{F}(w) = \frac{1}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^n} (-1)^{w\cdot x} F(x)$ where $x \cdot y := \sum_{i=1}^n x_i y_i$. The set $\{\hat{F}(w)\colon w \in \mathbb{Z}_2^n\}$ is called the *Fourier spectrum* of F.

To define the Fourier transform of a Boolean function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$, we identify f with a real-valued function $F\colon \mathbb{Z}_2^n \to \mathbb{R}$ in a canonical way: $F(x) := (-1)^{f(x)}/\sqrt{2^n}$. Note that F is normalized: $\sum_{x\in\mathbb{Z}_2^n} |F(x)|^2 = 1$. Now we can abuse Definition 3 as follows:

▶ **Definition 4.** The *Fourier transform* of $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ is $\hat{F}(w) = \frac{1}{2^n} \sum_{x\in\mathbb{Z}_2^n} (-1)^{w\cdot x + f(x)}$.

To avoid confusion, we use lower case letters for $\mathbb{Z}_2$-valued functions and capital letters for $\mathbb{R}$-valued functions.

▶ **Definition 5.** The *convolution* of functions $F, G\colon \mathbb{Z}_2^n \to \mathbb{R}$ is a function $(F * G)\colon \mathbb{Z}_2^n \to \mathbb{R}$ defined as $(F * G)(x) := \sum_{y\in\mathbb{Z}_2^n} F(y)G(x-y)$. The *t-fold convolution* of $F\colon \mathbb{Z}_2^n \to \mathbb{R}$ is a function $F^{*t}\colon \mathbb{Z}_2^n \to \mathbb{R}$ defined as

$$F^{*t}(w) := (\underbrace{F * \cdots * F}_{t})(w) = \sum_{y_1,\ldots,y_{t-1}\in\mathbb{Z}_2^n} F(y_1)\cdots F(y_{t-1})F\big(w-(y_1+\cdots+y_{t-1})\big). \tag{1}$$

▶ **Fact.** Let $F, G, H\colon \mathbb{Z}_2^n \to \mathbb{R}$ denote arbitrary functions. The Fourier transform and convolution have the following basic properties:

1. The Fourier transform is linear: $\widehat{F+G} = \hat{F} + \hat{G}$.
2. The Fourier transform is self-inverse: $\hat{\hat{F}} = F$.
3. Since $H^{\otimes n}$ is unitary, the Plancherel identity $\sum_{w\in\mathbb{Z}_2^n}|\hat{F}(w)|^2 = \sum_{x\in\mathbb{Z}_2^n}|F(x)|^2$ holds.
4. Convolution is commutative ($F * G = G * F$) and associative ($(F * G) * H = F * (G * H)$).
5. The Fourier transform and convolution are related through the following identities: $(\hat{F} * \hat{G})/\sqrt{2^n} = \widehat{FG}$ and $\widehat{(F * G)}/\sqrt{2^n} = \hat{F}\hat{G}$, where $FG\colon \mathbb{Z}_2^n \to \mathbb{C}$ is the entry-wise product of functions F and G: $(FG)(x) := F(x)G(x)$.
6. By induction, the t-fold convolution satisfies the identity $\big[\hat{F}/\sqrt{2^n}\big]^{*t} = \widehat{F^t}/\sqrt{2^n}$.

The following t-fold generalization of the Fourier spectrum plays a key role:

▶ **Definition 6.** For $t \geq 1$, the *t-fold Fourier coefficient* of $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ at $w \in \mathbb{Z}_2^n$ is $\mathcal{F}^t(w) := \sqrt{\big[\hat{F}^2\big]^{*t}(w)}$. In particular, for $t = 1$ we have $\mathcal{F}^1(w) = |\hat{F}(w)|$.

We can express $\mathcal{F}^t(w)$ in many equivalent ways using the identities listed above:

$$\big[\mathcal{F}^t(w)\big]^2 = \big[\hat{F}^2\big]^{*t}(w) = \left[\frac{1}{\sqrt{2^n}}\big(\widehat{F * F}\big)\right]^{*t}(w) = \frac{1}{\sqrt{2^n}}\,(\widehat{F * F})^t\,(w). \tag{2}$$

3 Characterization of extreme cases

In this section we explore the set of functions for which the quantum query complexity of the BHSP is extreme. Recall that the BHSP can be solved with one query for bent functions and with $\Theta(\sqrt{2^n})$ queries for delta functions. Here we prove that BHSP_f can be solved exactly with one query only if f is bent, and with $O(\sqrt{2^n})$ queries (with bounded error) for any f.

3.1 Easy functions are bent

In general, the quantum query complexity of the BHSP for an arbitrary function is unknown. However, the problem becomes particularly easy for *bent functions*, where a single query suffices to solve the problem exactly [30]. In fact, bent functions are the only functions with this property, as we show here.

Bent functions can be characterized in many equivalent ways [40, 41]. The standard definition is that bent functions have a "flat" Fourier spectrum:

▶ **Definition 7.** A Boolean function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ is *bent* if all its Fourier coefficients $\hat{F}(w)$ (see Definition 4) have the same absolute value: $|\hat{F}(w)| = 1/\sqrt{2^n}$ for all $w \in \mathbb{Z}_2^n$.

While many examples of bent functions have been constructed (e.g., see [42, 43, 44]), no complete classification is known. As an example, the *inner product* of two n-bit strings (modulo two) is a bent function [41, 42]: $\mathrm{IP}_n(x_1,\ldots,x_n,y_1,\ldots,y_n) := \sum_{i=1}^n x_i y_i$.

We make a few simple observations about bent functions. Recall from Sect. 2 that the Fourier spectrum of f is normalized as $\sum_{w\in\mathbb{Z}_2^n}|\hat{F}(w)|^2 = 1$, so the spectrum is "flat" only when $|\hat{F}(w)| = 1/\sqrt{2^n}$ for all $w \in \mathbb{Z}_2^n$. Recall from Definition 4 that $\hat{F}(w)$ is always an integer multiple of $1/2^n$. Thus an n-variable function can only be bent if n is even [43, 42]. Moreover, from $|\hat{F}(0)| = 1/\sqrt{2^n}$ we get that $|\sum_{w\in\mathbb{Z}_2^n}(-1)^{f(x)}| = \sqrt{2^n}$, so a bent function f is close to being balanced: $|f| = (2^n \pm \sqrt{2^n})/2$ where $|f| := |\{x \in \mathbb{Z}_2^n \colon f(x) = 1\}|$ is the *Hamming weight* of f.

Our main result regarding bent functions is as follows.

▶ **Theorem 8.** *Let $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ be a Boolean function with $n \geq 2$. A quantum algorithm can solve BHSP_f exactly with a single query to O_{f_s} if and only if f is bent.*

The proof is based on a characterization of an exact one-query quantum algorithm using a system of linear equations. This system can be analyzed in terms of the autocorrelation of f, which in turn characterizes whether f is bent. The proof appears in Appendix A.

3.2 Hard functions

In this section we study hard instances of the BHSP. First, we observe that the quantum query complexity of solving BHSP_f for any function f is $O(\sqrt{2^n})$.

▶ **Theorem 9.** *For any $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$, the bounded-error quantum query complexity of BHSP_f is $O(\sqrt{2^n})$.*

If we view f as a 2^n-bit string indexed by $x \in \mathbb{Z}_2^n$, this is a special case of the oracle identification problem considered by Ambainis et al. [45, Theorem 3], who show the following.

▶ **Theorem 10** (Oracle Identification Problem)**.** *Given oracle access to an unknown N-bit string with the promise that it is one of N known strings, the bounded-error quantum query complexity of identifying the unknown string is $O(\sqrt{N})$.*

In the BHSP, we have $N := 2^n$. By Theorem 9, the hardest functions are those with query complexity $\Omega(\sqrt{N})$. We know that delta functions have this query complexity, but are there any other functions that are as hard? The delta functions have $|f| = 1$ (recall that $|f|$ denotes the Hamming weight of f). Next we show that as $|f|$ increases, the query complexity strictly decreases at first, until $|f| = \Theta(\sqrt{N})$. For example, functions with $|f| = 2$ have strictly smaller query complexity than the delta functions. However, as we approach $|f| = \Omega(N)$, our upper bound is $\Theta(\sqrt{N})$ again. Without loss of generality, we assume that $|f| \leq N/2$; otherwise we can simply negate the function to obtain a function with $|f| \leq N/2$ that has exactly the same query complexity. Formally, we show the following refinement of Theorem 9.

▶ **Theorem 11.** *For any $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ with $1 \leq |f| \leq N/2$, the bounded-error quantum query complexity of BHSP_f is at most $\frac{\pi}{4}\sqrt{N/|f|} + O(\sqrt{|f|})$.*

Proof. The algorithm has two parts. First we look for a "1" in the bit string contained in the oracle, i.e., an x such that $f(x) = 1$. This can be done by a variant of Grover's algorithm that finds a "1" in a string of length N using at most $\frac{\pi}{4}\sqrt{N/|f|}$ queries [46]. Now we have an x such that $f_s(x) = 1$ for some unknown s. Note that there can be at most $|f|$ shifts s

with this property, because each corresponds to a distinct solution to $f(x+s)=1$ and there are only $|f|$ solutions to this equation.

We are now left with $|f|$ candidates for the black-box function. Viewing this as an oracle identification problem, we have oracle access to an N-bit string that could be one of $|f|$ possible candidates. Although the string has length N, there are only $|f|$ potential candidates, so intuitively it seems like we should be able to restrict the strings to length $|f|$ and apply Theorem 10 to obtain the desired result.

Formally, it can be shown that given $k \geq 2$ distinct Boolean strings of length N, there is a subset of indices, S, of size at most $k-1$, such that all the strings are distinct when restricted to S. We show this by induction. The base case is easy: we can choose any index that differentiates the two distinct strings. Now say we have m distinct strings $y_1, y_2, \ldots, y_m$ and a subset of indices S of size at most $m-1$, such that the m strings are distinct on S. We want to add another string y_{m+1} and increase the size of S by at most 1. If y_{m+1} differs with $y_1, y_2, \ldots, y_m$ on S, then we do not need to add any more indices to S and we are done. If y_{m+1} agrees with one of $y_1, y_2, \ldots, y_m$ on all of S, first note that it can only agree with one such string; to differentiate between these two, we add any index at which they differ to S, which must exist since they are distinct. ◀

This shows that a function can be hard—i.e., can have query complexity $\Theta(\sqrt{N})$—only if $|f|$ is $O(1)$ or $\Theta(N)$.

Note that there do exist hard functions with $|f| = \Theta(N)$. For example, consider the following function: $f(x) = 1$ if the first bit of x is 1 or if x is the all-zero string. This essentially embeds a delta function on the last $n-1$ bits, and thus requires $\Theta(\sqrt{N})$ queries. This function has $|f| = N/2 + 1$. However, there are also easy functions with $|f| = \Theta(N)$, namely the bent functions. Thus the Hamming weight does not completely characterize the hardness of the BHSP at high Hamming weight. However, it precisely characterizes the quantum query complexity at low Hamming weight:

▶ **Theorem 12.** *For any $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ with no undetectable shifts, the bounded-error quantum query complexity of* BHSP_f *is* $\Omega(\sqrt{N/|f|})$.

This follows from a simple application of the quantum adversary argument, with the adversary matrix taken to be the all ones matrix with zeroes on the diagonal. It also follows from Theorem 4 of [45].

4 The PGM approach

We now present an approach to the Boolean hidden shift problem based on the pretty good measurement (PGM) [47]. In particular, this approach shows that the Boolean hidden shift problem for random functions has small query complexity (see Sect. 4.3.3).

The main idea of the PGM approach is as follows. We apply the oracle on the uniform superposition and prepare t independent copies of the resulting state (see Sect. 4.1). Then we use knowledge of the function f to perform the PGM in order to extract the hidden shift s (see Sect. 4.2). A similar strategy was used to efficiently solve the hidden subgroup problem for certain semidirect product groups, including the Heisenberg group [48], and was subsequently applied to a hidden polynomial problem [49].

4.1 Performing t queries in parallel

In this section we describe a quantum circuit that prepares a state with $w \cdot s$ encoded in the phase, where s is the hidden shift and w is the label of the corresponding standard basis

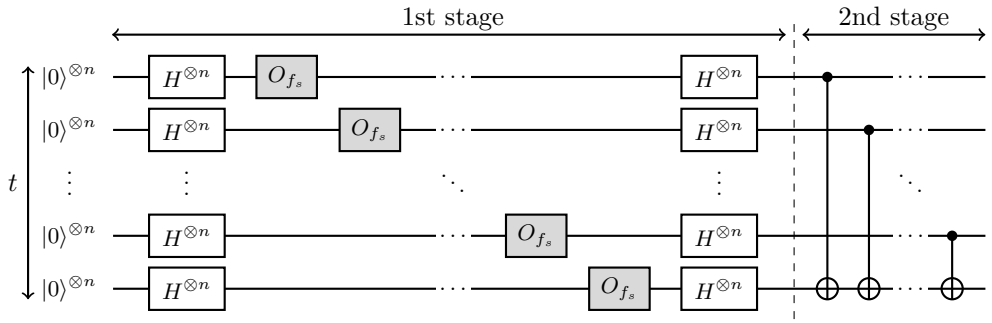


Figure 1 Quantum algorithm for preparing the t-fold Fourier sate $|\Phi^t(s)\rangle$ in Eq. (8). The state on any register at the end of the first stage is given in Eq. (4).

vector. We use this circuit t times in parallel, followed by a sequence of CNOTs, to prepare a certain state $|\Phi^t(s)\rangle$. In the next section we perform a PGM on these states for different values of s.

4.1.1 Circuit

The circuit for preparing $|\Phi^t(s)\rangle$ appears in Fig. 1. It consists of two stages. The first stage prepares t identical copies of the same state by using one oracle call between two quantum Fourier transforms on each register independently. Recall from Definition 1 that the oracle acts on n qubits and encodes the function in the phase: $O_{f_s}\colon |x\rangle \mapsto (-1)^{f(x+s)}|x\rangle$. The second stage entangles the states by applying a sequence of transversal controlled-NOT gates acting as $|x\rangle|y\rangle \mapsto |x\rangle|y+x\rangle$ for $x, y \in \mathbb{Z}_2^n$.

Note that all unitary post-processing after the oracle queries can be omitted since it does not affect the distinguishability of the states. We include it only to simplify the analysis.

4.1.2 Analysis

During the first stage of the circuit, the first register evolves under $H^{\otimes n} O_{f_s} H^{\otimes n}$ (see Fig. 1):

$$|0\rangle^{\otimes n} \mapsto \frac{1}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^n} |x\rangle \mapsto \frac{1}{\sqrt{2^n}} \sum_{x\in\mathbb{Z}_2^n} (-1)^{f(x+s)}|x\rangle \mapsto \frac{1}{2^n} \sum_{x,y\in\mathbb{Z}_2^n} (-1)^{f(x+s)+x\cdot y}|y\rangle. \tag{3}$$

We can rewrite the resulting state as follows:

$$\sum_{y\in\mathbb{Z}_2^n} (-1)^{s\cdot y} \left(\frac{1}{2^n} \sum_{x\in\mathbb{Z}_2^n} (-1)^{f(x)+x\cdot y} \right) |y\rangle = \sum_{y\in\mathbb{Z}_2^n} (-1)^{s\cdot y} \hat{F}(y)|y\rangle. \tag{4}$$

The overall state after the first stage is just the t-fold tensor product of the above state:

$$\sum_{y_1,\ldots,y_t\in\mathbb{Z}_2^n} (-1)^{s\cdot(y_1+\cdots+y_t)} \bigotimes_{i=1}^{t} \hat{F}(y_i)|y_i\rangle. \tag{5}$$

In the second stage of the algorithm, the controlled-NOT gates transform this state into

$$\sum_{y_1,\ldots,y_t\in\mathbb{Z}_2^n} (-1)^{s\cdot(y_1+\cdots+y_t)} \left[\bigotimes_{i=1}^{t-1} \hat{F}(y_i)|y_i\rangle\right] \hat{F}(y_t)|y_1+\cdots+y_t\rangle \tag{6}$$

$$= \sum_{y_1,\ldots,y_t\in\mathbb{Z}_2^n} (-1)^{s\cdot y_t} \left[\bigotimes_{i=1}^{t-1} \hat{F}(y_i)|y_i\rangle\right] \hat{F}\big(y_t-(y_1+\cdots+y_{t-1})\big)|y_t\rangle. \tag{7}$$

We can rewrite this state as

$$|\Phi^t(s)\rangle := \sum_{w\in\mathbb{Z}_2^n} (-1)^{s\cdot w}|\mathcal{F}_w^t\rangle|w\rangle, \tag{8}$$

where the non-normalized state $|\mathcal{F}_w^t\rangle$ on $(t-1)n$ qubits is given by

$$|\mathcal{F}_w^t\rangle := \sum_{y_1,\ldots,y_{t-1}\in\mathbb{Z}_2^n} \hat{F}(y_1)\cdots\hat{F}(y_{t-1})\hat{F}\big(w-(y_1+\cdots+y_{t-1})\big)|y_1\rangle\cdots|y_{t-1}\rangle. \tag{9}$$

Its norm is just the t-fold Fourier coefficient: $\||\mathcal{F}_w^t\rangle\| = \mathcal{F}^t(w)$ (see Definition 6).

4.2 The pretty good measurement

Let $\{\rho_s^{(t)}\colon s\in\mathbb{Z}_2^n\}$ be a set of mixed states where $\rho_s^{(t)}$ is given with probability p_s. The *pretty good measurement* (PGM) [47] for discriminating these states is a POVM with operators $\{E_s\colon s\in\mathbb{Z}_2^n\}\cup\{E_*\}$ where

$$E_s := E^{-1/2}\, p_s\rho_s^{(t)}\, E^{-1/2}, \qquad E := \sum_{s\in\mathbb{Z}_2^n} p_s\rho_s^{(t)}, \qquad E_* := I - \sum_{s\in\mathbb{Z}_2^n} E_s. \tag{10}$$

In our case, $\rho_s^{(t)} := |\Phi^t(s)\rangle\langle\Phi^t(s)|$ and $p_s := 1/2^n$ where $|\Phi^t(s)\rangle$ is defined in Eq. (8).

To find the operators E_s, we compute

$$E = \sum_{s\in\mathbb{Z}_2^n} \frac{1}{2^n} \sum_{w,w'\in\mathbb{Z}_2^n} (-1)^{(w+w')\cdot s}|\mathcal{F}_w^t\rangle\langle\mathcal{F}_{w'}^t| \otimes |w\rangle\langle w'| \tag{11}$$

$$= \sum_{w\in\mathbb{Z}_2^n} \||\mathcal{F}_w^t\rangle\|^2 \cdot \frac{|\mathcal{F}_w^t\rangle\langle\mathcal{F}_w^t|}{\||\mathcal{F}_w^t\rangle\|^2} \otimes |w\rangle\langle w|. \tag{12}$$

From now on we use the convention that terms with $\||\mathcal{F}_w^t\rangle\| = 0$ are omitted from all sums. As E is a sum of mutually orthogonal rank-1 operators with eigenvalues $\||\mathcal{F}_w^t\rangle\|^2$, we find

$$E^{-1/2} = \sum_{w\in\mathbb{Z}_2^n} \frac{1}{\||\mathcal{F}_w^t\rangle\|} \cdot \frac{|\mathcal{F}_w^t\rangle\langle\mathcal{F}_w^t|}{\||\mathcal{F}_w^t\rangle\|^2} \otimes |w\rangle\langle w|. \tag{13}$$

Note that $E_s = |E_s\rangle\langle E_s|$ where $|E_s\rangle := E^{-1/2}\sqrt{p_s}|\Phi^t(s)\rangle$. We can express $|E_s\rangle$ as follows:

$$|E_s\rangle = \left(\sum_{w\in\mathbb{Z}_2^n} \frac{|\mathcal{F}_w^t\rangle\langle\mathcal{F}_w^t|}{\||\mathcal{F}_w^t\rangle\|^3} \otimes |w\rangle\langle w|\right) \frac{1}{\sqrt{2^n}} \left(\sum_{w\in\mathbb{Z}_2^n} (-1)^{w\cdot s}|\mathcal{F}_w^t\rangle|w\rangle\right) \tag{14}$$

$$= \frac{1}{\sqrt{2^n}} \sum_{w\in\mathbb{Z}_2^n} (-1)^{w\cdot s} \frac{|\mathcal{F}_w^t\rangle}{\||\mathcal{F}_w^t\rangle\|} \otimes |w\rangle. \tag{15}$$

Notice that the vectors $|E_s\rangle$ are orthonormal, so the PGM is just an orthogonal measurement in this basis (with another outcome corresponding to the orthogonal complement). Therefore the measurement is unambiguous: if it outputs a value of s (rather than the inconclusive outcome $*$) then it is definitely correct. The corresponding zero-error algorithm can be summarized as follows:

PGM(f,t)
1. Prepare $|\Phi^t(s)\rangle$ using the circuit shown in Fig. 1.
2. Recover s by performing an orthogonal measurement with projectors $\{|E_s\rangle\langle E_s| : s \in \mathbb{Z}_2^n\} \cup \{E_*\}$.

▶ **Lemma 13.** *The t-query algorithm* **PGM**(f,t) *solves* BHSP$_f$ *with success probability*

$$p_f(t) := \left(\frac{1}{\sqrt{2^n}} \sum_{w\in\mathbb{Z}_2^n} \mathcal{F}^t(w)\right)^2, \tag{16}$$

where $\mathcal{F}^t(w) = \||\mathcal{F}_w^t\rangle\|$ denotes the t-fold Fourier spectrum of $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ (see Definition 6).

Proof. Recall that the PGM for discriminating the states $|\Phi^t(s)\rangle = \sum_{w\in\mathbb{Z}_2^n}(-1)^{s\cdot w}|\mathcal{F}_w^t\rangle|w\rangle$ from Eq. (8) is an orthogonal measurement on $|E_s\rangle$ (defined in Eq. (15)) and the orthogonal complement. Thus, given the state $|\Phi^t(s)\rangle$, the success probability to recover the hidden shift s correctly is $\left|\langle E_s|\Phi^t(s)\rangle\right|^2$. This is equal to the expression in Eq. (16). Moreover, it does not depend on s, so $p_f(t)$ is the success probability even if s is chosen adversarially as in the definition of BHSP$_f$ (Problem 3). Note that the convention of omitting terms with $\||\mathcal{F}_w^t\rangle\| = 0$ is consistent since such terms do not appear in Eq. (16). ◀

We can use Eq. (2) to write the success probability as

$$p_f(t) = \frac{1}{2^n}\left(\sum_{w\in\mathbb{Z}_2^n} \sqrt{\frac{1}{\sqrt{2^n}}\,\widehat{(F*F)^t}\,(w)}\right)^2. \tag{17}$$

Recall from Sect. 2 that $\mathcal{F}^1(w) = |\hat{F}(w)|$, so for $t = 1$ we have

$$p_f(1) = \frac{1}{2^n}\left(\sum_{w\in\mathbb{Z}_2^n} |\hat{F}(w)|\right)^2. \tag{18}$$

4.3 Performance analysis

In this section we analyze the performance of the PGM algorithm described above on several different classes of Boolean functions. For delta functions our algorithm performs worse than Grover's algorithm. On the other hand, for bent and random functions it needs only one and two queries, respectively.

4.3.1 Delta functions

Let us check how our algorithm performs when f is a delta function, i.e., $f(x) = \delta_{x,x_0}$ for some $x_0 \in \mathbb{Z}_2^n$. A simple calculation using the Fourier spectrum of a delta function shows that the success probability of **PGM**(f,t) is

$$p_f(t) = \frac{1}{2^{2n}}\left((2^n-1)\sqrt{1-\left(\frac{2^n-4}{2^n}\right)^t} + \sqrt{1+(2^n-1)\left(\frac{2^n-4}{2^n}\right)^t}\right)^2. \tag{19}$$

Unfortunately, if we choose $t = \sqrt{2^n}$, then the success probability goes to 0 as $n \to \infty$. In fact, the same happens even if $t = c^n$ for any $c < 2$. Only if we take $t = 2^n$ does the success probability approach a positive constant $1 - 1/e^4 \approx 0.98$ as $n \to \infty$. This means that the PGM algorithm does not give us the quadratic speedup of Grover's algorithm. (Indeed, this follows from the more general fact that quantum speedup for unstructured search cannot be parallelized [50].) Thus the PGM algorithm is not optimal in general.

4.3.2 Bent functions

Let f be a Bent function. Recall from Sect. 3.1 that its Fourier spectrum is "flat", i.e., $|\hat{F}(w)| = 1/\sqrt{2^n}$ for all $w \in \mathbb{Z}_2^n$. In this case, Eq. (18) gives $p_f(1) = 1$, so we can find the hidden shift with certainty by measuring $|\Phi^1(s)\rangle$ with the pretty good measurement (recall that preparing $|\Phi^1(s)\rangle$ requires only one query to O_{f_s}), reproducing a result of Rötteler.

▶ **Theorem 14** ([30]). *If f is a bent function then a quantum algorithm can solve* BHSP_f *exactly using a single query to O_{f_s}.*

4.3.3 Random functions

For random Boolean functions, our algorithm performs almost as well as for bent functions. For random f, we are only able to show that the expected success probability of the one-query algorithm $\mathbf{PGM}(f, 1)$ is at least $2/\pi + o(1)$ for large n (see Theorem 19 in Appendix B), so the algorithm only succeeds with constant probability, which cannot easily be boosted. However, the expected success probability of the two-query algorithm $\mathbf{PGM}(f, 2)$ is exponentially close to 1.

▶ **Theorem 15.** *Let f be an n-argument Boolean function chosen uniformly at random and suppose that a hidden shift for f is chosen adversarially. Then* $\mathbf{PGM}(f, 2)$ *solves* BHSP_f *with expected success probability $\bar{p} \geq 1 - \frac{3}{64} \cdot 2^{-n}$.*

The proof uses the second moment method to lower bound the expected success probability. We compute the variance of the 2-fold Fourier spectrum by relating it to the combinatorics of pairings. The proof appears in Appendix C.

Theorem 15 implies that our algorithm can determine the hidden shift with near certainty as $n \to \infty$. This is surprising since some functions, such as delta functions (see Sect. 3.2), require $\Omega(\sqrt{2^n})$ queries. Furthermore, a randomly chosen function could have an undetectable shift (see Sect. D.1), in which case it is not possible in principle to completely determine an adversarially chosen shift with success probability more than 1/2.

At first glance, Theorem 15 may appear to be a strengthening of the main result of [31], which shows that $O(n)$ queries suffice to solve a version of the Boolean hidden shift problem for a random function. However, while our approach uses dramatically fewer queries, the results are not directly comparable: Ref. [31] considers a weaker model in which the unshifted function is given by an oracle rather than being known explicitly. In particular, while the result of [31] gives an average-case exponential separation between classical and quantum query complexity, such a result is not possible in the model where the function is known explicitly. In this model, there cannot be a super-polynomial speedup for quantum computation. This follows from general results from learning theory discussed at the end of Sect. 1. In particular, it follows that if the quantum query complexity of the problem for a function f is Q, then the deterministic classical query complexity of the problem for the same function is at most $O(nQ^2)$ [37].

5 Quantum rejection sampling with parallel queries

In this section we explain a hybrid approach that combines the Quantum Rejection Sampling (QRS) algorithm for the BHSP [28] with the PGM approach. The resulting algorithm does not require an extra amplification step for boosting the success probability, unlike the original QRS algorithm.

5.1 Original quantum rejection sampling approach

▶ **Theorem 16** ([28]). *For a given Boolean function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$, define unit vectors $\boldsymbol{\pi}, \boldsymbol{\sigma} \in \mathbb{R}^{2^n}$ as $\pi_w := |\hat{F}(w)|$ and $\sigma_w := 1/\sqrt{2^n}$ for $w \in \mathbb{Z}_2^n$. Moreover, let*

$$p_{\min} := (\boldsymbol{\sigma}^{\mathsf{T}} \cdot \boldsymbol{\pi})^2 = \frac{1}{2^n}\Bigg(\sum_{w\in\mathbb{Z}_2^n} |\hat{F}(w)|\Bigg)^2, \quad p_{\max} := \sum_{k\colon \pi_k>0} \sigma_k^2 = \frac{1}{2^n}|\{w\colon \hat{F}(w)\neq 0\}|. \tag{20}$$

For any desired success probability $p \in [p_{\min}, p_{\max}]$, the quantum rejection sampling algorithm solves BHSP_f *with $O(1/\|\boldsymbol{\varepsilon}^p_{\boldsymbol{\pi}\to\boldsymbol{\sigma}}\|)$ queries, where the "water-filling" vector $\boldsymbol{\varepsilon}^p_{\boldsymbol{\pi}\to\boldsymbol{\sigma}} \in \mathbb{R}^{2^n}$ is defined in [28].*

In particular, if $p_{\max} = 1$ then the QRS algorithm can achieve any success probability arbitrarily close to 1 with $O\big(1/(\sqrt{2^n}\hat{F}_{\min})\big)$ queries, where $\hat{F}_{\min} := \min_w |\hat{F}(w)|$. However, if $\hat{F}(w) = 0$ for some w, then from Eq. (20) we see that $p_{\max} < 1$. In this case one needs an additional amplification step to boost the success probability (a method based on SWAP test was proposed in [28]). We show that this step can be avoided by using t parallel queries in the original QRS algorithm for some $t \leq n$.

5.2 Non-degenerate functions with almost vanishing spectrum

Before explaining our hybrid approach, let us verify that there exist non-trivial functions with a large fraction of their Fourier spectrum equal to zero, so the issue discussed above applies.

It is easy to construct degenerate functions with the desired property. For example, if a function is shift-invariant, i.e., $f(x+s) = f(x)$ for some $s \in \mathbb{Z}_2^n$, then at least half of the Fourier spectrum of f is guaranteed to be zero. The same also happens if $f(x+s) = f(x)+1$ (see Lemma 24 in Sect. D.1). However, such examples are not interesting, since a shift-invariant n-argument Boolean function is equivalent to an $(n-1)$-argument Boolean function (see Sect. D.1 for more details).

Instead, we consider Boolean functions defined using decision trees. A *decision tree* is a binary tree whose vertices are labeled by arguments of f and whose leaves contain the values of f. An example of such tree and the rules for evaluating the corresponding function are given in Fig. 2.

Without loss of generality, we can consider only decision trees where on each path from the root to a leaf no argument appears more than once (otherwise some parts of the tree would not be reachable). The length of a longest path from the root to a leaf is the *height* of the tree. If a Boolean function is defined by a decision tree of height h, then all its Fourier coefficients with Hamming weight larger than h are zero (see Lemma 25 in Sect. D.2). This observation can be used to construct non-degenerate Boolean functions with almost vanishing Fourier spectrum.

▶ **Example.** The 10-argument Boolean function f_{10} whose decision tree is shown in Fig. 2 has no shift invariance, yet 928 (out of $2^{10} = 1024$) of its Fourier coefficients are zero.

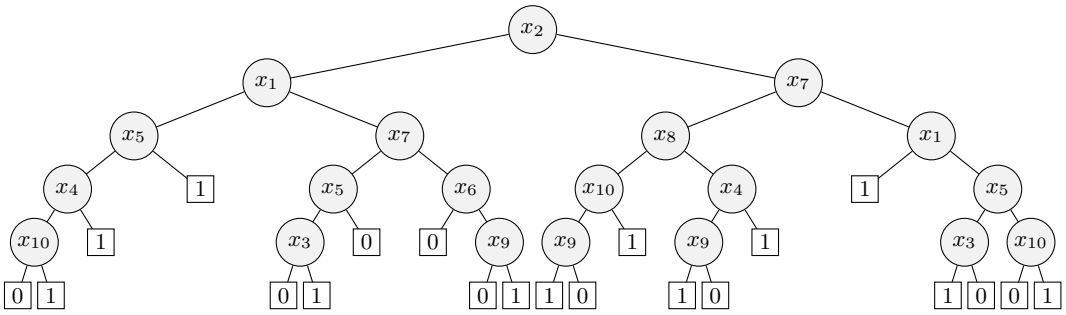


Figure 2 Decision tree for a 10-argument Boolean function f_{10}. To compute the value of the function for given input $x_1, \ldots, x_{10} \in \mathbb{Z}_2^n$, proceed down the tree starting from the root; move left if the corresponding argument is equal to 0 or right if it is equal to 1. Once a leaf is reached, its label is the value of the function for the given input. For example, $f_{10}(x_1, \ldots, x_{10})$ evaluates to zero when $x_2 = x_1 = x_5 = x_4 = x_{10} = 0$, since the leftmost leaf has label zero. This tree has height five.

5.3 The t-fold Fourier spectrum as t increases

Let us now show how to deal with the zero Fourier coefficients. The main idea stems from the following observation: if $S_t := \{w \in \mathbb{Z}_2^n : \mathcal{F}^t(w) \neq 0\}$ then $S_{t+1} = S_t + S_1$ (see Prop. 26 in Sect. D.3). If S_1 spans $\mathbb{Z}_2^n$, we can apply this recursively and eliminate all zeroes from the t-fold Fourier spectrum $\mathcal{F}^t$. In particular, it suffices to take $t \leq n$ (see Lemma 27 in Sect. D.3). For example, for f_{10} the fraction of non-zero values of $\mathcal{F}^t$ for $t = 1, 2, 3, 4$ is 0.09, 0.61, 0.94, 1, respectively. In particular, $\mathcal{F}^4$ is non-zero everywhere.

5.4 Quantum rejection sampling with t-fold queries

We can use quantum rejection sampling with t queries in parallel to solve the BHSP. Suppose we transform the t-fold Fourier state $|\Phi^t(s)\rangle$ from Eq. (8) into the PGM basis vector $|E_s\rangle$ defined in Eq. (15) using QRS. This corresponds to setting $\pi_w = \mathcal{F}^t(w)$ and $\sigma_w = 1/\sqrt{2^n}$. Since the circuit from Fig. 1 can be used to prepare $|\Phi^t(s)\rangle$ with t queries, Theorem 16 still holds if $|\hat{F}(w)|$ is replaced by $\mathcal{F}^t(w)$ and the query complexity is multiplied by t. This observation together with Lemma 27 implies that as long as f is not shift invariant, we can recover the hidden shift s with success probability arbitrarily close to 1 using quantum rejection sampling with some $t \leq n$.

▶ **Theorem 17.** *Let $f : \mathbb{Z}_2^n \to \mathbb{Z}_2$ be a Boolean function and let p be sufficiently large. Then BHSP_f can be solved with success probability p using $O(t/\|\boldsymbol{\varepsilon}^p_{\boldsymbol{\pi}\to\boldsymbol{\sigma}}\|)$ queries for some $t \in \{1, \ldots, n\}$ where $\pi_w := \mathcal{F}^t(w)$, $\sigma_w := 1/\sqrt{2^n}$, and the "water-filling" vector $\boldsymbol{\varepsilon}^p_{\boldsymbol{\pi}\to\boldsymbol{\sigma}} \in \mathbb{R}^{2^n}$ is defined in [28].*

6 Conclusions

A comparison of quantum query complexity bounds for solving the BHSP for different classes of functions is given in Table 1. If the QRS algorithm works for random functions with $O(1)$ queries, then it is optimal up to constant factors in all three cases listed in the table. However, from Sect. 5.1 we know that the basic QRS algorithm without amplification performs poorly when f has many zero Fourier coefficients (which is the case, e.g., for the decision trees considered in Sect. D.2). This suggests that the basic (unamplified) QRS algorithm is likely not optimal in general.

Table 1 Summary of quantum query complexity upper and lower bounds for BHSP. We do not know the query complexity of the QRS algorithm for random functions.

Approach	Functions			Comments
	delta	bent	random	
PGM	$O(2^n)$	1	2	zero error
QRS [28]	$O(\sqrt{2^n})$	1	?	
"Simon" [31]	$O(n\sqrt{2^n})$	$O(n)$	$O(n)$	zero error, black-box f okay
Learning theory [38]	$O(n \log n\sqrt{2^n})$	$O(n \log n)$	$O(n \log n)$	optimal up to log factors $\forall f$
Lower bounds:	$\Omega(\sqrt{2^n})$	1	1	

The "Simon"-type approach due to [31] always has an overhead of a factor $O(n)$, reflecting the fact that at least n linearly independent equations are needed to solve a linear system in n variables. (Note that this approach works in the weaker model where the unshifted function is given by an oracle, so it still provides an upper bound when the function is known explicitly.) The learning theory approach [38] also has logarithmic overhead. Finally, the PGM approach performs very well in the easy cases, the bent and random functions, but fails to provide any speedup for delta functions. As mentioned in Sect. 4.3.1, this can be attributed to the fact that Grover's algorithm is intrinsically sequential.

In summary, none of the algorithms listed in Table 1 is optimal. However, by combining these algorithms and possibly adding some new ideas, one might obtain an algorithm that is optimal for all Boolean functions. In particular, the QRS approach with t-fold queries appears promising.

We conclude by mentioning some open questions regarding the Boolean hidden shift problem:

1. Find a query-optimal quantum algorithm for general functions (recall that the learning theory algorithm is only optimal up to logarithmic factors [37, 38]).
2. Identify natural classes of Boolean functions lying between the two extreme cases of bent and delta functions (say, the decision trees considered in Sect. D.2) and characterize the quantum query complexity of the BHSP for these functions.
3. Determine the number of queries required by the QRS algorithm for random functions.
4. What is the query complexity of verifying a given shift? (A quantum procedure with one-sided error, based on the swap test, was given in [28].)
5. What is the quantum query complexity of extracting one bit of information about the hidden shift?
6. What is the classical query complexity of the Boolean hidden shift problem?
7. Can we say anything non-trivial about the time complexity of the Boolean hidden shift problem, either classically or quantumly?
8. Can the BHSP for random functions be solved with a single query? Our approach based on the PGM only gives a lower bound on the expected success probability that approaches $2/\pi$ for large n (see Theorem 19), whereas we require a success probability that approaches 1 as $n \to \infty$. It might be fruitful to consider querying the oracle with non-uniform amplitudes.

Finally, it might be interesting to consider the generalization of the Boolean hidden shift problem to the case of functions $f\colon \mathbb{Z}_d^n \to \mathbb{Z}_d$.

Acknowledgements. We thank Jérémie Roland for useful discussions and Dmitry Gavinsky for suggesting to use decision trees to construct non-degenerate functions with many zero Fourier coefficients. Part of this work was done while AC and MO were visiting NEC Labs, and during the Quantum Cryptanalysis seminar (No. 11381) at Schloss Dagstuhl. This work was supported in part by NSERC, the Ontario Ministry of Research and Innovation, and the US ARO/DTO. MO acknowledges additional support from the DARPA QUEST program under contract number HR0011-09-C-0047.

References

1 Andrew M. Childs and Wim van Dam. Quantum algorithms for algebraic problems. *Rev. Mod. Phys.*, 82(1):1–52, Jan 2010. arXiv:0812.0380, doi:10.1103/RevModPhys.82.1.

2 David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London. Series A: Mathematical and Physical Sciences*, 439(1907):553–558, 1992. doi:10.1098/rspa.1992.0167.

3 Daniel R. Simon. On the power of quantum computation. In *Proceedings of the 35th Annual Symposium on Foundations of Computer Science (FOCS 1994)*, pages 116–123, Nov 1994. doi:10.1109/SFCS.1994.365701.

4 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. Earlier version in FOCS 1994, pp. 124–134. arXiv:quant-ph/9508027, doi:10.1137/S0097539795293172.

5 Alexei Kitaev. Quantum measurements and the Abelian Stabilizer Problem. 1995. arXiv:quant-ph/9511026.

6 Richard Jozsa. Quantum algorithms and the Fourier transform. *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences*, 454(1969):323–337, 1998. arXiv:quant-ph/9707033, doi:10.1098/rspa.1998.0163.

7 Michele Mosca and Artur Ekert. The hidden subgroup problem and eigenvalue estimation on a quantum computer. In *Quantum Computing and Quantum Communications*, volume 1509 of *Lecture Notes in Computer Science*, pages 174–188. Springer, 1999. arXiv:quant-ph/9903071, doi:10.1007/3-540-49208-9_15.

8 Richard Jozsa. Quantum factoring, discrete logarithms, and the hidden subgroup problem. *Computing in Science Engineering*, 3(2):34–43, Mar/Apr 2001. arXiv:quant-ph/0012084, doi:10.1109/5992.909000.

9 Sean Hallgren. Polynomial-time quantum algorithms for Pell's equation and the principal ideal problem. *Journal of the ACM*, 54(1):4:1–4:19, Mar 2007. doi:10.1145/1206035.1206039.

10 Dan Boneh and Richard Lipton. Quantum cryptanalysis of hidden linear functions. In *Advances in Cryptology – CRYPTO 1995*, volume 963 of *Lecture Notes in Computer Science*, pages 424–437. Springer, 1995. doi:10.1007/3-540-44750-4_34.

11 Robert Beals. Quantum computation of Fourier transforms over symmetric groups. In *Proceedings of the 29th Annual ACM Symposium on Theory of Computing (STOC 1997)*, pages 48–53. ACM, 1997. doi:10.1145/258533.258548.

12 Peter Høyer. Efficient quantum transforms. 1997. arXiv:quant-ph/9702028.

13 Mark Ettinger and Peter Høyer. A quantum observable for the graph isomorphism problem. 1999. arXiv:quant-ph/9901029.

14 Andris Ambainis, Loïck Magnin, Martin Roetteler, and Jérémie Roland. Symmetry-assisted adversaries for quantum state generation. In *Proceedings of the 26th Annual IEEE Conference on Computational Complexity (CCC'11)*, pages 167–177. IEEE Computer Society, 2011. arXiv:1012.2112, doi:10.1109/CCC.2011.24.

15 Oded Regev. Quantum computation and lattice problems. *SIAM Journal on Computing*, 33(3):738–760, 2004. arXiv:cs/0304005, doi:10.1137/S0097539703440678.

16 Greg Kuperberg. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem. *SIAM Journal on Computing*, 35(1):170–188, 2005. arXiv:quant-ph/0302112, doi:10.1137/S0097539703436345.

17 Oded Regev. A subexponential time algorithm for the dihedral hidden subgroup problem with polynomial space. 2004. arXiv:quant-ph/0406151.

18 Greg Kuperberg. Another subexponential-time quantum algorithm for the dihedral hidden subgroup problem. 2011. arXiv:1112.3333.

19 Andrew M. Childs, David Jao, and Vladimir Soukharev. Constructing elliptic curve isogenies in quantum subexponential time. 2010. arXiv:1012.4019.

20 Mark Ettinger and Peter Høyer. On quantum algorithms for noncommutative hidden subgroups. *Advances in Applied Mathematics*, 25(3):239–251, 2000. arXiv:quant-ph/9807029, doi:10.1006/aama.2000.0699.

21 Wim van Dam, Sean Hallgren, and Lawrence Ip. Quantum algorithms for some hidden shift problems. *SIAM Journal on Computing*, 36(3):763–778, 2006. arXiv:quant-ph/0211140, doi:10.1137/S009753970343141X.

22 Katalin Friedl, Gábor Ivanyos, Frédéric Magniez, Miklos Santha, and Pranab Sen. Hidden translation and orbit coset in quantum computing. In *Proceedings of the 35th Annual ACM Symposium on Theory of Computing (STOC 2003)*, pages 1–9. ACM, 2002. arXiv:quant-ph/0211091, doi:10.1145/780542.780544.

23 Cristopher Moore, Daniel Rockmore, Alexander Russell, and Leonard J. Schulman. The power of strong Fourier sampling: Quantum algorithms for affine groups and hidden shifts. *SIAM Journal on Computing*, 37(3):938–958, Jun 2007. arXiv:quant-ph/0503095, doi:10.1137/S0097539705447177.

24 Andrew M. Childs and Pawel Wocjan. On the quantum hardness of solving isomorphism problems as nonabelian hidden shift problems. *Quantum Information and Computation*, 7(5):504–521, Jul 2007. URL: http://www.rintonpress.com/journals/qiconline.html#v7n56, arXiv:quant-ph/0510185.

25 Andrew M. Childs and Wim van Dam. Quantum algorithm for a generalized hidden shift problem. In *Proceedings of the 18th ACM-SIAM Symposium on Discrete Algorithms (SODA 2007)*, pages 1225–1232. SIAM, 2007. URL: http://dl.acm.org/citation.cfm?id=1283383.1283515, arXiv:quant-ph/0507190.

26 Gábor Ivanyos. On solving systems of random linear disequations. *Quantum Information and Computation*, 8(6&7):579–594, 2008. URL: http://www.rintonpress.com/journals/qiconline.html#v8n67, arXiv:0704.2988.

27 Ivan B. Damgård. On the randomness of Legendre and Jacobi sequences. In *Advances in Cryptology – CRYPTO 1988*, volume 403 of *Lecture Notes in Computer Science*, pages 163–172. Springer, 1990. doi:10.1007/0-387-34799-2_13.

28 Maris Ozols, Martin Roetteler, and Jérémie Roland. Quantum rejection sampling. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference (ITCS 2012)*, pages 290–308. ACM, 2012. arXiv:1103.2774, doi:10.1145/2090236.2090261.

29 Martin Rötteler. Quantum algorithms to solve the hidden shift problem for quadratics and for functions of large Gowers norm. In *Proceedings of the 34st International Symposium on Mathematical Foundations of Computer Science (MFCS 2009)*, volume 5734 of *Lecture Notes in Computer Science*, pages 663–674. Springer, 2009. arXiv:0911.4724, doi:10.1007/978-3-642-03816-7_56.

30 Martin Rötteler. Quantum algorithms for highly non-linear Boolean functions. In *Proceedings of the 21st ACM-SIAM Symposium on Discrete Algorithms (SODA 2010)*, pages

448–457. SIAM, 2010. URL: http://dl.acm.org/citation.cfm?id=1873601.1873638, arXiv:0811.3208.

31 Dmitry Gavinsky, Martin Roetteler, and Jérémie Roland. Quantum algorithm for the Boolean hidden shift problem. In *Computing and Combinatorics*, volume 6842 of *Lecture Notes in Computer Science*, pages 158–167. Springer, 2011. arXiv:1103.3017, doi:10.1007/978-3-642-22685-4_14.

32 Mirmojtaba Gharibi. Reduction from non-injective hidden shift problem to injective hidden shift problem. *Quantum Information and Computation*, 13(3&4):0221–0230, 2013. URL: http://www.rintonpress.com/journals/qiconline.html#v13n34, arXiv:1207.4537.

33 Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the 28th Annual ACM Symposium on Theory of Computing (STOC 1996)*, pages 212–219. ACM, 1996. arXiv:quant-ph/9605043, doi:10.1145/237814.237866.

34 Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997. arXiv:quant-ph/9701001, doi:10.1137/S0097539796300933.

35 Wim van Dam. Quantum algorithms for weighing matrices and quadratic residues. *Algorithmica*, 34(4):413–428, 2008. arXiv:quant-ph/0008059, doi:10.1007/s00453-002-0975-4.

36 Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. Earlier version in STOC 1993, pp. 11–20. doi:10.1137/S0097539796300921.

37 Rocco A. Servedio and Steven J. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM Journal on Computing*, 33(5):1067–1092, 2004. doi:10.1137/S0097539704412910.

38 Alp Atıcı and Rocco A. Servedio. Improved bounds on quantum learning algorithms. *Quantum Information Processing*, 4(5):355–386, 2005. arXiv:quant-ph/0411140, doi:10.1007/s11128-005-0001-2.

39 Ronald de Wolf. A brief introduction to Fourier analysis on the Boolean cube. *Theory of Computing Library – Graduate Surveys*, 1:1–20, 2008. doi:10.4086/toc.gs.2008.001.

40 Thomas W. Cusick and Pantelimon Stănică. *Cryptographic Boolean Functions and Applications*. Academic Press/Elsevier, 2009. URL: http://books.google.ca/books?id=OAkhkLSxxxMC&pg=PA73.

41 John F. Dillon. A survey of bent functions. *The NSA technical journal*, pages 191–215, 1972.

42 Jessie F. MacWilliams and Neil J.A. Sloane. *The theory of error-correcting codes: Part 2*. North-Holland, 1977. URL: http://books.google.ca/books?id=nv6WCJgcjxcC&pg=PA426.

43 John F. Dillon. Elementary Hadamard difference sets. In *Proceedings of the 6th Southeastern Conference on Combinatorics, Graph Theory, and Computing*, pages 237–249. Utilitas Mathematica Pub., 1975.

44 Hans Dobbertin. Construction of bent functions and balanced Boolean functions with high nonlinearity. In *Fast Software Encryption*, volume 1008 of *Lecture Notes in Computer Science*, pages 61–74. Springer, 1995. doi:10.1007/3-540-60590-8_5.

45 Andris Ambainis, Kazuo Iwama, Akinori Kawachi, Hiroyuki Masuda, Raymond H. Putra, and Shigeru Yamashita. Quantum identification of Boolean oracles. In *Proceedings of the 21st Annual Symposium on Theoretical Aspects of Computer Science (STACS 2004)*, volume 2996 of *Lecture Notes in Computer Science*, pages 105–116. Springer, 2004. arXiv:quant-ph/0403056, doi:10.1007/978-3-540-24749-4_10.

46 Michel Boyer, Gilles Brassard, Peter Høyer, and Alain Tapp. Tight bounds on quantum searching. *Fortschritte der Physik*, 46(4-5):493–505, 1998. arXiv:quant-ph/9605034, doi:10.1002/(SICI)1521-3978(199806)46:4/5<493::AID-PROP493>3.0.CO;2-P.

47 Paul Hausladen and William K. Wootters. A 'pretty good' measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994. doi:10.1080/09500349414552221.

48 Dave Bacon, Andrew M. Childs, and Wim van Dam. From optimal measurement to efficient quantum algorithms for the hidden subgroup problem over semidirect product groups. In *Proceedings of the 46th Annual Symposium on Foundations of Computer Science (FOCS 2005)*, pages 469–478, Oct 2005. arXiv:quant-ph/0504083, doi:10.1109/SFCS.2005.38.

49 Thomas Decker, Jan Draisma, and Pawel Wocjan. Efficient quantum algorithm for identifying hidden polynomials. *Quantum Information and Computation*, 9(3-4):215–254, 2009. URL: http://www.rintonpress.com/journals/qiconline.html#v9n34, arXiv:0706.1219.

50 Christof Zalka. Grover's quantum searching algorithm is optimal. *Physical Review A*, 60:2746–2751, 1999. arXiv:quant-ph/9711070, doi:10.1103/PhysRevA.60.2746.

51 Thomas Koshy. *Catalan Numbers with Applications*. Oxford University Press, 2008. URL: http://books.google.ca/books?id=MqPLSivdBDAC&pg=PA48.

A Converse for bent functions

The goal of this appendix is to prove Theorem 8. First we need an alternative characterization of bent functions.

▶ **Proposition 18.** A Boolean function f is bent if and only if $(F * F)(x) = \delta_{x,0}$.

Proof. If $(F * F)(x) = \delta_{x,0}$, then using identities from Sect. 2, we find

$$\hat{F}^2(w) = \frac{1}{\sqrt{2^n}}\widehat{(F * F)}(w) = \frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n}(-1)^{w\cdot x}(F * F)(x) = \frac{1}{2^n} \tag{21}$$

so f is bent. Conversely, if f is bent then

$$(F * F)(w) = \sqrt{2^n}\widehat{\hat{F}^2}(w) = \sum_{x\in\mathbb{Z}_2^n}(-1)^{w\cdot x}\hat{F}^2(x) = \sum_{x\in\mathbb{Z}_2^n}(-1)^{w\cdot x}\frac{1}{2^n} = \delta_{w,0} \tag{22}$$

and the result follows. ◀

▶ **Theorem 8.** *Let $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ be a Boolean function with $n \geq 2$. A quantum algorithm can solve* BHSP_f *exactly with a single query to O_{f_s} if and only if f is bent.*

Proof. The most general one-query algorithm for solving BHSP_f using a controlled phase oracle (or equivalently, an oracle that computes the function in a register) performs a query on some superposition of all binary strings $x \in \mathbb{Z}_2^n$ and an extra symbol "$\varnothing$" that allows for the possibility of not querying the oracle. Without loss of generality, the initial state is

$$\alpha_\varnothing|\varnothing\rangle + \sum_{x\in\mathbb{Z}_2^n}\alpha_x|x\rangle \tag{23}$$

for some amplitudes $\alpha_\varnothing \in \mathbb{C}$ and $\alpha_x \in \mathbb{C}$ for $x \in \mathbb{Z}_2^n$ such that $|\alpha_\varnothing|^2 + \sum_{x\in\mathbb{Z}_2^n}|\alpha_x|^2 = 1$. The oracle acts trivially on $|\varnothing\rangle$, so the state after the query is

$$|\phi_s\rangle := \alpha_\varnothing|\varnothing\rangle + \sum_{x\in\mathbb{Z}_2^n}\alpha_x(-1)^{f(x+s)}|x\rangle \tag{24}$$

where $s \in \mathbb{Z}_2^n$ is the hidden shift. For an exact algorithm, we must have

$$\forall\, s \neq s' :\ 0 = \langle \phi_s | \phi_{s'} \rangle = |\alpha_\varnothing|^2 + \sum_{x \in \mathbb{Z}_2^n} |\alpha_x|^2 (-1)^{f(x+s)+f(x+s')}. \tag{25}$$

We can describe Eq. (25) as a linear system of equations. Define $p_\varnothing := |\alpha_\varnothing|^2$ and let p be a sub-normalized probability distribution on $\mathbb{Z}_2^n$ defined by $p_x := |\alpha_x|^2$. Let M be a rectangular matrix with rows labeled by elements of $A := \{(s, s') \in \mathbb{Z}_2^n \times \mathbb{Z}_2^n : s \neq s'\}$ and columns labeled by $x \in \mathbb{Z}_2^n$, with entries

$$M_{ss';x} := (-1)^{f(x+s)+f(x+s')}. \tag{26}$$

Then Eq. (25) is equivalent to

$$Mp = -p_\varnothing u \tag{27}$$

where u is the all-ones vector indexed by elements of A. In other words, there exists an exact one-query quantum algorithm for solving BHSP_f if and only if Eq. (27) holds for some $p_\varnothing$ and p that together form a probability distribution on $\{\varnothing\} \cup \mathbb{Z}_2^n$.

If f is bent, there is an exact one-query quantum algorithm corresponding to $p_\varnothing = 0$ and $p = \mu$, the uniform distribution (i.e., $\mu_x := 1/2^n$ for all $x \in \mathbb{Z}_2^n$). Notice that the entries of the vector $M\mu$ are

$$\begin{aligned}
(M\mu)_{ss'} &= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} M_{ss';x} && (28)\\
&= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x+s)+f(x+s')} && (29)\\
&= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)+f(x+s+s')} && (30)\\
&= (F * F)(s + s'). && (31)
\end{aligned}$$

Prop. 18 implies that $(F * F)(x) = \delta_{x,0}$, so $(Mp)_{ss'} = 0$ for all $s \neq s'$. Since $p_\varnothing = 0$, Eq. (27) holds and the algorithm is exact.

To prove the converse, assume there is an exact one-query quantum algorithm that solves BHSP_f. Then Eq. (27) holds for some $p_\varnothing$ and p that form a probability distribution on $\{\varnothing\} \cup \mathbb{Z}_2^n$.

First, we claim that without loss of generality, the probabilities p_x can be set equal for all $x \in \mathbb{Z}_2^n$. More precisely, we set $\bar{p} := (1 - p_\varnothing)\mu$ and show that Eq. (27) still holds if we

replace p by $\bar{p}$. Note that $1 - p_\varnothing = \sum_{y \in \mathbb{Z}_2^n} p_{x+y}$ for any $x \in \mathbb{Z}_2^n$, so

$$(M\bar{p})_{ss'} = \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} M_{ss',x} \, (1 - p_\varnothing) \tag{32}$$

$$= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x+s)+f(x+s')} \sum_{y \in \mathbb{Z}_2^n} p_{x+y} \tag{33}$$

$$= \frac{1}{2^n} \sum_{y \in \mathbb{Z}_2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x+y+s)+f(x+y+s')} p_x \tag{34}$$

$$= \frac{1}{2^n} \sum_{y \in \mathbb{Z}_2^n} \sum_{x \in \mathbb{Z}_2^n} M_{(y+s,y+s'),x} \, p_x \tag{35}$$

$$= \frac{1}{2^n} \sum_{y \in \mathbb{Z}_2^n} (Mp)_{(y+s,y+s')} \tag{36}$$

$$= -p_\varnothing \tag{37}$$

where the last equality follows since p is a solution of Eq. (27). We conclude that $\bar{p}$ is also a solution of Eq. (27), i.e.,

$$(1 - p_\varnothing) M\mu = -p_\varnothing u. \tag{38}$$

Recall from Eqs. (28) to (31) that $(M\mu)_{ss'} = (F * F)(s + s')$, which together with Eq. (38) implies that $(1 - p_\varnothing)(F * F)(s + s') = -p_\varnothing$ for all $s \neq s'$. Clearly, there is no solution with $p_\varnothing = 1$. Thus we have

$$(F * F)(w) = -\frac{p_\varnothing}{1 - p_\varnothing} \leq 0 \tag{39}$$

for any $w \neq 0$. Observe that $(F * F)(w) = \sum_{x \in \mathbb{Z}_2^n} \frac{1}{2^n} (-1)^{f(x)+f(x+w)}$ is an integer multiple of $1/2^n$ and $(F * F)(0) = 1$ for any f. Thus, we can rewrite Eq. (39) as

$$(F * F)(w) = \begin{cases} 1 & \text{if } w = 0, \\ -k/2^n & \text{otherwise} \end{cases} \tag{40}$$

for some integer $k \geq 0$. Therefore

$$\sum_{w \in \mathbb{Z}_2^n} (F * F)(w) = 1 - \frac{2^n - 1}{2^n} k. \tag{41}$$

On the other hand,

$$\sum_{w \in \mathbb{Z}_2^n} (F * F)(w) = \sum_{w \in \mathbb{Z}_2^n} \sum_{x \in \mathbb{Z}_2^n} \frac{1}{2^n} (-1)^{f(x)+f(x+w)} \tag{42}$$

$$= \left[\frac{1}{\sqrt{2^n}} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} \right]^2 \tag{43}$$

$$= \frac{1}{2^n} \left[\sum_{x \in \mathbb{Z}_2^n} \big(1 - 2f(x)\big) \right]^2 \tag{44}$$

$$= \frac{1}{2^n} \big(2^n - 2|f|\big)^2. \tag{45}$$

Putting this together with Eq. (41) gives

$$\left(2^n - 2|f|\right)^2 = 2^n - (2^n - 1)k. \tag{46}$$

This equation has no solutions for $k \geq 2$ since the right-hand side is negative (for $n \geq 2$). Similarly, there are no solutions for $k = 1$ since the left-hand side is even and the right-hand side is odd. Therefore $k = 0$ (and hence $p_\varnothing = 0$), which implies that f is bent by Eq. (40) and Prop. 18. ◀

Note that there is a solution to Eq. (46) with $k = 2$ and $n = 1$, provided $|f| = 1$. This trivial case involves the one-argument Boolean functions $f(x) = x$ and $f(x) = \mathrm{NOT}(x)$. For these functions we can choose $p_\varnothing = 1/2$ and $p_0 = p_1 = 1/4$ to determine the hidden shift exactly with one query. A deterministic classical algorithm can also solve BHSP_f with one query for these functions.

B Success probability of one-query PGM for random functions

In this appendix, we show that for one query, the expected success probability of $\mathbf{PGM}(f, 1)$ approaches a constant less than 1 for large n. This suggests that one query might not be enough to solve the problem with success probability arbitrarily close to 1. However, we do not know if the PGM algorithm has optimal success probability in the one-query case.

▶ **Theorem 19.** *Let f be an n-argument Boolean function chosen uniformly at random and suppose that a hidden shift for f is chosen adversarially. Then $\mathbf{PGM}(f, 1)$ solves BHSP_f with one query to O_{f_s} and expected success probability $\bar{p} \geq 1/2$ over the choice of f. Indeed, $\bar{p} \geq 2/\pi - o(1)$ as $n \to \infty$.*

Proof. Recall from Eq. (16) in Lemma 13 that $\mathbf{PGM}(f, t)$ recovers the hidden shift of f correctly after t queries with success probability $p_f(t)$. If the function f is chosen uniformly at random, then the expected success probability after t queries is

$$\bar{p}(t) := \frac{1}{2^{2^n}} \sum_f p_f(t) = \frac{1}{2^{2^n}} \sum_f \frac{1}{2^n} \left(\sum_{w \in \mathbb{Z}_2^n} \mathcal{F}^t(w) \right)^2. \tag{47}$$

We can obtain a lower bound on $\bar{p}(t)$ using the Cauchy-Schwarz inequality:

$$\bar{p}(t) \geq \frac{1}{2^n} \frac{1}{(2^{2^n})^2} \left(\sum_f \sum_{w \in \mathbb{Z}_2^n} \mathcal{F}^t(w) \right)^2 = 2^n \left(\frac{1}{2^n} \sum_{w \in \mathbb{Z}_2^n} \frac{1}{2^{2^n}} \sum_f \mathcal{F}^t(w) \right)^2 =: \tilde{p}(t). \tag{48}$$

Taking $t = 1$, this gives

$$\bar{p} \geq \frac{1}{2^n} \frac{1}{(2^{2^n})^2} \left(\sum_f \sum_{w \in \mathbb{Z}_2^n} |\hat{F}(w)| \right)^2 \tag{49}$$

$$= \frac{1}{2^n} \left(\frac{1}{2^{2^n}} \sum_{w \in \mathbb{Z}_2^n} \sum_f \left| \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{w \cdot x + f(x)} \right| \right)^2. \tag{50}$$

For each w we can define $f'(x) := w \cdot x + f(x)$ and change the order of summation by summing over f' instead of f. The value of this sum does not depend on w, so we get

$$\bar{p} \geq \frac{1}{2^n} \left(\frac{1}{2^{2^n}} \sum_f \left| \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)} \right| \right)^2 = \frac{L(2^n)^2}{2^n} \tag{51}$$

where

$$L(N) := \frac{1}{2^N} \sum_{z \in \{1,-1\}^N} \left| \sum_{i=1}^{N} z_i \right| \tag{52}$$

is the expected distance traveled by N steps of a random walk on a line (where each step is of size one and is to the left or the right with equal probability). It remains to lower bound $L(N)$.

Let $N = 2m$ for some integer $m \geq 1$. Using standard identities for sums of binomial coefficients, we compute

$$L(2m) = \frac{1}{2^{2m}} \cdot 2 \sum_{k=0}^{m} (2m - 2k) \binom{2m}{k} \tag{53}$$

$$= \frac{1}{2^{2m}} \cdot 2m \binom{2m}{m}. \tag{54}$$

Since the central binomial coefficient satisfies [51, p. 48]

$$\binom{2m}{m} \geq \frac{4^m}{\sqrt{4m}}, \tag{55}$$

we find

$$L(2m) \geq \sqrt{m}. \tag{56}$$

For $N = 2^n$ this gives $L(2^n) \geq \sqrt{2^n/2}$. We plug this in Eq. (51) and get $\bar{p} \geq 1/2$. In fact, according to Stirling's formula $\binom{2m}{m} \sim 4^m/\sqrt{\pi m}$ as $m \to \infty$. This means that $L(N) \sim \sqrt{2N/\pi}$ as $N \to \infty$ and our lower bound on $\bar{p}$ approaches $2/\pi$ as $n \to \infty$. ◀

C Two queries suffice for random functions

In this appendix we prove the following:

▶ **Theorem 15.** *Let f be an n-argument Boolean function chosen uniformly at random and suppose that a hidden shift for f is chosen adversarially. Then* $\mathbf{PGM}(f, 2)$ *solves* BHSP_f *with expected success probability* $\bar{p} \geq 1 - \frac{3}{64} \cdot 2^{-n}$.

C.1 Strategy

Our goal is lower bound $\tilde{p}(t)$, as defined in Eq. (48). Let us define a random variable X over Boolean functions $f \colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ and binary strings $w \in \mathbb{Z}_2^n$, whose value is

$$X := \left[\mathcal{F}^t(w)\right]^2 = \left[\hat{F}^2\right]^{*t}(w), \tag{57}$$

where f and w are chosen uniformly at random. Notice from Eq. (48) that

$$\tilde{p}(t) = 2^n \left(\mathbb{E}[\sqrt{X}]\right)^2. \tag{58}$$

Clearly, for any $x \geq 0$ we have

$$\mathbb{E}[\sqrt{X}] \geq \sqrt{x} \Pr(X \geq x). \tag{59}$$

Our strategy is to use a one-sided version of Chebyshev's inequality, known as Cantelli's inequality, to lower-bound $\Pr(X \geq x)$, and then choose a value of x that maximizes our lower bound on $\tilde{p}(t)$.

▶ **Fact** (Cantelli's inequality)**.** Let $\mu := \mathbb{E}[X]$ and $\sigma^2 := \mathbb{E}[X^2] - \mu^2$ be the mean and variance of X, respectively. Then $\Pr(X - \mu \geq k\sigma) \geq \frac{1}{1+k^2}$.

Alternatively, if we substitute X by $-X$ and reverse the inequality then

$$\Pr(X \geq \mu - k\sigma) \geq \frac{k^2}{1+k^2}. \tag{60}$$

If we substitute $x := \mu - k\sigma$ in Eq. (59), then according to the above inequality,

$$\mathbb{E}[\sqrt{X}] \geq \sqrt{\mu - k\sigma}\,\frac{k^2}{1+k^2}. \tag{61}$$

Using Eq. (48), Eq. (58), and Eq. (61) gives

$$\bar{p}(t) \geq \tilde{p}(t) = 2^n \big(\mathbb{E}[\sqrt{X}]\big)^2 \geq 2^n(\mu - k\sigma)\left(1 + \frac{1}{k^2}\right)^{-2}. \tag{62}$$

It remains to lower bound μ (Sect. C.2), upper bound σ (Sect. C.3), and make a reasonable choice of the deviation parameter k (Sect. C.4).

C.2 Computing the mean

Let us compute the mean

$$\mu = \mathbb{E}[X] = \frac{1}{2^{2^n}} \sum_f \frac{1}{2^n} \sum_{w \in \mathbb{Z}_2^n} \big[\hat{F}^2\big]^{*t}(w) \tag{63}$$

for any integer $t \geq 1$. Notice that

$$\begin{aligned}
\sum_{w \in \mathbb{Z}_2^n} \big[\hat{F}^2\big]^{*t}(w) &= \sum_{w, y_1, \ldots, y_{t-1} \in \mathbb{Z}_2^n} \hat{F}(y_1)^2 \cdots \hat{F}(y_{t-1})^2 \hat{F}\big(w - (y_1 + \cdots + y_{t-1})\big)^2 && (64)\\
&= \sum_{y_1, \ldots, y_t \in \mathbb{Z}_2^n} \hat{F}(y_1)^2 \cdots \hat{F}(y_{t-1})^2 \hat{F}(y_t)^2 && (65)\\
&= \left(\sum_{y \in \mathbb{Z}_2^n} \hat{F}(y)^2\right)^t && (66)\\
&= 1 && (67)
\end{aligned}$$

by unitarity of the Fourier transform (see Plancherel's identity in Sect. 2). We conclude that

$$\mu = \frac{1}{2^n} \tag{68}$$

independent of t.

C.3 Computing the variance

Next we compute the variance

$$\mathbb{E}[X^2] = \frac{1}{2^{2^n}} \sum_f \frac{1}{2^n} \sum_{w \in \mathbb{Z}_2^n} \Big(\big[\hat{F}^2\big]^{*t}(w)\Big)^2. \tag{69}$$

Note that from Eq. (2) and Plancherel identity we have

$$\sum_{w\in\mathbb{Z}_2^n}\left(\left[\hat{F}^2\right]^{*t}(w)\right)^2 = \sum_{w\in\mathbb{Z}_2^n}\left(\frac{1}{\sqrt{2^n}}\left(\widehat{F*F}\right)^t(w)\right)^2 = \frac{1}{2^n}\sum_{w\in\mathbb{Z}_2^n}(F*F)^{2t}(w). \tag{70}$$

We substitute this in Eq. (69) and get

$$\mathbb{E}[X^2] = \frac{1}{2^{2^n}}\sum_f \frac{1}{2^n}\left(\frac{1}{2^n}\sum_{w\in\mathbb{Z}_2^n}(F*F)^{2t}(w)\right) \tag{71}$$

$$= \frac{1}{2^{2n}}\sum_{w\in\mathbb{Z}_2^n}\frac{1}{2^{2^n}}\sum_f\left(\frac{1}{2^n}\sum_{x\in\mathbb{Z}_2^n}(-1)^{f(x)+f(w+x)}\right)^{2t}. \tag{72}$$

C.3.1 Counting pairings

Let us introduce some combinatorial ideas that will help us to evaluate the sum in Eq. (72).

▶ **Definition 20.** Let S be a finite set and let $l \geq 1$ be an integer. We say that $a_1, a_2, \ldots, a_{2l} \in S$ are *paired* if there exists a permutation π of $\{1, 2, \ldots, 2l\}$ such that $a_{\pi(2i-1)} = a_{\pi(2i)}$ for all $i \in \{1, 2, \ldots, l\}$. Define $\Delta\colon S^{2l} \to \mathbb{Z}_2$ as

$$\Delta(a_1, a_2, \ldots, a_{2l}) := \begin{cases} 1 & \text{if } a_1, a_2, \ldots, a_{2l} \text{ are paired,} \\ 0 & \text{otherwise.} \end{cases} \tag{73}$$

Notice that for $l = 2$ we have $\Delta(a, b, c, d) = \delta_{a,b}\delta_{c,d} + \delta_{a,c}\delta_{b,d} + \delta_{a,d}\delta_{b,c} - 2\delta_{a,b,c,d}$, so the number of ways to pair four elements of S is

$$\sum_{a,b,c,d\in S}\Delta(a,b,c,d) = 3\sum_{a,b,c,d\in S}\delta_{a,b}\delta_{c,d} - 2\sum_{a,b,c,d\in S}\delta_{a,b,c,d} = 3|S|^2 - 2|S|. \tag{74}$$

▶ **Proposition 21.** Let $S = \{0,1\}^n$. Then for any $a_1, a_2, \ldots, a_{2l} \in S$,

$$\frac{1}{2^{2^n}}\sum_f (-1)^{f(a_1)+f(a_2)+\cdots+f(a_{2l})} = \Delta(a_1, a_2, \ldots, a_{2l}) \tag{75}$$

where the sum is over all Boolean functions $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$.

Proof. Clearly, if $a_1, a_2, \ldots, a_{2l}$ are paired, then the exponent of -1 is even and the sum is 1. Otherwise, we can omit the paired arguments, and all remaining a_i are distinct. Since we are averaging over all f and the values that f takes at distinct points are independent, the sum vanishes. ◀

We can use this observation to rewrite Eq. (72) as follows:

$$\mathbb{E}[X^2] = \frac{1}{2^{2(t+1)n}}\sum_{w\in\mathbb{Z}_2^n}\sum_{a_1,\ldots,a_{2t}\in\mathbb{Z}_2^n}\Delta(a_1, a_1+w, a_2, a_2+w, \ldots, a_{2l}, a_{2l}+w). \tag{76}$$

C.3.2 Evaluating the variance at $t = 2$

In general, the variance depends on t. However, we are interested only in the $t = 2$ case, so from now on we will assume that $t = 2$ and do not write the dependence on t explicitly. For $t = 2$, Eq. (76) reads

$$\mathbb{E}[X^2] = \frac{1}{2^{6n}}\sum_{w\in\mathbb{Z}_2^n}\sum_{a,b,c,d\in\mathbb{Z}_2^n}\Delta(a, a+w, b, b+w, c, c+w, d, d+w). \tag{77}$$

We consider two cases. First, when $w = 0$, the eight arguments of Δ are always paired, so the inner sum in Eq. (77) evaluates to

$$\sum_{a,b,c,d \in \mathbb{Z}_2^n} \Delta(a,a,b,b,c,c,d,d) = 2^{4n}. \tag{78}$$

Now suppose $w \neq 0$. Then $w_i = 1$ for some $i \in \{1,\ldots,n\}$ and thus either $a_i = 0$ or $a_i + w_i = 0$ (and similarly for b, c, and d). In total there are $2^4 = 16$ cases. Since Δ is invariant under permutations of arguments, we can substitute a by $a + w$, which effectively swaps the arguments a and $a + w$. By performing a similar operation for b, c, and d, we can ensure that $a_i = b_i = c_i = d_i = 0$. Among the eight arguments of Δ in Eq. (77), arguments a, b, c, and d can be paired only among themselves since $w_i = 1$. Moreover, once a and b are paired, then so are $a + w$ and $b + w$. Thus, we can restrict the ith bit of w to be 1 and ignore the four extra arguments of Δ. Then the inner sum in Eq. (77) becomes

$$16 \sum_{a,b,c,d \in \mathbb{Z}_2^{n-1}} \Delta(a,b,c,d) = 16 \cdot \left(3 \cdot 2^{2n-2} - 2 \cdot 2^{n-1}\right) = 12 \cdot 2^{2n} - 16 \cdot 2^n, \tag{79}$$

where the first equality follows from Eq. (74) with $S = \mathbb{Z}_2^{n-1}$.

By combining Eq. (78) and Eq. (79), we can rewrite Eq. (77) as

$$\mathbb{E}[X^2] = \frac{1}{2^{6n}}\left(2^{4n} + (2^n - 1) \cdot (12 \cdot 2^{2n} - 16 \cdot 2^n)\right) \tag{80}$$

$$= \frac{1}{2^{2n}} + \frac{12}{2^{3n}} - \frac{28}{2^{4n}} + \frac{16}{2^{5n}}. \tag{81}$$

Using the value of μ from Eq. (68), we see that for $n \geq 1$ the variance is

$$\sigma^2 = \mathbb{E}[X^2] - \mu^2 = \frac{12}{2^{3n}} - \frac{28}{2^{4n}} + \frac{16}{2^{5n}} \geq \frac{1}{2^{3n}}. \tag{82}$$

C.4 Choosing the deviation

To complete the lower bound on the success probability, recall from Eq. (62) that

$$\bar{p} \geq 2^n(\mu - k\sigma)\left(1 + \frac{1}{k^2}\right)^{-2}. \tag{83}$$

Substituting the bounds on μ and σ from Eq. (68) and Eq. (82), respectively, gives

$$\bar{p} \geq \left(1 - \frac{k}{\sqrt{2^n}}\right)\left(1 + \frac{1}{k^2}\right)^{-2}. \tag{84}$$

Notice that $\left(1 + \frac{1}{k^2}\right)^{-2} \geq 1 - \frac{2}{k^2}$ for any k, so

$$\bar{p} \geq \left(1 - \frac{k}{\sqrt{2^n}}\right)\left(1 - \frac{2}{k^2}\right) \geq 1 - \frac{k}{\sqrt{2^n}} - \frac{2}{k^2}. \tag{85}$$

It remains to make a good choice for k. Let $\alpha = \sqrt{2^n}$ and $k = \alpha^c$ for some $c > 0$. Then

$$\bar{p} \geq 1 - \alpha^{c-1} - 2\alpha^{-2c}. \tag{86}$$

Choosing $c = 1/3$ (i.e., $k = 2^{n/6}$) gives

$$\bar{p} \geq 1 - \frac{3}{64} \cdot 2^{-n}. \tag{87}$$

This concludes the proof of Theorem 15.

D Zeroes in the Fourier spectrum

D.1 Undetectable shifts and anti-shifts

In some cases the Boolean hidden shift problem cannot be solved exactly in principle. For example, if the function f is invariant under some shift, then the hidden shift cannot be uniquely determined, as the oracle does not contain enough information (an extreme case of this is a constant function which is invariant under all shifts). In this section we consider such degenerate functions and analyze their Fourier spectra.

▶ **Definition 22.** Let $b \in \mathbb{Z}_2$. We say that s is a *b-shift* for a function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ if f has the following property: $\forall x \in \mathbb{Z}_2^n\colon f(x+s) = f(x) + b$. We refer to 0-shifts as *undetectable shifts* since they cannot be distinguished from the trivial shift $s = 0$. We also refer to 1-shifts as *anti-shifts* since they negate the truth table of f.

The following result provides an alternative characterization of b-shifts. It relates the maximal and minimal autocorrelation value of F to undetectable shifts and anti-shifts of f, respectively (see Definition 5 for the definition of convolution).

▶ **Proposition 23.** The string $s \in \mathbb{Z}_2^n$ is a b-shift for function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ if and only if $(F * F)(s) = (-1)^b$, where $F(x) := (-1)^{f(x)}/\sqrt{2^n}$ for all $x \in \mathbb{Z}_2^n$.

Proof. Let s be a b-shift of f. Then

$$(F * F)(s) = \sum_{x \in \mathbb{Z}_2^n} F(x)F(x+s) \tag{88}$$

$$= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{f(x)}(-1)^{f(x)+b} \tag{89}$$

$$= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{b} \tag{90}$$

$$= (-1)^b. \tag{91}$$

For the converse, note that all terms on the right-hand side of Eq. (88) have absolute value equal to $1/2^n$. In total there are 2^n terms, so $|(F * F)(s)| \leq 1$. If this bound is saturated, then all terms in Eq. (88) must have the same phase. Thus, s is a b-shift for some $b \in \mathbb{Z}_2$. ◀

If s' and s'' are undetectable shifts of f then so is $s'+s''$, since $f(x+s'+s'') = f(x+s') = f(x)$ for any x. Hence the set of all undetectable shifts forms a linear subspace of $\mathbb{Z}_2^n$. Also, if a' and a'' are anti-shifts, then $a' + a''$ is an undetectable shift. In particular, a Boolean function with no undetectable shifts can have at most one anti-shift.

If we want to solve the hidden shift problem for a function f that has an undetectable shift s, we can apply an invertible linear transformation A on the input variables such that $A \cdot 0 \ldots 01 = s$. Thus we simulate the oracle for the function $f'(x) := f(A \cdot x)$ such that $f'(x + 0 \ldots 01) = f'(x)$. Notice that f' is effectively an $(n-1)$-argument function, since it does not depend on the last argument. Similarly, if f has a k-dimensional subspace of undetectable shifts, it is effectively an $(n-k)$-argument function. Solving the hidden shift problem for such a function is equivalent to solving it for the reduced $(n-k)$-argument function f' and picking arbitrary values for the remaining k arguments. In this sense, Boolean functions with undetectable shifts are degenerate and we can consider only functions with no undetectable shifts without loss of generality.

Similarly, if f has an anti-shift, we can use the same construction to show that it is equivalent to a function f' such that $f'(x_1, \ldots, x_{n-1}, x_n) = f''(x_1, \ldots, x_{n-1}) \oplus x_n$ where f'' is an $(n-1)$-argument function. To solve the hidden shift problem for f', we first solve it for f'' and then learn the value of the remaining argument x_n via a single query. In this sense, Boolean functions with anti-shifts are also degenerate. Thus, without loss of generality we can consider the hidden shift problem only for *non-degenerate* functions, i.e., ones that have no b-shifts for any $b \in \mathbb{Z}_2$.

Finally, let us show that Boolean functions with b-shifts have at least half of their Fourier coefficients equal to zero. Let $\mathcal{S}$ be an $(n-1)$-dimensional subspace of $\mathbb{Z}_2^n$, and let us denote the two cosets of $\mathcal{S}$ in $\mathbb{Z}_2^n$ by $\mathcal{S}_b := \mathcal{S} + br$, where $b \in \mathbb{Z}_2$ and $r \in \mathbb{Z}_2^n \setminus \mathcal{S}$ is any representative of the coset for $b = 1$. The following result relates the property of having a b-shift to the property of having zero Fourier coefficients with special structure.

▶ **Lemma 24.** *A function $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ has a non-zero b-shift if and only if there is an $(n-1)$-dimensional subspace $\mathcal{S} \subset \mathbb{Z}_2^n$ such that $\hat{F}(w) = 0$ when $w \notin \mathcal{S}_b$.*

Proof. Assume that s is a b-shift of f. Then

$$\hat{F}(w) = \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{w \cdot x + f(x)} \tag{92}$$

$$= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{w \cdot (x+s) + f(x+s)} \tag{93}$$

$$= \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{w \cdot (x+s) + f(x) + b} \tag{94}$$

$$= (-1)^{w \cdot s + b} \frac{1}{2^n} \sum_{x \in \mathbb{Z}_2^n} (-1)^{w \cdot x + f(x)} \tag{95}$$

$$= (-1)^{w \cdot s + b} \hat{F}(w). \tag{96}$$

Thus, $\hat{F}(w) = 0$ when $w \cdot s \neq b$. Let $\mathcal{S}$ be the $(n-1)$-dimensional subspace of $\mathbb{Z}_2^n$ orthogonal to s. Then $w \in \mathcal{S}_b \Leftrightarrow w \cdot s = b$ and thus $\hat{F}(w) = 0$ when $w \notin \mathcal{S}_b$.

For the converse, assume that $\mathcal{S}$ is an $(n-1)$-dimensional subspace of $\mathbb{Z}_2^n$ and $\hat{F}(w) = 0$ when $w \notin \mathcal{S}_b$. Let $s \in \mathbb{Z}_2^n$ be the unique non-zero vector orthogonal to $\mathcal{S}$. Then $\mathcal{S}_b = \{w\colon w \cdot s = b\}$ and we have

$$F(x+s) = \hat{\hat{F}}(x+s) \tag{97}$$

$$= \frac{1}{\sqrt{2^n}} \sum_{w \in \mathbb{Z}_2^n} (-1)^{(x+s) \cdot w} \hat{F}(w) \tag{98}$$

$$= \frac{1}{\sqrt{2^n}} \sum_{w \in \mathcal{S}_b} (-1)^{(x+s) \cdot w} \hat{F}(w) \tag{99}$$

$$= (-1)^b \frac{1}{\sqrt{2^n}} \sum_{w \in \mathcal{S}_b} (-1)^{x \cdot w} \hat{F}(w) \tag{100}$$

$$= (-1)^b F(x). \tag{101}$$

Hence $f(x+s) = f(x) + b$ and thus s is a b-shift of f. ◀

D.2 Decision trees

In the previous section we discussed degenerate cases of Boolean functions that have many zero Fourier coefficients. In this section we explain how to construct non-degenerate examples.

▶ **Lemma 25.** *If f is a Boolean function defined by a decision tree of height h then $\hat{F}(w) = 0$ when $|w| > h$.*

Proof. Since the Boolean function f is given by a decision tree, let $\{P_1, \ldots, P_m\}$ be the set of all paths that start at the root of this tree and end at a parent of a leaf labeled by 1. For example, $P_1 = \{x_2, x_1, x_5, x_4, x_{10}\}$ and $P_2 = \{x_2, x_7, x_1\}$ are two such paths for the tree shown in Fig. 2. We can write the disjunctive normal form of f as

$$f(x) = \bigvee_{i=1}^{m} \bigwedge_{j \in P_i} (b_j^{(i)} \oplus x_j) \tag{102}$$

where "$\vee$" and "$\wedge$" represent logical OR and AND functions, respectively, and $b_j^{(i)} \in \mathbb{Z}_2$ is equal to 1 if and only if variable x_j has to be negated on path P_i. For example, x_{10} is negated on P_1, and x_2 and x_7 are negated on P_2.

To prove the desired result about the Fourier coefficients of f, we switch from Boolean functions to (± 1)-valued functions with (± 1)-valued variables. In particular, we replace $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ by a function $\tilde{F}\colon \{1,-1\}^n \to \{1,-1\}$ in variables $X_i \in \{1,-1\}$ such that

$$\tilde{F}\big((-1)^x\big) = (-1)^{f(x)} \tag{103}$$

for all $x \in \mathbb{Z}_2^n$.

Notice that the (± 1)-valued versions of logical NOT, AND, and OR functions are given by the following polynomials:

$$\mathrm{NOT}(X) := -X, \tag{104}$$

$$\mathrm{AND}(X_1, \ldots, X_k) := \quad 1 - 2\prod_{i=1}^{k} \frac{1 - X_i}{2}, \tag{105}$$

$$\mathrm{OR}(X_1, \ldots, X_k) := -1 - 2\prod_{i=1}^{k} \frac{1 + X_i}{2}. \tag{106}$$

We can use these polynomials and Eq. (102) to write $\tilde{F}$ as

$$\tilde{F}(X) = \mathrm{OR}_{i=1}^{m} \mathrm{AND}_{j \in P_i} (-1)^{b_j^{(i)}} X_j, \tag{107}$$

where $\mathrm{OR}_{i=1}^{m} X_i$ stands for $\mathrm{OR}(X_1, \ldots, X_m)$ and a similar convention is used for AND.

When we determine the value of f using a decision tree, each input $x \in \mathbb{Z}_2^n$ leads to a unique leaf of the tree. Thus, when $f(x) = 1$, there is a unique value of i in Eq. (102) for which the corresponding term in the disjunction is satisfied. With this promise we can simplify Eq. (106) to

$$\mathrm{OR}(X_1, \ldots, X_k) := \sum_{i=1}^{k} (X_i - 1) + 1. \tag{108}$$

If we use this in Eq. (107), we get

$$\tilde{F}(X) = \sum_{i=1}^{m} \Big(\mathrm{AND}_{j \in P_i} (-1)^{b_j^{(i)}} X_j - 1 \Big) + 1, \tag{109}$$

$$= 1 - 2\sum_{i=1}^{m} \prod_{j \in P_i} \frac{1 - (-1)^{b_j^{(i)}} X_j}{2}. \tag{110}$$

Notice that this polynomial has degree at most $\max_i |P_i| \leq h$, the height of the tree. On the other hand, the Fourier transform is self-inverse (see Sect. 2), so

$$(-1)^{f(x)} = \sqrt{2^n} F(x) = \sqrt{2^n} \hat{\hat{F}}(x) = \sum_{w \in \mathbb{Z}_2^n} (-1)^{x \cdot w} \hat{F}(w). \tag{111}$$

The (± 1)-valued equivalent of this equation is

$$\tilde{F}(X) = \sum_{w \in \mathbb{Z}_2^n} \hat{F}(w) \prod_{i \colon w_i = 1} X_i. \tag{112}$$

By comparing this with Eq. (110) we conclude that $\hat{F}(w) = 0$ when $|w| > h$. ◀

According to this lemma, we can use the following strategy to construct Boolean functions with a large fraction of their Fourier coefficients equal to zero. We pick a random decision tree with many variables but small height, i.e., large n and small h (notice that $n \leq 2^h - 1$). Then we are guaranteed that the fraction of non-zero Fourier coefficients does not exceed

$$\frac{1}{2^n} \sum_{k=0}^{h} \binom{n}{k} \leq \frac{2^{H(\frac{h}{n})n}}{2^n} = \left(\frac{1}{2^n}\right)^{1 - H(\frac{h}{n})} \tag{113}$$

where $H(p) := -p \log_2 p - (1-p) \log_2 (1-p)$ is the binary entropy function. In particular, if $h \sim \log_2 n$ then this fraction vanishes as n goes to infinity, i.e., $\hat{F}$ is zero almost everywhere.

However, notice that when the number of zero Fourier coefficients is large, it is also more likely to pick a degenerate Boolean function (i.e., one that has a b-shift for some $b \in \mathbb{Z}_2$); we would like to avoid this. Recall from Lemma 24 that f has a b-shift only if all its non-zero Fourier coefficients lie in a coset $\mathcal{S}_b$ of some $(n-1)$-dimensional subspace $\mathcal{S} \subset \mathbb{Z}_2^n$. Unfortunately, we do not know the probability that a random decision tree with n variables and height $\log_2 n$ corresponds to a Boolean function with this property.

D.3 Zeroes in the t-fold Fourier spectrum

In this section we study the fraction of zeroes in the t-fold Fourier spectrum $\mathcal{F}^t$ of f as a function of t. The main observation is Lemma 27, which shows that unless f has an undetectable shift, $\mathcal{F}^t$ becomes non-zero everywhere when t is sufficiently large. This means that even for functions with a high density of zeroes in the Fourier spectrum, one can boost the success probability of the basic quantum rejection sampling approach discussed in Sect. 5.1 by using the t-fold generalization from Sect. 5.4.

▶ **Proposition 26.** Let $S_t := \{w \in \mathbb{Z}_2^n \colon \mathcal{F}^t(w) \neq 0\}$ be the set of strings for which $\mathcal{F}^t$ is non-zero. Then $S_{t+1} = S_t + S_1$ where $A + B := \{a + b \colon a \in A, b \in B\}$.

Proof. Note that $\left[\mathcal{F}^{t+1}\right]^2 = \left[\mathcal{F}^t\right]^2 * \left[\mathcal{F}^1\right]^2$ from Definition 6. Also, $\mathcal{F}^t(w) \geq 0$ for any $t \geq 1$ and $w \in \mathbb{Z}_2^n$. Assume that $w_0 \in S_t$ and $w_1 \in S_1$. Then $\mathcal{F}^t(w_0) > 0$ and $\mathcal{F}^1(w_1) > 0$, so

$$\left[\mathcal{F}^{t+1}\right]^2 (w_0 + w_1) = \sum_{x \in \mathbb{Z}_2^n} \left[\mathcal{F}^t\right]^2 (x) \cdot \left[\mathcal{F}^1\right]^2 (w_0 + w_1 - x) \tag{114}$$

$$\geq \left[\mathcal{F}^t\right]^2 (w_0) \cdot \left[\mathcal{F}^1\right]^2 (w_0 + w_1 - w_0) > 0. \tag{115}$$

Thus $w_0 + w_1 \in S_{t+1}$ and hence $S_t + S_1 \subseteq S_{t+1}$. Conversely, if w cannot be written in the form $w_0 + w_1$ for some $w_0 \in S_t$ and $w_1 \in S_1$ then $\mathcal{F}^{t+1}(w) = 0$, since all terms of the sum in Eq. (114) vanish. ◀

▶ **Lemma 27.** *If $f\colon \mathbb{Z}_2^n \to \mathbb{Z}_2$ does not have an undetectable shift, then there exists $t \in \{1,\ldots,n\}$ such that $\mathcal{F}^t$ is non-zero everywhere.*

Proof. If S_1 spans the whole space $\mathbb{Z}_2^n$, we can inductively apply Prop. 26 to conclude that $S_t = \mathbb{Z}_2^n$ for some sufficiently large t. In particular, it suffices to take $t \leq n$ (say, if S_1 is the standard basis). On the other hand, if S_1 spans only a proper subspace of $\mathbb{Z}_2^n$, then it is contained in some $(n-1)$-dimensional subspace $\mathcal{S}_0$. Since $\mathcal{F}^1 = |\hat{F}|$ vanishes outside of $\mathcal{S}_0$, we conclude by Lemma 24 that f has an undetectable shift. ◀

Dequantizing Read-once Quantum Formulas

Alessandro Cosentino, Robin Kothari, and Adam Paetznick

David R. Cheriton School of Computer Science
Institute for Quantum Computing
University of Waterloo

Abstract

Quantum formulas, defined by Yao [FOCS'93], are the quantum analogs of classical formulas, i.e., classical circuits in which all gates have fanout one. We show that any read-once quantum formula over a gate set that contains all single-qubit gates is equivalent to a read-once classical formula of the same size and depth over an analogous classical gate set. For example, any read-once quantum formula over Toffoli and single-qubit gates is equivalent to a read-once classical formula over Toffoli and NOT gates. We then show that the equivalence does not hold if the read-once restriction is removed. To show the power of quantum formulas without the read-once restriction, we define a new model of computation called the one-qubit model and show that it can compute all boolean functions. This model may also be of independent interest.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases formulas, dequantization, circuit complexity

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.80

1 Introduction

It is widely believed that quantum computers can outperform classical computers for certain problems. Two prominent examples of such problems are factoring, solved by Shor's algorithm [18], and simulation of quantum systems [7, 13, 2]. Many restricted versions of quantum computers also outperform classical models. Studying the power of restricted quantum models can help identify the "quantum" features that are required for computational speedups.

Some restricted models are also practically motivated, and could be available before we are able to build unrestricted quantum computers. The "one clean qubit" model [12], for example, can solve some problems that are not known to have efficient classical algorithms [19]. Similarly, log-depth quantum circuits can implement Shor's algorithm with the aid of a classical computer [5].

On the other hand, if enough restrictions are placed on a quantum model, it may be efficiently simulable by a classical model. In analogy with derandomization, we call this *dequantization*. For example, the simulation of a polynomial-time quantum computer by an exponential-time classical computer is an example of dequantization, albeit a very weak one. On the other hand, if the quantum model is *equivalent* to a classical model, we call this *strong* dequantization. For example, if it were shown that a polynomial-time quantum computer can be simulated by a polynomial-time classical computer, this would be an example of strong dequantization. In this paper, we strongly dequantize read-once quantum formulas by showing that they are equivalent to classical read-once formulas.

Several past results can be viewed as dequantizing quantum models of computation. For example, Valiant introduced and dequantized a restricted model of quantum computation [21] that was later shown to be equivalent to a classical model [22]. Additionally,

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 80–92


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

quantum circuits containing only Clifford gates are equivalent to classical circuits of CNOT and NOT gates [8, 4]. Other examples of dequantization appear in [6, 20].

Classical formulas are a well-studied restriction of circuits in which gates have a single output wire and each gate is a function from $k \geq 1$ bits to one bit. Compared to general circuits, the power of formulas is much better understood and several lower bound techniques are known for explicit functions [23, 11]. The study of formulas has lead to a better understanding of the difficulty of proving lower bounds for general circuits. Similarly, studying quantum formulas may lead to a better understanding of the power of quantum circuits.

Indeed, little is known about quantum formulas. They were defined and examined by Yao [25] in 1993. But, other than additional results by Roychowdhury and Vatan [16], they are largely unstudied.

In this paper, we ask whether it is possible to dequantize quantum formulas. Informally, the question is whether, for a quantum gate set G, there exists a classical gate set $\hat{G}$ of roughly equivalent power, such that any quantum formula over G can be written as a classical formula over $\hat{G}$. We discuss this question in Section 3 and define an appropriate classical gate set $\hat{G}$ (Definition 1) corresponding to any quantum gate set G. Our main results (Theorem 2 and Theorem 3) essentially resolve this question for read-once quantum formulas by showing that read-once quantum formulas over a gate set that includes all single-qubit gates can always be dequantized.

One utility of our classical gate set $\hat{G}$ is that, in some cases of interest, it corresponds to the gate set that one would naturally expect. For the set of all k-qubit channels (for some constant k), which is the gate set used in previous papers on quantum formulas [25, 16], $\hat{G}$ is the set of all k-bit gates. The set of arbitrary fanin Toffoli gates and all single-qubit gates is a commonly used gate set in quantum circuit complexity (see, e.g., [9]). We show in Theorem 4 that, for this gate set, $\hat{G}$ is the set of classical arbitrary fanin Toffoli gates and the NOT gate.

It is natural to ask whether the read-once constraint is required for dequantization to hold. In Section 4 we show that Theorem 2 and Theorem 3 are false if the read-once constraint is dropped. In particular, we show that there exist quantum formulas over a gate set G that cannot be simulated by classical formulas over $\hat{G}$. To show this we define a model of computation that we call the one-qubit model. Our model is similar to the "one clean qubit" model of Knill and Laflamme [12], but we do not have any mixed states in addition to the one clean qubit. We show that this model can compute any boolean function (Theorem 5). However, the one-qubit model is contained in quantum formulas over a gate set G for which $\hat{G}$ contains only the NOT gate and the PARITY gate. Even classical *circuits* of arbitrary size over NOT and PARITY cannot compute, for example, the AND function on two bits. Thus our dequantization theorems are false without the read-once constraint.

2 Preliminaries

We start with an introduction to classical formulas and then extend the definition to quantum formulas. We refer the reader to textbooks on circuit complexity [23, 11] or quantum computing [15] for further information.

A gate set F is a set of functions from $k \geq 1$ bits to one bit. A classical formula over a gate set F is a circuit composed of gates from F in which the output of each gate is connected to the input of at most one other gate—i.e., gates have fanout one. Note, however, that input bits may have arbitrary fanout, i.e., more than one gate can use the same bit x_i as an input. The formula outputs a single bit and is said to exactly compute a boolean function

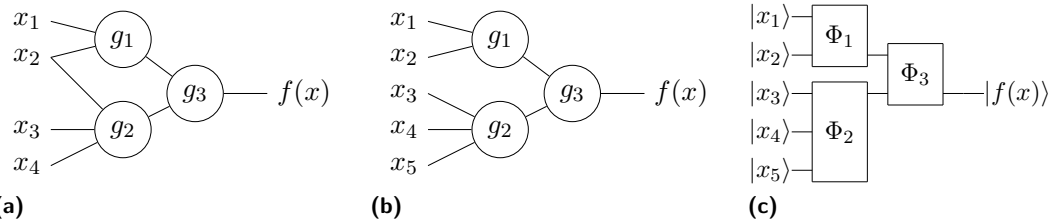


Figure 1 (a) A classical (read-many) formula may use input bits multiple times to compute a function $f(x)$. Here the input x_2 is used twice and the formula is over the gate set $\hat{G} = \{g_1, g_2, g_3\}$. (b) A classical *read-once* formula may use each input bit only once. (c) A quantum read-once formula has the same structure as a classical read-once formula, but may contain quantum channels with single qubit outputs. Here the quantum formula is over the gate set $G = \{\Phi_1, \Phi_2, \Phi_3\}$.

f if the output of the formula is $f(x)$ on input x. The restriction that each gate has fanout one makes the circuit look like a tree in which the output gate is the root, non-output gates are internal nodes, and leaves are labeled by input bits. A *read-once* formula is one in which each input bit appears at most once.

Yao defined a quantum formula as a single-output quantum circuit composed of unitary gates in which the path connecting any input to the output is unique. Equivalently, a quantum circuit is a quantum formula if every gate has at most one output that is used as an input to a subsequent gate. The other outputs of a gate are never used again and can be discarded (traced out). We can regard the unitary and discard step as one operation and call that a quantum gate. This makes the analogy with classical formulas clearer. In this paper we use the phrases "quantum channel" and "quantum gate" interchangeably; a quantum gate need not be unitary. We will sometimes talk about a formula over a set of unitaries, which may have multiple output qubits. In this case, the formula may use any single-output channel obtained by applying one of the unitaries in the set and then tracing out all but one of the output qubits.

We define a quantum gate set to be a set G of quantum channels from $k \geq 1$ qubits to one qubit. A quantum channel is a completely-positive trace-preserving map. In the case that $k = 1$ we call the channel a *single-qubit* channel, otherwise we call the channel a *k-qubit* channel. Note that a quantum formula may read input bits multiple times, just like a classical formula. A read-once quantum formula is one in which each input bit appears at most once. Figure 1 shows an example of a classical formula, a read-once classical formula and a read-once quantum formula. Similar to the classical case, one gate is designated as the output gate. A quantum formula is said to exactly compute a boolean function f if the output of the formula is $|f(x)\rangle$ on input x. In Section 3.2 we discuss how to extend this definition to bounded-error quantum formulas.

The *size* of a formula is defined as the number of gates, excluding single-bit or single-qubit gates, following standard convention in classical circuit complexity. The *depth* of a formula is the maximum number of multi-bit or multi-qubit gates on any path from the output to an input. We say that a formula accepts a language $L \subseteq \{0,1\}^*$ if on input x it outputs 1 if and only if $x \in L$.

We now state previously known results about quantum formulas. Yao first showed that the majority function needs super-linear sized bounded-error quantum formulas over the set of all three-qubit unitaries [25]. Later, Roychowdhury and Vatan [16] showed that a classical formula-size lower bound technique due to Nechiporuk [23, 11] also extends to bounded-error quantum formulas over the set of all k-qubit unitary matrices, for any constant k.

Roychowdhury and Vatan also showed that any function computed by a bounded-error quantum formula over k-qubit unitaries can be computed by a classical *circuit* of slightly larger size and depth. Our result applies to a wider range of gate sets and dequantizes to the more limited model of classical read-once formulas.

3 Dequantization of read-once formulas

Our main objective is to determine the conditions under which it is possible to dequantize quantum formulas. More precisely, for a quantum gate set G we seek a classical gate set $\hat{G}$ for which a language L is accepted by a quantum formula over G if and only if it is accepted by a classical formula over $\hat{G}$.

Note that we want the two classes to be exactly equal in power, thus proving a strong dequantization result. If we only require that all functions computed by quantum formulas over G can be computed by classical formulas, then the result is trivial since we could just allow $\hat{G}$ to be the set of all boolean functions.

In this section, we show that read-once quantum formulas, for gate sets which include all single-qubit gates, can be strongly dequantized to classical formulas of the same size and depth. We then explicitly construct the classical gates sets corresponding to some particular quantum gate sets of interest. For example, any read-once quantum formula over all k-qubit channels is equivalent to a classical read-once formula over all k-bit functions. Similarly, read-once quantum formulas over Toffoli gates and all single-qubit gates can be dequantized to classical read-once formulas over Toffoli and NOT.

3.1 Dequantization of exact read-once formulas

We first prove the claim for exact read-once formulas. The proof is simpler in the exact case, and contains all of the essential ideas. Extension to bounded-error is discussed in Section 3.2.

Before attempting to prove the theorem, let us discuss the correspondence between the quantum gate set G and its classical counterpart $\hat{G}$. Call a quantum channel *classical* if the output is classical whenever the input is classical. If G contains a classical channel, it is clear that $\hat{G}$ should contain a gate that performs the same classical operation.

Some gates in G may be non-classical, but may be composed with other gates in a quantum formula to form a classical gate. Consider, for example, the depth-one quantum formula in Figure 2. Given a classical input, this formula outputs a classical bit and therefore computes a classical function f. For strong dequantization, we require that $\hat{G}$ admit a depth-one classical formula that computes f.

We therefore define $\hat{G}$ to contain only those gates that can be obtained by depth-one quantum formulas over G. Informally, this is the smallest gate set that would suffice to prove a strong dequantization result.

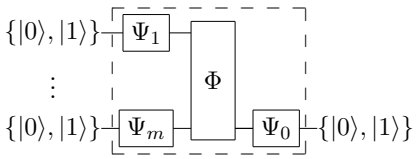


Figure 2 A depth-one quantum formula computes a classical function if on classical inputs it outputs a classical bit.

▶ **Definition 1.** Let G be a set of quantum channels. Define $\hat{G}$ to be the set of all classical gates that are computable by a *depth-one* quantum formula over G.

We will also need the following fact about density matrices and quantum channels.

▶ Fact 1. Let Φ be a single-qubit channel and ρ, σ be 2×2 density matrices. If $||\Phi(\rho) - \Phi(\sigma)||_1 = 2$ then ρ and σ are pure and orthogonal.

Proof. We use the fact that the trace distance is monotone non-increasing under the action of quantum channels [24, (9.72)] to conclude that $||\rho - \sigma||_1 \geq 2$. Thus $||\rho - \sigma||_1 = 2$, since the trace distance of two qubits cannot be larger than two [24, (9.11)]. Furthermore, the trace distance is maximum only when ρ and σ have support on orthogonal subspaces [24, (9.12)]. Since these are 2×2 matrices which act on a two-dimensional vector space, their supports can be at most one-dimensional if the supports are orthogonal. Thus they can be written in the form $\rho = |\psi\rangle\langle\psi|$, $\sigma = |\phi\rangle\langle\phi|$. Finally, since they have support on orthogonal subspaces, $\langle\psi|\phi\rangle = 0$, as claimed. ◀

We are now ready to state the main theorem.

▶ **Theorem 2.** *Let G be a set of quantum channels that includes all single-qubit channels. Then a language L is accepted by an exact read-once quantum formula of depth d and size s over G if and only if L is accepted by a read-once classical formula of depth d and size s over $\hat{G}$, where $\hat{G}$ is given by Definition 1.*

Proof. One direction of the "if and only if" is obvious; we only prove the other direction. The proof is by induction on the depth of the quantum formula. The claim is clearly true for depth-one since $\hat{G}$ contains all classical functions computable by depth-one quantum formulas over G.

Consider the final multi-qubit gate $\Phi \in G$ of a depth-d quantum formula that accepts language L. Φ may be followed by a single-qubit channel Ψ, the output of which is a classical state. Let m be the number of inputs to Φ.

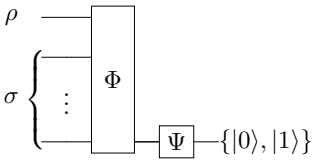


Figure 3 The final gates Φ and Ψ of a quantum formula.

Each input qubit to Φ is the output of a quantum sub-formula of depth at most $d - 1$ on a distinct subset of the original input bits. We cannot invoke the induction hypothesis directly on the depth-$(d-1)$ sub-formulas, however, because the outputs may be quantum states. We will show that each sub-formula can be replaced with different sub-formula of the same size and depth that outputs a classical bit.

Without loss of generality, let the first input of Φ depend on the input bits $\{x_1, \ldots, x_l\}$ for some $1 \leq l \leq n$. Let the state of the first input qubit, which is a function of $\{x_1, \ldots, x_l\}$, be called ρ. Let σ be the state of the remaining input qubits. Thus σ is a function of the remaining inputs $\{x_{l+1}, \ldots, x_n\}$.

Let P be the set of all distinct states ρ obtained by enumerating over all inputs $\{x_1, \ldots, x_l\}$. Our goal is to show that either P contains exactly two orthogonal pure states or the output Φ is independent of the first input qubit. If P contains exactly two orthogonal pure states, then there is a unitary channel Ψ_1 such that $\Psi_1(\rho) \in \{|0\rangle\langle 0|, |1\rangle\langle 1|\}$ for all $\rho \in P$. By composing Ψ_1 and Φ we may construct a new channel that accepts a classical bit as its first input. On the other hand, if the output is independent of the first input, we can consider a channel Ψ_1 that always outputs a fixed state independent of its input.

First, we look for two density matrices in P, which we will call ρ_1 and ρ_2, and a bit string $x_{l+1} \ldots x_n$ such that the induced state σ satisfies $\Phi(\rho_1 \otimes \sigma) \neq \Phi(\rho_2 \otimes \sigma)$. Since the

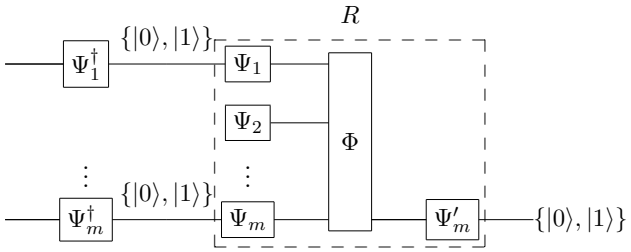


Figure 4 Single qubit channels are prepended to input qubits. If the input qubit depends on the input to the formula then both the channel and its adjoint are prepended. Otherwise, the input qubit is fixed and is replaced by a single channel. The channel Φ is then replaced by a classical gate $R \in \hat{G}$.

formula is read-once, bits $\{x_1, \ldots, x_l\}$ are independent of bits $\{x_{l+1}, \ldots, x_n\}$. Thus if there are no ρ_1, ρ_2 and σ that satisfy the above condition, then the output of Φ is independent of $\{x_1, \ldots, x_l\}$ and we may replace ρ by a channel Ψ_1 that on any input (and classical inputs in particular) outputs some fixed element of P. Note that this is the only place in the proof in which the read-once condition is required.

Assume that such a ρ_1, ρ_2 and σ exist. For this fixed σ, define $\Phi'(\rho) := \Psi(\Phi(\rho \otimes \sigma))$. The output of Φ' is a classical state and therefore ρ_1 and ρ_2 induce orthogonal classical outputs. Thus $||\Phi'(\rho_1) - \Phi'(\rho_2)||_1 = 2$ and by Fact 1, we know that ρ_1 and ρ_2 are pure and orthogonal. Let Ψ_1 be a unitary channel such that $\Psi_1(\rho_1) = |0\rangle\langle 0|$ and $\Psi_1(\rho_2) = |1\rangle\langle 1|$.

We now show that $|P| = 2$. That is, ρ_1 and ρ_2 are the *only* states in P. Assume that $|P| > 2$, and let $\rho_3 \in P$ be distinct from ρ_1, ρ_2 and such that $\Phi'(\rho_3) \neq \Phi'(\rho_1)$. Since $\Phi'(\rho)$ is classical for any $\rho \in P$, using Fact 1 we again have that ρ_1 and ρ_3 are pure and orthogonal. But since ρ_2 is the unique state orthogonal to ρ_1, we conclude that $\rho_3 = \rho_2$, a contradiction. Assuming that $\Phi'(\rho_3) \neq \Phi'(\rho_2)$ similarly leads to a contradiction.

A similar argument applies for the set of possible states on the remaining input qubits of Φ. For each qubit k there are two possible (pure) states ρ_0 and ρ_1 and a unitary channel Ψ_k such that $\Psi_k(\rho_0) = |0\rangle\langle 0|$ and $\Psi_k(\rho_1) = |1\rangle\langle 1|$, or the action of Φ is independent of qubit k and we may replace it with an input-independent channel Ψ_k. We now add the gates Ψ_k before gate Ψ on input qubit k. If Ψ_k was an input-independent channel, we have not changed the output of the circuit. If Ψ_k was a unitary channel that changes basis, we now need to add $\Psi_k^\dagger$ before it to ensure that the output is unchanged. The channel formed by the Ψ_ks, Φ and the output gate Ψ'_m is now classical and has the same action as a classical gate $R \in \hat{G}$ as shown in Figure 4.

The inputs to R are quantum sub-formulas of depth at most $d-1$, each of which outputs a classical bit. By induction, each sub-formula may be replaced by a classical formula of the same depth over $\hat{G}$. The resulting circuit is a depth-d classical formula over $\hat{G}$ that accepts language L. ◀

3.2 Extension to the bounded-error case

We now extend the main result to the bounded-error setting. The proof is essentially the same as that of Theorem 2, but does not require that the formula outputs be orthogonal states. This section can be safely skipped without loss on continuity.

There are several natural definitions of bounded-error quantum formulas. For example, we could say that a quantum formula over a gate set G computes a function f with error ϵ, if on input x, the output of the formula is ϵ-close to $|f(x)\rangle$ in trace distance. However, for

bounded-error, most reasonable definitions will be equivalent (up to constant factors). So we use the following definition which is more convenient for our proofs: A quantum formula over G computes f with error δ if all the output states corresponding to $f(x) = 0$ are at least $2 - \delta$ away in trace distance from all the output states corresponding to $f(x) = 1$.

This definition is equivalent to the previous one, up to constants, by noting that if one state is close to $|0\rangle$ and the other is close to $|1\rangle$, then they must necessarily by far apart, which can be proved using the triangle inequality. In the other direction, using a result of Gutoski and Watrous [10], it can be shown that two sets of states with a lower bound on the minimum pairwise distance can be distinguished with high probability. Note that when $\delta = \epsilon = 0$, the definitions coincide, except that the second definition only requires the output states to be orthogonal, not necessarily $|0\rangle$ and $|1\rangle$. Since our gate sets contain all single-qubit gates, this distinction is not important. The definition of $\hat{G}$ will also have to be similarly modified to incorporate bounded-error quantum formulas over G.

In the exact case, the output of a formula is always either $|0\rangle$ or $|1\rangle$. Furthermore, we showed that even the output of sub-formulas is classical, up to a change of basis. We now want to prove a similar claim for bounded-error sub-formulas. We wish to show that if the function depends on the output of a sub-formula, then the set of output states of the sub-formula can be partitioned into two non-empty sets, S_0 and S_1, such that the trace distance between any pair of states in S_0 and S_1 is at least $2 - \delta$.

Now we can state the bounded-error analog of Theorem 2. Note that the definition of $\hat{G}$ used in this theorem is different from that in Theorem 2, since it allows functions to be computed with bounded error.

▶ **Theorem 3.** *Let G be a set of quantum channels that includes all single-qubit channels. If a language is accepted by a bounded-error read-once quantum formula over G then it is also accepted by an exact read-once classical formula over $\hat{G}$, of the same size and depth, where $\hat{G}$ is defined as the set of all classical gates that can be computed by a depth-one bounded-error read-once quantum formula.*

Proof. The proof proceeds like the proof of Theorem 2. We use the same notation as in the other proof. As before, we consider the first input of the last multi-qubit gate, and assume it depends on the input bits $\{x_1, x_2, \ldots, x_l\}$, and let P denote the set of all states obtained by enumerating over the the input bits $\{x_1, x_2, \ldots, x_l\}$. Denote the state on the rest of the inputs to the last gate as σ.

First we look for two states in P, ρ and ρ', and a bit string $x_{l+1} \ldots x_n$ which induces a state σ on the other inputs such that $\rho \otimes \sigma$ and $\rho' \otimes \sigma$ lead to different outputs of the formula. If no such ρ, ρ' and σ exist, then the output of the last gate is independent of $\{x_1, \ldots, x_l\}$ and we may replace the first input qubit by a channel that on any input (and classical inputs in particular) outputs some fixed element of P.

Assume that such a ρ, ρ' and σ exist. As before, for this fixed σ, define $\Phi'(\rho) := \Psi(\Phi(\rho \otimes \sigma))$. Without loss of generality, assume that ρ leads to output 0 and ρ' leads to output 1. Let $S_b := \{\rho \in P : \rho \text{ leads to output } b\}$. Then for any states $\rho_0 \in S_0$ and $\rho_1 \in S_1$ we must have $||\Phi'(\rho_0) - \Phi'(\rho_1)||_1 \geq 2 - \delta$. Using the monotonicity of trace distance under quantum channels [24, (9.72)], this implies $||\rho_0 - \rho_1||_1 \geq 2 - \delta$, which means all the states in S_0 are far from all the states in S_1. Thus the first sub-formula satisfies the definition of a bounded-error quantum formula. By the induction hypothesis, there is a classical formula over $\hat{G}$ of the same size and depth that outputs 0 when the quantum sub-formula would have output some state in S_0 and outputs 1 when it would have output some state in S_1.

Now we wish to show that all the states in S_b are equivalent from the perspective of the last gate. More precisely, let ρ_0 and ρ_0' be two different states in S_0, and let σ be some

valid state induced by $x_{l+1}, \ldots x_n$. Let us also partition the output states of the last gate into two sets T_0 and T_1, for which the trace distance between any pair of states in the two sets is at least $2 - \delta$. We wish to show that if the output corresponding to $\rho_0 \otimes \sigma$ is in T_b, then so is the output corresponding to $\rho_0' \otimes \sigma$. States $\rho_0 \otimes \sigma$ and $\rho_0' \otimes \sigma$ arise due to valid input strings x and x', respectively. Thus the corresponding outputs must be in T_0 or T_1. Since the output of $\rho_0 \otimes \sigma$ is in T_b, and since ρ_0 cannot be too far from ρ_0', the output of $\rho_0' \otimes \sigma$ is also in T_b. More precisely, since both ρ_0 and ρ_0' are at least $2 - \delta$ away from some state $\rho_1 \in S_1$, we know that they can be at most 2δ distance apart. Thus, by monotonicity of trace distance under quantum channels, their corresponding outputs cannot be too far apart.

Now we know that all the states in S_b are equivalent from the perspective of the last gate. We also know that there is a classical formula over $\hat{G}$ of the same size and depth that outputs 0 when the quantum sub-formula would have output some state in S_0 and outputs 1 when it would have output some state in S_1. We can just add a single-qubit gate to the output of this sub-formula, which on input $|0\rangle$ outputs some fixed state ρ_0 from S_0 and on input $|1\rangle$ outputs a fixed state ρ_1 from S_0. This formula also computes the same function $f(x)$, as shown in the previous paragraph.

Continuing this on all the inputs to the last gate, we obtain a depth-one quantum formula which accepts as inputs the outputs of a classical depth $d-1$ formula. Using the definition of $\hat{G}$, there is a classical gate that computes the same function, which gives us a classical depth-d formula for the entire function. ◀

3.3 Application to concrete gate sets

A simple corollary of Theorem 2 is that L is accepted by a read-once quantum formula over the set of all k-qubit channels (for some constant k) if and only if it is accepted by a read-once classical formula over the set of all k-bit functions. This is the gate set used in the previous studies of quantum formulas [25, 16].

Another gate set of interest is the set of arbitrary fanin Toffoli gates and all single-qubit gates. This gate set is commonly used in the study of quantum circuits [9]. We now explicitly construct $\hat{G}$ for this gate set and show that it only contains classical arbitrary fanin Toffoli gates and the NOT gate.

Let us define Φ_m^{Tof} to be the quantum channel obtained from a Toffoli gate with $m \geq 2$ qubits by tracing out all the $m-1$ control qubits. We can assume that all Toffoli gates that appear in the formula always output the target qubit and trace out the control qubits because by conjugating the target qubit and a control qubit with the Hadamard matrix, H, it is possible to exchange the roles of the target and that control qubit.

To compute $\hat{G}$ for the set of all single-qubit gates and Φ_m^{Tof} for $m \geq 2$ it suffices to list all possible classical gates that can be obtained from Φ_m^{Tof} by placing single-qubit channels before and after it. The classical m-bit Toffoli gate computes the function $f(x_1, x_2, \ldots, x_m) = (x_1 \wedge x_2 \wedge \ldots \wedge x_{m-1}) \oplus x_m$. Let F_m^{Tof} be the set of all functions obtained from the classical m-bit Toffoli by placing single-bit gates before and after it. The only single-bit gates are NOT gates, and channels that output a constant bit.

▶ **Theorem 4.** *Let f be any function on $x \in \{0,1\}^m$ that can be obtained by placing single-qubit gates before and after Φ_m^{Tof}. Then $f \in F_m^{Tof}$.*

Proof. Let the classical function f be defined by single-qubit channels $\Psi_1, \ldots, \Psi_m$ on the inputs followed by Φ_m^{Tof} and a single-qubit quantum channel Ψ_m' on the output qubit.

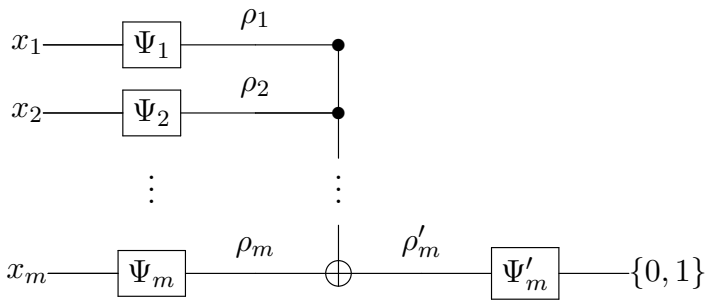


Figure 5 A depth-one quantum formula over single-qubit gates and an m-bit Toffoli gate.

Let $\{\rho_1, \ldots, \rho_m\}$ be the inputs to Φ_m^{Tof} (i.e., the outputs of $\Psi_1, \ldots, \Psi_m$) induced by a particular choice of x, and ρ'_m be the output of the channel (i.e., the input to Ψ'_m). See Figure 5.

First let us handle the trivial cases. If Ψ_m is a channel which always outputs an eigenstate of X, i.e., $X\rho_m X = \rho_m$ for $x_m = 0$ and $x_m = 1$, then the Toffoli gate leaves it unaffected. Thus the output ρ'_m is only a function of x_m, and since the output of the gate is classical, it is some classical one-bit function of x_m, all of which are contained in F_m^{Tof}. We may now assume this is not the case. Let us also assume that f outputs 0 on some input and 1 on some input, otherwise it is a constant function.

For any $x, x' \in \{0,1\}^{m-1}$, define $\alpha(x, x') = \prod_{i=1}^{m-1} \rho_i(x_i, x'_i)$. Furthermore, let $0 \leq a \leq 1$ be defined as $a = \alpha(11\ldots1, 11\ldots1)$. Then the input to Φ_m^{Tof} is $\rho_1 \otimes \rho_2 \otimes \ldots \otimes \rho_m = \sum_{x,x'} \alpha_{xx'} |x\rangle\langle x'| \otimes \rho_m$. The output of Φ_m^{Tof} is given by

$$\begin{aligned}
\rho'_m &= \text{Tr}_{1,\ldots,m-1}\left(\sum_{x,x'} \alpha_{xx'} |x\rangle\langle x'| \otimes X^{\text{AND}(x)} \rho_m X^{\text{AND}(x')}\right) \\
&= \sum_x \alpha_{xx} X^{\text{AND}(x)} \rho_m X^{\text{AND}(x)} \\
&= \left[\prod_{i=1}^{m-1} \rho_i(1,1)\right] X\rho_m X + \left[\sum_{x \neq 1\ldots1} \prod_{i=1}^{m-1} \rho_i(x_i, x_i)\right] \rho_m \\
&= aX\rho_m X + (1-a)\rho_m. \qquad (1)
\end{aligned}$$

By assumption, $\Psi'_m(\rho'_m)$ is classical and both outputs 0 and 1 are possible, thus by Fact 1, ρ'_m must always be pure. Since we have also assumed that $X\rho_m X \neq \rho_m$, (1) implies that ρ'_m can be pure only if $a = 0$ or $a = 1$. But a takes values 0 or 1 on all inputs, so the following short argument implies that each ρ_i must be classical (i.e., $\rho_i(0,0) = 0$ or $\rho_i(0,0) = 1$) for every input x. Toward a contradiction, assume this is not the case. Thus, for some i, there exists an input x_i so that ρ_i is not classical and, in particular, $\rho_i(1,1) \neq 0$ and $\rho_i(1,1) \neq 1$. On this input x, $\prod_{i=1}^{m-1} \rho_i(1,1)$ cannot be 0 or 1. But this quantity is a, which could only be 0 or 1. Thus we have reached a contradiction, and all the ρ_i must be classical for every input x_i. But this means every channel Ψ_i is a classical one-bit function of x_i. Thus the inputs to Φ_m^{Tof} are always classical and so Φ_m^{Tof} can be replaced with its classical equivalent, which completes the proof. ◀

This theorem shows that the only classical gates obtainable from Φ_m^{Tof} by placing single-qubit channels before and after it are the classical Toffoli gate of size m and gates derived

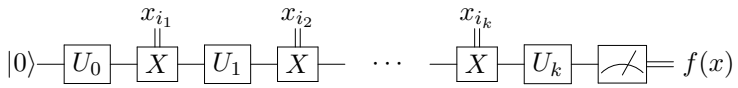


Figure 6 A one-qubit program of length k.

from it using single-bit gates. In particular, this means that if the gate set G contains all single-qubit channels and the CNOT gate, which is the Toffoli gate for $m = 2$, then $\hat{G}$ contains only the CNOT gate and the NOT gate. We use this result in Section 4.

4 One-qubit model

Informally, Theorem 2 and Theorem 3 show that read-once quantum formulas are equivalent to classical read-once formulas. We now show that the claim is no longer true if we drop the read-once constraint. To this end, we introduce a new model, which we call the *one-qubit* model. This model is independent of the previous sections, and may be of interest outside of the context of quantum formulas.

The one-qubit model consists of a single qubit initialized to $|0\rangle$ followed by an alternating sequence of single-qubit unitaries and CNOT (controlled-X) gates. The control of each CNOT is a bit x_i of the input x. The output is determined by a measurement in the standard basis. See Figure 6. We call an algorithm of this kind a one-qubit program. We say that a one-qubit program exactly computes a function f if the output of the measurement is $f(x)$ with probability one on input x.

The *length* of a one-qubit program is defined as the total number of CNOT gates. For instance, the program in Figure 6 has length k. Notice that the model is unchanged if, instead of CNOT, we choose any controlled-V gate such that $UVU^\dagger = X$ for some single-qubit unitary U. Note also that input bits may be re-used as many times as desired.

We prove that the one-qubit model is universal, that is, it can compute all boolean functions. Specifically, if a boolean function has a depth-d circuit over fanin-2 AND and OR, and NOT gates, then it has a one-qubit program of length 4^d. Here, the depth of a circuit is defined as the maximum number of AND or OR gates on any path from the output to an input. Note that all boolean functions can be expressed as circuits of polynomial depth, thus one-qubit programs of exponential length can compute all boolean functions. Moreover, one-qubit programs with polynomial length can exactly compute any function in NC^1, the set of functions computed by log-depth poly-size bounded-fanin circuits over AND, OR and NOT. Our proof resembles the original proof of Barrington's theorem [3], a surprising result in complexity theory, which showed that bounded-width branching programs can compute any function in NC^1.

▶ **Theorem 5.** *The one-qubit model can compute all boolean functions. More precisely, any function that has a depth-d circuit over fanin-2* AND *and* OR*, and* NOT *gates can be computed exactly by a one-qubit program of length 4^d.*

Proof. Use the notation X^x to denote a CNOT gate that is controlled by the variable x. Let C be a circuit of depth d that computes a function $C(x)$. Starting from the circuit C, we construct a one-qubit program F of length 4^d that computes the same function. That is, $F = X^{C(x)}$ so that $F|0\rangle = |C(x)\rangle$.

We prove the claim by induction on the structure of the circuit. Circuit C can be seen as a binary tree with the input variables as the leaves of the tree and the gates as the internal nodes. Given a gate at the root of the circuit (AND or NOT), we assume that the induction

hypothesis holds for the subcircuits that produce the inputs of the gate and we prove the claim for the entire circuit.

Let us start with the base case. If C is a single variable x_i, the corresponding one-qubit program is a single CNOT gate controlled by x_i, i.e., $F = X^{x_i}$.

The induction step consists of two cases. Suppose that C is composed of a NOT gate and a subcircuit C', and let F' be the one-qubit program of length 4^d that simulates C'. Then we can append a Pauli-X gate to F' and obtain the program F of the same length 4^d that computes C, i.e., $F = XX^{C'(x)} = X^{\bar{C}'(x)} = X^{C(x)}$. Note that the depth of circuits C and C' is the same, since NOT gates do not contribute to the depth. Also note that single-qubit unitaries do not contribute to the length of the program, and thus the length is unchanged.

Now suppose C is composed of an AND gate that connects two subcircuits C' and C'' and let F' and F'' be the programs of length 4^{d-1} that compute C' and C'', respectively. Then consider the following program F described as an equation:

$$F = VX^{C'(x)}H^{C''(x)}X^{C'(x)}H^{C''(x)}V = V(iY)^{C'(x)\wedge C''(x)}V = iX^{C(x)}, \tag{2}$$

where $V = \frac{1}{\sqrt{2}}(X+Y)$ is the unitary that satisfies $V^2 = \mathbb{1}$, $VYV = X$ and $VXV = Y$.

This program effectively computes the AND of two sub-programs. It starts with a V gate, followed by the subprogram F', which is equivalent to $X^{C'(x)}$. Then we need a subprogram that performs $H^{C''(x)}$. By the induction hypothesis, we have a subprogram F'' that performs $X^{C''(x)}$. Since the Hadamard matrix and Pauli X gate are unitarily equivalent, there is a unitary matrix R such that $RXR = H$ and $RHR = X$. Conjugating $X^{C''(x)}$ with R gives us $RX^{C''(x)}R$, which is the same as $(RXR)^{C''(x)}$, which is $H^{C''(x)}$. The other gates in F are constructed similarly.

Thus the program F requires some single qubit gates, two copies of the program for F' and two copies of the program for F''. Since F' and F'' have length at most 4^{d-1} and single qubit unitaries do not count towards the length, we get a program F of length 4^d, which performs $X^{C(x)}$ up to an irrelevant global phase.

These two cases suffice to prove the theorem since we can replace the OR gates in the circuit by AND and NOT gates, without increasing its size or depth. ◀

In the case of NC^1, the depth of the circuit is at most $O(\log n)$, therefore the length of the resulting one-qubit program is polynomial.

▶ **Corollary 6.** *Any function in* NC^1 *can be computed exactly by a one-qubit program of polynomial length.*

Now that we have shown that the one-qubit model can compute all boolean functions, it remains to show that our dequantization claim is false without the read-once constraint. First observe that the one-qubit model is a restricted model of quantum formulas over the gate set G consisting of all single-qubit gates and the CNOT gate.

We know from Theorem 4 (case $m = 2$) that, in this case, the classical gate set $\hat{G}$ consists of only the NOT gate and PARITY gate. But no formula and, indeed, no *circuit* of any size over $\hat{G}$ can compute the AND of two bits, since NOT and PARITY do not form a complete basis. In particular, circuits over $\hat{G}$ can only compute functions expressible as degree-one polynomials over $\mathbb{F}_2$. Thus we conclude that Theorem 2 and Theorem 3 are false if the read-once constraint is dropped.

Readers familiar with quantum branching programs [14, 17, 1] may notice that the one-qubit model is contained in exact width-two quantum branching programs. However, in our model the only input-dependent gate is a Pauli-X, whereas in quantum branching programs, we can apply arbitrary input-controlled unitaries. By conjugating the X gate with

arbitrary single qubit gates the one-qubit model can obtain any unitary matrix with the same eigenvalues as X, but not an arbitrary unitary matrix. Thus it is not clear that results about width-two quantum branching programs can be ported over to our model.

5 Conclusions and open problems

We have shown that read-once quantum formulas over any gate set are only as powerful as read-once classical formulas over a related gate set. As a concrete example, we showed that read-once quantum formulas over Toffoli and single-qubit gates dequantize to the natural analog, read-once classical formulas over Toffoli and NOT gates. Perhaps our results may be extended to constant-depth quantum circuit classes, many of which are defined over Toffoli and single-qubit gates, e.g., [9]. Our proof technique fails when the formula restriction is lifted, but for constant-depth it may be possible to reuse some of the same ideas.

Another obvious open problem is to dequantize all quantum formulas, not just read-once formulas. Although we show that our classical gate set is insufficient to do so, there might be another classical gate set that works.

Finally, our dequantization result implies that lower bounds on read-once quantum formulas may be obtained from analogous classical lower bounds. For related models, including general (i.e., not read-once) quantum formulas and constant-depth quantum circuits, the problem of finding non-trivial lower bounds remains.

Acknowledgements. The authors would like to thank Andrew Childs, Ben Reichardt and John Watrous for insightful comments.

This work was supported by Canada's NSERC, MITACS, the Ontario Ministry of Research and Innovation, the U.S. ARO, and the Mprime Network.

References

1 Farid Ablayev, Aida Gainutdinova, Marek Karpinski, Cristopher Moore, and Christopher Pollett. On the computational power of probabilistic and quantum branching program. *Information and Computation*, 203(2):145–162, 2005.

2 Dorit Aharonov and Amnon Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge. In *Proceedings of the 35th annual ACM symposium on Theory of computing*, STOC '03, pages 20–29, 2003.

3 David A. Mix Barrington. Bounded-width polynomial-size branching programs recognize exactly those languages in NC^1. *Journal of Computer and System Sciences*, 38:150–164, 1989.

4 Harry Buhrman, Richard Cleve, Monique Laurent, Noah Linden, Alexander Schrijver, and Falk Unger. New limits on fault-tolerant quantum computation. In *Proceedings of the 47th Annual Symposium on Foundations of Computer Science*, FOCS '06, pages 411–419, 2006.

5 Richard Cleve and John Watrous. Fast parallel circuits for the quantum fourier transform. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, FOCS '00, pages 526–, 2000.

6 Stephen Fenner, Frederic Green, Steven Homer, and Yong Zhang. Bounds on the power of constant-depth quantum circuits. In *Proceedings of the 15th international conference on Fundamentals of Computation Theory*, FCT'05, pages 44–55, 2005.

7 Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6):467–488, 1982.

8 Daniel Gottesman. The Heisenberg Representation of Quantum Computers. *arXiv preprint quant-ph/9807006*, 1998.

9 Frederic Green, Steven Homer, Cristopher Moore, and Christopher Pollett. Counting, fanout and the complexity of quantum ACC. *Quantum Information & Computation*, 2(1):35–65, December 2002.

10 Gus Gutoski and John Watrous. Quantum interactive proofs with competing provers. In *STACS 2005*, volume 3404 of *Lecture Notes in Computer Science*, pages 605–616. Springer Berlin Heidelberg, 2005.

11 Stasys Jukna. *Boolean Function Complexity: Advances and Frontiers.* Algorithms and Combinatorics. Springer, 2012.

12 Emanuel Knill and Raymond Laflamme. Power of one bit of quantum information. *Physical Review Letters*, 81:5672–5675, 1998.

13 Seth Lloyd. Universal quantum simulators. *Science*, 273(5278):1073–1078, 1996.

14 Masaki Nakanishi, Kiyoharu Hamaguchi, and Toshinobu Kashiwabara. Ordered quantum branching programs are more powerful than ordered probabilistic branching programs under a bounded-width restriction. In *Proceedings of the 6th Annual International Conference on Computing and Combinatorics*, COCOON '00, pages 467–476, 2000.

15 Michael A. Nielsen and Isaac L. Chuang. *Quantum computation and quantum information.* Cambridge University Press, 2000.

16 Vwani P. Roychowdhury and Farrokh Vatan. Quantum formulas: A lower bound and simulation. *SIAM Journal on Computing*, 31(2):460–476, 2002.

17 Martin Sauerhoff and Detlef Sieling. Quantum branching programs and space-bounded nonuniform quantum complexity. *Theoretical Computer Science*, 334(1-3):177–225, April 2005.

18 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.

19 Peter W. Shor and Stephen P. Jordan. Estimating Jones polynomials is a complete problem for one clean qubit. *Quantum Information & Computation*, 8(8):681–714, September 2008.

20 Barbara M. Terhal and David P. DiVincenzo. Adptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *Quantum Information & Computation*, 4(2):134–145, 2004.

21 Leslie G. Valiant. Quantum circuits that can be simulated classically in polynomial time. *SIAM Journal on Computing*, 31(4):1229–1254, 2002.

22 Maarten Van den Nest. Quantum matchgate computations and linear threshold gates. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, 467(2127):821–840, 2011.

23 Ingo Wegener. *The complexity of Boolean functions.* John Wiley & Sons, Inc., 1987.

24 Mark M. Wilde. From classical to quantum Shannon theory. *arXiv preprint 1106.1445*, 2011.

25 Andrew C. Yao. Quantum circuit complexity. In *Proceedings of the 34th Annual Foundations of Computer Science*, SFCS '93, pages 352–361, 1993.

The Minimum Size of Qubit Unextendible Product Bases

Nathaniel Johnston

Institute for Quantum Computing, University of Waterloo
Waterloo, Ontario N2L 3G1, Canada
nathaniel.johnston@uwaterloo.ca

Abstract

We investigate the problem of constructing unextendible product bases in the qubit case – that is, when each local dimension equals 2. The cardinality of the smallest unextendible product basis is known in all qubit cases except when the number of parties is a multiple of 4 greater than 4 itself. We construct small unextendible product bases in all of the remaining open cases, and we use graph theory techniques to produce a computer-assisted proof that our constructions are indeed the smallest possible.

1998 ACM Subject Classification G.2.3 Applications

Keywords and phrases unextendible product basis, quantum entanglement, graph factorization

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.93

1 Introduction

Unextendible product bases play a rather diverse and important role in quantum information theory [7]. While their original motivation was for the construction of bound entangled states [5, 12, 13], they have also been used to build indecomposible positive maps [14], to demonstrate Bell inequalities without a quantum violation [3], and demonstrate the existence of nonlocality without entanglement [4].

Furthermore, in the qubit case (i.e., the case where each local space has dimension 2), it has been shown that unextendible product bases can be used to construct tight Bell inequalities with no quantum violation [2] and subspaces of small dimension that are locally indistinguishable [8]. It is the qubit case that we focus on in the present paper. In particular, we consider the question of how small a qubit unextendible product basis can be.

The minimum cardinality of a qubit unextendible product basis on p qubits is well-known to equal $p+1$ when p is odd [1]. When p is even, however, the problem is more difficult. It was shown in [9] that the minimum cardinality equals $p+2$ when $p=4$ or $p \equiv 2 \pmod 4$. Our contribution is to solve the remaining cases (i.e., when $p \geq 8$ and $p \equiv 0 \pmod 4$ – more specifically, we show that the minimum cardinality is $p+3$ when $p=8$ and $p+4$ in all other cases.

Our approach is as follows: we formally introduce the mathematical preliminaries and graph theory techniques that we make use of in Section 2. We construct unextendible product bases of the claimed cardinality in Section 3. Finally, Section 4 is devoted to the proof that there does not exist a smaller unextendible product basis in these cases.

2 Unextendible Product Bases and Orthogonality Graphs

A pure quantum state is represented by a unit vector $|v\rangle \in \mathbb{C}^{d_1} \otimes \cdots \otimes \mathbb{C}^{d_p}$ (and in our setting, $d_1 = \cdots = d_p = 2$ always). We say that $|v\rangle$ is a *product state* if we can write it in

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 93–105
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

the form

$$|v\rangle = |v_1\rangle \otimes \cdots \otimes |v_p\rangle \quad \text{with} \quad |v_j\rangle \in \mathbb{C}^2 \ \forall\, j.$$

An *unextendible product basis (UPB)* is an orthonormal set $\mathcal{S} \subseteq (\mathbb{C}^2)^{\otimes p}$ of product states such that there is no product state orthogonal to every member of $\mathcal{S}$. It is clear that every UPB in $(\mathbb{C}^2)^{\otimes p}$ contains at least $p+1$ states – if it contained only p product states $|v_0\rangle, \ldots, |v_{p-1}\rangle$ then we could construct another product state that is, for each $0 \leq j < p$, orthogonal to $|v_j\rangle$ on the $(j+1)$-th party and thus violate unextendibility.

It turns out that the trivial lower bound of $p+1$ states can be attained when p is odd, and can almost be attained when p is even, as indicated by our main result:

▶ **Theorem 1.** *Let $f(p)$ be the smallest possible number of states in a UPB in $(\mathbb{C}^2)^{\otimes p}$. Then:*
(a) *if p is odd then $f(p) = p+1$;*
(b) *if $p = 4$ or $p \equiv 2\,(mod\ 4)$ then $f(p) = p+2$;*
(c) *if $p = 8$ then $f(p) = p+3$;*
(d) *otherwise, $f(p) = p+4$.*

Case (a) of Theorem 1 is demonstrated by the "GenShifts" UPB constructed in [7]. Case (b) of Theorem 1 was proved in [9], and in general our techniques and presentation are similar to those of that paper. Our contribution is to prove cases (c) and (d) and hence complete the characterization. It is worth pointing out that cases (c) and (d) of Theorem 1 are the first known cases (qubit or otherwise) where the minimum cardinality of a UPB exceeds the trivial lower bound $1 + \sum_j (d_j - 1)$ by more than 1 (see [6, 9] for several examples where the trivial lower bound is exceeded by exactly 1).

Orthogonality graphs provide a very useful tool when dealing with unextendible product bases, particularly in the qubit case. Given a set of product states $\mathcal{S} = \{|v_0\rangle, \ldots, |v_{s-1}\rangle\} \subseteq (\mathbb{C}^2)^{\otimes p}$ with $|\mathcal{S}| = s$, we say that the *orthogonality graph of $\mathcal{S}$* is the graph on s vertices $V := \{v_0, \ldots, v_{s-1}\}$ such that there is an edge (v_i, v_j) of color ℓ if and only if $|v_i\rangle$ and $|v_j\rangle$ are orthogonal to each other on party ℓ. Rather than actually using p colors to color the edges of the orthogonality graph, for ease of visualization we instead draw p different graphs on the same set of vertices – one for each party (see Figure 1).

The requirement that $\mathcal{S}$ is an orthonormal set is equivalent to requiring that every edge is present on at least one party in its orthogonality graph. In order to help us visualize the unextendibility requirement, we make a few more observations. In particular,

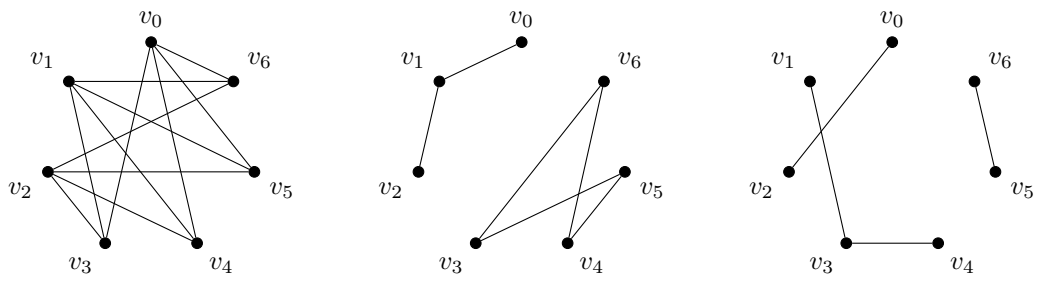


Figure 1 The orthogonality graph of a set of 7 product states in $(\mathbb{C}^2)^{\otimes 3}$. This set of states is a product basis, since every edge is present in at least one of the three graphs, but it is extendible, since we can find a product state that is orthogonal to the states associated with v_3, v_4, v_5, v_6 on the first subsystem, v_0, v_2 on the second subsystem, and v_1 on the third subsystem.

if $|w_0\rangle, |w_1\rangle, |w_2\rangle \in \mathbb{C}^2$ are such that $\langle w_0|w_1\rangle = \langle w_0|w_2\rangle = 0$, then it is necessarily the case that $|w_1\rangle = |w_2\rangle$ (up to irrelevant complex phase). It follows that the orthogonality graph associated with any qubit in a product basis is the disjoint union of complete bipartite graphs. For example, in Figure 1 the left graph is $K_{3,4}$, the center graph is the disjoint union of $K_{1,2}$ and $K_{2,2}$, and the right graph is the disjoint union of $K_{1,2}$ and two copies of $K_{1,1}$.

Furthermore, not only does every set of product states have an orthogonality graph that can be decomposed into the disjoint union of complete bipartite graphs, but the converse is also true: every graph that is built from complete bipartite graphs in this way is the orthogonality graph of some set of product states. To see this, on each party assign to each complete bipartite graph a distinct basis of $\mathbb{C}^2$ in the obvious way. For example, one set of product states giving rise to the orthogonality graph depicted in Figure 1 is as follows:

$$|v_0\rangle := |0\rangle \otimes |0\rangle \otimes |0\rangle, \quad |v_1\rangle := |0\rangle \otimes |1\rangle \otimes |+\rangle, \quad |v_2\rangle := |0\rangle \otimes |0\rangle \otimes |1\rangle,$$
$$|v_3\rangle := |1\rangle \otimes |+\rangle \otimes |-\rangle, \quad |v_4\rangle := |1\rangle \otimes |+\rangle \otimes |+\rangle, \quad |v_5\rangle := |1\rangle \otimes |-\rangle \otimes |b\rangle,$$
$$|v_6\rangle := |1\rangle \otimes |-\rangle \otimes |b^\perp\rangle,$$

where $|+\rangle := \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$, $|-\rangle := \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, and $\{|b\rangle, |b^\perp\rangle\}$ is any orthonormal basis of $\mathbb{C}^2$ not equal to $\{|0\rangle, |1\rangle\}$ or $\{|+\rangle, |-\rangle\}$.

It is often useful to draw orthogonality graphs of sets of qubit product states in a form that makes their decomposition in terms of complete bipartite graphs more transparent – we draw shaded regions indicating which vertices are equal to each other (up to complex phase) on the given party, and lines between shaded regions indicate that all states in one of the regions are orthogonal to all states in the other region on that party (see Figure 2).

It now becomes straightforward to see whether or not a product basis is unextendible just by looking at its orthogonality graph. A set of product states is unextendible if and only if there is no way to choose one shaded region on each party such that every vertex $v_0, v_1, \ldots, v_{s-1}$ is contained within at least one of the shaded regions. For example, the set of product states described by Figure 2 is extendible because we can choose the shaded region containing v_3, v_4, v_5, v_6 on the first subsystem, v_0, v_2 on the second subsystem, and v_1, v_4 on the third subsystem.

The following simple lemma shows that, in an orthogonality graph of a UPB, every shaded region must be connected to exactly one other shaded region via an edge.

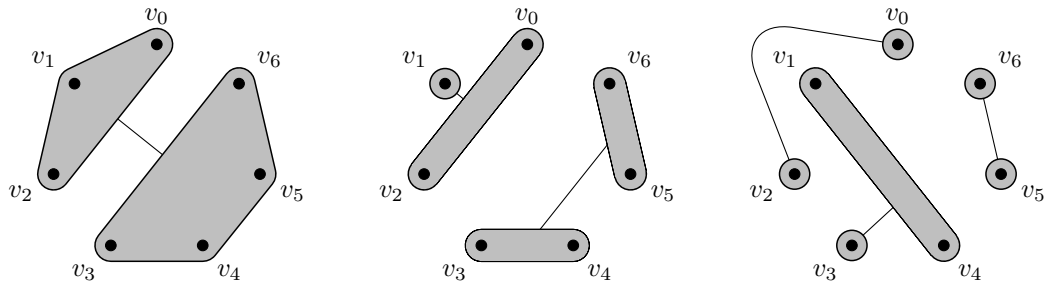


Figure 2 A representation of the same orthogonality graph as that of Figure 1. Vertices within the same shaded region represent states that are equal to each other on that party. Lines between shaded regions indicate that every state within one of the regions is orthogonal to every state within the other region.

▶ **Lemma 2.** *If $\mathcal{S} \subseteq (\mathbb{C}^2)^{\otimes p}$ is a UPB, then for all $|v\rangle \in \mathcal{S}$ and all integers $1 \leq j \leq p$ there is another product state $|w\rangle \in \mathcal{S}$ such that $|v\rangle$ and $|w\rangle$ are orthogonal on the j-th subsystem.*

Proof. Suppose that there exists $1 \leq j \leq p$ and $|v\rangle := |v_{(1)}\rangle \otimes \cdots \otimes |v_{(p)}\rangle \in \mathcal{S}$ such that $|v\rangle$ is not orthogonal to any other member of $\mathcal{S}$ on the j-th subsystem. Because $\mathcal{S}$ is a product basis, $|v\rangle$ must be orthogonal to every member of $\mathcal{S}$ on the remaining $p-1$ subsystems. It follows that if $|v_{(j)}^{\perp}\rangle$ is orthogonal to $|v_{(j)}\rangle$ then the product state $|v_{(1)}\rangle \otimes \cdots |v_{(j-1)}\rangle \otimes |v_{(j)}^{\perp}\rangle \otimes |v_{(j+1)}\rangle \otimes \cdots \otimes |v_{(p)}\rangle$ is orthogonal to every element of $\mathcal{S}$, which shows that $\mathcal{S}$ is extendible. ◀

An obvious corollary of Lemma 2 is that, in the orthogonality graph of a UPB, every party must have an even number of distinct shaded regions – a fact that will be very useful in Section 4.

3 Construction of Small UPBs

Recall that our goal is to show that the smallest UPB in $(\mathbb{C}^2)^{\otimes 8}$ consists of 11 states and the smallest UPB in $(\mathbb{C}^2)^{\otimes 4k}$ consists of $4k+4$ states when $k \geq 3$. Our first step toward this goal is to construct a UPB of the desired size in these cases.

▶ **Lemma 3.** *There exists a UPB in $(\mathbb{C}^2)^{\otimes 8}$ consisting of 11 states.*

Proof. The result follows simply from demonstrating an orthogonality graph on 11 vertices that satisfies the product basis and unextendibility requirements described in Section 2. Such an orthogonality graph is provided in Figure 3.

Indeed, it is straightforward (albeit tedious) to check that the 8 graphs depicted in Figure 3 contain all 55 possible edges between 11 vertices, so the corresponding product states are mutually orthogonal. Unextendibility follows from the (also straightforward but tedious) fact that there is no way to choose a shaded region containing 2 vertices on 3 different parties without at least 2 of them containing the same vertex. ◀

We note that the UPB of Lemma 3 was found by a combination of computer search and tweaking by hand, and it does not seem to generalize to other values of p in any natural way. On the other hand, the UPBs that we now construct of cardinality $4k+4$ are much "tidier".

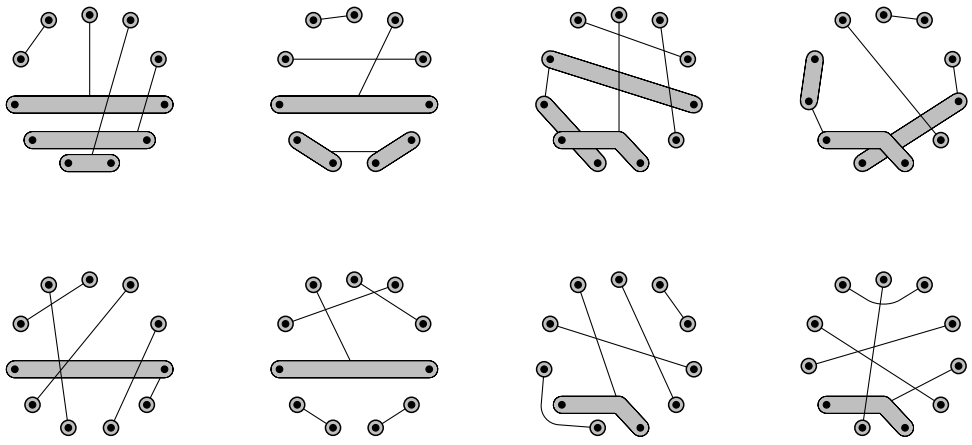


Figure 3 Orthogonality graphs demonstrating that there exists an 11-state UPB in $(\mathbb{C}^2)^{\otimes 8}$.

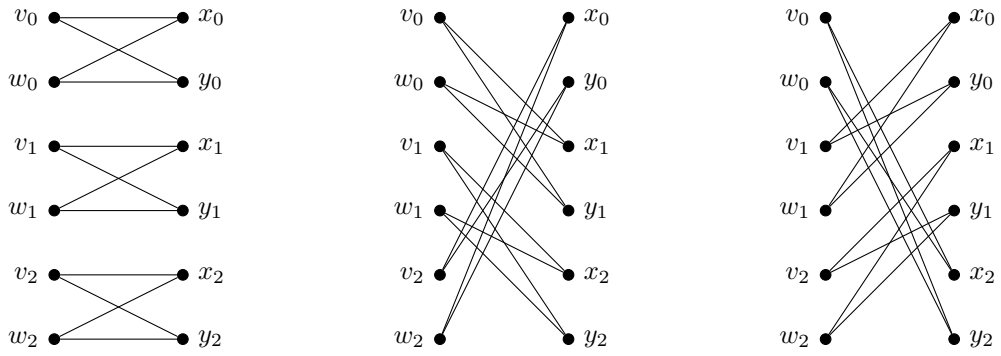


Figure 4 The graphs $B_{0,2}$ (left), $B_{1,2}$ (center), and $B_{2,2}$ (right), used in the construction of a UPB of size 12 in $(\mathbb{C}^2)^{\otimes 8}$.

▶ **Lemma 4.** *If $k \geq 2$ then there exists a UPB in $(\mathbb{C}^2)^{\otimes 4k}$ consisting of $4k+4$ states.*

Proof. We begin by defining a family of $k+1$ graphs $B_{j,k} := (V, E_j)$ for $0 \leq j \leq k$, each on the same set of $4k+4$ vertices $V := \{v_i, w_i, x_i, y_i, : 0 \leq i \leq k\}$. The set of edges E_j in the graph $B_{j,k}$ is defined as follows:

$$E_j := \big\{(v_i, x_{(i+j)(\text{mod } (k+1))}), (v_i, y_{(i+j)(\text{mod } (k+1))}),$$
$$(w_i, x_{(i+j)(\text{mod } (k+1))}), (w_i, y_{(i+j)(\text{mod } (k+1))}) : 0 \leq i \leq k\big\}.$$

The three graphs $B_{0,2}$, $B_{1,2}$, and $B_{2,2}$ in the $k=2$ case are depicted in Figure 4. It is clear that the graph obtained by taking the union of all edges in all sets $B_{j,k}$ for $0 \leq j \leq k$ is $K_{2k+2,2k+2}$, the complete bipartite graph on two sets of $2k+2$ vertices.

We now define three sets of states $S^{(j)} = \{|v_i^{(j)}\rangle, |w_i^{(j)}\rangle, |x_i^{(j)}\rangle, |y_i^{(j)}\rangle : 0 \leq i \leq k\} \subseteq \mathbb{C}^2$ that have orthogonality graphs $B_{j,k}$ for $0 \leq j \leq 2$ respectively. To this end, let $\{|b_i\rangle, |b_i^\perp\rangle\}_{i=0}^{2k+1}$ be distinct orthonormal bases of $\mathbb{C}^2$ (i.e., $\langle b_i | b_i^\perp\rangle = 0$ for all i, but $|\langle b_i | b_j\rangle|, |\langle b_i | b_j^\perp\rangle|, |\langle b_i^\perp | b_j^\perp\rangle| \notin \{0, 1\}$ whenever $i \neq j$). Then let

$$|v_i^{(j)}\rangle := |w_i^{(j)}\rangle := |b_i\rangle \quad \text{and} \quad |x_i^{(j)}\rangle := |y_i^{(j)}\rangle := |b^\perp_{(i-j)(\text{mod } (k+1))}\rangle,$$

for $0 \leq j \leq 2$, which clearly results in the desired orthogonality graphs. Furthermore, each set $S^{(j)}$ has the property that any state $|z\rangle \in \mathbb{C}^2$ can be orthogonal to at most two elements of $S^{(j)}$ – a fact that we will use later when discussing unextendibility.

For each of the remaining $k-2$ graphs $B_{j,k}$ $(3 \leq j \leq k)$, we construct sets of product states $S^{(2j-3,2j-2)} = \{|v_i^{(2j-3,2j-2)}\rangle, |w_i^{(2j-3,2j-2)}\rangle, |x_i^{(2j-3,2j-2)}\rangle, |y_i^{(2j-3,2j-2)}\rangle : 0 \leq i \leq k\} \subseteq \mathbb{C}^2 \otimes \mathbb{C}^2$ that have orthogonality graphs $B_{j,k}$ for $3 \leq j \leq k$. To this end, define

$$|v_i^{(2j-3,2j-2)}\rangle := |b_i\rangle \otimes |b_i\rangle$$
$$|w_i^{(2j-3,2j-2)}\rangle := |b_{i+(k+1)}\rangle \otimes |b_{i+(k+1)}\rangle$$
$$|x_i^{(2j-3,2j-2)}\rangle := |b^\perp_{(i-j)(\text{mod } (k+1))}\rangle \otimes |b^\perp_{(i-j)(\text{mod } (k+1))+(k+1)}\rangle$$
$$|y_i^{(2j-3,2j-2)}\rangle := |b^\perp_{(i-j)(\text{mod } (k+1))+(k+1)}\rangle \otimes |b^\perp_{(i-j)(\text{mod } (k+1))}\rangle,$$

which results in the desired orthogonality graphs.

We now turn our attention to the complement graph of $K_{2k+2,2k+2}$, which is simply the disjoint union of two disjoint copies of K_{2k+2}, the complete graph on $2k+2$ vertices. We

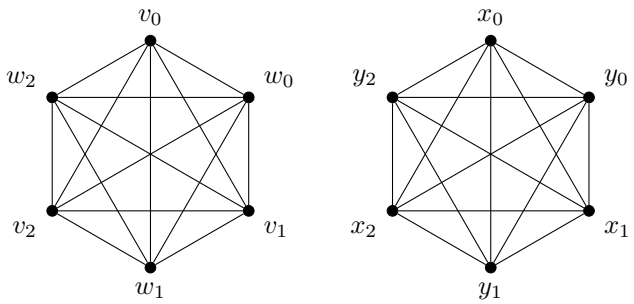


Figure 5 The graph K_6^2 that is the disjoint union of two copies of K_6.

denote this graph by K_{2k+2}^2, and it is depicted in the $k = 2$ case in Figure 5. The graph K_{2k+2}^2 will be the orthogonality graph of the remaining $4k - (3 + 2(k-2)) = 2k + 1$ parties.

Our goal now is to define sets of states $S^{(j)} = \{|v_i^{(j)}\rangle, |w_i^{(j)}\rangle, |x_i^{(j)}\rangle, |y_i^{(j)}\rangle : 0 \leq i \leq k\} \subseteq \mathbb{C}^2$ for $2k - 1 \leq j \leq 4k - 1$ such that their orthogonality graphs, when taken together, contain all edges of K_{2k+2}^2. To this end, we recall that it is well-known that K_{2k+2} always has a 1-factorization [10, Theorem 9.1], so K_{2k+2}^2 clearly has a 1-factorization as well (see Figure 6). This 1-factorization decomposes K_{2k+2}^2 into $2k + 1$ distinct 1-regular spanning subgraphs, and any such graph is clearly the orthogonality graph of the set of states $\{|b_0\rangle, |b_0^\perp\rangle, \ldots, |b_{2k+1}\rangle, |b_{2k+1}^\perp\rangle\} \subset \mathbb{C}^2$ (under an appropriate labelling of the vertices).

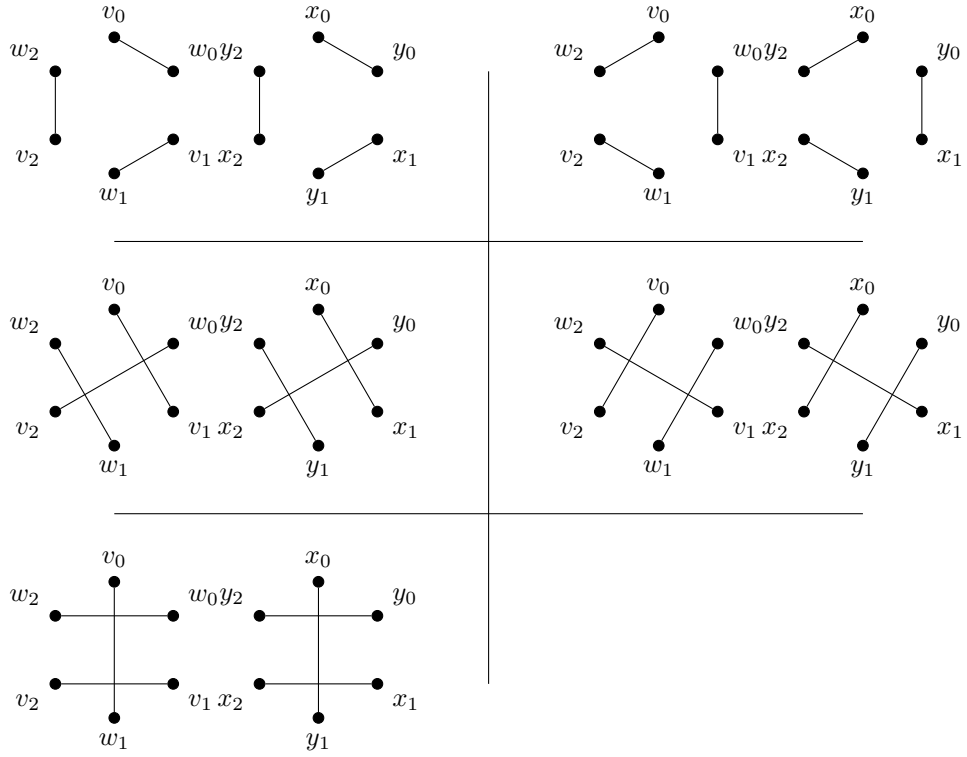


Figure 6 A 1-factorization of K_6^2, which is useful for constructing a UPB of size 12 in $(\mathbb{C}^2)^{\otimes 8}$.

Since the union of the sets of edges present in all of the graphs considered so far is the complete graph K_{4k+4}, we know that the states in the set

$$\mathcal{S} := \left\{ \bigotimes_{j=1}^{4k} |v_i^{(j)}\rangle, \bigotimes_{j=1}^{4k} |w_i^{(j)}\rangle, \bigotimes_{j=1}^{4k} |x_i^{(j)}\rangle, \bigotimes_{j=1}^{4k} |y_i^{(j)}\rangle : 0 \leq i \leq k \right\}$$

are mutually orthogonal. To see why this set is unextendible, recall that any non-zero product state can be orthogonal to at most 2 states on each of the first 3 subsystems, and at most 1 state on each of the remaining $4k-3$ subsystems. It follows that any nonzero product state can be orthogonal to at most $2 \cdot 3 + 1 \cdot (4k-3) = 4k+3$ of these product states. Since no nonzero product state can be orthogonal to all $4k+4$ members of $\mathcal{S}$, it is unextendible, which completes the proof. ◀

4 Proof of Minimality

We now turn our attention to the problem of proving that the UPBs constructed in Section 3 are the smallest possible. Because the main result of [1] tells us that the minimum cardinality of a UPB in $(\mathbb{C}^2)^{\otimes 4k}$ is at least $4k+2$, we only have to prove that there is no UPB of cardinality $4k+2$ when $k \geq 2$ and no UPB of cardinality $4k+3$ when $k \geq 3$. While the proof that there is no UPB of cardinality $4k+2$ is relatively straightforward, the proof that there is no UPB of cardinality $4k+3$ is more involved and consists of many cases and sub-cases. We make use of a C script to solve some of the messier cases, while we solve the simpler cases by hand.

For the entirety of this section, we make use of *partial orthogonality graphs*, which are the same as orthogonality graphs, except perhaps with some conditions unspecified. For example, in Figure 7 the lack of lines indicating orthogonality between shaded regions does not signify that there are no regions orthogonal to each other, but rather that we just don't care *which* regions are orthogonal to each other. Similarly, in Figure 8 there are vertices that are drawn outside of any shaded region. This is intended to mean that we don't care what the shaded region involving that vertex looks like. In general, we only specify the pieces of the orthogonality graphs that are relevant for our proofs.

It will be convenient for us to let $P_1, \ldots, P_{4k}$ denote the $4k$ different parties. We also let M_j denote the maximum number of vertices contained within a single shaded region on party P_j (which is equal to the maximum number of states in the UPB that are equal to each other on party P_j), and let $C_{n,j}$ denote the number of distinct shaded regions containing exactly n vertices on party j (i.e., $C_{n,j}$ is the number of distinct group of exactly n states in the UPB that are equal to each other on party P_j). For example, in Figure 2, if the graphs correspond to parties P_1, P_2 and P_3, then $M_1 = 4$, $M_2 = M_3 = 2$, $C_{3,1} = 1$, $C_{4,1} = 1$, $C_{1,2} = 1$, $C_{2,2} = 3$, $C_{1,3} = 5$, and $C_{2,3} = 1$.

▶ **Lemma 5.** *There is no UPB in* $(\mathbb{C}^2)^{\otimes 4k}$ *of cardinality* $4k+2$ *when* $k \geq 2$.

Proof. Suppose for a contradiction that there exists a UPB of cardinality $4k+2$ in $(\mathbb{C}^2)^{\otimes 4k}$. If it were the case that $M_j \geq 3$ for some j, then we could find a product state that is orthogonal to the 3 corresponding states on that party and to any 1 of the product states on each of the remaining $4k-1$ parties, for a total of all $4k+2$ elements of the UPB, which violates unextendibility. Hence $M_j \leq 2$ for all $1 \leq j \leq 4k$. We now split into two cases.

Case 1: There is at most one party P_j with $M_j = 2$.

Between the $4k$ parties, there must be a total of $(4k+2)(4k+1)/2 = 8k^2 + 6k + 1$ edges in the union of their orthogonality graphs. The $4k-1$ parties other than P_j must be the

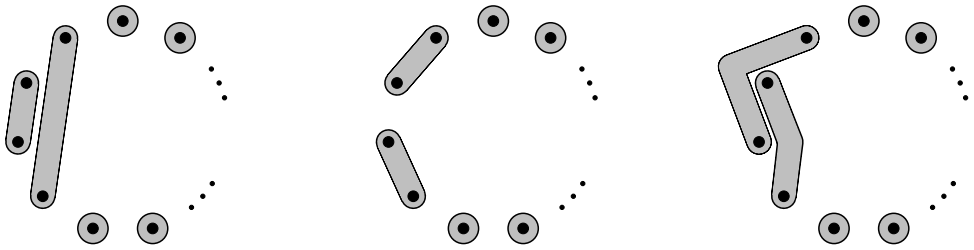


Figure 7 Partial orthogonality graphs of three parties that each have two sets of two equal states, used in the proof of case 2 of Lemma 5. There is no way to add another pair of equal states on any party without violating unextendibility.

disjoint union of $2k+1$ copies of $K_{1,1}$, for a total of at most $(4k-1)(2k+1) = 8k^2+2k-1$ edges. The remaining party P_j then needs at least $(8k^2+6k+1)-(8k^2+2k-1) = 4k+2$ edges. It is easily seen, however, that the largest number of edges that the orthogonality graph of party P_j can have is obtained when it is the disjoint union of k copies of $K_{2,2}$ and one copy of $K_{1,1}$, which results in only $4k+1$ edges, which gives the desired contradiction.
Case 2: There are two (or more) parties $P_i \neq P_j$ with $M_i = M_j = 2$.

It is not difficult to see that $C_{2,\ell} \in \{0, 2\}$ for all ℓ or else either Lemma 2 or unextendibility is violated. Furthermore, it is not difficult to see that the unique (up to repositioning vertices and parties) way to have $C_{2,\ell} = 2$ for 3 distinct values of ℓ is given in Figure 7, and there is no way to have $C_{2,\ell}$ for a fourth value of ℓ without violating unextendibility. A simple calculation reveals that the maximum number of edges that can be obtained from the orthogonality graphs of these 3 parties is $(2k+3)+2(2k+2) = 6k+7$. The orthogonality graphs of the remaining $4k-3$ parties are the disjoint union of $2k+1$ copies of $K_{1,1}$, so they each have $2k+1$ edges. Thus the total number of edges among the orthogonality graphs of all $4k$ parties is at most $(6k+7)+(4k-3)(2k+1) = 8k^2+4k+4$. This quantity is smaller than the $8k^2+6k+1$ required edges when $k \geq 2$, which gives the desired contradiction.

◀

Note that the hypothesis of Lemma 5 that $k \geq 2$ really is required, since we have $8k^2+4k+4 \geq 8k^2+6k+1$ in case 2 of the proof of the lemma when $k = 1$, so it may be possible to fit all of the required edges into the orthogonality graphs. Indeed, it was shown in [9] that a UPB consisting of $4k+2$ states in $(\mathbb{C}^2)^{\otimes 4k}$ exists in the $k = 1$ case.

We now turn our attention to proving that there is no UPB of cardinality $4k+3$ when $k \geq 3$. The idea and techniques used in the proof of this statement are quite similar to the $4k+2$ case, but there are more cases to consider.

▶ **Lemma 6.** *There is no UPB in* $(\mathbb{C}^2)^{\otimes 4k}$ *of cardinality* $4k+3$ *when* $k \geq 3$.

Proof. Suppose for a contradiction that there exists a UPB of cardinality $4k+3$ in $(\mathbb{C}^2)^{\otimes 4k}$. If there exists $1 \leq j \leq p$ such that $M_j \geq 4$, then we can find a product state that is orthogonal to at least 4 corresponding states on party P_j and to 1 of the product states on each of the remaining $4k-1$ parties, for a total of $4k+3$ elements of the UPB, which violates unextendibility. Hence $M_j \leq 3$ for all j. Furthermore, this same argument shows that if there exists $i \geq 1$ such that we can choose a single shaded region on each of i parties so that together they contain at least $i+3$ vertices, then unextendibility will be violated. Finally, note that since $4k+3$ is odd, Lemma 2 implies that $M_j \geq 2$ for all j.

We now split into 4 cases, depending on the value of $\max_j\{C_{3,j}\}$ (i.e., the maximum number of sets of 3 equal states on any party).

Case 1: $\max_j\{C_{3,j}\} \geq 3$.

Because $M_j \geq 2$ for all j, it easily follows that we can find shaded regions on two parties that contain $3 + 2 = 5$ distinct vertices, which contradicts unextendibility.

Case 2: $\max_j\{C_{3,j}\} = 2$.

Suppose without loss of generality that party P_1 is such that $C_{3,1} = 2$. Unextendibility immediately implies that $C_{3,j} = 0$ for $j \geq 2$. Since there are $4k - 3$ left over vertices on party P_1, which is odd, there must be a copy of $K_{2,1}$ on this party, as in Figure 8. Since v_1 is connected to only one other state on party P_1, it must be connected to 2 states on each of 2 other parties. These sets of 2 vertices must be disjoint and must each contain one of v_2, v_3, v_4 and one of v_5, v_6, v_7. Thus parties P_2 and P_3, without loss of generality, are as in Figure 8, which clearly implies extendibility and rules out this case.

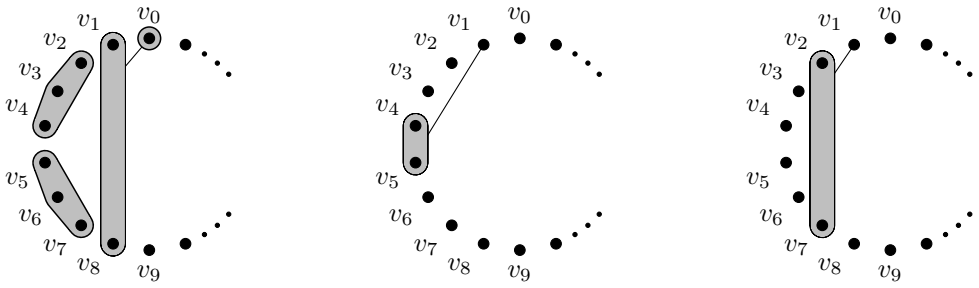


Figure 8 The (essentially unique) partial orthogonality graphs of parties P_1 (left), P_2 (center) and P_3 (right) in case 2 of Lemma 6. Such a product basis is necessarily extendible, as we can find a product state that is orthogonal to the states corresponding to v_1 and v_8 on party P_1, v_4 and v_5 on party P_2, v_2 and v_7 on party P_3, and one of the $4k - 3$ remaining states on each of the remaining $4k - 3$ parties.

Case 3: $\max_j\{C_{3,j}\} = 0$.

Since $M_j = 2$ for all j, simple parity arguments show that $C_{2,j} \in \{1, 3, 5, \ldots\}$ for every j. We now split into two sub-cases, depending on the value of $\max_j\{C_{2,j}\}$ (i.e., the maximum number of sets of 2 equal states on any party).

Case 3(a): $\max_j\{C_{2,j}\} \geq 5$.

Suppose that party P_1 has $C_{2,1} \geq 5$. We first argue that there must be at least one other party P_2 with $C_{2,2} \geq 3$. To see this, suppose the contrary – suppose that $C_{2,j} = 1$ for all $j \geq 2$. Then each of these $4k-1$ parties contributes at most $2k+2$ edges to the orthogonality graph, for a total of $(4k - 1)(2k + 2) = 8k^2 + 6k - 2$ edges. The party P_1 contributes no more than $4k + 2$ edges, for a total of $8k^2 + 10k$ edges among all $4k$ parties. However, the complete graph on $4k + 3$ vertices has $(4k + 3)(4k + 2)/2 = 8k^2 + 10k + 3$ edges, so there are at least 3 pairs of non-orthogonal product states in our set, which contradicts the assumption that we are working with a UPB.

We now pick an arbitrary party $P_3 \neq P_1, P_2$. Because $C_{2,3} \geq 1$, we are now able to choose one shaded region on each of parties P_1, P_2, P_3 such that 6 vertices are contained within these regions, which shows that unextendibility is violated. To this end, we choose any shaded region on party P_3 that contains two vertices, then we pick any shaded region on party P_2 that is disjoint from the two vertices we chose on party P_3, and finally we choose any shaded region on party P_1 that is disjoint from all four of the previously-chosen vertices (see Figure 9).

Case 3(b): $\max_j\{C_{2,j}\} \leq 3$.

We begin by noting that the brute-force computer search shows that there can be no more

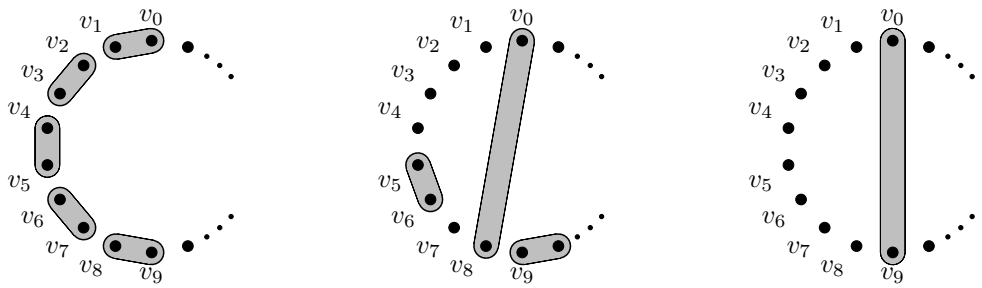


Figure 9 An example of a partial orthogonality graph in case 3(a) of Lemma 6. Such a product basis is necessarily extendible, as we can choose the shaded region containing v_0 and v_9 on party P_3, the disjoint shaded region (i.e., the one containing v_5 and v_6) on party P_2, and the disjoint shaded region (i.e., the one containing v_2 and v_3) on party P_1, for a total of 6 vertices on 3 parties.

than 4 distinct parties P_j for which $C_{2,j} \geq 3$ [11]. Each of these four parties has at most $2k+4$ edges in its orthogonality graph, and each of the remaining $4k-4$ parties has at most $2k+2$ edges on its orthogonality graph, for a total of at most $4(2k+4)+(4k-4)(2k+2) = 8k^2+8k+8$ edges. The complete graph on $4k+3$ vertices has $(4k+3)(4k+2)/2 = 8k^2+10k+3$ edges, so when $k \geq 3$ there are not enough edges in the orthogonality graph, so the set of states does not form a product basis, which contradicts our assumption that we are working with a UPB. Note that this is the case in which the UPB of Lemma 3 arises if $k = 2$, so the fact that we require $k \geq 3$ here is not surprising.

Case 4: $\max_j\{C_{3,j}\} = 1$.

By parity arguments, we see that every party P_j with $C_{3,j} = 1$ must also have $C_{2,j} \in \{1, 3, 5, \dots\}$. Furthermore, if there exist two (or more) parties P_1, P_2 such that $M_1 = M_2 = 3$, then unextendibility is violated unless $C_{2,j} = 1$ whenever $M_j = 3$.

Case 4(a): There exist three (or more) parties P_1, P_2, P_3 such that $M_1 = M_2 = M_3 = 3$.

Because there must exist a shaded region containing exactly 2 vertices on each party P_1, P_2, P_3, it is easily verified that the only possible configuration of shaded regions on those parties (up to repositioning vertices and parties) that doesn't break unextendibility is the one depicted in Figure 10.

The parties P_1, P_2, P_3 can have no more than $(2k+5)+2(2k+3) = 6k+11$ distinct edges among them (since there will be a lot of overlap at the left edge of the graphs if we make each group of 3 equal states orthogonal to the group of 2 equal states). It is straightforward to see that none of the remaining $4k-3$ parties P_j can have $M_j \geq 3$ or $C_{2,j} \geq 2$ without breaking unextendiblity. Thus those $4k-3$ parties can produce no more than $2k+2$ edges each, for a total of $6k+11+(4k-3)(2k+2) = 8k^2+8k+5$ edges. Since $8k^2+8k+5 < 8k^2+10k+3$ when

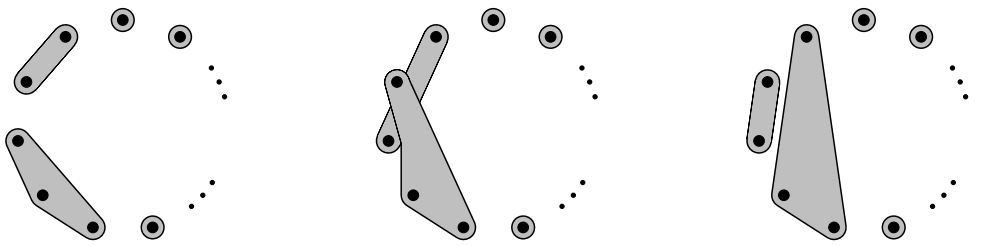


Figure 10 The (essentially unique) partial orthogonality graph that does not violate unextendibility in case 4(a).

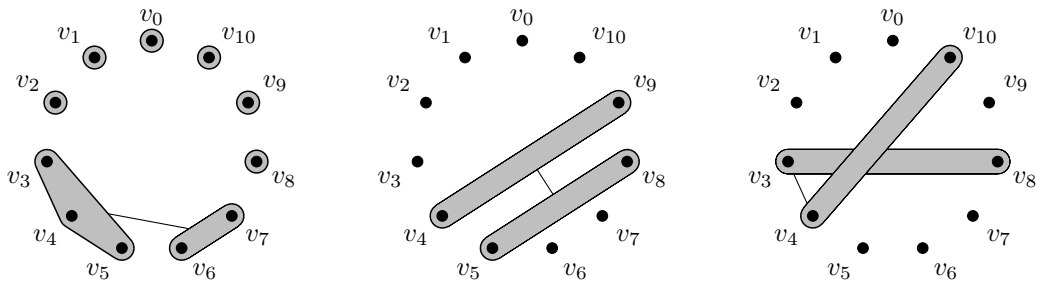


Figure 11 An example of a partial orthogonality graph in case 4(b).

$k \geq 2$, there are some edges missing from the orthogonality graphs, which is a contradiction.

Case 4(b): There exists a party P_1 such that $M_1 = 3$, but $M_j \leq 2$ for $j \geq 2$.

Party P_1 contributes at most $2k + 5$ edges to the orthogonality graph, and the unextendibility requirement implies that $C_{2,j} \leq 3$ for $j \geq 2$. Suppose that there are m indices $2 \leq j_1, j_2, \ldots, j_m \leq 4k$ such that $C_{2,j_i} = 3$ for $1 \leq i \leq m$ and $C_{2,j} = 1$ for all other values of j. Then there are at most $(2k + 5) + m(2k + 4) + (4k - m - 1)(2k + 2) = 8k^2 + 8k + 2m + 3$ total edges between all $4k$ parties. As in the previous cases, we need a total of $8k^2 + 10k + 3$ edges, which implies that $m \geq k$. We already saw via brute-force search in case 3(b) that we can't have $m \geq 5$, so we only need to rule out the $3 \leq k \leq 4$ cases.

If the group of 3 identical states on party P_1 is represented by vertices v_3, v_4, and v_5 (see Figure 11), then each one of the 3 groups of 2 identical states on the other parties must contain exactly one of v_3, v_4, or v_5. By refining our brute-force computer search to take this restriction into account, we find that there is no configuration of shaded regions that does not violate unextendibility when $m \geq 3$ [11], so no such UPB exists when $k \geq 3$.

Case 4(c): There exist two parties P_1, P_2 such that $M_1 = M_2 = 3$, but $M_j \leq 2$ for $j \geq 3$.

In this case, there are (up to relabelling vertices and parties) only two possible configurations of parties P_1 and P_2, which are depicted in Figures 12 and 13. Notice that in Figure 12, the shaded region on party P_1 that contains exactly two vertices *does not* share any common vertices with the shaded region on party P_2 that contains exactly two vertices, while in Figure 13 those two regions contain the common vertex v_1.

Suppose for now that parties P_1 and P_2 have a total of at most $4k + 8$ distinct edges on their orthogonality graphs. If there are m parties P_j $(j \geq 3)$ for which $C_{2,j} = 3$, then we have a total of at most $(4k + 8) + m(2k + 4) + (4k - m - 2)(2k + 2) = 8k^2 + 8k + 2m + 4$ edges. In all of these m parties, we require that one of the shaded regions contains v_2 and v_3

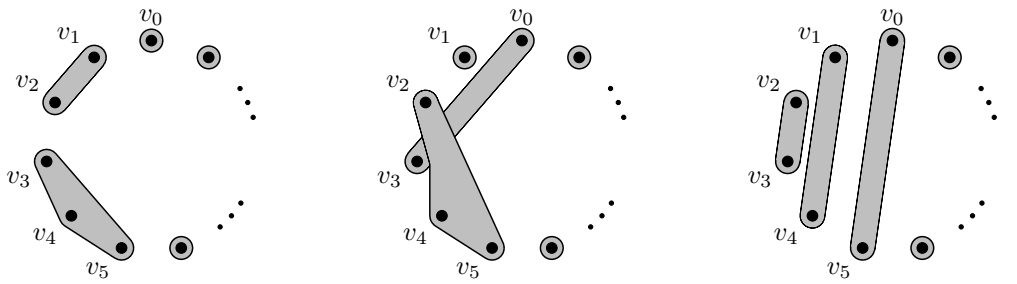


Figure 12 One of two possible partial orthogonality graphs of parties P_1, P_2, and P_3 that does not violate unextendibility in case 4(c).

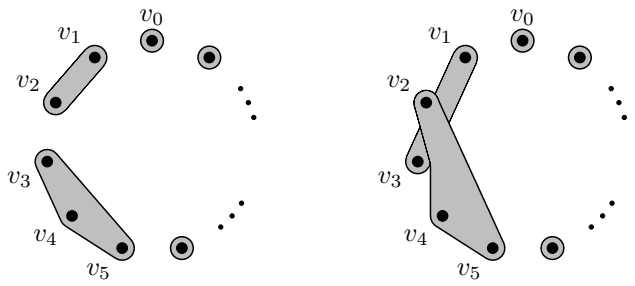


Figure 13 The other possible partial orthogonality graph of parties P_1, P_2 that does not violate unextendibility in case 4(c).

and the other shaded regions containing two vertices each contain one of v_4 or v_5. Thus, the brute-force search described in case 4(b) applies here as well and shows that $m \leq 2$. However, when $m = 2$ we have $8k^2 + 8k + 2m + 4 = 8k^2 + 8k + 8 < 8k^2 + 10k + 3$ when $k \geq 3$, which shows that there can not possibly be enough edges on the orthogonality graphs in this case.

The only remaining possibility is that the parties P_1 and P_2 have a total of at least $4k+9$ distinct edges (and hence *exactly* $4k + 9$ distinct edges). In this case, parties P_1 and P_2 must be as in Figure 12, and on both of the parties P_1 and P_2 the set of 3 equal states must be orthogonal to the set of 2 equal states. Furthermore, it is not difficult to show that in this case, any party P_j with $C_{2,j} = 3$ can have at most $2k + 4$ edges, but if it has $2k + 4$ edges then at least one of those edges must already be present on either party P_1 or P_2. It follows that each party P_j $(j \geq 3)$ can introduce at most $2k + 3$ new edges that have not already been counted. Thus, if there are m parties P_j $(j \geq 3)$ for which $C_{2,j} = 3$, we have a total of at most $(4k + 9) + m(2k + 3) + (4k - m - 2)(2k + 2) = 8k^2 + 8k + m + 5$ edges. Since $m \leq 2$ (as before) and $k \geq 3$, it follows that $8k^2 + 8k + m + 5 < 8k^2 + 10k + 3$, which again shows that there can not possibly be enough edges on the orthogonality graphs in this case. ◀

Acknowledgements. Thanks are extended to Gus Gutoski for suggesting a computer search to fill in the gaps in the proof of Lemma 6. The author was supported by the Natural Sciences and Engineering Research Council of Canada and the Mprime Network.

References

1 N. Alon and L. Lovász. Unextendible product bases. *J. Combinatorial Theory, Ser. A*, 95:169–179, 2001.

2 R. Augusiak, T. Fritz, M. Kotowski, M. Kotowski, M. Pawłowski, M. Lewenstein, and A. Acín. Tight Bell inequalities with no quantum violation from qubit unextendible product bases. *Phys. Rev. A*, 85:042113, 2012.

3 R. Augusiak, J. Stasinska, C. Hadley, J. K. Korbicz, M. Lewenstein, and A. Acín. Bell inequalities with no quantum violation and unextendible product bases. *Phys. Rev. Lett.*, 107:070401, 2011.

4 C. H. Bennett, D. P. DiVincenzo, C. A. Fuchs, T. Mor, E. Rains, P. W. Shor, J. A. Smolin, and W. K. Wootters. Quantum nonlocality without entanglement. *Phys. Rev. A*, 59:1070–1091, 1999.

5 C. H. Bennett, D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal. Unextendible product bases and bound entanglement. *Phys. Rev. Lett.*, 82:5385–5388, 1999.

6 J. Chen and N. Johnston. The minimum size of unextendible product bases in the bipartite case (and some multipartite cases). E-print: arXiv:1301.1406 [quant-ph], 2013.

7 D. P. DiVincenzo, T. Mor, P. W. Shor, J. A. Smolin, and B. M. Terhal. Unextendible product bases, uncompletable product bases and bound entanglement. *Commun. Math. Phys.*, 238:379–410, 2003.

8 R. Duan, Y. Xin, and M. Ying. Locally indistinguishable subspaces spanned by three-qubit unextendible product bases. *Phys. Rev. A*, 81:032329, 2010.

9 K. Feng. Unextendible product bases and 1-factorization of complete graphs. *Discrete Appl. Math.*, 154:942–949, 2006.

10 F. Harary. *Graph Theory.* Addison-Wesley, Reading, Mass., 1969.

11 N. Johnston. Code for proving that no UPB of size $4k + 3$ exists on $4k$ qubits. Published electronically at `http://www.njohnston.ca/publications/qubit-upbs/code/`, 2013.

12 J. M. Leinaas, P. Ø. Sollid, and J. Myrheim. Unextendible product bases and extremal density matrices with positive partial transpose. E-print: arXiv:1104.1318 [quant-ph], 2011.

13 Ł. Skowronek. Three-by-three bound entanglement with general unextendible product bases. *J. Math. Phys.*, 52:122202, 2011.

14 B. M. Terhal. A family of indecomposable positive linear maps based on entangled quantum states. *Linear Algebra Appl.*, 323:61–73, 2001.

Robust Online Hamiltonian Learning

Christopher E. Granade*[,1,2], Christopher Ferrie[1,3], Nathan Wiebe[1,3], and D. G. Cory[1,4,5]

1 **Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario, Canada**
2 **Department of Physics, University of Waterloo, Waterloo, Ontario, Canada**
3 **Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada**
4 **Department of Chemistry, University of Waterloo, Waterloo, Ontario, Canada**
5 **Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada**

Abstract

In this work we combine two distinct machine learning methodologies, sequential Monte Carlo and Bayesian experimental design, and apply them to the problem of inferring the dynamical parameters of a quantum system. The algorithm can be implemented *online* (during experimental data collection), avoiding the need for storage and post-processing. Most importantly, our algorithm is capable of learning Hamiltonian parameters even when the parameters change from experiment-to-experiment, and also when additional noise processes are present and unknown. The algorithm also numerically estimates the Cramer-Rao lower bound, certifying its own performance. We further illustrate the practicality of our algorithm by applying it to two test problems: (1) learning an unknown frequency and the decoherence time for a single–qubit quantum system and (2) learning couplings in a many–qubit Ising model Hamiltonian with no external magnetic field.

1998 ACM Subject Classification G.3 Probability and Statistics

Keywords and phrases Quantum information, sequential Monte Carlo, Bayesian, experiment design, parameter estimation

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.106

1 Introduction

The problem of characterizing quantum systems is of fundamental importance to quantum information science. Without an accurate understanding, for example, of the noise processes that a quantum computer experiences, error correction may be quite difficult; furthermore, certification of quantum dynamics is essential for determining whether the predictions made by a quantum simulator can be trusted. This latter problem is especially timely since quantum simulation experiments are approaching a complexity where classical computers are unable to simulate their evolution [1, 2, 3]. Natural solutions to this problem, such as tomographic methods [4, 5, 6, 7, 8, 9, 10], are often impractical for learning parameters for large quantum systems, as well as for learning parameters such as T_2. This prompts the question of whether there exists a practical error robust technique that can be used to characterize quantum systems with unknown decoherence processes.

* Corresponding author (cgranade@cgranade.com).

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 106–125


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

We make this learning process tractable by utilizing information about a system, rather than starting from worst-case assumptions such as those made in traditional quantum process and state tomography. In practice, we often have knowledge about the dynamical model that describes a system of interest, and wish to improve that knowledge by estimating specific model parameters. Thus, practical Hamiltonian finding can often be achieved via a suitable parameterization of the Hamiltonian, $H(x_1, \ldots, x_d)$, reducing the problem to estimating the vector of parameters $\boldsymbol{x} = (x_1, \ldots, x_d)$. The task we consider is the design of experiments for the purpose of deducing these parameters in the smallest number of experiments possible. Our algorithm also provides a *region estimation* for the Hamiltonian parameters that encloses some fixed volume of parameter space in which the mean or the variance of the Hamiltonian parameters are expected to be found with high–probability. We also generalize this concept to allow the algorithm to learn *hyperparameters*, which describe the distribution of the Hamiltonian parameters in cases where the parameters randomly drift between experiments. Pseudocode for all of our algorithms is given in Appendix B.

Our algorithm achieves this by combining sequential Monte Carlo methods [11] with Bayesian experiment design [12] to choose experiments that maximize the expected reduction in the uncertainty in the unknown parameters based on the results of prior experiments. We call such derived strategies *adaptive* or *online*. This approach not only reduces the number of experiments needed to learn the unknown parameters within a fixed error tolerance, but it also makes the learning process more robust. In addition to robustness, Bayesian updating provides a natural estimate of the uncertainty in the unknown parameters in the form of the width of the prior distribution. In contrast, it can be difficult to quantify the uncertainty in the estimated Hamiltonian using traditional methods based on inversion or tomography.

It is worth noting that approaches that are similar to our own have been considered very recently in a wide variety of classical contexts [13, 14, 15, 16, 17], and also for measurement adaptive quantum state tomography [18]. Other machine learning ideas have also been generalized to the quantum domain [19, 20, 21, 22, 23, 24, 25]; however, to the best of our knowledge, no method based on ideas from machine learning has been proposed for learning unknown Hamiltonian parameters that is as broadly applicable or as robust to noise as our method.

This paper is organized as follows. In section 2, we review the formalism of Bayesian experimental design. Section 3 introduces the sequential Monte Carlo algorithm. In section 4, we discuss the application of our algorithm to region estimation and hyperparameter estimation. We then explore the implications of the numerical benchmarking results in sections 5 and 6 before concluding.

2 Experimental Design Formalism

The essence of our experimental design process is that we choose experiments not according to a pre–determined sequence, but rather our algorithm adaptively chooses experiments that are expected to be very informative (given the current state of knowledge of the unknown parameters). We model a sequence of experiments as a sequence of experimental controls $\{c_1, \ldots, c_N\}$ and a corresponding sequence of acquired data $\{d_1, \ldots, d_N\}$. Bayesian updating is then used to formalize how the acquired data impacts our current state of knowledge about the unknown Hamiltonian, where we connect to the theory of parameter estimation by using the predictions of quantum mechanics as a probabilistic model, called a likelihood function.

To clarify, suppose we have performed an experiment with control settings c_1. We are then ultimately interested in the *posterior distribution* $\Pr(\boldsymbol{x}|d_1; c_1)$, the probability distri-

bution of the model parameters $\boldsymbol{x}$ given this data. By Bayes' rule and the conditional independence of each datum, the posterior distribution is given by

$$\Pr(\boldsymbol{x}|d_1;c_1) = \frac{\Pr(d_1|\boldsymbol{x};c_1)}{\Pr(d_1|c_1)}\Pr(\boldsymbol{x}),$$

where $\Pr(\boldsymbol{x})$ is the *prior*, which encodes any *a priori* knowledge of the model parameters. $\Pr(d_1|\boldsymbol{x};c_1)$ is the likelihood, which can be computed using Born's rule. The total likelihood $\Pr(d_1|c_1)$ can simply be thought as a normalization factor. Subsequent experiments update the prior according to the following iterative rule

$$\Pr(\boldsymbol{x}|d_{j+1},\dots;c_{j+1},\dots) = \frac{\Pr(d_{j+1}|\boldsymbol{x};c_{j+1})}{\Pr(d_{j+1}|c_{j+1})}\Pr(\boldsymbol{x}|d_j,\dots;c_j,\dots).$$

The idea of adaptive experiment design can be formalized in various ways, the most natural for our purposes being called *Bayesian experimental design* [12]. For this, we conceive of possible future data d_{N+1} obtained from a, possibly different, set of experimental controls c_{N+1}. The probability of obtaining this data can be computed from the distributions at hand via marginalizing over model parameters

$$\Pr(d_{j+1}|d_j,\dots;c_{j+1}) = \int \Pr(d_{j+1}|\boldsymbol{x};c_{j+1})\Pr(\boldsymbol{x}|d_j,\dots;c_j,\dots)\,\mathrm{d}\boldsymbol{x}.$$

Note, in the remainder of this work, we will use the following abbreviated notation for expectation values:

$$\Pr(d_{j+1}|d_j,\dots;c_{j+1},\dots) = \mathbb{E}_{\boldsymbol{x}|d_{j+1},\dots;c_{j+1},\dots}[\Pr(d_{j+1}|\boldsymbol{x};c_{j+1})], \tag{1}$$

where the subscript on $\mathbb{E}$ denotes the variable for the expectation to be taken over.

The expectation value in (1) can be used to inform the algorithm about the choices of experimental parameters that are more useful than others. This usefulness is quantified, for a given choice of a *utility function* $U(d_{j+1},c_{j+1})$, by the expected *utility* of an experiment

$$U(c_{j+1}) = \mathbb{E}_{d_{j+1}|d_j,\dots;c_{j+1},\dots}[U(d_{j+1},|c_{j+1})],$$

where $U(d_{j+1},c_{j+1})$ is the utility we would derive if experiment c_{j+1} yielded result d_{j+1}. The choice of the utility function is motivated by the figure of merit that we want to optimize, and will be considered in Appendix A.

3 Sequential Monte Carlo Algorithm

A major drawback of using Bayesian inference for Hamiltonian learning stems from the fact that the parameter space is continuous. This means that the prior will have in general support over an infinite number of possible Hamiltonians, which in turn makes applying Bayes' rule and sampling from the resultant posterior intractable. We address this problem by using *sequential Monte Carlo* (SMC) methods, such as those described in the recent tutorial by Doucet and Johansen [11].

At each step of the SMC algorithm, the current distribution is approximated by a weighted sum of Dirac-delta functions, so that $\Pr(\boldsymbol{x}|D) \approx \sum_{k=1}^{n} w_k(D)\delta(\boldsymbol{x}-\boldsymbol{x}_k)$, where $w_k(D)$ is a *weight* that describes the relative plausibility of the hypothesis $\boldsymbol{x}_k$, having observed the data record D. Each term in this sum is referred to as a *particle*.

Since the Bayes update rule for many observations $\{d_1,d_2,\dots,d_N\}$ can be processed sequentially by updating the weights: $w_k(d_{j+1}\cup D) = \Pr(d_{j+1}|\boldsymbol{x}_k)w_k(d_j)/\mathcal{N}$, where $\mathcal{N}$ is found by the normalization constraint that $\sum_k w_k(D) = 1$.

The particle approximation can be made arbitrarily accurate by increasing the number of particles, and will be a good approximation at every update provided we feed in, at the initial stage, the appropriate weights $\{w_k\}$ and support points $\{\boldsymbol{x}_k\}$. Since both the weights and support points of the particles carry information about distributions over the model parameters $\boldsymbol{x}$, we can without loss of generality choose the initial weights to be uniform, $w_k = 1/n$ for all k, and the initial support points to be samples from the correct prior $\Pr(\boldsymbol{x})$. Having made the particle approximation, we perform Bayes updates using the algorithm below.

Sequential Monte Carlo techniques require careful effort to avoid introducing errors due to limited numerical precision. The first problem any SMC algorithm runs into is zero weights. This is doubly painful since we are effectively operating with fewer particles but using the same amount of computational resources. Since the support of our approximate distribution is a measure-zero set according to the correct distribution, all the weights will eventually be zero; we cannot avoid this but it can be postponed by using *resampling* techniques.

Generally, the idea behind resampling is to adaptively change the location of the particles to those which are most likely. This works because a particle approximation to a probability distribution can be equally well approximated using constant weight particles with variable density, or variable weight particles with constant density. Hence, we can "resample" the distribution by using constant weight particles to approximate to the prior distribution to alleviate problems caused by the weights of the particles becoming small enough to impact the numerical stability of the methods. The simplest of these types of algorithm chooses n particles (the original number), with replacement, according to the distribution of weights then reset the weights of all particles to $1/n$. Thus, zero weight particles are "moved" to higher weight locations. To determine when to resample, we shall compare the effective sample size $n_{\text{ess}} = 1/\sum_i w_i^2$ to a threshold `resample_threshold`, which is the effective ratio of the original number of particles n. We use `resample_threshold` $= 0.5$, as suggested by [26].

The resampling algorithm we use was first proposed in [26] and is given explicitly in Algorithm 2. The idea behind the algorithm conforms to the intuition given above but it incorporates randomness to search larger volumes of the parameter space. This randomness is inserted in the resampling algorithm by applying a random perturbation to the location of each particle that is introduced during the resampling process. Thus, the new particles are randomly spread around the previous locations of the old. More formally, we model this by randomly choosing a particle location $\boldsymbol{x}_i$, then perturbing it by a normally distributed vector $\boldsymbol{\epsilon} \sim \mathcal{N}(0, \Sigma)$ (we will come back to how to choose the mean and covariance). The new particles are thus samples of the convolved distribution

$$p(\boldsymbol{x}') = \sum_i w_i \frac{1}{\sqrt{(2\pi)^k |\Sigma|}} \exp\left(-\frac{1}{2}(\boldsymbol{x}' - \boldsymbol{\mu}_i)^{\mathrm{T}} \Sigma^{-1} (\boldsymbol{x}' - \boldsymbol{\mu}_i)\right), \tag{2}$$

where k is the number of model parameters. A distribution of this form is known as a *mixture distribution*, and can be efficiently sampled by first choosing a particle, then choosing a perturbation vector.

To choose the mean $\boldsymbol{\mu}_i$ of each term in the resampling mixture distribution, we choose a vector that is a convex combination of the original particle location $\boldsymbol{x}_i$ and the expected model $\boldsymbol{\mu} = \mathbb{E}[\boldsymbol{x}]$, so that $\boldsymbol{\mu}_i = a\boldsymbol{x}_i + (1-a)\boldsymbol{\mu}$, where a is a tunable parameter of the resampling algorithm. We will use $a = 0.98$, as suggested by [26]. The covariance of each perturbation is then given by $\boldsymbol{\Sigma} = (1 - a^2)\operatorname{Cov}[\boldsymbol{x}]$. Our resampling algorithm then involves drawing n new particles from the distribution given by (2) and setting the weight of each new particle to $1/n$.

We combine these prior algorithms to obtain Algorithm 4, which is our complete algorithm for adaptively designing experiments using the SMC approximation. Note that we have left unspecified here the choice of local optimizer; in practice, this will be chosen depending on what works for a given experimental model. Due to the simulation cost of optimization, Algorithms 3 and 4 allow for the setting of an additional parameter, `approx_ratio`, that controls the quality with which the utility function is calculated.

4 Region and Hyperparameter Estimation

In addition to providing an accurate estimate of the true model parameters for the system, it is important to be able to quantify the uncertainty in the estimated model parameters. This task can be achieved by finding a region $\hat{X}$ of the space of models such that $\Pr(\boldsymbol{x}_0 \in \hat{X})$ is maximized and such that the volume $\mathrm{Vol}(\hat{X})$ is minimized.

We make the problem of region estimate amenable to analysis by SMC by reducing it to a problem of estimating an expectation value. In particular, the probability of the true model being within a region can be expressed as $\Pr(\boldsymbol{x}_0 \in \hat{X}) = \mathbb{E}[1_{\hat{X}}]$, where $1_{\hat{X}}$ is the *indicator function* for $\hat{X}$. The expectation value of this indicator function can then be computed using SMC as $\mathbb{E}[1_{\hat{X}}] \approx \sum_i w_i 1_{\hat{X}}(\boldsymbol{x}_i) = \sum_{i,\boldsymbol{x}_i \in \hat{X}} w_i$.

Thus, by construction, any region containing particles of total weight at least r will have an approximate probability mass of at least r. We formalize this intuition by introducing a *probability mass* function $m(R)$ on regions R such that $m(R) = \mathbb{E}[1_R]$. Similarly, let $\tilde{m}(R) = \sum_{i,\ i \in R} w_i$ be an approximation of $m(R)$ using the SMC algorithm.

We thus seek a region $\hat{X}$ such that $\mathrm{Vol}(\hat{X})$ is small, $m(\hat{X})$ is large and such that $\hat{X}$ is an efficiently computable property of the current SMC state. We achieve the latter two properties by choosing some appropriate geometric function of a set of particles X_r whose weight is above some threshold weight r; for example, the convex hull or the minimum-volume enclosing ellipse of X_r both satisfy $\tilde{m}(X_r) \geq r$ and may be computed using well-known classical algorithms [27, 28].

In practice, the covariance matrix of the posterior distribution will often suffice as a region estimate because the posterior distribution will often be approximately normally distributed. This assumption holds when the Fisher information is non–singular. More generally, under the assumption of a normally distributed posterior, the error ellipse of points $\boldsymbol{x}$ satisfying

$$(\boldsymbol{x} - \boldsymbol{\mu})^{\mathrm{T}} \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{\mu}) \leq Z^2 \tag{3}$$

for some $Z > 0$ will contain a ratio $(\mathrm{cdf}_{\mathcal{N}}(Z) - \mathrm{cdf}_{\mathcal{N}}(-Z))^d = \mathrm{erf}\left[\frac{Z}{\sqrt{2}}\right]^d$ of the particle weight, where $\mathrm{cdf}_{\mathcal{N}}(Z)$ is the cumulative distribution function for the normal distribution, evaluated at Z. Thus, if the assumption of a normal posterior is a good approximation, then the covariance matrix of the posterior distribution as approximated by SMC can be used as a region estimator.

We can generalize further by considering the fact that quantum systems seldom have consistent Hamiltonians from experiment to experiment, due to experimental errors. Thus, we would like to form a region estimate for such Hamiltonians that encompasses experiment-to-experiment variation, but that expands that region as little as possible. Hyperparameters allow us to address this by switching from the problem of estimating Hamiltonian parameters to one that involves learning the parameters that describe the *distribution* of Hamiltonian parameters.

We denote the hyperparameters for a model Hamiltonian as $\boldsymbol{y}$ to avoid subtle conceptual differences between the hyperparameters and the distributions on $\boldsymbol{x}$ that they describe. The

probability distribution for $\boldsymbol{x}$ can then be written as $\Pr(\boldsymbol{x}|\boldsymbol{y})$. Despite interpretational differences, the hyperparameters can also be learned using Algorithm 4 in exactly the same way that $\boldsymbol{x}$ is learned. The region estimates yielded by the algorithm are region estimations for $\boldsymbol{y}$ and, as we will show shortly, can easily be converted into region estimates for $\boldsymbol{x}$.

The drawback to this approach is that computations of the likelihood function can become much more expensive because it typically will have to be computed by sampling from the parameterized distribution. In some important special cases, this drawback can be avoided by analytically performing the marginalization over $\boldsymbol{x}$,

$$\Pr(D|\boldsymbol{y}) = \int \mathrm{d}\boldsymbol{x} \Pr(D|\boldsymbol{x}) \Pr(\boldsymbol{x}|\boldsymbol{y}).$$

In Section 5, we discuss a particular case where the marginalization is analytically tractable.

The resulting means and covariance matrices for $\boldsymbol{y}$ can be readily converted to the corresponding quantities for $\boldsymbol{x}$ by using the chain rule for expectation values,

$$\mathbb{E}_{\boldsymbol{x},\boldsymbol{y}}[\boldsymbol{x}] = \mathbb{E}_{\boldsymbol{y}}[\mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}}[\boldsymbol{x}]]. \tag{4}$$

This expectation value can be computed using the posterior distribution $\Pr(\boldsymbol{y}|D)$ and the intermediate model distribution $\Pr(\boldsymbol{x}|\boldsymbol{y})$, which will typically be easy to compute from the definition of the hyperparameters. The covariance matrix for $\boldsymbol{x}$ is slightly more complicated. It is straightforward to verify that

$$\mathrm{Cov}_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{y}}\left[\mathrm{Cov}_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x})\right] + \mathrm{Cov}_{\boldsymbol{y}}\left(\mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}}[\boldsymbol{x}]\right). \tag{5}$$

For the special case that $\boldsymbol{x}$ is a single parameter, the covariance can be replaced with the variance to obtain that

$$\mathrm{Var}_{\boldsymbol{x},\boldsymbol{y}}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{y}}\left[\mathrm{Var}_{\boldsymbol{x}|\boldsymbol{y}}(\boldsymbol{x})\right] + \mathrm{Var}_{\boldsymbol{y}}\left(\mathbb{E}_{\boldsymbol{x}|\boldsymbol{y}}[\boldsymbol{x}]\right). \tag{6}$$

Using the covariance ellipse region estimate given by (3) to estimate a hyperparameter region thus translates to a region estimator for the model parameters $\boldsymbol{x}$, if the distribution over hyperparameters $\boldsymbol{y}$ is approximately Gaussian near its peak. In the limit of many experiments, we find that this is a good assumption, as is discussed in Section 5.3.

5 Single-Qubit Test Case

We will now proceed to apply our techniques to learning unknown parameters in a single qubit system. Our model has a qubit that evolves under an internal Hamiltonian of the form $H(\omega) = \frac{\omega}{2}\sigma_z$. Here ω is an unknown parameter whose value we want to estimate. An experiment consists of preparing a single known input state $\psi_{\mathrm{in}} = |+\rangle$, the $+1$ eigenstate of σ_x, evolving under H for time t and performing a measurement in the σ_x basis.

We will slightly generalize this model by allowing noise sources which lead to a decay in the information extractable from any measurement. This can manifest from, for example, a T_2 dephasing process which leads to the following likelihood function:

$$\Pr(0|\omega;t) = e^{-\frac{t}{T_2}} \cos^2\left(\frac{\omega}{2}t\right) + \frac{1 - e^{-\frac{t}{T_2}}}{2}, \tag{7}$$

where ω is the unknown parameter to be estimated, t is the controllable parameter and T_2 is a known constant. This model was studied in references [29, 30, 31] where analytical solutions based on Fisher information and the Cramer-Rao bound were given.

We consider the seemingly simple generalization of this model where both ω *and* T_2 are unknown. Even for such a simple generalization as this, the methods discussed in [29, 30, 31] are not adequate for this more general problem. In particular, the Fisher matrix of any one measurement is singular and hence the standard Cramer-Rao bound does not hold – nor is it possible to utilize standard asymptotic approximations to normal distributions.

This generalization is closely related to the case in which T_2 is infinite, but where the "true" precession frequency ω is itself distributed according to a Gaussian distribution of mean μ and variance σ^2. In this case, following the discussion of Section 4, the probability of data conditioned on the *hyperparameters* $\boldsymbol{y} = (\mu,\ \sigma)$ can be found by marginalizing over the intermediate random variable ω, so that

$$\Pr(d|\mu,\sigma;t) = \int \Pr(d|\omega)\Pr(\omega|\mu,\sigma)\mathrm{d}\omega. \tag{8}$$

For the specific example of the Gaussian distribution,

$$\Pr(0|\mu,\sigma;t) = \frac{1}{\sigma\sqrt{2\pi}}\int \cos^2\left(\frac{\omega t}{2}\right) e^{-\frac{(\omega-\mu)^2}{\sigma^2}}\mathrm{d}\omega = \frac{1}{2}\left(1 + e^{-2\sigma^2 t^2}\cos(2\mu t)\right). \tag{9}$$

At this point, we have entirely removed ω from the problem, leaving a two-parameter model, where we wish to estimate the mean and variance of an unknown normal distribution.

As another example, instead of marginalizing against a Gaussian distribution, we consider the case that the intermediate model parameter ω is drawn from a Lorentz distribution. A Lorentz distribution is completely determined by its location and scale parameters ω_0 and γ, respectively, and so we use these hyperparameters to derive a new model,

$$\Pr(0|\omega_0,\gamma;t) = \int \cos^2(\omega t/2)\frac{1}{\pi\gamma\left(\frac{(\omega-\omega_0)^2}{\gamma^2}+1\right)}\mathrm{d}\omega = \frac{1}{2}\left(1 + e^{-t\gamma}\cos\left(t\omega_0\right)\right). \tag{10}$$

Note that if we identify $\gamma = T_2^{-1}$, then the Lorentz hyperparameter model is the identical to that of Equation (7). This illustrates the relationship between decoherence processes and the lack of knowledge formalized by a hyperparameter model. In a similar fashion, (9) is also model of decoherence. Due to the t^2 dependence of the Gaussian-hyperparameter model, (9) represents a decoherence process that cannot be written in Lindblad form [32] because it cannot be drawn from a quantum dynamical semigroup.

5.1 Results for Unknown T_2

Here we report on the performance of our algorithm for the comparatively challenging task of learning Hamiltonian parameters without a precise estimate of T_2. These calculations were performed using the true distributions $\omega \sim \mathcal{N}(0.5, 0.0025)$ and $1/T_2 \sim \mathcal{N}(0.001, 0.00025^2)$, and with the scale matrix $\boldsymbol{Q} = \mathrm{diag}(1, 0.0025/0.00025^2) = \mathrm{diag}(1, 100)$.

The guess heuristic that we focus on chooses times randomly from an exponential distribution with mean 1 000, corresponding to the mean value of T_2 according to the initial prior. This choice of guess function is motivated by the fact that the most informative experiments (as measured by Fisher information) tend to occur at $t \approx T_2$ [31]. A secondary benefit is that the guess function is certainly not optimal for the problem, and will allow us to illustrate that a sub–optimal guess function can be used in concert with local optimization (in our case Newton conjugate gradient optimization (NCG) is used) to find near optimal experiments given the current state of knowledge about the unknown Hamiltonian.

We examine the variation of the MSE with the number of guesses used in Figure 1. The figure shows that, in the absence of local optimization of experiment times, the MSE for both

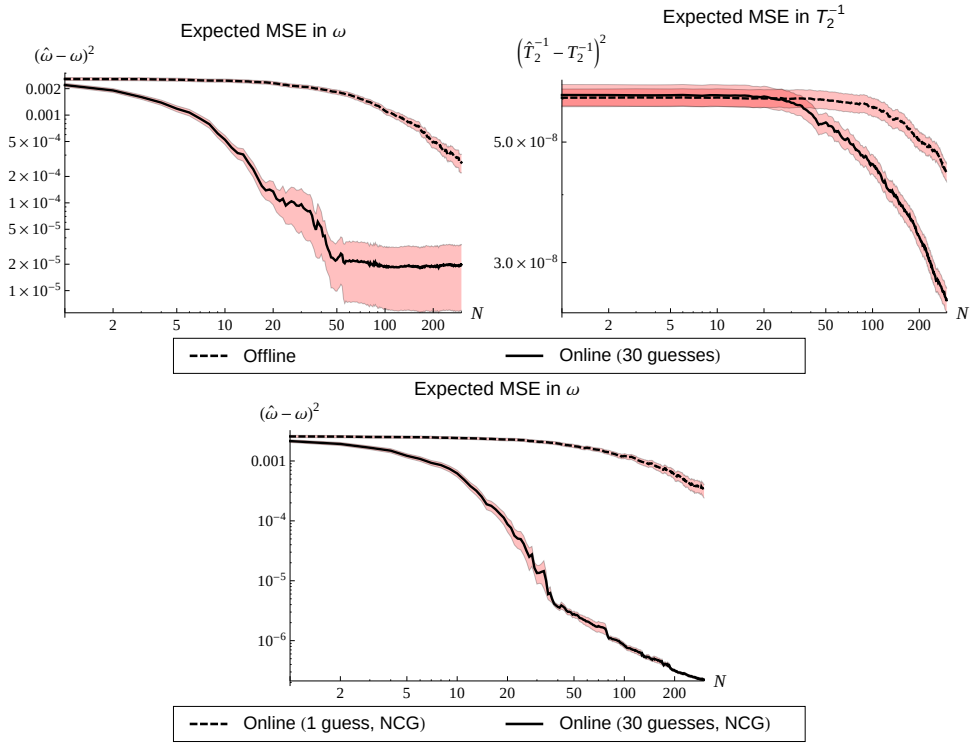


Figure 1 Benchmarking of the "unknown-T_2"model using $n = 5\,000$ particles and random initial guesses without local optimization. Data indicated dashed lines correspond to trials where a single initial guess was used for each experiment, while data indicated by solid lines were collected using 30 guesses per experiment. Errors in estimating performance are indicated by red shaded regions about each curve.

ω and $1/T_2$ is significantly improved by using an increased number of guesses. In particular, we find that if 30 guesses are used, then only 50 experiments are required on average to learn ω within a 0.9% error, even without a well characterized T_2. The improvement is much more substantial for ω than it is for $1/T_2$ because the contrast on T_2 is much less significant.

Figure 1 examines the effect of increasing the number of guesses for strategies that use NCG. The most significant qualitative difference between the data collected using NCG and that of Figure 1 is that the MSE for ω shows no evidence of saturating and instead continues to shrink as the number of experiments are increased (as seen most clearly in Figure 2). This implies that our randomized guess heuristic is unlikely to randomly guess very informative experiments after a fixed number of experiments, but the landscape is sufficiently devoid of local optima that NCG optimization finds informative experiments in the vicinity of our uninformed guesses. We also observe that NCG does not substantially improve the MSE if 1 guess is used. This suggests that the landscape is not sufficiently convex that local optimization about an individual guess is likely to find experiments that are substantially more informative. We therefore conclude that increasing number of guesses used and using NCG substantially improves the MSE for ω and has a much more subtle effect on the knowledge of T_2 if local optimization is used.

It is useful to benchmark the performance of our algorithm against the Bayesian Cramer-Rao bound (BCRB—see appendix), which gives a lower bound on the MSE. Figure 3 provides

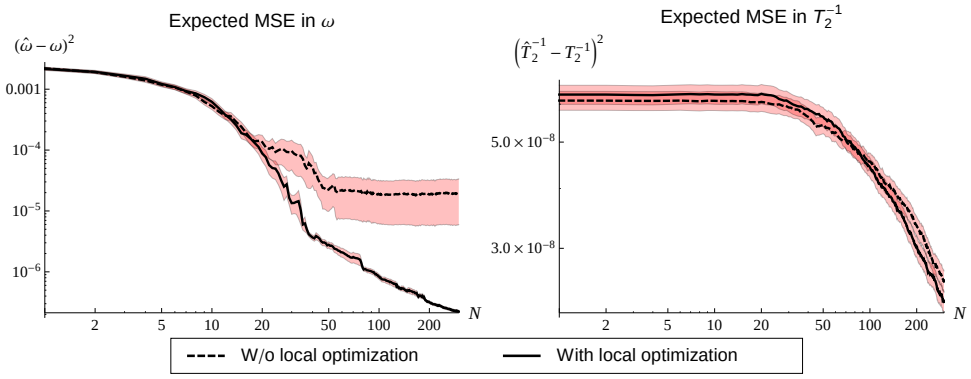


Figure 2 Benchmarking of the "unknown-T_2"model using $n = 5\ 000$ particles and 30 random initial guesses. Data indicated dashed lines correspond to trials where a each initial guess was used without local optimization, while data indicated by solid lines were collected using NCG optimization for each guess. The unoptimized data is averaged over 1,109 trials while the optimized data is averaged over 930 trials. Errors in estimating performance are indicated by red shaded regions about each curve.

a comparison of the MSE, the estimate of the MSE given by the variance of the posterior and the BCRB for ω, T_2^{-1} and $\mathrm{Tr}(\Sigma \cdot \boldsymbol{Q})$. We see that the expected posterior variance is typically within statistical error of the MSE for all three of these quantities, suggesting that the posterior variance can be used as a very good estimate of the MSE for this model. We also note that the MSE is very close to the MSE for the T_2^{-1} data and $\mathrm{Tr}(\Sigma \cdot \boldsymbol{Q})$. The MSE for ω is within a constant multiple of the BCRB. We do not, in fact, expect that the MSE in ω should approach the BCRB because the algorithm chooses experiments to optimize $\mathrm{Tr}(\Sigma \cdot \boldsymbol{Q})$ rather than the error for either ω or T_2^{-1} individually.

5.2 Region Estimation

One of the most substantial contributions of our algorithm is its ability to provide region estimates for the location of the true Hamiltonian, which allow us to quantify our uncertainty in the true model parameters. We compare the probability mass enclosed by the covariance region estimator described in Section 4. A simplifying assumption is made in our analysis: we assume that the posterior distribution is approximately Gaussian. Although difficult to justify theoretically, we have find for the examples that we consider that the posterior appears Gaussian locally around our estimate after a sufficiently large number of experiments. We expect this behavior to be generic, although region estimators such as the convex hull or the minimum-volume enclosing ellipse may be used even if the posterior is not approximately normal.

Under the Gaussian model of the posterior distribution, we expect the true model parameters to be within an ellipse described by the covariance matrix whose volume is then described by the Z–score used. For example, in the one–dimensional case approximately 95% of the probability mass is located within 2–standard deviations, which corresponds to $Z = 2$. We choose $Z = 3$ standard deviations from the mean for these examples which correspond to probability masses of $\tilde{m}(\mathrm{Cov}(\hat{\boldsymbol{x}})^{-1}/Z^2) \approx 0.9973$ and $\tilde{m}(\mathrm{Cov}(\hat{\boldsymbol{x}})^{-1}/Z^2) \approx 0.9946$ for the one– and two–parameter cases respectively.

We show in Figure 4, and [33], that the approximate probability mass $\tilde{m}$ approaches the probability mass we would expect for a normal distribution for the known-T_2 model in

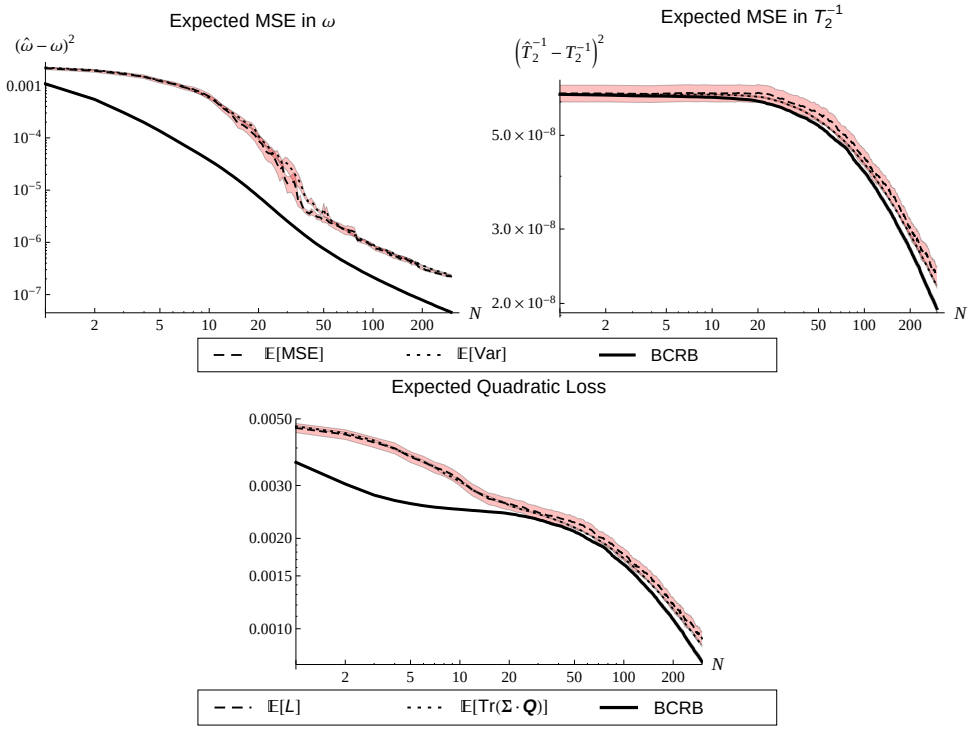


Figure 3 The actual and estimated performance, as a function of the number of measurements N, of the sequential Monte Carlo algorithm for $n = 5\,000$ particles. The model is that of equation (7) with unknown T_2 (which is estimated as $\Gamma = 1/T_2$ for numerical precision considerations). The dotted curve is the posterior variance of the particles; dashed is the actual mean squared error and solid is numerically calculated Bayesian Cramer-Rao lower bound. In the upper subfigures, the MSE and variances are those of the individual parameters ω and T_2^{-1}, respectively, while the lower subfigure shows the actual and estimated quadratic losses scaled using $\boldsymbol{Q} = \mathrm{diag}(1, \sigma_\omega^2/\sigma_{T_2^{-1}}^2)$, where σ_ω^2 and $\sigma_{T_2^{-1}}^2$ are the variances in ω and T_2^{-1} according to the initial prior π.

the limit of large N, providing evidence in favor of our use of the covariance ellipse as a region estimator on the posterior. In particular, we note that the value of $\tilde{m}$ approaches 0.9973, such that the quality of the Gaussian approximation improves as we collect data. The transient behavior for small experiment numbers occurs because insufficient experiments have been considered for the posterior to approach a Gaussian. In this specific example, the average differences in enclosed probability mass after each experiment are on the order of 0.01%, and thus may not be of practical significance.

5.3 Hyperparameter Region Estimation Performance

Having demonstrated the effectiveness of our region estimation algorithm, it remains to show that the generalization to hyperparameter regions works as described in Section 4. The objective here is to analyze the robustness of our algorithm in the presence of fluctuating "true" parameters of the Hamiltonian. We do so by using the Gaussian hyperparameter model as discussed in Section 5, then comparing the model parameter region volume and probability mass for the region estimated from Equation (5) to the volume and probability mass of the corresponding "true" model parameter region. We benchmark this model by

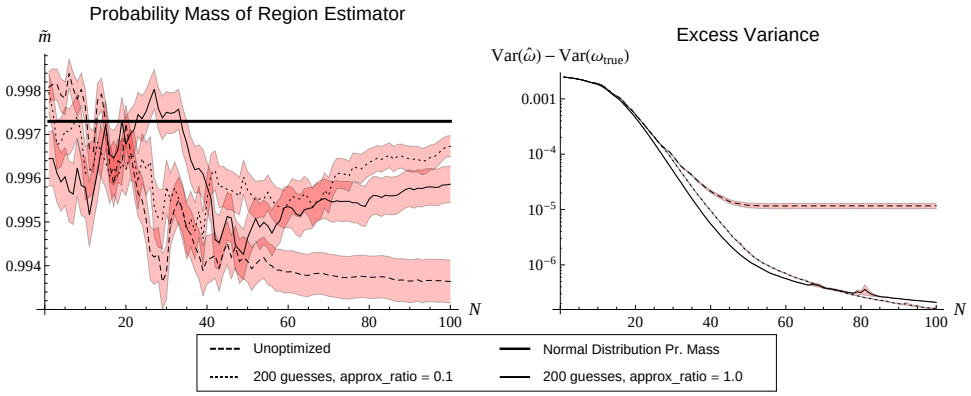


Figure 4 Benchmarking region estimators for Gaussian hyperparameter model using $n = 2\,000$ particles, $\omega \sim \mathcal{N}(\mu, \sigma^2)$ where $\mu \sim \mathcal{N}(0.5, 0.001^2)$ and $\sigma^2 \sim \mathcal{N}(0.0025, 0.0025^2)$.

choosing "true" hyperparameters μ and σ^2 for ω according to the normal distribution

$$\mu, \sigma^2 \sim \mathcal{N}\left[(\mu_\mu, \mu_{\sigma^2}), \mathrm{diag}(\sigma_\mu^2, \sigma_{\sigma^2}^2)\right]. \tag{11}$$

Recall that the unknown frequency is distributed as $\omega \sim \mathcal{N}(\mu, \sigma^2)$. In particular, this true distribution does not admit any correlation between the mean and variance hyperparameters. We then use the true distribution as our prior distribution.

In Figure 4, we find that the probability mass contained within our estimated region for the Hamiltonian agrees well with our theoretical expectations. In particular, we assume a Gaussian posterior and use a Z-score of 3 which implies that we should anticipate that 99.7% of the probability mass will lie within the region estimation of $\mathbb{E}[\hat{\omega}] \pm 3\sqrt{\mathrm{Var}(\hat{\omega})}$. We find very good agreement with this assumption, and find that at worst 99.4% of the probability mass for the hyperparameters lies within the estimated region. The data also suggests that these small differences vanish for the optimized data sets, which appear to approach the ideal enclosed probability mass of 99.7% in the limit of large N.

Hyperparameters are not typically a quantity of interest by themselves. They usually are of relevance because they parameterize a distribution of the unknown parameter. Following Equation (6), we calculate $\mathrm{Var}(\hat{\omega})$ as $\mathrm{Var}(\hat{\omega}) = \mathrm{Var}(\hat{\mu}) + \mathbb{E}[\hat{\sigma}^2]$. We find that, as the number of experiments grows, our region estimator for ω slightly overestimates the "true" variance of ω (on average). This bias vanishes as the number of experiments increases. We can therefore conclude that we can use the method of hyperparameters to robustly estimate the distribution of an unknown frequency, even in the presence of noise.

5.4 Computational Cost

Another way that we can assess the cost of inferring the Hamiltonian of a system is in terms of the classical computing time needed to learn the Hamiltonian parameters to within a fixed error tolerance (as measured by the number of likelihood calls made). Our previous discussion found that the experimental time (measured by the number of experiments) can be minimized by choosing measurements that minimize the risk, and showed that increasingly sophisticated heuristics for generating these guesses tended to reduce the experimental time. This suggests that a trade-off may be present between the experimental time and the classical processing time needed to learn the parameter. This tradeoff will become increasingly relevant as the size of the quantum system grows, since existing quantum simulation

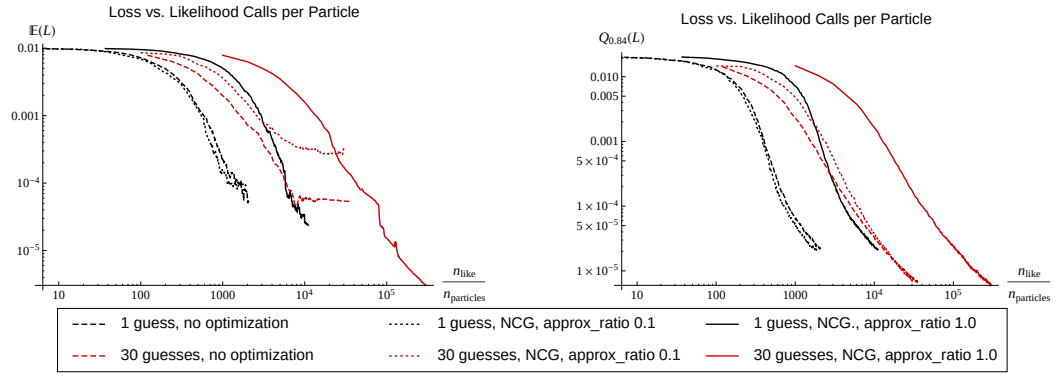


Figure 5 This figure compares the mean–square error as a function of the computational time for the known T_2 model with $T_2 = 100$, 5 000 particles, `approx_ratio` $= 1$ and guessed experimental times chosen randomly from an exponential distribution with mean T_2. The expected loss incurred by each optimization strategy is shown in the left figure and the figure on the right shows the 84^{th} percentile $Q_{0.84}$ of the loss, such that no more than 16% of trials incur loss greater than the shown percentile.

techniques do not scale efficiently with the number of particles in the system and thus the cost of performing a likelihood call may asymptotically become much more expensive than performing an experiment.

If computational time is of primary importance (rather than experimental time), then the relative merits of the experimental design heuristics changes. In total, our data sets in Figure 5 required (on average) a number of likelihood calls that fell within the range $[1.05 \times 10^7, 1.5 \times 10^9]$. A likelihood call required the evaluation of $\exp(-t/T_2)\cos^2\left(\frac{\omega}{2}t\right) + (1 - \exp(-t/T_2))/2$, which required time on the order of 10^{-7} seconds on our computers and lead to total computational times that were on the order of a second to a minute. If the rate at which experiments can be performed were much faster than 200 Hz then the utility of our algorithm as a means to speed up data collection may be lost. If the two rates are approximately comparable, then interesting trade-offs appear between the computational time needed and the total experimental time.

These trade-offs become apparent by plotting the scaling of the MSE as a function of the computational time for the randomized guess heuristic in Figure 5. The first feature that is obvious from the plot is that the strategies which yielded the lowest MSE per experiment tend to yield the highest MSE per likelihood call; although several of these strategies cause the expected loss (mean–square error) to saturate after a finite number of experiments. In particular, this causes the strategy with 30 guesses and no optimization as well as the strategy with 30 guesses, NCG optimization and `approx_ratio` $= 0.1$ to intersect the curve for the cases with NCG optimization and `approx_ratio` $= 1$. Here the approximation ratio is the ratio of the particles that are used in the updating (see Algorithm 3). On the surface, this seems to indicate that the more expensive heuristics may have an advantage if small loss is desired; but this is misleading and to get a complete picture we need to look at more than just the expected performance of the strategies.

We can get a better understanding of this saturation by looking at the plot of the 84^{th} percentile of the loss in Figure 5, which shows that all of these strategies continue to provide improved estimates of ω even into this regime of saturation for at least 84% of the trials considered. This shows that there were a few trials where very poor guesses were chosen and the algorithm became stuck at a large MSE. The data also suggests that the use of NCG and

a large value of the approximation ratio can mitigate these problems, causing the learning algorithm to become more stable at the price of requiring more computational time.

6 Multi–Qubit Test Case

We will now focus on an example that shows the viability of our algorithm in cases where the Hamiltonian acts on many qubits rather than just one. The model that we consider is the Ising Model with no external magnetic field with a complete graph of interactions on n qubits:

$$H = \sum_{i>j} x_{i,j} \sigma_i^z \sigma_j^z, \tag{12}$$

where $x_{i,j}$ are real valued coefficients. In these examples, we choose these coefficients randomly from the interval $[0, 1]$ (the absence of frustrations does not affect the difficulty of the learning problem). The goal of the learning problem is to learn each $x_{i,j}$. We represent these parameters for the Hamiltonian H_k using the vector $\boldsymbol{x}_k$.

In direct analogy to the single qubit examples, an experiment involves setting the initial state to be $|+\rangle^{\otimes n}$, evolving the state under the Hamiltonian in (12) for some evolution time t and finally applying the Hadamard transform to each qubit and measuring the result in the computational basis. The evolution time is the only control parameter in these experiments.

We also use a different guess heuristic to choose t for this problem than the exponential random guess heuristic that was used in previous experiments since, most notably, we are not considering decoherence processes. We choose the times by drawing two Hamiltonians H_j and $H_{j'}$ from the current prior $\Pr(H_k|d_n, \ldots; c_n, \ldots)$ and then choose $t = 1/|\boldsymbol{x}_j - \boldsymbol{x}_{j'}|_2$, rather than choosing the times randomly and using NCG to find a locally optimal experiment near that guess. We call this strategy the "particle guess heuristic." Although it may not seem it, this strategy is adaptive. In particular, the particle guess heuristic will tend to choose experiments that have short evolution times when the posterior distribution is broad, and longer evolution times when the distribution is narrow. Long evolution times are needed to distinguish dynamics of nearby Hamiltonians, thus the heuristic adaptively chooses experiments that will be informative based on the current uncertainty in the unknown parameters. We pick this strategy because it outperforms the exponential guessing strategy for such problems, especially in absentia of local optimization.

We avoid local optimization in these numerical experiments because the effects of incorporating local optimization have been well discussed in previous examples and the improvements brought about by using local optimization in these cases is qualitatively similar to the single qubit case. Furthermore, the cost of computing the likelihood in these cases is substantially higher so including local optimization would only restrict the range of numerical examples that we could provide.

We examine the scaling of the quadratic loss that occurs when using SMC to learn Hamiltonian parameters for (12), using $\mathbf{Q} = \mathbb{1}$, in Figure 6. It is clear that our algorithm is capable of learning parameters of many–qubit Hamiltonians. The scaling of the quadratic loss is, similar to the single qubit case, exponential in the number of experiments taken. The slower rate of learning for the $n = 5$ and $n = 6$ cases is due largely to the fact that these cases have 10 and 15 unknown parameters that must be learned, in contrast to the 6 that must be learned in the $n = 4$ case. This shows that our algorithm is capable of learning Hamiltonian parameters in not just single qubit cases, but also in multi–parameter estimation problems that are relevant in real world applications such as characterizing superconducting quantum devices or certifying an analog quantum simulator.

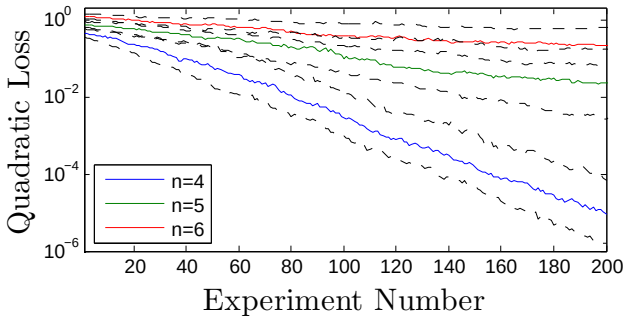


Figure 6 This figure compares the quadratic loss as a function of the computational time for the Ising model with 20 000 particles, `approx_ratio` = 1 and guessed experimental times chosen using the particle guess heuristic. The dashed lines represent the 25^{th} and 75^{th} percentile of the quadratic loss, whereas the solid lines represent the median.

7 Conclusions

Our work provides a simple algorithm that applies Bayesian inference to learn a Hamiltonian in an online fashion; that is to say, that our algorithm learns the Hamiltonian parameters as the experiment proceeds rather than collecting data and inferring the Hamiltonian through post–processing. This eliminates the need to store and process gigabytes of data that are recovered from even relatively short experiments. Our work has several advantages over existing approaches to learning Hamiltonian parameters. First, it can be used to estimate the optimal parameterization of the dynamics of an arbitrary quantum system within a space of model Hamiltonians. Second, it can be used to provide a region estimatate of the Hamiltonian parameters. The importance of this is obvious: it allows us to not only learn the unknown parameters but also quantify our uncertainty in them. Third, our analysis of the algorithm shows a clear trade off between the experimental time and the computational time needed to parameterize the Hamiltonian.

We note a natural extension of our algorithm to include classical simulators which do not deterministically compute the likelihood function but generate random samples according to it [34]. The distinction between *strong* and *weak* simulation has been a topic of recent interest in computational complexity [35, 36]. The present work and that of [34] add to the discussion of this distinction by clarifying the relationship between simulating a physical model classically and estimating the parameters in it.

An extension of our work would be to consider more advanced optimization heuristics than conjugate gradient searches (such as particle swarm optimization algorithms). Similarly, more advanced resampling techniques may lead to substantial reductions in the number of particles which in turn would reduce the computational cost of the algorithm. Finally, estimates of how the number of experiments required to achieve a specific mean–square error scales with the number of unknown parameters would be an important extension of this work since it would assess the viability of these techniques for controlling and characterizing larger quantum systems.

Acknowledgements. This work was financially supported by the Canadian government through NSERC and CERC and by the United States government through DARPA. NW would like to acknowledge funding from USARO-DTO.

References

1 B. P. Lanyon, C. Hempel, D. Nigg, M. M¨ uller, R. Gerritsma, F. Zähringer, P. Schindler, J. T. Barreiro, M. Rambach, G. Kirchmair, M. Hennrich, P. Zoller, R. Blatt, and C. F. Roos. Universal digital quantum simulation with trapped ions. *Science* **334** 57, 2011.

2 R. Gerritsma, B. P. Lanyon, G. Kirchmair, F. Zähringer, C. Hempel, J. Casanova, J. J. Garcia-Ripoll, E. Solano, R. Blatt, and C. F. Roos. Quantum simulation of the klein paradox with trapped ions. *Physical Review Letters* **106** 060503, 2011.

3 K. Kim, M.-S. Chang, S. Korenblit, R. Islam, E. E. Edwards, J. K. Freericks, G.-D. Lin, L.-M. Duan, and C. Monroe. Quantum simulation of frustrated Ising spins with trapped ions. *Nature*, **465** 590, 2010.

4 Matteo Paris and Jaroslav Rehacek, editors. *Quantum State Estimation*, volume 649 of *Lecture Notes in Physics*. Springer, 2004.

5 Ariel Bendersky, Fernando Pastawski, and Juan Pablo Paz. Selective and efficient estimation of parameters for quantum process tomography. *Physical Review Letters* **100** 190403, 2008.

6 Ariel Bendersky, Fernando Pastawski, and Juan Pablo Paz. Selective and efficient quantum process tomography. *Physical Review A* **80** 032116, 2009.

7 M. Mohseni and A. T. Rezakhani. Equation of motion for the process matrix: Hamiltonian identification and dynamical control of open quantum systems. *Physical Review A* **80** 010101, 2009.

8 M P A Branderhorst, J Nunn, I A Walmsley, and R L Kosut. Simplified quantum process tomography. *New Journal of Physics* **11** 115010, 2009.

9 Steven T. Flammia and Yi K. Liu. Direct Fidelity Estimation from Few Pauli Measurements. *Physical Review Letters* **106** 230501, 2011.

10 Marcus P. da Silva, Olivier L. Cardinal, and David Poulin. Practical Characterization of Quantum Devices without Tomography. *Physical Review Letters* **107** 210404, 2011.

11 Arnaud Doucet and Adam M. Johansen. A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later. In *The Oxford Handbook of Nonlinear Filtering*. Oxford University Press, 2009.

12 Thomas J. Loredo. Bayesian Adaptive Exploration. *AIP Conference Proceedings* **707** 330, 2004.

13 Hendrik Kuck, Nando de Freitas, and Arnaud Doucet. SMC Samplers for Bayesian Optimal Nonlinear Design. In *Nonlinear Statistical Signal Processing Workshop*. IEEE, 2006.

14 Bruno Scarpa and David B. Dunson. Bayesian methods for searching for optimal rules for timing intercourse to achieve pregnancy. *Statistics in Medicine* **26** 1920, 2007.

15 D. R. Cavagnaro, M. A. Pitt, and J. I. Myung. Adaptive Design Optimization in Experiments with People. *Advances in Neural Information Processing Systems* **22** 234, 2010.

16 N. Kantas, A. Lecchini-Visintini, and J. M. Maciejowski. Simulation-based Bayesian optimal design of aircraft trajectories for air traffic management. *International Journal of Adaptive Control and Signal Processing* **24** 882, 2010.

17 Xun Huan and Youssef M. Marzouk. Simulation-based optimal Bayesian experimental design for nonlinear systems. *Journal of Computational Physics* **232** 288, 2013.

18 F. Huszár and N. M. T. Houlsby. Adaptive Bayesian quantum tomography. *Physical Review A* **85** 052120, 2012.

19 Rocco A. Servedio and Steven J. Gortler. Equivalences and Separations Between Quantum and Classical Learnability. *SIAM Journal on Computing* **33** 1067, 2004.

20 Esma Aïmeur, Gilles Brassard, and Sébastien Gambs. Machine Learning in a Quantum World Advances in Artificial Intelligence. volume 4013 of *Lecture Notes in Computer Science*. Springer Berlin / Heidelberg, 2006.

21 Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Science*, **463** 3089, 2007.

22 Alexander Hentschel and Barry C. Sanders. Machine Learning for Precise Quantum Measurement. *Physical Review Letters*, **104** 063603, 2010.

23 Kristen L. Pudenz and Daniel A. Lidar. Quantum adiabatic machine learning. *Quantum Information Processing* **12** 2027, 2013.

24 Alexander Hentschel and Barry C. Sanders. Efficient algorithm for optimizing adaptive quantum metrology processes. *Physical Review Letters*, **107** 233601, 2011.

25 Alexandr Sergeevich and Stephen D. Bartlett. Optimizing qubit Hamiltonian parameter estimation algorithms using PSO. *Proceedings of 2012 IEEE Conference on Evolutionary Computation* 1, 2012.

26 J. Liu and M. West. *Combined parameter and state estimation in simulation-based filtering.* Springer-Verlag, 2000.

27 Michael J. Todd and E. Alper Yıldırım. On Khachiyan's algorithm for the computation of minimum-volume enclosing ellipsoids. *Discrete Applied Mathematics* **155** 1731, 2007.

28 C. Bradford Barber, David P. Dobkin, and Hannu Huhdanpaa. The quickhull algorithm for convex hulls. *ACM Transactions on Mathematical Software* **22** 469, 1996.

29 Alexandr Sergeevich, Anushya Chandran, Joshua Combes, Stephen Bartlett, and Howard Wiseman. Characterization of a qubit Hamiltonian using adaptive measurements in a fixed basis. *Physical Review A* **84** 052315, 2011.

30 Christopher Ferrie, Christopher E. Granade, and D. G. Cory. Adaptive hamiltonian estimation using bayesian experimental design. *AIP Conference Proceedings* **1443** 165, 2012.

31 Christopher Ferrie, Christopher Granade, and D. Cory. How to best sample a periodic probability distribution, or on the accuracy of Hamiltonian finding strategies. *Quantum Information Processing* **12** 611, 2013.

32 G. Lindblad. On the generators of quantum dynamical semigroups. *Communications in Mathematical Physics* **48** 119, 1976.

33 Christopher E Granade, Christopher Ferrie, Nathan Wiebe, and D G Cory. Robust online Hamiltonian learning. *New Journal of Physics* **14** 103013, 2012.

34 Christopher Ferrie and Christopher E Granade. Likelihood-free quantum inference: tomography without the born rule, URL http://arxiv.org/abs/1304.5828, 2012.

35 Scott Aaronson and Alex Arkhipov. The Computational Complexity of Linear Optics, URL http://arxiv.org/abs/1011.3245, 2010.

36 M. Van den Nest. Simulating quantum computers with probabilistic methods. *Quantum Information & Computation* **11** 784, 2011.

37 Eric Jones, Travis Oliphant, Pearu Peterson, et al. SciPy: Open source scientific tools for Python, 2001–.

38 E. L. Lehmann and George Casella. *Theory of Point Estimation.* Springer, 2nd edition, 1998.

39 James O. Berger. *Statistical Decision Theory and Bayesian Analysis.* Springer, 2nd edition, 1985.

40 Robin Blume-Kohout and Patrick Hayden. Accurate quantum state estimation via "Keeping the experimentalist honest", URL http://arxiv.org/abs/quant-ph/0603116, 2006.

41 Richard D. Gill and Boris Y. Levit. Applications of the van Trees Inequality: A Bayesian Cramér-Rao Bound. *Bernoulli* **1** 59, 1995.

42 P. Tichavsky, C. H. Muravchik, and A. Nehorai. Posterior Cramer-Rao bounds for discrete-time nonlinear filtering. *IEEE Transactions on Signal Processing* **46** 1386, 1998.

A Utility Functions and the Cramer-Rao Lower Bound

Given a set of observed outcomes, the choice of subsequent experimental parameters that informs us most about the model parameters is given by the *utility function*. We test our method with a utility function that minimizes the expected variance in $\Pr(\boldsymbol{x}|d_{N+1}, D; c_{N+1}, C)$. We show that this choice is optimal for minimizing the the mean squared error of the protocol.

An *estimator* is a function $\hat{\boldsymbol{x}}$ that takes a set of observed data D collected from a set of experiments with controls C and produces an estimate for the unknown parameters $\boldsymbol{x}$. Here, we evaluate the quality of an estimator $\hat{\boldsymbol{x}}$ by using a generalization of the *squared error loss* called the *quadratic loss* as our figure of merit. The quadratic loss is defined for a vector of parameters $\boldsymbol{x}$, data D and experiment designs C, as

$$L_{\boldsymbol{Q}}(\boldsymbol{x}, \hat{\boldsymbol{x}}(D, C)) = (\boldsymbol{x} - \hat{\boldsymbol{x}}(D, C))^{\mathrm{T}} \boldsymbol{Q} (\boldsymbol{x} - \hat{\boldsymbol{x}}(D, C)), \tag{13}$$

where $\boldsymbol{Q}$ is a positive definite matrix on the space of unknown parameters that defines the relative scale between the various parameters of interest. The quadratic loss function is useful to us in that it is computationally inexpensive to calculate and may be analyzed by well-known statistical techniques. In particular, the Cramer-Rao bound can be used to lower-bound the mean quadratic loss incurred by an estimator, under the hypothesis of a given true model $\boldsymbol{x}$ [38].

Following a decision theoretic methodology [39], the *risk* of an estimator given a set of experiment designs C is its expected performance over all possible outcomes D with respect to the loss function:

$$R(\boldsymbol{x}, \hat{\boldsymbol{x}};\ C) = \mathbb{E}_{D|\boldsymbol{x};C}[L(\boldsymbol{x}, \hat{\boldsymbol{x}}(D; C))].$$

The Bayes risk is the average of this quantity with respect to a prior distribution on $\boldsymbol{x}$ (denoted π) and is given explicitly by

$$r(\pi; C) = \mathbb{E}_{\boldsymbol{x}}[R(\boldsymbol{x}, \hat{\boldsymbol{x}};\ C)] = \int \pi(\boldsymbol{x}) R(\boldsymbol{x}, \hat{\boldsymbol{x}};\ C) \mathrm{d}\boldsymbol{x}.$$

where $\hat{\boldsymbol{x}}$ is assumed to be a *Bayes estimator*, which means it is the one which minimizes the Bayes risk. When the loss function is taken to be squared error (in the single parameter case) or the quadratic loss (in the multi-parameter case), the Bayes risk is more familiarly known as *mean squared error* (MSE).

For quadratic loss (and many others [40]) the unique Bayes estimator is the mean of the posterior distribution $\hat{\boldsymbol{x}}(D; C) = \mathbb{E}_{\boldsymbol{x}|D;C}[\boldsymbol{x}]$. Minimizing the Bayes risk of a choice of parameters is equivalent to maximizing the negative Bayes risk for that set; therefore, it is reasonable to choose the negative Bayes risk as our utility function. It also has theoretical benefits in that it is easy to compare the performance of algorithms that take $U(c_{N+1}) = -r(\pi; c_{N+1}, C)$.

The question of how well can we estimator $\boldsymbol{x}$ becomes the question of how low can we make the Bayes risk $r(\pi; C)$. We lower bound the achievable risk via the Bayesian variant of the Cramer-Rao bound [41]. Both require finding the Fisher information:

$$\boldsymbol{I}(\boldsymbol{x}; C) = \mathbb{E}_{D|\boldsymbol{x};C}\left[\boldsymbol{\nabla}_{\boldsymbol{x}} \log\left(\Pr(D|\boldsymbol{x}; C)\right) \cdot \boldsymbol{\nabla}_{\boldsymbol{x}}^{\mathrm{T}} \log\left(\Pr(D|\boldsymbol{x}; C)\right)\right].$$

The Fisher information does not depend at all on the prior distribution, and thus is calculated in the same way regardless of how many experiments have already been performed.

The standard Cramer-Rao bound is then given by $\mathrm{Cov}(\hat{\boldsymbol{x}}) \geq \boldsymbol{I}(\boldsymbol{x}; C)^{-1}$, where $\boldsymbol{X} \geq \boldsymbol{Y}$ means that $\boldsymbol{X} - \boldsymbol{Y}$ is positive semi-definite. If we choose the matrix $\boldsymbol{Q}$ associated with the quadratic loss to be $\boldsymbol{Q} = \mathbb{1}$, then $R(\boldsymbol{x}, \hat{\boldsymbol{x}}; C) = \mathrm{Tr}(\mathrm{Cov}(\hat{\boldsymbol{x}})) \geq \mathrm{Tr}(\boldsymbol{I}(\boldsymbol{x}; C)^{-1})$. Clearly, this statement of the multivariate Cramer-Rao bound assumes that $\boldsymbol{I}$ is non-singular. Singular Fisher information matrices arise when there are experiments that provide no information about *at least one* of the experimental parameters. Unfortunately, that assumption is not met in general. We avoid this problem by considering the Bayesian information matrix $\boldsymbol{J}(\pi; C) = \mathbb{E}_{\boldsymbol{x}}[\boldsymbol{I}(\boldsymbol{x}; C)]$. Then, the *Bayesian Cramer-Rao bound* (BCRB) is given by [41]

$$r(\pi; C) \geq \boldsymbol{J}(\pi; C)^{-1}.$$

Lower bounds can be found for specific values of C using numerical integration. In practice, we calculate the BCRB using an iterative method, similar to [42].

B Pseudo–Code for Algorithms

Algorithm 1 Sequential Monte Carlo update algorithm.

Input: Particle weights $w_i(D)$, $i \in \{1, \ldots, n\}$, Particle locations $\boldsymbol{x}_i$, $i \in \{1, \ldots, n\}$, New datum d_{j+1}, obtained from an experiment with control c_{j+1}.
Output: Updated weights $w_i(D \cup d_{j+1})$.

```
function UPDATE({w_i(D)}, {x_i}, d_{j+1}, c_{j+1})
    for i ∈ 1 → n do
        w̃_i ← w_i(D) Pr(d_{j+1} | x_i, c_{j+1})
    end for
    return {w̃_j / Σ_i w̃_i}        ▷ We must normalize the updated weights before returning.
end function
```

Algorithm 2 Sequential Monte Carlo resampling algorithm.

Input: Particle weights w_i, $i \in \{1, \ldots, n\}$, Particle locations $\boldsymbol{x}_i$, $i \in \{1, \ldots, n\}$, Resampling parameter $a \in [0, 1]$.
Output: Updated weights w_i' and locations $\boldsymbol{x}_i'$.

```
function RESAMPLE({w_i}, {x_i}, a)
    μ ← MEAN({w_i}, {x_i}), Σ ← h² COV({w_i}, {x_i})
    h ← √(1 − a²)
    for i ∈ 1 → n do
        draw j with probability w_j        ▷ Choose a particle j to perturb.
        μ_i ← a x_j + (1 − a) μ            ▷ Find the mean for the new particle location.
        draw x_i' from N(μ_i, Σ)           ▷ Draw a perturbed particle location.
        w_i' ← 1/n                         ▷ Reset the weights to uniform.
    end for
    return {w_i'}, {x_i'}
end function
```

Algorithm 3 Reduced particle approximation for Sequential Monte Carlo utility functions.

Input: Particle weights w_i, $i \in \{1, \ldots, n\}$, Particle locations $\boldsymbol{x}_i$, $i \in \{1, \ldots, n\}$, Ratio `approx_ratio` of the particles to keep in the reduced approximation.
Output: Reduced sets of particle weights $\{\tilde{w}_i\}$ and locations $\{\tilde{\boldsymbol{x}}_i\}$.

```
function REAPPROX({w_i}, {x_i}, approx_ratio)
    ñ ← ⌊n · approx_ratio⌋
    draw π uniformly at random from Sym(n), the symmetric group acting on n elements
    {w̃_i} ← {w_{π(i)}}        ▷ Permute the elements to avoid patterns when sorting the weights.
    {x̃_i} ← {x_{π(i)}}
    {s_k} ← SORT({w̃_i})        ▷ Get a list of indices s_i such that w̃_{s_i} ≥ w̃_{s_j} for all i, j.
    return {w̃_i} ← {w̃_{s_i} : i ∈ 1 → ñ}, {x̃_i} ← {x̃_{s_i} : i ∈ 1 → ñ}
end function
```

Algorithm 4 Complete adaptive Bayesian experiment design algorithm, using sequential Monte Carlo approximations.

Input: A number of particles n to be used, A prior distribution π over models, A number of experiments N to perform, A resampling parameter $a \in [0,1]$, A threshold `resample_threshold` $\in [0,1]$ specifying how often to resample, An approximation ratio `approx_ratio`, An local optimization algorithm LocalOptimize, A heuristic GuessExperiment for choosing experiment controls, and a number n_{guesses} of potential experiments to consider in each iteration.

Output: An estimate $\hat{\boldsymbol{x}}$ of the true model $\boldsymbol{x}_0$.

function EstimateAdaptive(n, π, N, a, `resample_threshold`, `approx_ratio`, Optimize, n_{guesses}, GuessExperiment)

$w_i \leftarrow 1/n$ ▷ Start by initializing the SMC variables.
draw each $\boldsymbol{x}_i$ independently from π

for $i_{\text{exp}} \in 1 \to N$ **do** ▷ We now iterate through each experiment.
▷ If we are using a reduced particle set, populate that first.
if `approx_ratio` $\neq 1$ **then**
$\{\tilde{w}_i\}, \{\tilde{\boldsymbol{x}}_i\} \leftarrow$ Reapprox($\{w_i\}$, $\{\boldsymbol{x}_i\}$, `approx_ratio`)
else
$\{\tilde{w}_i\}, \{\tilde{\boldsymbol{x}}_i\} \leftarrow \{w_i\}, \{\boldsymbol{x}_i\}$
end if

▷ Heuristicly choose potential experiments, and optimize each independently.
for $i_{\text{guess}} \in 1 \to n_{\text{guesses}}$ **do**
$c_{i_{\text{guess}}} \leftarrow$ GuessExperiment(i_{exp})
$\hat{c}_{i_{\text{guess}}}, U_{i_{\text{guess}}} \leftarrow$ LocalOptimize(Utility, $c_{i_{\text{guess}}}$, $\{\tilde{w}_i\}$, $\{\tilde{\boldsymbol{x}}_i\}$)
end for

$i_{\text{best}} \leftarrow \text{argmax}_{i_{\text{guess}}} U_{i_{\text{guess}}}$ ▷ Pick the controls that maximize the optimized utility.
$\hat{c} \leftarrow \hat{c}_{i_{\text{best}}}$
$d_{i_{\text{exp}}} \leftarrow$ the result of performing $\hat{C}$ ▷ Perform the best experiment.
$\{w_i\}, \{\boldsymbol{x}_i\} \leftarrow$ Update($\{w_i\}$, $\{\boldsymbol{x}_i\}$, D, C) ▷ Find the new posterior distribution.

if $\sum_i w_i^2 < N \cdot$ `resample_threshold` **then** ▷ Resample if n_{ess} is too small.
$\{w_i\}, \{\boldsymbol{x}_i\} \leftarrow$ Resample($\{w_i\}$, $\{\boldsymbol{x}_i\}$, a)
end if
end for

▷ After all experiments have been performed, return the mean as an estimate.
return $\hat{\boldsymbol{x}} \leftarrow$Mean($\{w_i\}$, $\{\boldsymbol{x}_i\}$)

end function

Classical and Quantum Algorithms for Testing Equivalence of Group Extensions*

Kevin C. Zatloukal

Univerisity of Washington
Seattle, WA 98195
kevinz@cs.washington.edu

Abstract

While efficient algorithms are known for solving many important problems related to groups, no efficient algorithm is known for determining whether two arbitrary groups are isomorphic. The particular case of 2-nilpotent groups, a special type of central extension, is widely believed to contain the essential hard cases. However, looking specifically at central extensions, the natural formulation of being "the same" is not isomorphism but rather "equivalence," which requires an isomorphism to preserves the structure of the extension. In this paper, we show that equivalence of central extensions can be computed efficiently on a classical computer when the groups are small enough to be given by their multiplication tables. However, in the model of black box groups, which allows the groups to be much larger, we show that equivalence can be computed efficiently on a quantum computer but not a classical one (under common complexity assumptions). Our quantum algorithm demonstrates a new application of the hidden subgroup problem for general abelian groups.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases quantum computing, algorithms, computational group theory

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.126

1 Introduction

Finding an efficient algorithm for group isomorphism is one of the most notable open problems in computational group theory. While the problem is easily solved for abelian groups, the problem remains unsolved even for some very simple generalizations to non-abelian groups. In particular, the 2-nilpotent groups, which are central extensions of an abelian group by another abelian group, are widely believed to contain the essential hard cases (see e.g. [1]). Hence, the computational issues surrounding this type of group extension merit further study.

While isomorphism is the natural notion of what it means to be the same group, the natural notion of being the same extension is slightly different. Indeed, the theory of group extensions[1], whose study began near the start of the 20th century, defines two extensions to be the same or "equivalent" if there exists an isomorphism that preserves the structure of the extension. (We will define this precisely in the next section.)

Thus, it is interesting to consider whether there exists an efficient algorithm for testing equivalence of those extensions for which isomorphism appears difficult. In this paper, we will see that there is indeed an efficient algorithm.

* This work was partially supported by NSF grants CCF-0916400 and CCF-1111382.
[1] See the chapter in [2] for a nice introduction to the theory of group extensions.

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 126–145


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Group isomorphism has drawn particular interest from the quantum computing community due to its placement in the hierarchy of complexity classes. In particular, due to the work of [3], we know that the isomorphism problem for solvable groups is almost in the class NP ∩ coNP. This is the class that includes factoring and other problems for which quantum computers appear to give super-polynomial speedups. Hence, there is strong interest in determining whether the same is true of solvable group isomorphism. To date, however, no such quantum speedup is known even for the smaller class of 2-nilpotent groups.

Given the relationship between the conjectured hard cases of group isomorphism (2-nilpotent groups) and the problem of extension equivalence, it is natural to wonder whether the latter problem also could lead to a super-polynomial speedup of quantum algorithms over classical ones. As noted above, there is an efficient classical algorithm for testing equivalence. However, its efficiency depends on the fact that the given groups are small, in particular, small enough to write down their complete multiplication tables.

The usual setting for the group isomorphism problem has the input groups given by their multiplication tables. If one cannot solve the problem in this model, then other models are out of the question. However, it would be both interesting and useful to be able to test equivalence of larger groups, for which this model is inappropriate. In particular, for groups of matrices over finite fields (which includes, for example, simple groups of Lie type), individual matrices are small enough to multiply and invert efficiently, but writing out a multiplication table between all matrices in the group would often be infeasible. Yet, computational group theorists would still like to answer questions about such groups.

Matrix groups are often studied in the "black box group" model. (Indeed, this was the original motivation for the model.) Hence, it is natural for us to consider whether there exists an efficient algorithms for testing equivalence of group extensions in this model.

One case we will consider is extending a group given by a multiplication table by a black box group. In practical terms, this means extensions of a small group by a large one. Such extensions can already introduce substantial complexity. For example, the dihedral group D_{2N} is an extension of the tiny group $\mathbb{Z}_2$ by a potentially large cyclic group $\mathbb{Z}_N$. Considering that the hidden subgroup problem can be solved in quantum polynomial time for $\mathbb{Z}_N$ but not (currently) for D_{2N}, we can see that extensions of even constant-sized groups can introduce substantial computational difficulty.

In this paper, we show that there is an efficient quantum algorithm for testing equivalence of extensions of a small group by large abelian group or extensions of one large abelian group by another large abelian group. Furthermore, we will show that the existence of an efficient classical algorithm for either of these cases would break an existing cryptosystem.[2] Hence, under the hardness assumption of that cryptosystem, no efficient classical algorithm exists.

The quantum algorithm we present depends crucially on the ability to solve the hidden subgroup problem (HSP) for arbitrary abelian groups. (This is the essential quantum subroutine in our algorithm.) Interestingly, while some other problems in computational group theory that can be solved efficiently on a quantum computer can also be solved classically assuming the existence of oracles for factoring and/or discrete logarithm, our construction does not easily translate to that setting because there is no apparent way to solve abelian HSP classically, even with the help of such oracles. Hence, our work demonstrates a new and interesting application of efficient quantum algorithms for abelian HSP.

[2] Note that this cryptosystem depends on the hardness of factoring, so it is already known that quantum computers could break it. What was not known is the relationship of this to testing equivalence of extensions.

Related Work While we are aware of no prior work on the complexity of determining extension equivalence in these models, our motivation for this problem comes from the status of the group isomorphism problem for simple group extensions, and there, it is known that isomorphism can be determined efficiently on a quantum computer in certain special cases [4]. Interestingly, the groups to which this result applies have trivial equivalence classes[3], so the extension equivalence problem is trivial for such groups. (The answer is always "yes".) The fact that the one class of nonabelian solvable groups for which we have made progress on group isomorphism is one for which equivalence is trivial suggests that studying the extension equivalence problem may teach us something about the hard cases of group isomorphism.

2 Background

2.1 Computational Group Theory

The study of algorithms and complexity for problems in group theory is called *computational group theory*. In order to discuss these issues, we must first specify how the group will be given as input. Multiple approaches have been defined (see [5] for a nice review). We will need to use three of these in our later discussion.

The first approach is to describe a group G by its multiplication table (sometimes called the "Cayley table"). Multiplication of group elements can be performed by table lookup, inverses can be computed by scanning one row of the table, and so on. This is perhaps the most natural model. However, in order to use this approach, the group must be small enough that it is reasonable to write down a $|G| \times |G|$ table. This turns out to be too limiting for many computations that practitioners want to perform.

Another approach is the "black box group" model of Babai and Szemerédi [6]. In this model, group elements are identified by opaque strings (which need not be unique) and an oracle is provided that can perform the following group operations:

1. Given $g, h \in G$, compute gh.
2. Given $g \in G$, compute g^{-1}.
3. Given $g \in G$, determine whether $g = e$, the group identity [4].

Finally, we have to specify how the algorithm obtains the strings for some group elements in the first place. It is usual to assume that the input to algorithm will be a list of generators of the group (i.e., a list of strings identifying the generators).

While the black box model is restricted in terms of how it can work with the group, it is even more restricted in terms of what is considered efficient. Since a multiplication table has size $\tilde{O}(|G|^2)$,[5] any running time of $\mathrm{poly}(|G|)$ is efficient in the first model. On the other hand, a non-redundant[6] list of generators only has length $O(\log |G|)$,[7] so the input has size $O(\log^2 |G|)$. Hence, an algorithm is efficient in the second model only if it has running time $\mathrm{poly}(\log |G|)$, which is exponentially faster.

---

[3] This follows from the fact that the second cohomology groups (defined below) are trivial for semi-direct products.

[4] This also allows us to determine whether $g = h$ since this is equivalent to checking $gh^{-1} = e$.

[5] As is usual, $\tilde{O}(.)$ is the same as $O(.)$ but with suppressed terms that are logarithmically smaller than those included.

[6] This simply means that no proper subset of the generators still generates the group.

[7] This follows from the fact that each additional generator increase the size of the generated group by a factor equal to the index of the old group in the new one, and this index (an integer), since it is not 1, must be at least 2.

It should not be surprising then to find a large difference between which problems can be solved in the two models. In the first model, almost every natural group problem can be solved efficiently, the notable exception being the group isomorphism problem. In the second model, on the other hand, very few problems can be solved, at least classically. The main example of a problem that can be solved in this model is computing a derived series for a solvable group (that is, generators for each group in the series) or a central series for a nilpotent group.

Interestingly, it is known that quantum algorithms can do more in the black box model. In particular, for abelian or even solvable groups [7], a large number of problems can be solved, the most important example being computing the size of the group, $|G|$. We will show later on that the extension equivalence problem is another example.

The other approaches for specifying groups use representations of particular types. The most common of these, the third model we will need below, is to use a permutation representation. Specifically, we assume that the group is explicitly a subgroup of the symmetric group, $G \leq S_n$. The input is a set of generators of G, each of which is a permutation of the set $[n] \triangleq \{1, \ldots, n\}$.

As in the black box model, G can be specified by at most $O(\log |G|)$ generators. Each generator in the input has size $O(n \log n)$, so the input as a whole will have size $O(n \log n \log |G|)$. For an algorithm to be efficient then, its running time must be polynomial both in n and $\log |G|$. Furthermore, for this model to be useful, the size n of the set, called the "degree" of the representation, must be small. The fact that many groups have small-degree representations is one factor leading to the great success of this third approach. The other factor leading to its success is that many problems can be solved efficiently in this model. In fact, nearly all of the problems that are solvable with multiplication tables are efficiently solvable here as well. (See [5] for a long list of these problems.)

2.2 Group Extensions

A group E is said to be an *extension* of G by A if $A \lhd E$ and $E/A \cong G$. This is called a *central extension* if $A \leq Z(E)$. In particular, this means that A is abelian.

Central extensions are in some ways similar to semidirect products in that the elements can be thought of as pairs $(a, x) \in A \times G$ with a strange multiplication. Whereas multiplication in a semidirect product depends on a group homomorphism $G \to \mathrm{Aut}\, A$, multiplication in a central extension depends on a function $f : G \times G \to A$, where we have $(a, x)(b, y) = (abf(x, y), xy)$. The function f is called a "factor set." We will describe some of its properties below. In particular, we will show how to find f for a given extension E.

Central extensions are in some sense the other natural way to combine groups, aside from semidirect products. In particular, any group extension of G by A, where A is abelian but not necessarily central, is essentially a combination of a semidirect product and a central extension.[8] Hence, these two types represent the two extremes of extensions of abelian groups.

Finally, we can define the problem we are trying to solve. Two extensions, E_1 and E_2, of G by A are said to be *equivalent* if there exists an isomorphism $\gamma : E_1 \to E_2$ such that γ is the identity on A, $\gamma|_A = \mathrm{id}$, and gives rise to the identity on G, that is, $\pi_2 \circ \gamma = \pi_1$, where $\pi_i : E_i \to G$ is the canonical projection. This is the natural sense in which two extensions should be considered "the same".

[8] Any extension is identified, up to isomorphism, by a homomorphism from G to $\mathrm{Aut}\, A$ (the semi-direct product part) and a factor set (the central extension part). See [2] for details.

1-cochains	$C^1(G,A) = \{s : G \to A \mid s(e) = e\}$
2-cochains	$C^2(G,A) = \{f : G \times G \to A \mid f \text{ normalized}\}$
cocycles	$Z^2(G,A) = \{f : G \times G \to A \mid f \text{ normalized, cocycle condition}\} \subset C^2(G,A)$
$\partial : C^1 \to C^2$	homomorphism taking $s \in C^1(G,A)$ to $\partial s \in Z^2(G,A)$
coboundaries	$B^2(G,A) = \operatorname{Im} \partial \subset Z^2(G,A)$

Figure 1 The main objects in group cohomology.

On the other hand, it is possible for E_1 and E_2 to be isomorphic even if they are not equivalent extensions. (Indeed, this is not even a simple matter of dealing with isomorphisms of A and G: it is apparently possible for extensions of non-isomorphic groups to be isomorphic.) For this reason, equivalence is a more natural question to consider when looking specifically at group extensions: an equivalence is an isomorphism that respects the structure of the group extension.

2.3 Low Degree Group Cohomology

Cohomology groups are often defined in an abstract manner (via Ext functors, projective resolutions, etc.). However, in the case of group cohomology, the low degree cohomology groups also have concrete definitions that are equivalent but more useful for us.[9] (See [2] for a more detailed discussion.)

In this section, we will consider cohomology only of central extensions. Cohomology can be defined more generally, but this simpler case is all that we will need in later sections.

The key group for us is the second cohomology group, $H^2(G,A)$. In order to define this, however, we first need to define cocycles and coboundaries.

The 1-cocycles, $Z^1(G,A)$, are functions $f : G \to A$ that satisfy the identity $f(x) + f(y) - f(xy) = 0$, for all $x, y \in G$. These are simply group homomorphisms. (Note that we are using additive notation since A is abelian.) The 2-cocycles, $Z^2(G,A)$, are functions $f : G \times G \to A$ that satisfy the (admittedly odd-looking) identity $f(y,z) - f(xy,z) + f(x,yz) - f(x,y) = 0$, for all $x, y, z \in G$, and have $f(x,e) = f(y,e) = e$, for all $x, y \in G$.[10] These are precisely the factor sets mentioned earlier.

The 2-coboundaries, $B^2(G,A)$, are functions $G \times G \to A$ that arise by taking a function $s : G \to A$ only satisfying $s(e) = e$ (called a 1-cochain) by defining $\partial s \in B^2(G,A)$ by $\partial s(x,y) = s(x) + s(y) - s(xy)$. Note that, since s is not necessarily a homomorphism, we need not have $\partial s \not\equiv 0$. It is not hard to show that any function defined in such manner is also a 2-cocycle. In other words, we have $B^2(G,A) \leq Z^2(G,A)$. Furthermore, the function ∂ is in fact a (surjective) homomorphism $C^1(G,A) \to B^2(G,A)$, where $C^1(G,A)$ denotes the space of all cochains.

These definitions are summarized in Figure 1.

The sets $Z^2(G,A)$ and $B^2(G,A)$ are themselves groups with the group operation performed pointwise (i.e., $(f+g)(x,y) = f(x,y) + g(x,y)$). In fact, they are abelian groups since A is abelian. Hence, $B^2(G,A)$ is a normal subgroup of $Z^2(G,A)$, so we can consider the quotient group $H^2(G,A) \triangleq Z^2(G,A)/B^2(G,A)$. This is the *second cohomology group*.

[9] Historically, these were developed in the opposite order. The concrete definitions came first and the abstract later.

[10] Sometimes cocycles are defined only by the first condition. Then those that satisfy the second are called "normalized". We will assume throughout this paper that all cocycles, coboundaries, and cochains are properly normalized.

The most important fact for us is the relationship between $H^2(G, A)$ and group extensions.

▶ **Lemma 1.** *Elements of $H^2(G, A)$ are in 1-to-1 correspondence with equivalence classes of central extensions of G by A.*

Proof Sketch. While we need not go through this proof in detail (see [2] for full details), we do need describe how the correspondence works since our aim is to work in the group $H^2(G, A)$, using the elements corresponding to the two given extensions.

For an extension E of G by A, choose a representative of each coset of A in E (i.e., each element of $G \cong E/A$), where we require e to represent A itself. Encode these choices into a function $s : G \to E$. Then we can define a function $f : G \times G \to A$ by $f(x, y) \triangleq s(x)s(y)s(xy)^{-1}$. It is not hard to show that $f(x, y) \in A$ and that f is a factor set, i.e., $f \in Z^2(G, A)$.

This construction depends on the choice of representatives. Choosing a different set of representatives, we could get a different factor set $g : G \times G \to A$. However, if we do this, it will turn out $f - g$ is a 2-coboundary. Furthermore, the only other factor sets differing from f by a coboundary arise from other choices of representatives for the same extension. Hence, $f + B^2(G, A)$ uniquely represents this extension. ◀

3 Results

3.1 General Approach

With this background, the basic idea for computing equivalence of central extensions is simple. Given E_1 and E_2, two central extensions of G by A, we can compute the factor sets $f_1, f_2 \in Z^2(G, A)$ for these two extensions using any set of representatives. As described in Lemma 1, the factor sets correspond to the same extension iff $f_1 - f_2 \in B^2(G, A)$. Thus, the general approach is to reduce extension equivalence to testing membership in $B^2(G, A)$.

To make this concrete, we must specify what approach we use for representing groups. Below, we present two algorithms, one classical and one quantum, for implementing the outline just described. These algorithms differ in the approach used to specify the input groups, with the quantum algorithm using the more general approach of black box groups for A and E. Specifically, we have the following results.

▶ **Theorem 2.** *There exists a (classical) Monte Carlo algorithm for testing the equivalence of E_1 and E_2, two extensions of G by A, when all groups specified by multiplication tables, running in time $\tilde{O}(|G|^6\,|A|^3)$.*

▶ **Theorem 3.** *There exists a quantum algorithm for testing the equivalence of E_1 and E_2, two extensions of G by A, where A, E_1, and E_2 are given as black box groups and G is given by a multiplication table, running in time $O(|G|^6 \log^6 |A|)$.*

▶ **Theorem 4.** *There exists a quantum algorithm for testing the equivalence of E_1 and E_2, two extensions of G by A, where G is* abelian *and all groups are presented as black box groups running in time* $\operatorname{poly}\log|G| \operatorname{poly}\log|A|$.

For simplicity, we first prove the first two theorems, in subsections 3.2 and 3.3, respectively, assuming that E_1 and E_2 are *central* extensions. We discuss how to extend these two algorithms to non-central extensions in subsection 3.4. Theorem 4 is more complex and is treated in the appendix, in section A

In subsection 3.5, we show that the problem solved by the quantum algorithms are classically hard under the assumption of the Goldwasser–Micali cryptosystem [8] (that quadratic residuosity is classically hard).

▶ **Theorem 5.** *There exists a randomized polynomial time reduction from quadratic residuosity to testing equivalence of central extensions of G by A, where A is given as a black box group and either G is given as a multiplication table or G is abelian and given as a black box group. Hence, under the assumption that there is no efficient (classical) Monte Carlo algorithm for testing quadratic residuosity, there is no efficient Monte Carlo algorithm for testing equivalence of extensions of G by A in this model.*

Finally, in subsection 4, we use the machinery developed for these algorithms to show that we can also efficiently count the number of inequivalent extensions in the two models. Specifically, we have the following:

▶ **Theorem 6.** *There exists an efficient (classical) Monte Carlo algorithm for counting the number of equivalence classes of extensions of G by A when both groups are given by multiplication tables.*

▶ **Theorem 7.** *There exists an efficient quantum algorithm for counting the number of equivalence classes of extensions of G by A when A is given as a black box group and G is given by a multiplication table.*

3.2 Classical Algorithm

For the classical algorithm, we take the inputs A, G, and E_1 and E_2 as multiplication tables. This is the usual setup for the group isomorphism problem, and it is natural to consider extension equivalence in the same manner. However, we must also require that the isomorphism $E_i/A \cong G$ be provided explicitly so that we are not required to solve a group isomorphism problem in order to understand the relationship between E_i and G. This will be specified as a table of pairs (x, g), where each $x \in E_i$ appears exactly once along with the $g \in G$ such that $x + A \xrightarrow{\sim} g$.

Proof of Theorem 2. As described above, we will reduce to membership testing in $B^2(G, A)$. Since the group $B^2(G, A)$ has size $\sim |A|^{|G|}$, we cannot reduce to a membership test using a multiplication table because the time to write such a table is exponentially large in the input size. We also cannot reduce to a membership test using a black box model simply because there is no efficient classical algorithm known for membership testing in this model. Fortunately, we will see that we can reduce to a membership test using the third approach, a permutation representation. We can then perform the membership testing efficiently using the algorithm from [9].

First, note that we can represent A using the regular representation, that is, each $a \in A$ is represented as a permutation $\sigma(a)$ of the set A itself. The degree of this representation is $|A|$, which is small. And it is easy to see that this representation of A is faithful. (This is Cayley's theorem.)

Define $C^2(G, A)$ to be all maps $G \times G \to A$. These are simply vectors of $|G|^2$ elements of A. (Since $B^2(G, A) \leq Z^2(G, A) \leq C^2(G, A)$, we can think of elements of $B^2(G, A)$ and $Z^2(G, A)$ in the same way.) Put another way, $C^2(G, A)$ is a direct sum of $|G|^2$ copies of A. Hence, we can represent $f \in C^2(G, A)$ as the direct sum (as vector spaces) of $\sigma(f(g, h))$ for each $g, h \in G$. It is again clear that this representation is faithful: $\sigma(f)$ is the identity iff $\sigma(f(g, h))$ is identity for each $g, h \in G$ iff $f(g, h) = e$ for each $g, h \in G$ (since our representation of A is faithful) iff f is the identity in $C^2(G, A)$ (by definition).

In other words, our representation space is the set $\{a_{g,h} \mid a \in A,\ g, h \in G\}$ — elements of A labelled by pairs $(g, h) \in G \times G$. We can see that the degree of this representation is $n \triangleq |A|\,|G|^2$.

It is possible that A may have a permutation representation with smaller degree in special cases, but in the worst case, it must be $|A|$. In particular, any simple cyclic group requires this degree. It is also easy to see that any faithful representation of $C^2(G,A)$ must contain all $|G|^2$ copies of this representation. Hence, our degree of $|A|\,|G|^2$ cannot in general be improved.

In order to invoke a membership test for $B^2(G,A)$, we also need to provide a generating set. The easiest way to do this is to take a generating set for $C^1(G,A)$ and then push it forward to $B^2(G,A)$ by applying ∂. Any $f \in B^2(G,A)$ satisfies $f = \partial s$ for some $s \in C^1(G,A)$. So if $s_1, \ldots, s_k$ is a generating set for $C^1(G,A)$, then we have $s = s_1^{j_1} \ldots s_k^{j_k}$ for some $\{j_i\} \subset \mathbb{Z}_+$. And since ∂ is a homomorphism, we have $f = \partial(s_1^{j_1} \ldots s_k^{j_k}) = \partial(s_1)^{j_1} \ldots \partial(s_k)^{j_k}$. Thus, $\partial s_1, \ldots, \partial s_k$ is a generating set for $B^2(G,A)$.[11]

It is easy to find a minimal generating set for $C^1(G,A)$. Since this group is simply a direct sum of $|G|$ copies of A, a minimal generating set for $C^1(G,A)$ is given by $|G|$ copies of a minimal generating set for A. We can find a generating set for A with high probability simply by choosing $O(\log|A|)$ random elements [5]. And it is easy to see that we can choose random elements from A since we have an explicit list of its elements. Hence, we can construct a generating set for $C^1(G,A)$ of size $O(|G|\log|A|)$.

Finally, note that, since we have a simple formula for ∂, taking constant time to evaluate for each $(g,h) \in G \times G$, we can construct the generating set for $B^2(G,A)$ in $O(|G|^2)$ time for each element in the set. Since this set contains $O(|G|\log|A|)$ elements, we can construct the generating set in $O(|G|^3 \log|A|)$ time.

The other input to the membership test is the element $f_1 - f_2 \in Z^2(G,A)$. We can compute this easily in linear time once we construct a factor set f_i for each extension. To do this, we simply need to choose (arbitrarily) a representative $s_i(g) \in E_i$ for each $g \in G$, which we can do in one pass over the table providing the isomorphism $E_i/A \cong G$. (Also note that we must choose $e \in E$ to represent $e \in G$.) This takes $O(|E|) = O(|A|\,|G|)$ time. Next, we compute f_i for each $g,h \in G$ by $f_i(g,y) = s(g) + s(h) - s(gh)$. Finally, we subtract them pointwise to compute $f_1 - f_2$. All of the above can be done in $O(|A|\,|G| + |G|^2)$ time.

It remains to invoke a membership test for a permutation group. The fastest algorithms [5] apply to so-called "small-base groups", but unfortunately, this representation is not one.[12] For the general case, the fastest known algorithm is from [9] and runs in time $\tilde{O}(n^3)$.

All of the membership test algorithms for permutation groups work by first computing what is called a strong generating set. As noted in [9], Gaussian elimination is a special case of this construction, so the running time of $\tilde{O}(n^3)$ is in fact optimal for all algorithms that work in this manner.

We note that the time to run this membership test dominates the time required to prepare its inputs, so the overall running time will be $\tilde{O}(n^3) = \tilde{O}(|A|^3\,|G|^6)$. ◀

3.3 Quantum Algorithms

For classical algorithms, we excluded the possibility of using a membership test for black box groups because no efficient algorithm is known to exist. However, in the quantum case, we have such an algorithm [10]. As a result, it is natural to consider whether extension equivalence can also be solved in the black box model.

[11] Since ∂ is not an isomorphism, this generating set may be redundant. However, since its kernel is very small compared to $\left|C^1(G,A)\right|$, this increases the size of the generating set by a $1 - o(1)$ factor.

[12] The group $B^2(G,A)$ would be small-base if $\log\left|B^2(G,A)\right| = O(\text{poly}\log n) = O(\text{poly}(\log|G| + \log|A|))$, but we can see that $B^2(G,A)$ is much bigger than this.

Our quantum algorithm will take the inputs A and E as black box groups. That is, we are given a generating set for each and an oracle for performing the three operations listed earlier in the group E.[13]

For the group G, on the other hand, we first consider the case when G is given by a multiplication table. In this case, we can efficiently work with the group $B^2(G, A)$ since it has a generating set of size $O(|G|^2 \log |A|)$ and we only need a running time polynomial in $|G|$ in this model. Practically speaking, this means that we will be able to compute equivalence of extensions of a small group G by a large group A using this algorithm. Such extensions can still be quite complicated groups.

Finally, the isomorphism $E_i/A \cong G$ will be provided as an oracle since we cannot reasonably take a table with $|E|$ rows as input. Given an element $x \in E_i$, the oracle return the $g \in G$ corresponds to $x + A \in E_i/A$.

Proof of Theorem 3. As in the classical algorithm, we will apply the correspondence in Lemma 1 and reduce to a membership test in $B^2(G, A)$.

In order to use a membership test for $B^2(G, A)$, we must show how to construct an oracle for this group or a larger group containing it. We will work with $C^2(G, A)$. Since each element of $C^2(G, A)$ is a vector (or direct sum) of $|G|^2$ elements of A, we can identify elements of this group by strings containing $|G|^2$ strings for elements of A. We can perform multiplication and inverses pointwise, each using $|G|^2$ calls to the oracle for A. Similarly, the identity in $C^2(G, A)$ is simply $|G|^2$ copies of the identity in A, so we can also check for the identity with $|G|^2$ calls to the oracle for A.

One input to the membership test is a generating set for $B^2(G, A)$. We saw in the previous section that this can be constructed simply by making $|G|^2$ labelled copies of a generating set for A. In this case, we are given a generating set for A as input, and we can turn this into $|G|^2$ labelled copies in $O(|G|^2 \log |A|)$ time.[14]

The other input to the membership test is the element $f_1 - f_2$. As before, in order to compute these factor sets, we need to be able to choose a representative of each coset of A in E. However, note that our classical algorithm ran in $O(|E|)$ time, which is no longer efficient in this model. So we will need a slightly different approach.

Instead of enumerating E, we will select random elements from E and invoke the oracle we are given to find the projection in G. If $x \in E$ projects onto $g \in G$, then this gives us our representative $s(g) = x$ for g. We continue to select random elements until we have a representative for each $g \in G$ (aside from $e \in G$, which we set to $s(e) = e$).

Now, since we are only given a generating set for E, it is not possible to select uniformly random elements. However, we can compute nearly uniformly random elements as described in [11] in time linear in the size of the generating set for E (plus an $\tilde{O}(\log^5 |A|)$ additive term). The generated elements are nearly uniform in the sense that the probability of generating $x \in E$ is off by a $1 - o(1)$ factor, which we can choose to be arbitrarily small.

With this, the probability of producing any particular $g \in G$ will be $(1 \pm \epsilon)/|G|$. Hence, by standard calculations, we will produce a representative for each $g \in G$ with high probability after $O(|G| \log |G|)$ random choices. The overall time to compute these representatives if $\tilde{O}(|G| \log |A| + \log^5 |A|)$.

[13] This also works for A since $A \leq E$.

[14] This is assuming that we are given a generating set for A of size $O(\log |A|)$. We can easily reduce to a generating set of this size, if this is not what we are given, by using random subproducts as described in [5].

With choices of representatives s_i for each E_i, we can compute the factor sets f_i and their difference $f_1 - f_2$ in the same manner as in the classical algorithm. This takes time $O(|G|^2)$.

To perform the membership test, we apply the algorithm from [10], which can be used to compute the size of a subgroup. We call this once with the generating set for $B^2(G, A)$ and once with this generating set plus $f_1 - f_2$. If the latter subgroup is larger, then $f_1 - f_2 \notin B^2(G, A)$, and the extensions are not equivalent. Otherwise, they are equivalent.

As described in [12], the running time of the algorithm for computing group size depends on the size of the generating set, k, and the maximum order of any element in the group, q. As mentioned above, we have $k = O(|G|^2 \log |A|)$ for the first. For the second, the best bound we have in general is $q = |A|$.

The algorithm first performs $O(k \log q)$ group operations. Each of these translates into $|G|^2$ calls to the oracle for A. Thus, all together, it will perform $O(|G|^4 \log^2 |A|)$ calls to the oracle for A. The algorithm also performs $O(k^3 \log^2 q) = O(|G|^6 \log^5 |A|)$ other elementary operations as part of its post-processing, which dominates the running time.

There are a few other details about the running time of this algorithm that need to be considered. However, to keep this presentation simpler, we discuss those in the appendix, in section B. Here, it suffices here to say that the other necessary processing adds at most a $\log |A|$ factor to the running time, giving us a running time of $O(|G|^6 \log^6 |A|)$. ◀

As in the classical case, it turns out that the quantum algorithm needs to perform something like Gaussian elimination on a matrix.[15] This occurs within the post-processing steps of the algorithm for computing the size of the subgroup. The matrix in question has rows and columns indexed by generators, and since we have $O(|G|^2 \log |A|)$ generators, we get an $O(|G|^6)$ factor in the running time of the algorithm.

The dependence on $|A|$, on the other hand, is exponentially improved compared to the classical algorithm. Hence, if the group G is fairly small (i.e., $|G| = O(\log |A|)$) then the quantum algorithm is exponentially faster overall. As we will see in the next section, extensions of small groups (even constant sized) are complicated and interesting objects.

For the case where G is also presented as a black box group, the above approach does not work since we cannot efficiently write down a generating set for $B^2(G, A)$ or even a factor set $f \in Z^2(G, A)$. However, it is still possible to test equivalence provided that G is abelian. As this requires substantially more work, which is specific to this special case, we leave the proof of Theorem 4 to the appendix, in section A.

3.4 Algorithms for Non-Central Extensions

It is not hard to extend our algorithms to general extensions, i.e., without the assumption that A is central in E_1 and E_2.

The core fact needed by both algorithms is the correspondence between equivalence classes of extensions and elements of $H^2(G, A)$ given in Lemma 1. This relationship indeed holds for general extensions (i.e., under the assumption that A is abelian but not necessarily central). However, in the general setting, the definition of $H^2(G, A)$ is more complex.

If E is an extension of G by A and $t \in E$ is a representative of $g \in G$, then it does not hold that $t^{-1}at = a$ for all $a \in A$ if A is not central. It is easy to check that $t^{-1}at \in A$, however, and that any two representatives of $g \in G$ define the same action $a \mapsto a^t \triangleq t^{-1}at$. In fact, this defines a homomorphism $\varphi : G \to \operatorname{Aut} A$, as occurs in a semi-direct product.

[15] Specifically, computing the Smith normal form of a matrix. See [10] for details.

In the general case, extensions are identified not only by the groups G and A but also by $\varphi : G \to \operatorname{Aut} A$. Two extensions of G by A with action φ are equivalent if there exists a structure preserving isomorphism, as before. Lemma 1 then holds using a definition of $H^2(G, A)$ that changes the formula for ∂ to include φ.

In our algorithms, the only change is that we must use the new formula when constructing a generating set for $B^2(G, A)$. This new formula is $(\partial f)(x, y) \triangleq f(x)^y + f(y) - f(xy)$, where the action a^y of G on A is given by φ. Since this action is just conjugation by a representative and we have a representative for each $y \in G$, it is clear that we can compute this formula just as well. Hence, we can efficiently test equivalence of non-central group extensions of G by A, in both models, with the same running times.

3.5 Impossibility for Classical Algorithms in the Black Box Model

In this subsection, we show that the problem solved by our quantum algorithm is classically hard under the assumption of the Goldwasser–Micali cryptosystem that quadratic residuosity is classically hard. Our proof is a reduction from quadratic residuosity to testing equivalence of *central* extensions. Hence, this argues that the problem for black box groups is hard even for the simpler case of central extensions.

Proof of Theorem 5. The inputs to quadratic residuosity are a large number N and a $y \in \mathbb{Z}_N^*$, the group of multiplicative units modulo N. (We are also assured that the Jacobi symbol of y is $+1$, though that will play no part in the construction.) Both of these inputs are encoded in $O(\log N)$ bits, so an algorithm is only efficient if it runs in $O(\operatorname{poly} \log N)$ time.

The objective for this problem is to determine whether y has a square root in $\mathbb{Z}_N^*$, that is, whether there exists an $x \in \mathbb{Z}_N^*$ such that $y = x^2 \pmod N$. If such an x exists, y is called a "quadratic residue". Our reduction will construct two central extensions of $\mathbb{Z}_2$ by $\mathbb{Z}_N^*$ that are equivalent iff y is a quadratic residue. Since $\mathbb{Z}_2$ is both small and abelian, this is a special case of *both models* we considered for quantum algorithms. Hence, this one reduction will show that both problems are as hard as quadratic residuosity.

As mentioned above, we can create a group extension from any factor set $f : \mathbb{Z}_2 \times \mathbb{Z}_2 \to \mathbb{Z}_N^*$. If we know the values of this function, then we can perform multiplication by $(x, a)(y, b) = (xyf(a, b), a+b)$.[16] It is well-known that we can perform group operations in $\mathbb{Z}_N^*$ in $O(\operatorname{poly} \log N)$ time, and group operations in $\mathbb{Z}_2$ take constant time, so this computation can be performed efficiently. Likewise, the inverse of (x, a), given by $(x^{-1} f(a, -a)^{-1}, -a)$, can also be computed efficiently. Finally, we can easily check for the identity element, which is $(1, 0)$. This shows that we can efficiently provide an oracle for these extensions, once we have chosen their factor sets.

Each factor set provides only four outputs since $|\mathbb{Z}_2 \times \mathbb{Z}_2| = 4$. Furthermore, as noted in the definition, any factor set must also satisfy $f(a, e) = f(e, b) = e$ for all $a, b \in G$. In this case, that means that $f(0, 0) = f(0, 1) = f(1, 0) = 1$. Thus, each factor set is defined by the single value $f(1, 1)$. We will choose one extension to have $f(1, 1) = 1$ and the other to have $f(1, 1) = y$. Since y is provided in the input, it is clear that we can efficiently compute the value $f(a, b)$ for either of these extensions.

[16] Note that the group operation in $\mathbb{Z}_N^*$, while abelian, is usually written as multiplication, while that of $\mathbb{Z}_2$ is written as addition. We will follow those conventions in this section. Note, however, that we used the opposite conventions for A and G in earlier sections.

We should also note that, for an f so defined to be a 2-cocycle, it must satisfy the additional (odd-looking) condition provided in the definition. This condition ranges over three variables $a, b, c \in G$, and since $|G| = 2$ in this case, this provides 8 equations that must be satisfied. It is a simple matter to write these out for the two factor sets described above and verify that these always hold, regardless of the value of $f(1,1)$, so we have the freedom to choose $f(1,1) = y$ as above.

In addition to the oracle just described, our extension equivalence test requires descriptions of the groups A, G, and E. For $G = \mathbb{Z}_2$, we can be compute a multiplication table in constant time (for the first quantum model) or we can easily construct an oracle that computes group operations in $\mathbb{Z}_2$ in constant time (for the second quantum model). For $A = \mathbb{Z}_N^*$, we can produce a generating set (with high probability) by choosing $O(\log N)$ random elements. To do this, we simply choose random elements of $\mathbb{Z}_N$ and then check that they are in $\mathbb{Z}_N^*$ by computing the GCD with N. It is well-known that this can be done efficiently, and since there is only a $o(1)$ chance that this test fails, we can produce a generating set in $O(\text{poly} \log N)$ time. Finally, for the group E, we can again choose $O(\log N)$ random elements (since $|E| = 2\,|A|$), and since E as a set is simply $\mathbb{Z}_N^* \times \mathbb{Z}_2$, we can choose a uniformly random element of E by choosing $x \in \mathbb{Z}_N^*$ and $a \in \mathbb{Z}_2$ uniformly, then forming (x, a).

The last input we must provide for extension equivalence is the isomorphism $E_i/\mathbb{Z}_N^* \cong \mathbb{Z}_2$. This is simply the function that maps $(x, a) \mapsto a$. Obviously, this can be performed efficiently.

Let E_1 be the extension with factor set f_1 having $f_1(1,1) = y$ and E_2 be the extension with f_2 having $f_2(1,1) = 1$. Then we can see that $f_1 f_2^{-1} = f_1$. Thus, these extensions are equivalent iff there exists a cochain $s : \mathbb{Z}_2 \to \mathbb{Z}_N^*$ such that $\partial s = f$. By construction, any s will ensure that $\partial s(0,0) = \partial s(0,1) = \partial s(1,0) = 1$ (otherwise, they would not be valid factor sets), so we only need $\partial s(1,1) = f_1(1,1) = y$. Let $x = s(1)$.[17] Then $\partial s(1,1) = s(1)s(1)s(1+1)^{-1} = x \cdot x \cdot 1^{-1} = x^2$. Thus, we can see that the extensions are equivalent iff there exists an $x \in \mathbb{Z}_N^*$ such that $x^2 = y$, i.e., iff y is a quadratic residue. ◀

Note that this example shows that extending even a constant-sized group (in this case, $|G| = 2$) by a large group can introduce substantial difficulty.

4 Counting Equivalence Classes of Extensions

In this section, we show that it is possible to compute $\left|H^2(G, A)\right|$, the number of inequivalent extensions of G by A, using the machinery developed earlier for testing equivalence. The size $\left|H^2(G, A)\right|$ is another quantity that is sometimes computed by hand for extensions of small groups and would be interesting to compute for larger groups.

We start first with the quantum algorithm, which takes A as a black box group and G given by a multiplication table.

Proof of Theorem 7. Since $H^2(G, A) \cong Z^2(G, A)/B^2(G, A)$, we can compute the size of the former group from the sizes of the latter two. In fact, we computed $\left|B^2(G, A)\right|$ as part of our quantum algorithm for testing equivalence, so we know how this can be done.

To compute $\left|Z^2(G, A)\right|$, we use the fact that $Z^2(G, A) = \text{Ker}\,\partial^2$, where $\partial^2 : C^2(G, A) \to B^3(G, A)$ is similar to the map ∂ $(= \partial^1)$ we used above. This map is a surjection, so the first isomorphism theorem tells us that $B^3(G, A) \cong C^2(G, A)/Z^2(G, A)$, which means that $\left|Z^2(G, A)\right| = \left|C^2(G, A)\right| / \left|B^3(G, A)\right|$. From the definition, we have $\left|C^2(G, A)\right| = |A|^{|G|^2}$.

[17] Any (normalized) 1-cochain s must have $s(0) = 1$, so 1-cochains in this case are in 1-to-1 correspondence with the element of $\mathbb{Z}_N^*$ by the mapping $s \mapsto s(1)$.

To compute $\left|B^3(G,A)\right|$, we can use the same approach as for $B^2(G,A)$: we take a generating set for $C^2(G,A)$, which is simply $|G|^2$ copies of the generating set for A and has size $O(|G|^2 \log |A|)$; push this forward into $B^3(G,A)$ by applying the map ∂^2, which has a simple formula; and then invoke the algorithm for computing the size of an abelian black box group. With $\left|B^3(G,A)\right|$ in hand, we can compute $\left|Z^2(G,A)\right|$ and then $\left|H^2(G,A)\right|$ by arithmetic. All of these steps can be done in $O(\operatorname{poly}|G| \operatorname{poly}\log|A|)$ time, so this gives an efficient algorithm. ◀

Finally, we have a classical algorithm when A and G are given by multiplication tables.

Proof of Theorem 6. We repeat the same approach as just described for the quantum algorithm of computing $\left|B^2(G,A)\right|$ and $\left|B^3(G,A)\right|$. Now, our classical algorithm for testing equivalence did not compute $\left|B^2(G,A)\right|$ as part of its operation. However, we did show how to efficiently construct a permutation representation for $B^2(G,A)$, and it is well-known that we can compute the size of a permutation group efficiently [5], so we can compute the size of this group classically as well.

We can also efficiently construct a generating set for $B^3(G,A)$, just as we did above, by taking a generating set for $C^3(G,A)$ (in the same manner as we did for $C^2(G,A)$ in the classical case) and pushing it forward using ∂^2. We can compute the size of this group efficiently as well, using the algorithm mentioned above, and then perform the same arithmetic as above. ◀

5 Conclusion

In this paper, we considered the problem of testing whether two extensions of a group G by an abelian group A are the same or "equivalent." If both $|A|$ and $|G|$ are small, then we showed that there exists an efficient (classical) Monte Carlo algorithm for testing equivalence. On the other hand, if $|A|$ is so large that A can only be provided as a black box and either $|G|$ is small or $|G|$ is large and abelian, then there is still an efficient quantum algorithm for testing equivalence, whereas no efficient classical algorithm exists, under the assumption that there is no efficient classical algorithm for testing quadratic residuosity.

As mentioned in the introduction, one of the motivations for studying this problem is its relationship to the group isomorphism problem, an important open problem in computer science. Hence, it is worth considering what light these results shed on the group isomorphism problem.

While the isomorphism problem applies to arbitrary groups, it is widely believed that the case of 2-nilpotent groups contains the essential hard cases. Any such groups are central extensions, and hence, we can apply our classical algorithm above to test their equivalence. If the two extensions are equivalent, then they are isomorphic. However, the opposite does not hold.

We can conclude from this that, if it is the case that testing isomorphism of 2-nilpotent groups is hard, then the hardness must come from extensions that are isomorphic but inequivalent. Hence, it behooves us to understand further the computational complexity of distinguishing such extensions.

Acknowledgements. The author would like to thank Aram Harrow for many useful discussions, much encouragement, and careful feedback on earlier drafts of this paper. Funding was from NSF grants CCF-0916400 and CCF-1111382.

References

1 L. Babai, P. Codenotti, J. A. Grochow, and Y. Qiao. Code equivalence and group isomorphism. In *Proceedings of the Twenty-Second Annual Symposium on Discrete Algorithms*, pages 1395–1408, 2011.

2 J. Rotman. *An Introduction to the Theory of Groups*. Springer, 1994.

3 V. Arvind and J. Torán. Solvable group isomorphism is (almost) in NP ∩ coNP. *ACM Transactions on Computation Theory*, 2(2):4:1–4:22, March 2011.

4 F. Le Gall. An efficient quantum algorithm for some instances of the group isomorphism problem. In *Proceedings of the 27th International Symposium on Theoretical Aspects of Computer Science*, pages 549–560, 2010. `arXiv:1001.0608`.

5 Á. Seress. *Permutation Group Algorithms*. Cambridge University Press, 2003.

6 L. Babai and E. Szemerédi. On the complexity of matrix group problems, i. In *Proceedings of the 25th IEEE Symposium on Foundations of Computer Science*, pages 229–240, 1984.

7 J. Watrous. Quantum algorithms for solvable groups. In *Proceedings of the 33rd ACM Symposium on Theory of Computing*, pages 60–67, 2001. `arXiv:quant-ph/0011023`.

8 S. Goldwasser and S. Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, April 1984.

9 L. Babai, G. Cooperman, L Finkelstein, E. Luks, and A. Seress. Fast Monte Carlo algorithms for permutation groups. *Journal of Computer and System Sciences*, 50(2):296–308, April 1995.

10 K. Cheung and M. Mosca. Decomposing finite abelian groups. *Quantum Information & Computation*, 1(3):26–32, October 2001. `arXiv:cs/0101004`.

11 L. Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. In *Proceedings of the 23rd ACM Symposium on Theory of Computing*, pages 164–174, 1991.

12 M. Mosca. *Quantum Computer Algorithms*. PhD thesis, University of Oxford, 1999.

13 P. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Scientific and Statistical Computing*, 26(5):1484–1509, October 1997. `arXiv:quant-ph/9508027`.

A Quantum Algorithm for Large, Abelian G

As mentioned in subsection 3.3, when G is a black box group, we have little hope of working with the group $B^2(G, A)$ since we cannot efficiently write down a generating set. Worse, we cannot even write down an $f \in Z^2(G, A)$ corresponding to our extension because this requires $|G|$ numbers in the general case. Hence, it is clear that we will need to put some restrictions on the form of f if we are to work with it efficiently. Below, we will see that this can be done without loss of generality in the case where G is abelian.

By the structure theorem for abelian groups, we know that $G \cong \mathbb{Z}_{d_1} \times \cdots \times \mathbb{Z}_{d_m}$ for some integers $d_1 \mid d_2 \mid \cdots \mid d_m$, which means $m = O(\log |G|)$. We can use the algorithm of [10] to efficiently decompose G into a product of this form on a quantum computer, so we can assume that we have G in this form.

As usual, we will have $f = \partial s$ for some $s : G \to E$. In particular, for $\{x_i \in \mathbb{Z}_{d_i}\}_{i \in [m]}$, we will choose $s(x_1, \ldots, x_m) = s_1^{x_1} \ldots s_m^{x_m}$ for some $\{s_i \in E\}$ such that s_i is a representative of $e_i \triangleq (0, \ldots, 0, 1, 0, \ldots, 0) \in G$ (where the 1 is in the i-th place). We can check that this s is a valid set of representatives for G. Since $\pi : E \to G$ is a homomorphism, we can see that $\pi(s(x_1, \ldots, x_m)) = (\pi s_1)^{x_1} \ldots (\pi s_m)^{x_m} = e_1^{x_1} \ldots e_m^{x_m} = (x_1, 0, \ldots, 0) \ldots (0, \ldots, 0, x_m) = (x_1, \ldots, x_m)$.

Most importantly, it is clear that we can write down the numbers $s_1, \ldots, s_m$ efficiently in terms of our generators for A, so this gives us an efficient way to represent s and $f = \partial s$.

Let us define $\mathcal{F}(G, E)$ to be the set of functions $G \to E$ of the above form, i.e., $s \in \mathcal{F}(G, E)$ iff $s(x_1, \ldots, x_m) = s_1^{x_1} \cdots s_m^{x_m}$ for some $s_1, \ldots, s_m \in E$. Note that we have $s(0, \ldots, 0) = 0$, so these functions are normalized. Since $s(x_1, \ldots, x_m)$ is always a representative of $(x_1, \ldots, x_m) \in G$, as we saw in the proof of Lemma 1, we then always have $\partial s \in Z^2(G, A)$, that is, $\partial\mathcal{F}(G, E) \subset Z^2(G, A)$. Likewise, if we consider the functions $\mathcal{F}(G, A)$ (with codomain A rather than E), we see that these are a subset of $C^1(G, A)$ — every $s \in \mathcal{F}(G, A)$ is a 1-cochain, but not every 1-cochain is in this concise form (defined in terms of some $s_1, \ldots, s_m$) — so we define $B^2_{\mathcal{F}}(G, A) \triangleq \partial\mathcal{F}(G, A) \subset \partial C^1(G, A) = B^2(G, A)$. (It may be helpful to refer back to Figure 1 for the definitions of C^2, B^2, Z^2, etc.)

The following lemma shows that it will be sufficient to work with $B^2_{\mathcal{F}}(G, A)$.

▶ **Lemma 8.** *Suppose that $f \in \partial\mathcal{F}(G, E_1)$ and $g \in \partial\mathcal{F}(G, E_2)$, then $f - g \in B^2(G, A)$ iff $f - g \in B^2_{\mathcal{F}}(G, A)$.*

Proof. Since $B^2_{\mathcal{F}}(G, A) \subset B^2(G, A)$, the reverse direction is immediate.

For the forward direction, suppose that $f - g \in B^2(G, A)$. We know that $f = \partial s$ for some $s \in \mathcal{F}(G, E_1)$. Since g differs from f by a coboundary, E_1 and E_2 are equivalent extensions. This means, in particular, that there exists an isomorphism $\tau : E_2 \to E_1$ respecting A and G. Now, let $u \in \mathcal{F}(G, E_2)$ be such that $g = \partial u$. Then we can see that

$$\tau(g(x, y)) = \tau(\partial u(x, y)) = \tau(u(x)u(y)u(x + y)^{-1}) = \tau u(x)\tau u(y)(\tau u(x + y))^{-1}.$$

Since $g(x, y) \in A$ and τ restricts to identity on A, we see that $g(x, y) = \tau g(x, y) = (\partial\tau u)(x, y)$. Thus, g can be realized as ∂t for some $t : G \to E_1$, namely, $t = \tau u$. Futhermore, since u is of the form $u(x_1, \ldots, x_m) = u_1^{x_1} \ldots u_m^{x_m}$, we see that $t(x_1, \ldots, x_m) = \tau u(x_1, \ldots, x_m) = (\tau u_1)^{x_1} \ldots (\tau u_m)^{x_m}$, which shows that $t \in \mathcal{F}(G, E_1)$ with $t_i \triangleq \tau u_i$ the representative of e_i for each $i \in [m]$.

The above shows that we can restrict our attention to considering $f - g = \partial s - \partial t$, where $s, t \in \mathcal{F}(G, E_1)$. In this case, we can compute

$$\begin{aligned} f(x) - g(y) &= s(x)s(y)s(x + y)^{-1}(t(x)t(y)t(x + y)^{-1})^{-1} \\ &= s(x)s(y)s(x + y)^{-1}t(x + y)t(y)^{-1}t(x)^{-1}. \end{aligned}$$

Now, note that $s(x + y)^{-1}t(x + y) \in A$ since

$$\pi(s(x + y)^{-1}t(x + y)) = (\pi s(x + y))^{-1}\pi t(x + y) = -(x + y) + (x + y) = 0$$

in G. Since A is central in E, we can move $s(x + y)^{-1}t(x + y)$ to the end. This leaves $s(y)t(y)^{-1}$ adjacent. Since this is in A for the same reason, we can rearrange this as well. Thus, we have $f(x) - g(y) = s(x)t(x)^{-1}s(y)t(y)^{-1}s(x + y)^{-1}t(x + y)$. This is close, but not identical, to

$$\partial(st^{-1})(x, y) = s(x)t(x)^{-1}s(y)t(y)^{-1}(s(x + y)t(x + y)^{-1})^{-1},$$

the only difference being the order of the last two factors.

We can show, however, that these two terms commute. In particular, let $x = (x_1, \ldots, x_m)$. Then we have $s(x_1, \ldots, x_m) = s_1^{x_1} \ldots s_m^{x_m}$ and $t(x_1, \ldots, x_m) = t_1^{x_1} \ldots t_m^{x_m}$ so that $s(x)t(x)^{-1} = s_1^{x_1} \ldots s_m^{x_m} t_m^{-x_m} \ldots t_1^{-x_1}$. Since s_m and t_m are both representatives of $e_m \in G$, we know that $s_m^{x_m} t_m^{-x_m} \in A$, which means we can move this term to the end. Repeating this as above, we have $s(x)t(x)^{-1} = s_1^{x_1} t_1^{-x_1} \ldots s_m^{x_m} t_m^{-x_m}$. Now, since s_m and t_m are both representatives

of e_m, they must differ by a factor of some $a_m \in A$, so we have $t_m = s_m a_m$, which means that $s_m^{x_m} t_m^{-x_m} = s_m^{x_m} s_m^{-x_m} a_m^{-x_m}$, and more generally, $s(x)t(x)^{-1} = a_1^{-x_1} \dots a_m^{-x_m}$. Now, if we compute the product in the other order, we have $t(x)^{-1}s(x) = t_m^{-x_m} \dots t_1^{-x_1} s_1^{x_1} \dots s_m^{x_m} = t_1^{-x_1} s_1^{x_1} \dots t_m^{-x_m} s_m^{x_m}$ by the same rearranging as before, and since $t_1^{-x_1} s_1^{x_1} = s_1^{-x_1} a_1^{-x_1} s_1^{x_1} = a_1^{-x_1}$ (using the fact that A is central in E_1), we can see that $t(x)^{-1}s(x) = a_m^{-x_m} \dots a_1^{-x_1}$. This is equal to what we computed for $s(x)t(x)^{-1}$ since A is abelian, so we have shown that $f(x) - g(y) = \partial(st^{-1})(x, y)$.

If we let $v : G \to E_1$ be defined by $v(x) = s(x)t(x)^{-1}$, then we have shown above that $f - g = \partial v$. In particular, we showed $v(x_1, \dots, x_m) = a_1^{-x_1} \dots a_m^{-x_m}$, which means that $v \in \mathcal{F}(G, A)$ with $v_i = a_i^{-1}$. Thus, we have seen that $f - g \in \partial\mathcal{F}(G, A) = B^2_{\mathcal{F}}(G, A)$. ◀

The following two lemmas tell us more about what elements in these groups look like.

▶ **Lemma 9.** *If $h \in B^2_{\mathcal{F}}(G, A)$, then there exist $\alpha_1, \dots, \alpha_m \in A$ such that $h(x, y) = \prod_{i=1}^m \alpha_i^{\delta_i}$, where $\delta_i = 1$ if $x_i + y_i \geq d_i$ and 0 otherwise and $\alpha_i = a_i^{d_i}$ for some a_i.*

Proof. If h is as above, we know that $h = \partial v$ for some $v \in C^1_{\mathcal{F}}(G, A)$, where v is of the form $v(x_1, \dots, x_m) = a_1^{x_1} \dots a_m^{x_m}$ for some $\{a_i \in A\}$. Since A is abelian, we can see that

$$h(x, y) = v(x_1, \dots, x_m)v(y_1, \dots, y_m)v(x_1 + y_1, \dots, x_m + y_m)^{-1} = \prod_{i=1}^m a_i^{x_i} a_i^{y_i} a_i^{-(x_i+y_i) \bmod d_i}$$

because $x_i + y_i$ in G is computed mod d_i. If $x_i + y_i < d_i$, then the mod has no effect, and we see that $h(x, y) = e$. On the other hand, if $x_i + y_i \geq d_i$, then $-(x_i + y_i) \bmod d_i = -x_i - y_i + d_i$. This means that $a_i^{x_i} a_i^{y_i} a_i^{-(x_i+y_i) \bmod d_i} = a_i^{d_i}$, so we can see that $h(x, y) = \prod_{i=1}^m a_i^{d_i \delta_i}$, where each δ_i is defined as in the statement of the lemma. We get the form in the statement by defining $\alpha_i = a_i^{d_i}$. ◀

▶ **Lemma 10.** *If $f \in Z^2_{\mathcal{F}}(G, A)$, so that $f = \partial s$ for some $s \in \mathcal{F}(G, E)$, then there exist $\{\alpha_i \in A\}_{1 \leq i \leq m}$ and $\{\beta_{i,j} \in A\}_{1 \leq i < j \leq m}$ such that $f(x, y) = \prod_{1 \leq i \leq m} \alpha_i^{\delta_i} \prod_{1 \leq i < j \leq m} \beta_{i,j}^{y_i x_j}$, where δ_i is defined as in the previous lemma, $\alpha_i = s_i^{d_i}$, and $b_{i,j} = [s_i, s_j^{-1}]$.*

Proof. By definition, we have

$$f(x, y) = s(x)s(y)s(x + y)^{-1} = s_1^{x_1} \cdots s_m^{x_m} s_1^{y_1} \cdots s_m^{y_m} s_m^{-(x_m+y_m) \bmod d_m} \cdots s_1^{-(x_1+y_1) \bmod d_1}.$$

As in the previous lemma, we can rewrite this as

$$f(x, y) = s_1^{x_1} \cdots s_m^{x_m} s_1^{y_1} \cdots s_m^{y_m} s_m^{-x_m - y_m + d_m \delta_m} \cdots s_1^{-x_1 - y_1 + d_1 \delta_1}.$$

We can begin by using the fact that $s_i^{d_i} \in A$ for each i. This follows because $\pi(s_i^{d_i}) = \pi(s(e_i)^{d_i}) = (0, \dots, d_i, \dots, 0) = 0$ since the i-th part of G is $\mathbb{Z}_{d_i}$, meaning addition is modulo d_i.

Thus, we can define $\alpha_i \triangleq s_i^{d_i}$. Since A is abelian, we can pull all of these factors to the front. This puts f in the form

$$f(x, y) = \left(\prod_{i=1}^m \alpha_i^{\delta_i} \right) s_1^{x_1} \cdots s_m^{x_m} s_1^{y_1} \cdots s_m^{y_m} s_m^{-x_m - y_m} \cdots s_1^{-x_1 - y_1}.$$

In the middle of the latter product, we have $s_{m-1}^{y_{m-1}} s_m^{y_m} s_m^{-x_m - y_m} s_{m-1}^{-x_{m-1} - y_{m-1}}$. We can cancel $s_m^{y_m}$ and $s_m^{-y_m}$, leaving us with $s_{m-1}^{y_{m-1}} s_m^{-x_m} s_{m-1}^{-x_{m-1} - y_{m-1}}$. In order to cancel the $s_{m-1}^{y_{m-1}}$, we first have to move it past the $s_m^{-x_m}$. We can do this by introducing a commutator that

compensates for the order change. This allows the $s_{m-1}^{y_m}$ factor to cancel, leaving us with $[s_{m-1}^{y_{m-1}}, s_m^{x_m}] s_{m-1}^{-x_{m-1}}$.

More generally, we can consider $[s(u), s(v)]$ for any $u, v \in G$. We can see that

$$\pi[s(u), s(v)] = \pi(s(u)s(v)s(u)^{-1}s(v)^{-1}) = \pi s(u)\pi s(v)\pi s(u)^{-1}\pi s(v)^{-1} = u+v-u-v = 0,$$

which means that $[s(u), s(v)] \in A$. In particular, this means that we can move commutators to the front.

Hence, we can simplify $s_1^{x_1} \cdots s_m^{x_m} s_1^{y_1} \cdots s_m^{y_m} s_m^{-x_m-y_m} \cdots s_1^{-x_1-y_1}$ by introducing commutators to move each factor of $s_j^{-x_j}$ in front of each remaining factor of $s_i^{y_i}$. In the example above, we saw that there was no moving required for $i = m$, while $i = m-1$ only need to move past $j = m$. In general, will need to swap each pair of this form with $i < j$. Each such swap introduces a commutator, but since these are all in A, we can immediately move them to the front and continue swapping these factors and canceling the matching factors until nothing remains.

Finally, note that a swap of $s_i^{y_i}$ and $s_j^{-x_j}$ can be thought of as a number of swaps between s_i's and s_j^{-1}'s. Since each of the y_i copies of the first must move past each of the x_j copies of the second, we see that there are $y_i x_j$ swaps overall. Thus, we can write the commutator as $[s_i, s_j^{-1}]^{y_i x_j}$, giving us the form in the statement of the lemma. ◀

The following is the main result needed for our algorithm.

▶ **Lemma 11.** *Let $f, f' \in Z^2_{\mathcal{F}}(G, A)$. Write these in the form of the previous lemma with $\{\alpha_i\}, \{\beta_{i,j}\}$ for f and $\{\alpha'_i\}$ and $\{\beta'_{i,j}\}$ for f'. Then $f - f' \in B^2_{\mathcal{F}}(G, A)$ iff $\beta_{i,j} = \beta'_{i,j}$ for all $1 \le i < j \le m$ and $(\alpha_i)^{-1}\alpha'_i$ has a d_i-th root in A.*

Proof. We begin with the reverse direction. Let $a_i \in A$ be a d_i-th root of $(\alpha_i)^{-1}\alpha'_i$. Recall that $\alpha_i = s_i^{d_i}$. Replacing s_i with $s_i a_i$ gives another valid set of representatives and, hence, an extension equivalent to f'. Defining f'' using this set of representatives gives an $\alpha''_i = s_i^{d_i} a_i^{d_i} = \alpha_i(\alpha_i)^{-1}\alpha'_i = \alpha'_i$. Since f and f' agree on the $\beta_{i,j}$'s and since including extra factors from A does not change the $\beta_{i,j}$'s (as A is central and $\beta_{i,j}$ is a commutator), we see that f'' and f' agree on both the α_i's and $\beta_{i,j}$'s, so $f'' = f'$. Next, since f and f'' arise by choosing different representatives for the same extension, we know that $f - f'' \in B^2(G, A)$. However, since $f, f'' \in Z^2_{\mathcal{F}}(G, A)$, we have $f - f'' \in B^2_{\mathcal{F}}(G, A)$ by Lemma 8. Thus, we can see that $f - f' = (f - f'') + (f'' - f') = f - f'' \in B^2_{\mathcal{F}}(G, A)$.

For the forward direction, we will separately prove the two implications, that $f - f' \in B^2_{\mathcal{F}}(G, A)$ implies the condition on the $\beta_{i,j}$'s and that it implies the condition on the α_i's.

For the condition on the $\beta_{i,j}$'s, we will prove the contrapositive. First, suppose that $\beta_{i,j} \neq \beta'_{i,j}$ for some $i < j$. From the formula in Lemma 9, we can see that $h(e_i, e_j) = 0$ for any $h \in B^2_{\mathcal{F}}(G, A)$. On the other hand, from the formula in Lemma 10, we see that $f(e_i, e_j) = \beta_{i,j} \neq \beta'_{i,j} = f'(e_i, e_j)$. Since every coboundary is 0 on this pair, we conclude that $f - f' \notin B^2_{\mathcal{F}}(G, A)$.

Now, we prove the condition on the α_i's. Suppose that $h \triangleq f' - f \in B^2_{\mathcal{F}}(G, A)$. From the formula in Lemma 9, writing the constants for h as α''_i, we can see that $h(e_i, (d_i - 1)e_i) = \alpha''_i = a_i^{d_i}$. From the formula in Lemma 10, we see that $f(e_i, (d_i - 1)e_i) = \alpha_i$ and $f'(e_i, (d_i - 1)e_i) = \alpha'_i$. Taking $f' - f = h$ at the pair $(e_i, (d_i - 1)e_i)$ and writing with multiplicative notation, we see that $\alpha'_i(\alpha_i)^{-1} = \alpha''_i = a_i^{d_i}$. Since $(\alpha_i)^{-1}\alpha'_i = \alpha'_i(\alpha_i)^{-1}$ (both are in A), we see that the d_i-th root exists.

Thus, we have seen that, if the condition on the $\beta_{i,j}$'s and α_i's does not hold (so either the $\beta_{i,j}$ condition does not hold or the α_i condition does not hold), it is impossible to have $f - f' \in B^2_{\mathcal{F}}(G, A)$. ◀

We now have the necessary tools required to prove the theorem in this case.

Proof of Theorem 4. Assuming that we can compute a factor set in $Z^2_{\mathcal{F}}(G, A)$ for each extension, we only need to compute the α_i's and $\beta_{i,j}$'s from Lemma 11 for each factor set and check whether they satisfy the conditions of the last lemma.

We saw in the proof of the lemma that these constants can be found simply by evaluating the factor set at particular points. There are only $O(m^2) = O(\log^2 |G|)$ constants to compute. Given the simple form of each $f \in Z^2_{\mathcal{F}}(G, A)$, it is clear that we can perform these evaluations efficiently. Thus, we can efficiently determine the α_i's and $\beta_{i,j}$'s.

For the $\beta_{i,j}$'s, the conditions of Lemma 11 require us simply to check equality, which we can do for each (i, j) with one call to the oracle for A. For the α_i's, on the other hand, we need to determine whether the quotient of two α_i's is a d_i-th root.

Recalling that A is an abelian group, we can switch back to additive notation. Our goal is to determine whether there exists an $a \in A$ such that $d_i a = \alpha'_i - \alpha_i$. Since A is isomorphic to a product $\mathbb{Z}_{n_1} \times \cdots \times \mathbb{Z}_{n_k}$, this splits into k independent equations. For each $1 \leq j \leq k$, we want to find an a_j such that $d_i a_j = (\alpha'_i - \alpha_i)_j \pmod{n_k}$ or, equivalently, if there exist a_j and b_j such that $a_j d_i + b_j n_k = (\alpha'_i - \alpha_i)_j$. Let d be the greatest common denominator of d_i and n_k. We can solve this equation iff d divides $(\alpha'_i - \alpha_i)_j$.

Thus, for the α_i's, the conditions of Lemma 11 require us to compute the α_i's, split them into the parts of the direct product, and then check whether the difference in each component is divisible by the greatest common denominator of d_i and n_k. We get d_i by decomposing G into a direct product of cyclic groups using the algorithm of [10]. We apply the same algorithm to A to find n_k and the $(\cdot)_j$ components of $a'_i - \alpha_i$ needed above.[18] We then simply need to check divisibility for $O(\log |G|)$-bit numbers, which we can do efficiently on a classical computer. Since the quantum algorithm of [10] is efficient, we have seen that there is an efficient quantum algorithm for testing whether the difference of two factor sets is a coboundary.

It remains to describe how to compute each factor set or, more specifically, the representatives $s_1, \ldots, s_m$ for each of the direct factors (since we can efficiently evaluate a factor set given these numbers). As in our earlier quantum algorithm, we can produce nearly uniformly random elements from E and then apply the oracle to find the corresponding elements of G. This process gives us nearly uniformly random elements of G. As we have seen before, we need only $O(\log |G|)$ random elements to get a set that generates all of G. The key fact is that we have not only a generating set for G but rather a generating set for G with each generator coming from an element in E.

Since these generate G, we know that, for each $i \in [m]$, there exists a product that gives $e_i \in G$. The corresponding product of elements of E is thus a representative of e_i. To find this product, we apply the algorithm of [10] to express G as a direct product of cyclic groups and get the relations for converting from the generators we have to the standard generators for the direct factors. These relations come in the form of an $O(\log |G|) \times O(\log |G|)$ matrix. For each $i \in [m]$, one column of this matrix gives the relation for generating e_i as a product of powers of $O(\log |G|)$ of our random elements. Since we can compute powers efficiently and this matrix is small, we can efficiently compute this product to get e_i. More importantly, we

[18] The algorithm of [10] computes not only generators for the factors of the direct product but also formulas (the vectors $\mathbf{y}_i$) for converting from the original generators to the new ones. The map taking $\mathbf{e}_i \mapsto \mathbf{y}_i$ is invertible, so we can efficiently compute the reverse direction (from new generators to the original ones) as well.

can compute the product of the elements of E corresponding to these generators to produce a representative of e_i. This is a valid choice for s_i.

In summary, we find a set of representatives $\{s_i\}$ for each extension that allows us to efficiently compute a factor set in $Z^2_{\mathcal{F}}(G, A)$. Then, we can check whether their difference lies in $B^2_{\mathcal{F}}(G, A)$ by computing the α_i's and $\beta_{i,j}$'s for each extension and checking the conditions of the lemma. As we saw above, both of these steps can be performed efficiently on a quantum computer. ◀

B Quantum Algorithm for Computing Group Size

The quantum algorithm in subsection 3.3 requires a subroutine that computes the size of a black box group. Earlier, we cited the algorithm and analysis of [10, 12] but skipped some of the finer details of how the theorems from those papers translate into a running time for this subroutine in our algorithm. In this section, we fill in those missing details.

The algorithm of [10] is not explicitly for computing the size of the group. Rather, it is for decomposing the group into a direct product of cyclic groups. That is, it produces a set of generators, one for each of the direct factors. However, it is easy to compute the size of the group from this information.

In particular, the size of the group is simply the product of the sizes of the direct factors, and since each of these is a cyclic group, the size of each direct factor is simply the order of the generator. Hence, we can get the size of the group from the output of this algorithm by invoking an order finding subroutine.

Finding order is a special case of the algorithm for computing the period of a function, which is also described and analyzed in [12]. In our case, the function whose period we want to find is the map $n \mapsto g^n$, where $g \in A$ is the generator whose order we are computing. Since the order of g is bounded by $|A|$, the method of repeated squaring allows us to compute this map with $O(\log |A|)$ calls to the oracle for A.

The quantum period finding algorithm makes only one call to the function just described, taking $O(\log |A|)$ time. However, it must also perform $O(\log^2 |A|)$ post-processing, which dominates the running time.

To compute the size of our group, we need to find the order of all $O(|G|^2 \log |A|)$ generators, which we can see takes $O(|G|^2 \log^3 |A|)$ time. This is adds only a lower order term to the overall running time.

That completes the discussion of our own post-processing to compute the size of the group. However, we will also need to perform some pre-processing.

The algorithm described in [12] requires that all of the given generators have order that is p^k for some fixed prime p. This is done in order to reduce the amount of quantum computation that is needed because separation into different p-groups can be done classically, as we will now describe.

We start by finding the order of each generator. As noted above, this takes $O(|G|^2 \log^3 |A|)$ time. Next, we factor the order using Shor's algorithm [13], which takes $O(\log^3 |A|)$ time. Now, suppose that the order of g is $r = p_1^{j_1} \dots p_k^{j_k}$. Then, if we let $q_\ell = \prod_{i \neq \ell} p_i^{j_i}$, then we can see that the order of g^{q_ℓ} is $p_\ell^{j_\ell}$. Furthermore, we know from the Chinese remainder theorem that any $x \in \mathbb{Z}_r$ is uniquely determined by the values $x \bmod p_\ell^{j_\ell}$ for each ℓ. Hence, any power of g can be written uniquely as a product of powers of $g^{q_1}, \dots, g^{q_k}$.

We now have a generating set for which we know the prime power order of each element. Thus, we can separately pass the generators for each p-subgroup (those whose order is a power of p) to the algorithm from [10]. The structure theorem for finite abelian groups tells

us that our group is a direct product of the p-subgroups, so we can simply multiply their sizes to get the size of the whole group.

We can see that this pre-processing adds only a lower order term to the running time of the algorithm. While our generating set for the whole group may have grown, each generator adds at most a single generator to the set for each p-subgroup, so the running time of the group decomposition algorithm that we analyzed before is unchanged. The one difference is that we may need to invoke that algorithm as many as $\log |A|$ times, so this adds a factor of $\log |A|$ to our bound on the running time.

Finally, we should note that the decomposition algorithm described in [12] also mentions $O(k^2 \log q)$ classical group multiplications (meaning multiplication in the group $\mathbb{Z}_{|A|}$). This is dominated by the $O(k^3 \log q)$ part of the post-processing, which works in the same group, so it does not add to the overall running time.

Provable Advantage for Quantum Strategies in Random Symmetric XOR Games*

Andris Ambainis and Jānis Iraids

Faculty of Computing, University of Latvia
Raiņa bulvāris 19, Riga, LV-1586, Latvia
andris.ambainis@lu.lv, janis.iraids@gmail.com

Abstract

Non-local games are widely studied as a model to investigate the properties of quantum mechanics as opposed to classical mechanics. In this paper, we consider a subset of non-local games: symmetric XOR games of n players with 0-1 valued questions. For this class of games, each player receives an input bit and responds with an output bit without communicating to the other players. The winning condition only depends on XOR of output bits and is constant w.r.t. permutation of players.

We prove that for almost any n-player symmetric XOR game the entangled value of the game is $\Theta\left(\frac{\sqrt{\ln n}}{n^{1/4}}\right)$ adapting an old result by Salem and Zygmund on the asymptotics of random trigonometric polynomials. Consequently, we show that the classical-quantum gap is $\Theta(\sqrt{\ln n})$ for almost any symmetric XOR game.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Random Symmetric XOR games, Entanglement

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.146

1 Introduction

Non-local games provide a simple way to test the difference between quantum mechanics and the classical world. A prototypical example of a non-local game is the CHSH game [6] (based on the CHSH inequality of [5]). In the CHSH game, we have two players who cannot communicate between themselves but may share common random bits or a bipartite quantum state (which has been exchanged before the beginning of the game). A referee sends one uniformly random bit $a \in \{0,1\}$ to the 1^{st} player and an independent uniformly random bit $b \in \{0,1\}$ to the 2^{nd} player. Players respond by sending one-bit answers $x, y \in \{0,1\}$. They win in the following 2 cases:

(a) If at least one of a, b is equal to 0, players win if they produce x, y such that $x = y$;
(b) If $a = b = 1$, players win if they produce x, y such that $x \neq y$;

Classically, CHSH game can be won with probability at most 0.75. In contrast, if players use an entangled quantum state, they can win the game with probability $\frac{1}{2} + \frac{1}{2\sqrt{2}} = 0.85...$.

Other non-local games can be defined by changing the number of players, the number of possible questions and answers and the winning condition. Many non-local games have been studied and, in many cases, strategies that use an entangled quantum state outperform any classical strategy.

* Supported by EU FP7 projects QCS, QALGO and MQC. The presentation was supported by ERDF Project No. 2010/0202/2DP/2.1.1.2.0/10/APIA/VIAA/013

Recently [1], it has been shown that, for a large class of non-local games, quantum strategies are better than any classical strategy for almost all games in this class. Namely, [1] considered 2-player games in which the questions a, b are taken from the set $\{1, 2, \ldots, n\}$ and the winning condition is either $x = y$ or $x \neq y$, depending on a, b. (Games with a winning condition of such form are called XOR games.) [1] showed that, for $1 - o(1)$ fraction of all such games, the entangled value of the game is at least 1.2... times its classical value.

Then [2], it was discovered that a similar effect might hold for another class of games: n-player symmetric XOR games with binary questions. Namely, [2] showed a gap between entangled and classical values of order $\Omega(\sqrt{\log n})$ - assuming that a non-rigorous argument about the entangled value is correct.

In this paper, we make this gap rigorous, by proving upper and lower bounds on the entangled value of a random game in this class. We show that, with a high probability, the entangled value is equal to $\Theta\left(\frac{\sqrt{\log n}}{n^{1/4}}\right)$. The quantum-vs-classical gap of $\Theta(\sqrt{\log n})$ follows by combining this with the fact that the classical value is of the order $\Theta(\frac{1}{n^{1/4}})$ for almost any random game (shown in [2]).

To prove this result, we use an expression for the entangled value of a symmetric n-player XOR game with entangled answers from [3]. This expression reduces finding the entangled value to maximizing the absolute value of a polynomial in one complex variable. If conditions for the XOR game are chosen at random, this expression reduces to random trigonometric polynomials studied in [7].

Although maxima of random trigonometric polynomials have been studied in [7], they have been studied under different conditions. For this reason, we cannot apply the results from [7] directly. Instead, we adapt the ideas from [7] to prove a bound on maxima of random trigonometric polynomials that would be applicable in our setting.

2 Definitions

A non-local game with n players proceeds as follows:

1) Players are separated so that they cannot communicate – hence the name *non-local*,
2) The players receive inputs $x_1, x_2, \ldots, x_n \in I$ where I is the set of possible inputs. i-th player receives x_i,
3) The players respond with outputs $y_1, y_2, \ldots, y_n \in O$ where O is the set of possible outputs.
4) The winning condition $P(x_1, \ldots, x_n, y_1, \ldots, y_n)$ is consulted to determine whether the players win or lose. The condition is known to everyone at the start of the game.

The players are informed of the rules of the game and they can agree upon a strategy and exchange other information. In the classical case players may only use shared randomness. In the quantum case they can use an entangled quantum state which is distributed to the players before the start of the game.

We will restrict ourselves to the case when $I = O = \{0, 1\}$ and the vector of inputs $(x_1, \ldots, x_n)$ is chosen uniformly at random. In an *XOR* game, the winning condition $P(x_1, \ldots, x_n, y_1, \ldots, y_n)$ depends only on $x_1, \ldots, x_n$ and the parity of the output bits $\oplus_{j=1}^n y_j$. A game is *symmetric* if the winning condition does not change if $x_1, \ldots, x_n$ are permuted.

The winning conditions of a symmetric XOR game can be described by a list of $n+1$ bits: $G = (G_0, G_1, \ldots, G_n)$, where the players win if and only if $G_i = \oplus_{j=1}^n y_j$ when $\sum_{j=1}^n x_j = i$.

The *entangled value* of the game $Val_Q(G)$ is the probability of winning minus the probability of losing in the conditions that the players can use a shared quantum-physical system.

In this paper, we study the value of symmetric XOR games when the winning condition G is chosen randomly from the uniform distribution of all $(n+1)$-bit lists. We use the following lemma (which follows from a more general result by Werner and Wolf for non-symmetric XOR games [9]):

▶ **Lemma 1** (See [3]). *The entangled value of a symmetric XOR game [3] is*

$$Val_Q(G) = \max_{|\lambda|=1} \left| \sum_{j=0}^{n} (-1)^{G_j} p_j \lambda^j \right| \tag{1}$$

where p_j is the probability that players are given an input vector $(x_1, \ldots, x_n)$ with j variables $x_i = 1$.

In our case, since $(x_1, \ldots, x_n)$ is uniformly random, we have $p_j = \frac{\binom{n}{j}}{2^n}$.

In the following sections we introduce additional notation to keep the proofs more concise as well as to keep in line with the original proofs in [7]:

The *Rademacher system* is a set of functions $\{\varphi_m(t)\}$ for $m = 1, \ldots, n$ over $0 \le t \le 1$ such that $\varphi_m(t) = (-1)^k$, where k is the m-th digit after the binary point (in the fractional part of t) of the binary expansion of t. Rademacher system will turn out to be a convenient way to state that $\{G_j\}$ are random variables that follow a uniform distribution: if t is chosen randomly from a uniform distribution on $0 \le t \le 1$, then $\{\varphi_m(t)\}_{m=1}^{n+1}$ generates a uniformly random element from $\{+1, -1\}^{n+1}$. That in turn corresponds to coefficients $(-1)^{G_j}$ in eq. (1) being picked randomly.

Furthermore, we define

$$r_m = \binom{n}{m} \quad (n \text{ will be clear from context}),$$

$$R_n = \sum_{m=0}^{n} r_m^2,$$

$$T_n = \sum_{m=0}^{n} r_m^4,$$

$$P_n(x,t) = \sum_{m=0}^{n} r_m \varphi_{m+1}(t) \cos mx,$$

$$M_n(t) = \max_{0 \le x < 2\pi} |P_n(x,t)|.$$

3 Main Result

By adapting the work of Salem and Zygmund [7] on the asymptotics of random trigonometric polynomials, we show

▶ **Theorem 2.**

$$\lim_{n\to\infty} \Pr[M_n(t) \ge C_1 \sqrt{R_n \ln n}] = 1.$$

▶ **Theorem 3.**

$$\lim_{n\to\infty} \Pr[M_n(t) \le C_2 \sqrt{R_n \ln n}] = 1.$$

Our proof yields $C_1 = \frac{1}{4\sqrt{3}}$ and $C_2 = 2$.

We will now show how these two theorems lead to an asymptotic bound for the entangled value of a random game.

▶ **Corollary 4.** *For almost all n-player symmetric XOR games the entangled value of the game is $\Theta\left(\frac{\sqrt{\ln n}}{n^{1/4}}\right)$.*

Proof. From Lemma 1,

$$Val_Q(G) \geq \max_{|\lambda|=1} \left| \mathfrak{Re} \left(\sum_{j=0}^{n} \frac{(-1)^{G_j} \binom{n}{j} \lambda^j}{2^n} \right) \right| = \max_{\alpha \in [0;2\pi]} \left| \sum_{j=0}^{n} \frac{(-1)^{G_j} \binom{n}{j} \cos j\alpha}{2^n} \right|,$$

and

$$Val_Q(G) \leq \max_{|\lambda|=1} \left| \mathfrak{Re} \left(\sum_{j=0}^{n} \frac{(-1)^{G_j} \binom{n}{j} \lambda^j}{2^n} \right) \right| + \max_{|\lambda|=1} \left| \mathfrak{Im} \left(\sum_{j=0}^{n} \frac{(-1)^{G_j} \binom{n}{j} \lambda^j}{2^n} \right) \right| =$$
$$= \max_{\alpha \in [0;2\pi]} \left| \sum_{j=0}^{n} \frac{(-1)^{G_j} \binom{n}{j} \cos j\alpha}{2^n} \right| + \max_{\alpha \in [0;2\pi]} \left| \sum_{j=0}^{n} \frac{(-1)^{G_j} \binom{n}{j} \sin j\alpha}{2^n} \right|.$$

For a random game $\{(-1)^{G_j}\}$ follow the same distribution as $\{\varphi_{j+1}(t)\}$ for t uniformly distributed from interval $[0;1]$. Therefore Theorem 2 and Theorem 3 apply. Note that Theorem 3 is true for cosines as well as sines since we only use that $\cos^2 x \leq 1$, and so

$$\lim_{n\to\infty} \Pr\left[C_1 \frac{\sqrt{R_n \ln n}}{2^n} \leq Val_Q(G) \leq 2C_2 \frac{\sqrt{R_n \ln n}}{2^n} \right] = 1. \tag{2}$$

Finally,

$$\frac{\sqrt{R_n \ln n}}{2^n} = \frac{\sqrt{\binom{2n}{n} \ln n}}{2^n} \sim \frac{\sqrt{\frac{4^n}{\sqrt{\pi n}} \ln n}}{2^n} = \sqrt{\frac{\ln n}{\sqrt{\pi n}}}.$$

◀

4 Proof of Upper and Lower Bounds

We now proceed to prove theorems 2 and 3. Our proof is based on an old result by Salem and Zygmund [7], in which they prove bounds on the asymptotics of random trigonometric polynomials in a different setting (in which the coefficients r_m are not allowed to depend on n).

Due to the difference in the two settings, we cannot immediately apply the results from [7]. Instead, we prove corresponding theorems for our setting, re-using the parts of proof from [7] which also work in our case and replacing other parts with different arguments.

▶ **Lemma 5** (From [7]). *Let $f_n(t) = \sum_{m=0}^{n} c_m \varphi_{m+1}(t)$, where $\{\varphi_{m+1}(t)\}$ is the Rademacher system and c_m are real constants. Let $C_n = \sum_{m=0}^{n} c_m^2$, $D_n = \sum_{m=0}^{n} c_m^4$ and let λ be any real number. Then*

$$e^{\frac{1}{2}\lambda^2 C_n - \lambda^4 D_n} \leq \int_0^1 e^{\lambda f_n(t)}\, \mathrm{d}t \leq e^{\frac{1}{2}\lambda^2 C_n}.$$

▶ **Lemma 6** (From [7]). *Let $g(x,y)$, $a \le x \le b$, $c \le y \le d$, be a bounded real function. Suppose that*

$$|g(x,y)| \le A, \frac{\int_c^d \int_a^b g^2(x,y)\,\mathrm{d}x\,\mathrm{d}y}{(b-a)(d-c)} = B.$$

Then, for any positive number μ,

$$\frac{\int_c^d \int_a^b e^{\mu g(x,y)}\,\mathrm{d}x\,\mathrm{d}y}{(b-a)(d-c)} \le 1 + \mu\sqrt{B} + \frac{B}{A^2}e^{\mu A}.$$

Furthermore, when $\int_c^d \int_a^b g(x,y)\,\mathrm{d}x\,\mathrm{d}y = 0$,

$$\frac{\int_c^d \int_a^b e^{\mu g(x,y)}\,\mathrm{d}x\,\mathrm{d}y}{(b-a)(d-c)} \le 1 + \frac{B}{A^2}e^{\mu A}. \tag{3}$$

▶ **Lemma 7** (From [7]). *Let x be real and $P(x) = \sum_{m=0}^{n} \alpha_m \cos mx + \beta_m \sin mx$ be a trigonometric polynomial of order n, with real or imaginary coefficients. Let M denote the maximum of $|P(x)|$ and let θ be a positive number less than 1. Then there exists an interval of length not less than $\frac{1-\theta}{n}$ in which $|P(x)| \ge \theta M$.*

▶ **Lemma 8** (From [7]). *Let $\varphi(x) \ge 0$, and suppose that*

$$\int_0^1 \varphi(x)\,\mathrm{d}x \ge A > 0, \int_0^1 \varphi^2(x)\,\mathrm{d}x \le B$$

(clearly, $A^2 \le B$). Let $0 < \delta < 1$. Then

$$\Pr\left[\varphi(x) \ge \delta A \mid 0 \le x \le 1\right] \ge (1-\delta)^2 \frac{A^2}{B}.$$

▶ **Lemma 9.**

$$\frac{\sum_{i=0}^{n} \binom{n}{i}^4}{\left(\sum_{i=0}^{n} \binom{n}{i}^2\right)^2} \le \frac{4}{3} n^{-\frac{1}{2}}$$

Proof. If n is even:

$$\frac{\sum_{i=0}^{n} \binom{n}{i}^4}{\left(\sum_{i=0}^{n} \binom{n}{i}^2\right)^2} \le \frac{\sum_{i=0}^{n} \binom{n}{i}^2 \binom{n}{n/2}^2}{\left(\sum_{i=0}^{n} \binom{n}{i}^2\right)^2} = \frac{\binom{n}{n/2}^2}{\binom{2n}{n}} \le$$

$$\le \frac{\left(\frac{2^n}{\sqrt{3\frac{n}{2}+1}}\right)^2}{\frac{4^n}{\sqrt{4n}}} \le \frac{\sqrt{4n}}{3\frac{n}{2}+1} \le \frac{4}{3} n^{-\frac{1}{2}}$$

If n is odd:

$$\frac{\sum_{i=0}^{n} \binom{n}{i}^4}{\left(\sum_{i=0}^{n} \binom{n}{i}^2\right)^2} \le \frac{\sum_{i=0}^{n} \binom{n}{i}^2 \binom{n}{\lfloor n/2 \rfloor}^2}{\left(\sum_{i=0}^{n} \binom{n}{i}^2\right)^2} = \frac{\left(\frac{\binom{n+1}{\frac{n+1}{2}}}{2}\right)^2}{\binom{2n}{n}} \le$$

$$\le \frac{\left(\frac{2^{n+1}}{2\sqrt{3\frac{n+1}{2}+1}}\right)^2}{\frac{4^n}{\sqrt{4n}}} \le \frac{\sqrt{4n}}{3\frac{n+1}{2}+1} \le \frac{4}{3} n^{-\frac{1}{2}}$$

◀

Proof of Theorem 2. Set $I_n(t) = \frac{1}{2\pi}\int_0^{2\pi} e^{\lambda P_n(x,t)}\,\mathrm{d}x$. We proceed to give an upper bound for $\int_0^1 I_n(t)\,\mathrm{d}t$ and lower bound for $\int_0^1 I_n^2(t)\,\mathrm{d}t$ using Lemma 5. Then we will plug in these bounds in Lemma 8 for $\varphi = I_n$.

First, the lower bound clause of Lemma 5 applied to $I_n(t)$ gives for any real λ (we will assign its value later, at our convenience),

$$\begin{aligned}
&\int_0^1 I_n(t)\,\mathrm{d}t = \int_0^1 \left(\frac{1}{2\pi}\int_0^{2\pi} e^{\lambda P_n(x,t)}\,\mathrm{d}x\right)\mathrm{d}t = \frac{1}{2\pi}\int_0^{2\pi}\int_0^1 e^{\lambda P_n(x,t)}\,\mathrm{d}t\,\mathrm{d}x \geq \\
\geq &\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{2}\lambda^2\sum_{m=0}^n (r_m\cos mx)^2 - \lambda^4\sum_{m=0}^n (r_m\cos mx)^4}\,\mathrm{d}x \geq \\
\geq &\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{2}\lambda^2\sum_{m=0}^n (r_m\cos mx)^2 - \lambda^4 T_n}\,\mathrm{d}x = \\
= &\left(e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n}\right)\cdot\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{2}\lambda^2\sum_{m=0}^n (r_m\cos mx)^2 - \frac{r_m^2}{2}}\,\mathrm{d}x = \\
= &\left(e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n}\right)\cdot\frac{1}{2\pi}\int_0^{2\pi} e^{\frac{1}{4}\lambda^2\sum_{m=0}^n (r_m^2\cos 2mx)}\,\mathrm{d}x > \\
> &\left(e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n}\right)\cdot\frac{1}{2\pi}\int_0^{2\pi}\left(1 + \frac{1}{4}\lambda^2\sum_{m=0}^n (r_m^2\cos 2mx)\right)\mathrm{d}x \geq \\
\geq &\left(e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n}\right)
\end{aligned}$$

The second step is to establish an upper bound for $\int_0^1 I_n^2(t)\,\mathrm{d}t$. We start out in a similar fashion, by applying Lemma 5:

$$\begin{aligned}
&\int_0^1 I_n^2(t)\,\mathrm{d}t = \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi}\int_0^1 e^{\lambda(P_n(x,t)+P_n(y,t))}\,\mathrm{d}t\,\mathrm{d}x\,\mathrm{d}y \leq \\
&\leq \frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} e^{\frac{1}{2}\lambda^2\sum_{m=0}^n r_m^2(\cos mx+\cos my)^2}\,\mathrm{d}x\,\mathrm{d}y = \\
&= e^{\frac{1}{2}\lambda^2(R_n+r_0^2)}\cdot\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} e^{\frac{1}{2}\lambda^2 S_n(x,y)}\,\mathrm{d}x\,\mathrm{d}y
\end{aligned}$$

where

$$S_n(x,y) = \sum_{m=1}^n\left(\frac{1}{2}r_m^2\cos 2mx + \frac{1}{2}r_m^2\cos 2my + 2r_m^2\cos mx\cos my\right).$$

One can verify that

a)

$$\int_0^{2\pi}\int_0^{2\pi} S_n(x,y)\,\mathrm{d}x\,\mathrm{d}y = 0,$$

b)

$$\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} S_n(x,y)^2\,\mathrm{d}x\,\mathrm{d}y =$$

$$= \frac{1}{2\pi}\sum_{m=1}^{n}\int_0^{2\pi}\left(\frac{1}{2}r_m^2\cos 2mx\right)^2 \mathrm{d}x +$$
$$+ \frac{1}{2\pi}\sum_{m=1}^{n}\int_0^{2\pi}\left(\frac{1}{2}r_m^2\cos 2my\right)^2 \mathrm{d}y +$$
$$+ \frac{1}{(2\pi)^2}\sum_{m=1}^{n}\int_0^{2\pi}\int_0^{2\pi}\left(2r_m^2\cos mx\cos my\right)^2 \mathrm{d}x\,\mathrm{d}y =$$
$$= \frac{5}{4}T_n,$$

c)

$$|S_n(x,y)| \le 3R_n.$$

We apply eq. 3 from Lemma 6 with function $g = S_n$, $\mu = \frac{1}{2}\lambda^2$, $A = 3R_n$ and $B = \frac{5}{4}T_n$. We get

$$\frac{1}{(2\pi)^2}\int_0^{2\pi}\int_0^{2\pi} e^{\frac{1}{2}\lambda^2 S_n(x,y)}\,\mathrm{d}x\,\mathrm{d}y \le 1 + \frac{\frac{5}{4}T_n}{9R_n^2}e^{\frac{3}{2}\lambda^2 R_n} \le$$
$$\le 1 + \frac{T_n}{R_n^2}e^{\frac{3}{2}\lambda^2 R_n}.$$

And by Lemma 9,

$$1 + \frac{T_n}{R_n^2}e^{\frac{3}{2}\lambda^2 R_n} \le 1 + \frac{4}{3}n^{-\frac{1}{2}}e^{\frac{3}{2}\lambda^2 R_n}.$$

So far we have established the two prerequisites for Lemma 8:

1)

$$\int_0^1 I_n(t)\,\mathrm{d}t > e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n},$$

2)

$$\int_0^1 I_n^2(t)\,\mathrm{d}t \le e^{\frac{1}{2}\lambda^2(R_n + r_0^2)}\left(1 + \frac{4}{3}n^{-\frac{1}{2}}e^{\frac{3}{2}\lambda^2 R_n}\right).$$

The third step is to apply Lemma 8 with $\varphi = I_n$, $A = e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n}$, $B = e^{\frac{1}{2}\lambda^2(R_n + r_0^2)} \times$ $\times \left(1 + \frac{4}{3}n^{-\frac{1}{2}}e^{\frac{3}{2}\lambda^2 R_n}\right)$ and $\delta = n^{-\eta}$. This results in

$$\Pr[I_n(t) \ge n^{-\eta}e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n}] \ge (1 - n^{-\eta})^2 \frac{e^{\frac{1}{2}\lambda^2 R_n - 2\lambda^4 T_n}}{e^{\frac{1}{2}\lambda^2(R_n + r_0^2)}\left(1 + \frac{4}{3}n^{-\frac{1}{2}}e^{\frac{3}{2}\lambda^2 R_n}\right)} \ge$$
$$\ge (1 - n^{-\eta})^2 e^{-2\lambda^4 T_n - \frac{1}{2}\lambda^2 r_0^2}\left(1 - \frac{4}{3}n^{-\frac{1}{2}}e^{\frac{3}{2}\lambda^2 R_n}\right).$$

Finally we show that for suitably chosen λ, θ and η the claim follows. Set $\lambda = \theta\sqrt{\frac{\ln n}{R_n}}$ having θ such that $2\sqrt{\eta} < \theta < \sqrt{\frac{1}{3}}$. We deal with the two claims separately:

▶ **Lemma 10.**

$$I_n(t) \ge n^{-\eta}e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n} \implies M_n(t) \ge C_1\sqrt{R_n \ln n}.$$

Proof. Note that

$$e^{\lambda M_n(t)} \geq I_n(t) \geq e^{\frac{1}{4}\lambda^2 R_n - \lambda^4 T_n - \eta \ln n}.$$

Thus

$$\begin{aligned} M_n(t) &\geq \frac{1}{4}\lambda R_n - \lambda^3 T_n - \frac{\eta}{\lambda}\ln n = \\ &= \frac{\theta}{4}\sqrt{R_n \ln n} - \theta^3 \sqrt{R_n \ln n}\ln n \frac{T_n}{R_n^2} - \frac{\eta}{\theta}\sqrt{R_n \ln n} = \\ &= \sqrt{R_n \ln n}\left(\frac{\theta}{4} - \theta^3 \frac{4\ln n}{3\sqrt{n}} - \frac{\eta}{\theta}\right) \to \sqrt{R_n \ln n}\left(\frac{\theta}{4} - \frac{\eta}{\theta}\right). \end{aligned}$$

But $\frac{\theta}{4} - \frac{\eta}{\theta} = \text{constant} > 0$. We can choose θ arbitrarily close to $\sqrt{\frac{1}{3}}$ and η arbitrarily close to 0 to obtain $C_1 = \frac{1}{4\sqrt{3}}$. ◀

▶ **Lemma 11.**

$$\lim_{n\to\infty} (1 - n^{-\eta})^2 e^{-2\lambda^4 T_n - \frac{1}{2}\lambda^2 r_0^2}\left(1 - \frac{4}{3}n^{-\frac{1}{2}} e^{\frac{3}{2}\lambda^2 R_n}\right) = 1$$

Proof. Since η is positive, $n^{-\eta} \to 0$. Similarly,

$$e^{-2\lambda^4 T_n - \frac{1}{2}\lambda^2 r_0^2} = e^{-2\theta^4 (\ln n)^2 \frac{T_n}{R_n^2} - \frac{1}{2}\theta^2 r_0^2 \frac{\ln n}{R_n}} \geq e^{-\frac{8}{3\sqrt{n}}\theta^4 (\ln n)^2 - \frac{1}{2}\theta^2 r_0^2 \frac{\ln n}{R_n}} \to e^0 = 1,$$

and

$$\frac{4}{3}n^{-\frac{1}{2}} e^{\frac{3}{2}\lambda^2 R_n} = \frac{4}{3}n^{-\frac{1}{2}} e^{\frac{3}{2}\theta^2 \ln n} = \frac{4}{3}n^{\frac{3\theta^2 - 1}{2}} \to 0.$$

◀

◀

Proof of Theorem 3. We will examine $\int_0^1 \int_0^{2\pi} e^{\lambda |P_n(x,t)|}\,\mathrm{d}x\,\mathrm{d}t$. By Lemma 7 there exists $0 < \theta < 1$ such that:

$$\begin{aligned} &\int_0^1 \int_0^{2\pi} e^{\lambda |P_n(x,t)|}\,\mathrm{d}x\,\mathrm{d}t \geq \\ &\geq \int_0^1 \frac{1-\theta}{n} e^{\theta\lambda M_n(t)}\,\mathrm{d}t. \end{aligned}$$

On the other hand, by Lemma 5 we obtain:

$$\begin{aligned} &\int_0^1 \int_0^{2\pi} e^{\lambda |P_n(x,t)|}\,\mathrm{d}x\,\mathrm{d}t = \\ &= \int_0^{2\pi} \int_0^1 e^{\lambda |P_n(x,t)|}\,\mathrm{d}t\,\mathrm{d}x \leq \\ &\leq \int_0^{2\pi} \int_0^1 e^{\lambda P_n(x,t)} + e^{-\lambda P_n(x,t)}\,\mathrm{d}t\,\mathrm{d}x \leq \\ &\leq \int_0^{2\pi} \int_0^1 2e^{\frac{1}{2}\lambda^2 \sum_{m=0}^n r_m^2 \cos^2 mx}\,\mathrm{d}t\,\mathrm{d}x \leq \\ &\leq \int_0^{2\pi} \int_0^1 2e^{\frac{1}{2}\lambda^2 R_n}\,\mathrm{d}t\,\mathrm{d}x = \\ &= 4\pi e^{\frac{1}{2}\lambda^2 R_n}. \end{aligned}$$

Therefore,

$$\int_0^1 e^{\theta\lambda M_n(t)}\,\mathrm{d}t \le \frac{4\pi}{1-\theta} e^{\frac{1}{2}\lambda^2 R_n + \ln n}.$$

Have $\lambda = 2\sqrt{\frac{\ln n}{R_n}}$ and multiply both sides by $n^{-4-\eta}$, where $\eta > 0$. Then

$$\int_0^1 e^{\theta\lambda M_n(t) - (4+\eta)\ln n}\,\mathrm{d}t \le \frac{4\pi}{1-\theta} n^{-(1+\eta)}.$$

The sum over all n converges:

$$\sum_{n=1}^{\infty} \int_0^1 e^{\theta\lambda M_n(t) - (4+\eta)\ln n}\,\mathrm{d}t \le \sum_{n=1}^{\infty} \frac{4\pi}{1-\theta} n^{-(1+\eta)} < \infty.$$

Since the exponent function is non-negative and the whole sum converges, it is safe to interchange sum and integral:

$$\int_0^1 \sum_{n=1}^{\infty} e^{\theta\lambda M_n(t) - (4+\eta)\ln n}\,\mathrm{d}t < \infty.$$

Therefore, for almost all t

$$\sum_{n=1}^{\infty} e^{\theta\lambda M_n(t) - (4+\eta)\ln n} < \infty.$$

Hence, for almost all t there exists n_0 such that for all $n \ge n_0$

$$\theta\lambda M_n(t) - (4+\eta)\ln n < 0.$$

It follows that

$$\lim_{n\to\infty} \Pr\left[M_n(t) < \frac{(4+\eta)}{2\theta}\sqrt{R_n \ln n}\right] = 1.$$

◀

5 Conclusion

We have proven that the entangled value of almost any n-player symmetric XOR game is $\Theta\left(\frac{\sqrt{\ln n}}{n^{1/4}}\right)$ and therefore is by a factor of $\sqrt{\ln n}$ greater than its classical value.

In Fig. 1 we have plotted the sample mean value of the coefficient $\frac{\overline{Val_Q(n)}}{\frac{\sqrt{R_n \ln n}}{2^n}}$ over 10^5 random games for each n up to 100. We point out that the mean value of the coefficient is approaching ≈ 0.85. It would be interesting to determine if C_1 and $2C_2$ (see, eq. 2) can be further improved and whether the coefficients in fact tend to a common limit near 0.85.

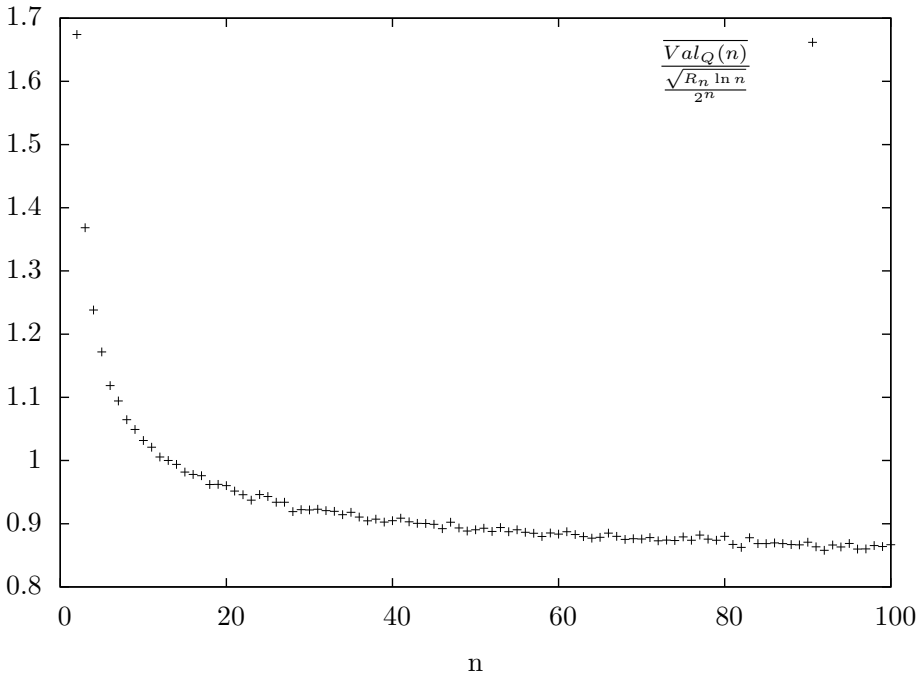


Figure 1 $\frac{\overline{Val_Q(n)}}{\frac{\sqrt{R_n \ln n}}{2^n}}$ for a random sample of n player games.

In this paper we have dealt with a small portion of non-local games. In particular, the case of random non-symmetric games is still open and there has been little progress in multiplayer XOR games with $m-ary$ input. The primary hurdle in the n-player $m-ary$ input setting is that at the moment it lacks a description in terms of algebraic and analytic expressions.

Recently Briët and Vidick have shown large quantum-classical gaps for some 3-player $m-ary$ input XOR games [4]. Despite being able to establish quantum-classical gaps for specific games, for a general 3-player $m-ary$ input XOR game calculating its the entangled value is difficult [8].

References

1 Andris Ambainis, Artūrs Bačkurs, Kaspars Balodis, Dmitrijs Kravčenko, Raitis Ozols, Juris Smotrovs, and Madars Virza. Quantum strategies are better than classical in almost any XOR game. In Artur Czumaj, Kurt Mehlhorn, Andrew Pitts, and Roger Wattenhofer, editors, *Automata, Languages, and Programming*, volume 7391 of *Lecture Notes in Computer Science*, pages 25–37. Springer Berlin Heidelberg, 2012.

2 Andris Ambainis, Jānis Iraids, Dmitry Kravchenko, and Madars Virza. Advantage of quantum strategies in random symmetric XOR games. In Antonín Kučera, Thomas A. Henzinger, Jaroslav Nešetřil, Tomáš Vojnar, and David Antoš, editors, *Mathematical and Engineering Methods in Computer Science*, volume 7721 of *Lecture Notes in Computer Science*, pages 57–68. Springer Berlin Heidelberg, 2013.

3 Andris Ambainis, Dmitry Kravchenko, Nikolajs Nahimovs, and Alexander Rivosh. Nonlocal quantum XOR games for large number of players. In Jan Kratochvíl, Angsheng Li, Jiří

Fiala, and Petr Kolman, editors, *Theory and Applications of Models of Computation*, volume 6108 of *Lecture Notes in Computer Science*, pages 72–83. Springer Berlin Heidelberg, 2010.

4 Jop Briët and Thomas Vidick. Explicit lower and upper bounds on the entangled value of multiplayer XOR games. *Communications in Mathematical Physics*, pages 1–27, 2012.

5 John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, Oct 1969.

6 R. Cleve, P. Høyer, B. Toner, and J. Watrous. Consequences and limits of nonlocal strategies. In *Proc. 19th IEEE Annual Conf. Computational Complexity*, pages 236–249, 2004.

7 R. Salem and A. Zygmund. Some properties of trigonometric series whose terms have random signs. *Acta Mathematica*, 91(1):245–301, 1954.

8 Thomas Vidick. Three-player entangled XOR games are NP-hard to approximate. *arXiv preprint arXiv:1302.1242*, 2013.

9 R. F. Werner and M. M. Wolf. All-multipartite Bell-correlation inequalities for two dichotomic observables per site. *Phys. Rev. A*, 64:032112, Aug 2001.

Towards Efficient Decoding of Classical-Quantum Polar Codes

Mark M. Wilde[1], Olivier Landon-Cardinal[2], and Patrick Hayden[1]

1 **School of Computer Science, McGill University**
3480 University Street, Montreal, Quebec H3A 2A7, Canada
mwilde@gmail.com; patrick@cs.mcgill.ca

2 **Département de Physique, Université de Sherbrooke**
Sherbrooke, Québec J1K 2R1, Canada
olivier.landon-cardinal@usherbrooke.ca

Abstract

Known strategies for sending bits at the capacity rate over a general channel with classical input and quantum output (a cq channel) require the decoder to implement impractically complicated collective measurements. Here, we show that a fully collective strategy is not necessary in order to recover all of the information bits. In fact, when coding for a large number N uses of a cq channel W, $N \cdot I(W_{\text{acc}})$ of the bits can be recovered by a non-collective strategy which amounts to coherent quantum processing of the results of product measurements, where $I(W_{\text{acc}})$ is the accessible information of the channel W. In order to decode the other $N(I(W) - I(W_{\text{acc}}))$ bits, where $I(W)$ is the Holevo rate, our conclusion is that the receiver should employ collective measurements. We also present two other results: 1) collective Fuchs-Caves measurements (quantum likelihood ratio measurements) can be used at the receiver to achieve the Holevo rate and 2) we give an explicit form of the Helstrom measurements used in small-size polar codes. The main approach used to demonstrate these results is a quantum extension of Arikan's polar codes.

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

Keywords and phrases classical-quantum channel, classical-quantum polar codes, quantum likelihood ratio, quantum successive cancellation decoder

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.157

1 Introduction

One of the most impressive recent developments in coding theory is the theory of polar codes [1]. These codes are provably capacity achieving, and their encoding and decoding complexities are both $O(N \log N)$, where N is the number of channel uses. Polar codes are based on the channel polarization effect, in which a recursive encoding induces a set of N synthesized channels from N instances of the original channel, such that some of the synthesized channels are nearly perfect and the others are nearly useless. The fraction of synthesized channels that is nearly perfect is equal to the capacity of the channel, and thus the coding scheme is simple: send the information bits through the synthesized channels that are nearly perfect.

An essential component of the polar coding scheme is Arikan's successive cancellation decoding algorithm [1]. This algorithm is channel dependent and operates as its name suggests: it decodes the information bits one after another, using previously decoded information to aid in constructing a test for decoding each bit in succession. In particular, the test for decoding

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 157–177


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

each information bit is a likelihood ratio test. Due to the structure in the polar encoder, there is a great deal of structure in the decoding tests, so much so that each likelihood ratio can be recursively computed. The upshot is that the complexity of the decoding algorithm is $O(N \log N)$.

Recently, there has been some effort in extending the theory of polar coding to the problem of transmission over quantum channels [23, 18, 26, 25]. In particular, these works developed the theory of polar coding for transmitting classical data over an arbitrary quantum channel [23], private classical data over an arbitrary quantum channel [25], quantum data over a quantum Pauli or erasure channel [18], and quantum data over an arbitrary quantum channel [26]. To prove that the polar coding schemes in Refs. [23, 26, 25] achieve communication rates equal to well-known formulas from quantum information theory, the authors of these works constructed a quantum successive cancellation decoder as a sequence of quantum hypothesis tests (in the spirit of Arikan [1]) and employed Sen's non-commutative union bound [20] in the error analysis. The major question left open from this effort is whether there exists an efficient implementation for a quantum successive cancellation decoder.[1,2]

In this paper, we detail our progress towards finding an efficient quantum successive cancellation decoder. The decoder outlined here is useful for decoding classical information transmitted over a channel with classical inputs and quantum outputs (known as a "classical-quantum channel" or "cq channel" for short). Since the schemes for private classical communication [25] and quantum communication [26] rely on the quantum successive cancellation decoder from Ref. [23], our results here have implications for these polar coding schemes as well. Our main result can be stated succintly as follows:
Claim: In order to achieve the symmetric Holevo capacity $I(W)$ of an arbitrary cq channel W, at most $N(I(W) - I(W_{\text{acc}}))$ of the bits require a fully collective strategy in order for them to be decoded reliably, while the other $N \cdot I(W_{\text{acc}})$ bits can be decoded efficiently and reliably in time $O(N^2)$ on a quantum computer using a product strategy that amounts to coherent quantum processing of the outcomes of product measurements.

Although the main result of this paper might be considered modest in light of reaching the full goal stated above, it still represents non-trivial progress beyond prior research and towards answering the efficient polar decoding question. Indeed, one might think that collective measurements would be necessary in order to recover any of the bits of a message when communicating at the Holevo capacity rate, as suggested by the original work of Holevo [15], Schumacher, and Westmoreland [19] and follow-up efforts on the pure-loss bosonic channel [6, 8]. Even the recent sequential decoding schemes suggest the same [7, 20] (see also [24] for the pure-loss bosonic case). As a side note, these sequential decoding schemes require a number of measurements exponential in the number of channel uses—thus, even though the physical realization of a single one of these measurements may be within experimental reach [17], the fact that these schemes require an exponential number of measurements

[1] By efficient, we mean that the decoder should run in $O(N^2)$ time on a quantum computer (or even better $O(N \log N)$). In computational complexity theory, "efficient" is often regarded to mean that an algorithm runs in time polynomial in the input length. However, for the demanding application of channel coding where delay should be minimized, we will consider a decoding algorithm to be "efficient" if it has a near-linear running time.

[2] Note that the scheme from Ref. [18] *does* provide an efficient $O(N \log N)$ implementation of a quantum successive cancellation decoder, essentially because sending classical states (encoded in some orthonormal basis) through a Pauli or erasure channel induces an effectively classical channel at the output (such that the resulting output states are commuting). One can then exploit a coherent version of Arikan's successive cancellation decoder to decode quantum information. Although this advance is useful, we would like to have an efficient decoder for an *arbitrary* quantum channel.

excludes them from ever being practical. The previous result in Ref. [23] suggests that only a linear number of collective measurements are required to achieve the Holevo rate, and our work here demonstrates that the number of collective measurements required is at most $N(I(W) - I(W_{\text{acc}}))$.

This paper contains other results of interest. First, we prove that collective Fuchs-Caves measurements (or quantum likelihood ratio measurements) [5] suffice for achieving the Holevo information rate with a cq polar coding scheme. It was already known from Ref. [23] that a sequence of Helstrom measurements suffices for achieving this rate, so this new result just adds to the ways in which one can achieve the Holevo rate of communication. We also plot the fraction of requisite collective measurements as a function of the mean photon number of the signaling states for the case of the pure-loss bosonic channel, in order to have a sense of the physical requirements necessary for high-rate communication over this channel. As one would expect, the fraction of collective measurements needed increases as the mean photon number of the signaling states decreases—we expect this to happen since the low photon-number regime is more quantum due to the non-orthogonality of the signaling states. Finally, we detail the explicit form of a polar decoder that uses Helstrom measurements—we do this for some simple two-, four-, and eight-bit polar codes. This final result should give a sense of how one can specify these tests for larger blocklength polar codes.

The paper is organized as follows. The next section reviews background material such as cq channels, the Holevo quantity, quantum fidelity, the accessible information, and the classical fidelity (Bhattacharya parameter). Section 3 reviews the Fuchs-Caves measurement from Ref. [5] and provides a useful upper bound on the error probability of a hypothesis test that employs this measurement as the decision rule. We review classical-quantum polar codes in Section 4.1. Our first simple observation is that collective Fuchs-Caves measurements suffice for achieving the Holevo rate of communication (Section 4.2). Our main result, a justification for Claim 1, appears in Section 4.3. In Section 5, we discuss the implications of Claim 1 for the pure-loss bosonic channel. Our last result on the explicit form of the Helstrom decoder for two-, four-, and eight-bit polar codes appears in Section 6. Finally, we conclude with a summary of our results and suggest that the Schur transform might be helpful in obtaining a general solution to the problem discussed in this paper.

2 Preliminaries

A classical-quantum channel (cq channel) has a classical input and a quantum output. In this work, we only consider cq channels with binary inputs, written as

$$W : x \to \rho_x, \tag{1}$$

where W labels the channel, the input $x \in \{0, 1\}$, and ρ_x is a density operator. The symmetric Holevo information of this channel is

$$I(W) \equiv H((\rho_0 + \rho_1)/2) - [H(\rho_0) + H(\rho_1)]/2, \tag{2}$$

where $H(\sigma) \equiv -\text{Tr}\{\sigma \log_2 \sigma\}$ is the von Neumann entropy. The symmetric Holevo information gives one way to characterize the quality of a cq channel for data transmission: it is equal to one if ρ_0 is orthogonal to ρ_1 and equal to zero if $\rho_0 = \rho_1$. The quantum fidelity $F(W)$ is another parameter that characterizes the quality of a cq channel:

$$F(W) \equiv F(\rho_0, \rho_1) \equiv \|\sqrt{\rho_0}\sqrt{\rho_1}\|_1, \tag{3}$$

where the trace norm $\|A\|_1$ of an operator A is defined as $\|A\|_1 \equiv \mathrm{Tr}\{\sqrt{A^\dagger A}\}$ [22, 16].[3] The quantum fidelity $F(W)$ is equal to one if $\rho_0 = \rho_1$ and equal to zero if ρ_0 is orthogonal to ρ_1. We have the following relationships between the symmetric Holevo information and the quantum fidelity:

$$I(W) \approx 1 \Leftrightarrow F(W) \approx 0, \tag{4}$$

$$I(W) \approx 0 \Leftrightarrow F(W) \approx 1, \tag{5}$$

which are made precise in Proposition 1 of Ref. [23].

From any cq channel, it is possible to induce a purely classical channel $p_{Y|X}(y|x)$ by having the receiver perform a quantum measurement at its output:

$$p_{Y|X}(y|x) \equiv \mathrm{Tr}\{\Lambda_y \rho_x\}, \tag{6}$$

where $\Lambda \equiv \{\Lambda_y\}$ is a positive operator-valued measure (POVM), a set of operators satisfying $\Lambda_y \geq 0$ and $\sum_y \Lambda_y = I$. Letting X be a uniform Bernoulli random variable and letting Y be the random variable corresponding to the outcome of the measurement, we can define the symmetric mutual information of the induced channel as

$$I(W, \Lambda) \equiv I(X;Y) \equiv H(X) + H(Y) - H(XY), \tag{7}$$

where H is the Shannon entropy of these random variables. The classical Bhattarcharya parameter is the statistical overlap between the resulting distributions:

$$Z(W, \Lambda) \equiv \sum_y \sqrt{p_{Y|X}(y|0)\ p_{Y|X}(y|1)}. \tag{8}$$

If one were to encode the conditional distribution $p_{Y|X}(y|x)$ along the diagonal of a matrix (so that it becomes a density operator), then it is clear that the symmetric Holevo information and fidelity of the resulting "cq channel" are equal to the symmetric mutual information and classical Bhattacharya parameter, respectively.

The symmetric accessible information is equal to the optimized symmetric mutual information:

$$I(W_{\mathrm{acc}}) \equiv \max_{\{\Lambda_y\}} I(W, \Lambda), \tag{9}$$

where the optimization is with respect to all POVMs $\Lambda = \{\Lambda_y\}$. As a consequence of the well-known Holevo bound, the symmetric Holevo information is an upper bound to the symmetric accessible information [14]:

$$I(W_{\mathrm{acc}}) \leq I(W). \tag{10}$$

3 The Fuchs-Caves Measurement

Rather than choosing a measurement to optimize the symmetric mutual information, one could also choose a measurement in such a way that it minimizes the statistical overlap

[3] Note that the quantum fidelity sometimes is defined as $\left\|\sqrt{\rho_0}\sqrt{\rho_1}\right\|_1^2$ in order for it to have the interpretation as a probability. We choose to remove the square in this work (as is often done) in order for it to reduce to the classical Bhattacharya parameter when the states are just probability distributions.

between the resulting distributions $p_{Y|X}(y|0)$ and $p_{Y|X}(y|1)$ [5]. We call such a measurement a "Fuchs-Caves" measurement since these authors proved that the minimum statistical overlap is equal to the quantum fidelity:

$$\min_{\{\Lambda_y\}} Z(W, \Lambda) = F(W). \tag{11}$$

Furthermore, they gave an explicit form for the measurement that achieves the minimum and interpreted it as a kind of "quantum likelihood ratio." Indeed, the measurement that achieves the minimum in (11) corresponds to a measurement in the eigenbasis of the following Hermitian operator:

$$\rho_0 \# \rho_1^{-1} \equiv \rho_1^{-1/2} \sqrt{\rho_1^{1/2} \rho_0 \rho_1^{1/2}} \rho_1^{-1/2}. \tag{12}$$

Diagonalizing $\rho_0 \# \rho_1^{-1}$ as

$$\rho_0 \# \rho_1^{-1} = \sum_y \lambda_y |y\rangle\langle y|, \tag{13}$$

Fuchs and Caves observed that the eigenvalues of $\rho_0 \# \rho_1^{-1}$ take the following form:

$$\lambda_y = \left(\frac{\langle y|\rho_0|y\rangle}{\langle y|\rho_1|y\rangle} \right)^{1/2}, \tag{14}$$

furthermore suggesting that this measurement is a good quantum analog of a likelihood ratio. In addition, Fuchs and Caves also observed that the operator

$$\rho_1 \# \rho_0^{-1} \equiv \rho_0^{-1/2} \sqrt{\rho_0^{1/2} \rho_1 \rho_0^{1/2}} \rho_0^{-1/2} \tag{15}$$

commutes with and is the inverse of $\rho_0 \# \rho_1^{-1}$. Thus, the eigenvectors of $\rho_1 \# \rho_0^{-1}$ are the same as those of $\rho_0 \# \rho_1^{-1}$ and its eigenvalues are the reciprocals of those of $\rho_0 \# \rho_1^{-1}$.

▶ **Lemma 1.** *When using the Fuchs-Caves measurement to distinguish ρ_0 from ρ_1, we have following upper bound on the probability of error $p_e(W)$ in terms of the quantum fidelity $F(W)$:*

$$p_e(W) \leq \tfrac{1}{2} F(W). \tag{16}$$

Proof. After performing the measurement specified by (13), the decision rule is as follows:

$$\text{decide } \rho_0 \text{ if } \lambda_y \geq 1, \tag{17}$$

$$\text{decide } \rho_1 \text{ if } \lambda_y < 1, \tag{18}$$

which corresponds to the projectors

$$\Pi_0 \equiv \sum_{y \,:\, \lambda_y \geq 1} |y\rangle\langle y|, \tag{19}$$

$$\Pi_1 = \sum_{y \,:\, \lambda_y < 1} |y\rangle\langle y|. \tag{20}$$

It is then easy to prove the bound in (16):

$$
\begin{aligned}
2\, p_e(W) &= \mathrm{Tr}\{\Pi_0 \rho_1\} + \mathrm{Tr}\{\Pi_1 \rho_0\} && (21)\\
&= \sum_{y \,:\, \lambda_y \geq 1} \langle y|\rho_1|y\rangle + \sum_{y \,:\, \lambda_y < 1} \langle y|\rho_0|y\rangle && (22)\\
&= \sum_{y \,:\, \lambda_y \geq 1} \langle y|\rho_1|y\rangle^{1/2} \langle y|\rho_1|y\rangle^{1/2} + \sum_{y \,:\, \lambda_y < 1} \langle y|\rho_0|y\rangle^{1/2} \langle y|\rho_0|y\rangle^{1/2} && (23)\\
&\leq \sum_{y \,:\, \lambda_y \geq 1} \langle y|\rho_1|y\rangle^{1/2} \langle y|\rho_0|y\rangle^{1/2} + \sum_{y \,:\, \lambda_y < 1} \langle y|\rho_0|y\rangle^{1/2} \langle y|\rho_1|y\rangle^{1/2} && (24)\\
&= \sum_{y} \langle y|\rho_1|y\rangle^{1/2} \langle y|\rho_0|y\rangle^{1/2} && (25)\\
&= F(\rho_0, \rho_1) && (26)
\end{aligned}
$$

where the last equality follows from (11). ◀

4 Decoding Classical-Quantum Polar Codes

4.1 Review

Ref. [23] demonstrated how to construct synthesized versions of W, by channel combining and splitting [1]. The synthesized channels $W_N^{(i)}$ are of the following form:

$$
\begin{aligned}
W_N^{(i)} &: u_i \to \rho_{(i),u_i}^{U_1^{i-1} B^N}, && (27)\\
\rho_{(i),u_i}^{U_1^{i-1} B^N} &\equiv \sum_{u_1^{i-1}} \frac{1}{2^{i-1}} \left|u_1^{i-1}\right\rangle\!\left\langle u_1^{i-1}\right|^{U_1^{i-1}} \otimes \overline{\rho}_{u_1^i}^{B^N}, && (28)\\
\overline{\rho}_{u_1^i}^{B^N} &\equiv \sum_{u_{i+1}^N} \frac{1}{2^{N-i}} \rho_{u^N G_N}^{B^N}, \qquad \rho_{x^N}^{B^N} \equiv \rho_{x_1}^{B_1} \otimes \cdots \otimes \rho_{x_N}^{B_N}, && (29)
\end{aligned}
$$

where G_N is Arikan's encoding circuit matrix built from classical CNOT and permutation gates. The registers labeled by U are classical registers containing the bits u_1 through u_{i-1}, and the registers labeled by B contain the channel outputs. If the channel is classical, then these states are diagonal in the computational basis, and the above states correspond to the distributions for the synthesized channels [1]. The interpretation of $W_N^{(i)}$ is that it is the channel "seen" by the input u_i if the previous bits u_1^{i-1} are available and if the future bits u_{i+1}^N are randomized. This motivates the development of a quantum successive cancellation decoder [23] that attempts to distinguish $u_i = 0$ from $u_i = 1$ by adaptively exploiting the results of previous measurements and quantum hypothesis tests for each bit decision.

The synthesized channels $W_N^{(i)}$ polarize, in the sense that some become nearly perfect for classical data transmission while others become nearly useless. To prove this result, one can model the channel splitting and combining process as a random birth process [1, 23], and then demonstrate that the induced random birth processes corresponding to the channel parameters $I(W_N^{(i)})$ and $F(W_N^{(i)})$ are martingales that converge almost surely to zero-one valued random variables in the limit of many recursions. The following theorem characterizes the rate with which the channel polarization effect takes hold [2, 23], and it is useful in proving statements about the performance of polar codes for cq channels:

▶ **Theorem 2.** *Given a binary input cq channel W and any $\beta < 1/2$, it holds that*

$$
\lim_{n \to \infty} \Pr_J\{F(W_{2^n}^{(J)}) < 2^{-2^{n\beta}}\} = I(W), \tag{30}
$$

where n indicates the level of recursion for the encoding, $W_{2^n}^{(J)}$ is a random variable characterizing the J^{th} split channel, and $F(W_{2^n}^{(J)})$ is the fidelity of that channel.

Assuming knowledge of the identities of the good and bad channels, one can then construct a coding scheme based on the channel polarization effect, by dividing the synthesized channels according to the following polar coding rule:

$$\mathcal{G}_N(W,\beta) \equiv \left\{i \in [N] : F(W_N^{(i)}) < 2^{-N^\beta}\right\}, \tag{31}$$

$$\mathcal{B}_N(W,\beta) \equiv [N] \setminus \mathcal{G}_N(W,\beta), \tag{32}$$

so that $\mathcal{G}_N(W,\beta)$ is the set of "good" channels and $\mathcal{B}_N(W,\beta)$ is the set of "bad" channels. The sender then transmits the information bits through the good channels and "frozen" bits through the bad ones. A helpful assumption for error analysis is that the frozen bits are chosen uniformly at random and known to both the sender and receiver.

One of the important advances in Ref. [23] was to establish that a quantum successive cancellation decoder performs well for polar coding over classical-quantum channels with equiprobable inputs. Corresponding to the split channels $W_N^{(i)}$ in (27) are the following projectors that attempt to decide whether the input of the i^{th} split channel is zero or one:

$$\Pi_{(i),0}^{U_1^{i-1}B^N} \equiv \left\{\rho_{(i),0}^{U_1^{i-1}B^N} - \rho_{(i),1}^{U_1^{i-1}B^N} \geq 0\right\}, \tag{33}$$

$$\Pi_{(i),1}^{U_1^{i-1}B^N} \equiv I - \Pi_{(i),0}^{U_1^{i-1}B^N}, \tag{34}$$

where $\{B \geq 0\}$ denotes the projector onto the positive eigenspace of a Hermitian operator B. After some calculations, one readily sees that

$$\Pi_{(i),0}^{U_1^{i-1}B^N} = \sum_{u_1^{i-1}} \left|u_1^{i-1}\right\rangle\!\left\langle u_1^{i-1}\right|^{U_1^{i-1}} \otimes \Pi_{(i),u_1^{i-1}0}^{B^N}, \tag{35}$$

where

$$\Pi_{(i),1}^{U_1^{i-1}B^N} = I - \Pi_{(i),0}^{U_1^{i-1}B^N}, \tag{36}$$

$$\Pi_{(i),u_1^{i-1}0}^{B^N} \equiv \{\overline{\rho}_{u_1^{i-1}0}^{B^N} - \overline{\rho}_{u_1^{i-1}1}^{B^N} \geq 0\}, \tag{37}$$

$$\Pi_{(i),u_1^{i-1}1}^{B^N} \equiv I - \Pi_{(i),u_1^{i-1}0}^{B^N}. \tag{38}$$

The observations above lead to a decoding rule for a successive cancellation decoder similar to Arikan's [1]:

$$\hat{u}_i = \begin{cases} u_i & \text{if } i \in \mathcal{A}^c \\ h\left(\hat{u}_1^{i-1}\right) & \text{if } i \in \mathcal{A} \end{cases}, \tag{39}$$

where $h\left(\hat{u}_1^{i-1}\right)$ is the outcome of the i^{th} collective measurement:

$$\{\Pi_{(i),\hat{u}_1^{i-1}0}^{B^N}, \Pi_{(i),\hat{u}_1^{i-1}1}^{B^N}\} \tag{40}$$

on the codeword received at the channel output (after $i-1$ measurements have already been performed). The set $\mathcal{A}$ labels the information bits. The measurement device outputs "0" if the outcome $\Pi_{(i),\hat{u}_1^{i-1}0}^{B^N}$ occurs and it outputs "1" otherwise. (Note that we can set $\Pi_{(i),\hat{u}_1^{i-1}u_i}^{B^N} = I$ if the bit u_i is a frozen bit.) The above sequence of measurements for the

whole bit stream u^N corresponds to a positive operator-valued measure (POVM) $\{\Lambda_{u^N}\}$ where

$$\Lambda_{u^N} \equiv \Pi^{B^N}_{(1),u_1} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(1),u_1}, \tag{41}$$

and $\sum_{u_{\mathcal{A}}} \Lambda_{u^N} = I^{B^N}$. The probability of error $P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c})$ for code length N, number K of information bits, set $\mathcal{A}$ of information bits, and choice $u_{\mathcal{A}^c}$ for the frozen bits is

$$P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}) = 1 - \frac{1}{2^K} \sum_{u_{\mathcal{A}}} \mathrm{Tr}\{\Lambda_{u^N} \rho_{u^N}\}, \tag{42}$$

where we are assuming a particular choice of the bits $u_{\mathcal{A}^c}$ in the sequence of projectors $\Pi^{B^N}_{(N),u_1^{N-1}u_N} \cdots \Pi^{B^N}_{(i),u_1^{i-1}u_i} \cdots \Pi^{B^N}_{(1),u_1}$ and setting $\Pi^{B^N}_{(i),u_1^{i-1}u_i} = I$ if u_i is a frozen bit. The formula also assumes that the sender transmits the information sequence $u_{\mathcal{A}}$ with uniform probability 2^{-K}. The probability of error averaged over all choices of the frozen bits is then

$$P_e(N, K, \mathcal{A}) = \frac{1}{2^{N-K}} \sum_{u_{\mathcal{A}^c}} P_e(N, K, \mathcal{A}, u_{\mathcal{A}^c}). \tag{43}$$

The following proposition from Ref. [23] determines an upper bound on the average ensemble performance of polar codes with a quantum successive cancellation decoder:
Proposition: For any classical-quantum channel W with binary inputs and quantum outputs and any choice of $(N, K, \mathcal{A})$, the following bound holds

$$P_e(N, K, \mathcal{A}) \leq 2 \sqrt{\sum_{i \in \mathcal{A}} \tfrac{1}{2} F(W_N^{(i)})}. \tag{44}$$

The proposition is proved by exploiting Sen's non-commutative union bound [20] and Lemma 3.2 of Ref. [11] (which upper bounds the probability of error in a binary quantum hypothesis test by the fidelity between the test states). The bound in (44) applies provided the sender chooses the information bits $U_{\mathcal{A}}$ from a uniform distribution. Thus, by choosing the channels over which the sender transmits the information bits to be in $\mathcal{A}$ and those over which she transmits agreed-upon frozen bits to be in $\mathcal{A}^c$, we obtain that the probability of decoding error satisfies $\Pr\{\widehat{U}_{\mathcal{A}} \neq U_{\mathcal{A}}\} = o(2^{-\frac{1}{2}N^\beta})$, as long as the code rate obeys $R = K/N < I(W)$.

A final point that will be useful is that Ref. [23] also proved that measurements consisting of the projections

$$\left\{ \sqrt{\rho^{U_1^{i-1}B^N}_{(i),0}} - \sqrt{\rho^{U_1^{i-1}B^N}_{(i),1}} \geq 0 \right\}, \tag{45}$$

rather than those in (33)-(34), also achieve the performance stated in Proposition 4.1.

4.2 Collective Fuchs-Caves Measurements Achieve the Holevo Rate

Our first observation is rather simple, just being that collective Fuchs-Caves measurements can also achieve the performance stated in Proposition 4.1. This result follows from Lemma 1's bound on the error probability of a Fuchs-Caves measurement and by performing an error analysis similar to that in the proof of Proposition 4 of Ref. [23] given in Section V of that paper. The explicit form of a Fuchs-Caves quantum successive cancellation decoder is given by projectors of the form in (35)-(38), with the Helstrom tests replaced by Fuchs-Caves projectors as given in (19)-(20).

This result also demonstrates that there are a variety of decoding measurements that one can exploit for achieving the Holevo information rate. However, the quantum successive cancellation decoder consisting of Helstrom measurements should outperform either the measurements in (45) or the Fuchs-Caves measurements when considering finite blocklength performance because the Helstrom measurement is the optimal test for distinguishing two quantum states.

4.3 Main Result

Our main observation is a bit more subtle than the above, but it is still elementary. Nevertheless, this observation has nontrivial consequences and represents a step beyond the insights in prior work regarding decoding of classical information sent over quantum channels [15, 19, 6, 8, 7, 20, 24, 23].

We begin by considering the "Fuchs-Caves" classical channel W_{FC} induced from W by performing the Fuchs-Caves measurement on every channel output:

$$W_{\text{FC}} : x \to p_{Y|X}(y|x) = \langle y|\rho_x|y\rangle, \tag{46}$$

where the orthonormal basis $\{|y\rangle\}$ is the same as that in (13). The specification of the polar code in the previous section specializes to this induced classical channel. The code consists of a set of "good" synthesized channels $\mathcal{G}_N(W_{\text{FC}}, \beta)$ and "bad" synthesized channels $\mathcal{B}_N(W_{\text{FC}}, \beta)$, where

$$\mathcal{G}_N(W_{\text{FC}}, \beta) \equiv \left\{i \in [N] : F(W^{(i)}_{\text{FC},N}) = Z(W^{(i)}_{\text{FC},N}) < 2^{-N^\beta}\right\}, \tag{47}$$

$$\mathcal{B}_N(W_{\text{FC}}, \beta) \equiv [N] \setminus \mathcal{G}_N(W_{\text{FC}}, \beta), \tag{48}$$

and the equality $F(W^{(i)}_{\text{FC},N}) = Z(W^{(i)}_{\text{FC},N})$ holds because the induced channels are classical. Furthermore, by Theorem 2, the number of good channels in the limit that N becomes large is as follows:

$$\lim_{N\to\infty} \frac{1}{N}|\mathcal{G}_N(W_{\text{FC}}, \beta)| = I(W_{\text{FC}}). \tag{49}$$

Finally, each bit of this classical polar code can be decoded in time $O(N)$ using a recursive calculation of likelihood ratios as given in (75)-(76) of Ref. [1].[4]

Now, our main observation is the following relationship between the good channels of W_{FC} and the good channels of W:

$$\mathcal{G}_N(W_{\text{FC}}, \beta) \subseteq \mathcal{G}_N(W, \beta). \tag{50}$$

This relationship holds because of the Fuchs-Caves formula from (11). For all i, we have that

$$F(W^{(i)}_N) = \min_{\{\Lambda_y\}} Z(W^{(i)}_N, \Lambda) \leq Z(W^{(i)}_{\text{FC},N}), \tag{51}$$

where the inequality follows because the tensor-product Fuchs-Caves measurement that induces the synthesized channel $W^{(i)}_{\text{FC},N}$ is a particular kind of measurement, and so its classical statistical overlap can only be larger than that realized by the optimal measurement

[4] Note that this is the "first decoding algorithm" of Arikan. A refinement implies that all of the bits can be decoded in time $O(N \log N)$, but the first decoding algorithm is what we will use in this work.

(which in general will be a collective measurement rather than a product measurement). Now, for all $i \in \mathcal{G}_N(W_{\text{FC}}, \beta)$, we have that

$$Z(W^{(i)}_{\text{FC},N}) < 2^{-N^\beta}. \tag{52}$$

This in turn implies that $F(W^{(i)}_N) < 2^{-N^\beta}$ by (51), and so for this i, we have that $i \in \mathcal{G}_N(W, \beta)$ and can conclude (50).

This observation has non-trivial implications for the structure of the polar decoder. For all of the bits in $\mathcal{G}_N(W_{\text{FC}}, \beta)$, the receiver can decode them with what amounts to an effectively "product" or "non-collective" strategy,[5] while for the bits in $\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\text{FC}}, \beta)$, we still require collective measurements in order for the receiver to decode them with the error probability guarantee given by (31). However, when decoding the bits in $\mathcal{G}_N(W_{\text{FC}}, \beta)$, the receiver should be careful to decode them in the least destructive way possible so that Sen's non-commutative union bound is still applicable and we obtain the overall error bound guaranteed by Proposition 4.1. In particular, the decoder should begin by performing an isometric extension of the Fuchs-Caves measurement on each channel output:

$$\sum_y |y\rangle\langle y| \otimes |\lambda_y\rangle, \tag{53}$$

where the orthonormal basis $\{|y\rangle\}$ is from the eigendecomposition in (13) and the basis $\{|\lambda_y\rangle\}$ encodes the eigenvalues to some finite precision. Such an operation coherently copies the likelihood ratios λ_y of the Fuchs-Caves measurement into an ancillary register. The receiver then performs a reversible implementation of Arikan's decoding algorithm to process these likelihood ratios according to (75)-(76) of Ref. [1]. Finally, the receiver coherently copies the value of a single decision qubit with a CNOT gate to an ancillary register, measures the decision qubit, and "uncomputes" these operations by performing the inverse of the Arikan circuit and the inverse of the operations in (53). Figure 1 depicts these operations. The effect of these operations is to implement a projection of the channel output onto a subspace spanned by eigenvectors $|y^N\rangle = |y_1\rangle \otimes \cdots \otimes |y_N\rangle$ of the Fuchs-Caves measurements such that

$$W^{(i)}_{\text{FC},N}\big(y^N, u_1^{i-1}|0\big) \geq W^{(i)}_{\text{FC},N}\big(y^N, u_1^{i-1}|1\big), \tag{54}$$

or onto the complementary subspace spanned by eigenvectors $|y^N\rangle$ such that

$$W^{(i)}_{\text{FC},N}\big(y^N, u_1^{i-1}|0\big) < W^{(i)}_{\text{FC},N}\big(y^N, u_1^{i-1}|1\big), \tag{55}$$

where y^N is the classical output of the Fuchs-Caves channel and u_1^{i-1} denotes the previously decoded bits. Thus, the fidelity bound from (52) is applicable and Sen's non-commutative union bound guarantees that the overall contribution of the error in decoding bit $i \in \mathcal{G}_N(W_{\text{FC}}, \beta)$ is no larger than 2^{-N^β}. The time that it takes to process each bit $i \in \mathcal{G}_N(W_{\text{FC}}, \beta)$ is $O(N)$, which is clear from the structure of the circuit and Arikan's "first decoding algorithm."

For all of the remaining bits $i \in \mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\text{FC}}, \beta)$, we still do not know whether there exists an efficient quantum algorithm for decoding them while having the error probability from Proposition 4.1. Thus, for now, we simply suggest for the receiver to use collective measurements to recover them.

[5] If a decoding strategy amounts to coherent implementations of product measurements followed by coherent processing of the outcomes, we still say that it is a product strategy rather than collective.

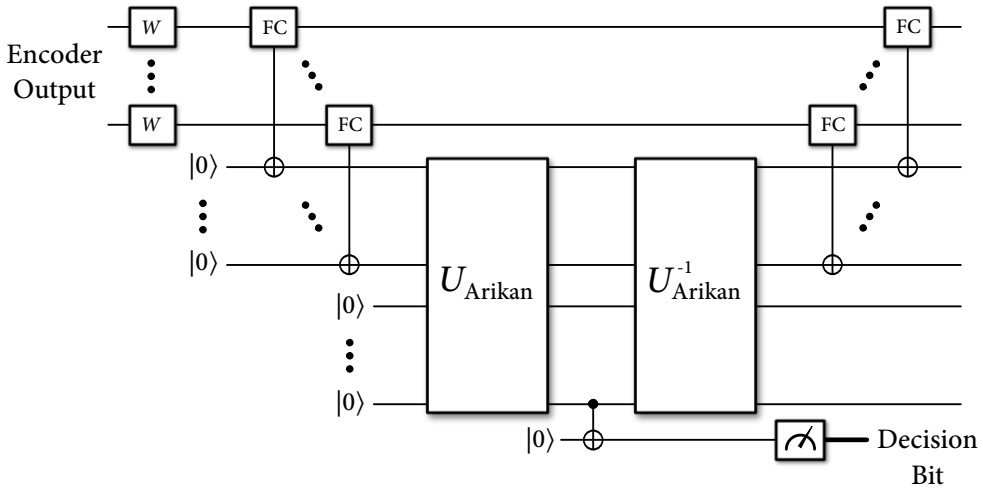


Figure 1 The circuit for recovering an information bit in the set $\mathcal{G}_N(W_{\mathrm{FC}}, \beta)$. The encoder output is fed into N instances of the channel W. The receiver acts with N of the unitaries in (53), labeled as "FC" boxes which coherently copy the likelihood ratios $\lambda_{y_1}, \ldots, \lambda_{y_N}$ into ancillary registers. The receiver then acts with a reversible implementation of Arikan's likelihood ratio computations, copies the decision bit into an ancillary register, and measures the decision bit to decode the i^{th} bit. The receiver finally performs the inverse of these operations to "clean up," i.e., to ensure that the next measurement can be performed, whether it be to decode a bit in the set $\mathcal{G}_N(W_{\mathrm{FC}}, \beta)$ or the set $\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$. The effect of this circuit is to perform the desired "gentle projection."

It should be clear from Proposition 2 and (49) that the size of the set $\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\mathrm{FC}}, \beta)$ in the limit is equal to

$$\lim_{N\to\infty} \frac{1}{N} |\mathcal{G}_N(W, \beta) \setminus \mathcal{G}_N(W_{\mathrm{FC}}, \beta)| = I(W) - I(W_{\mathrm{FC}}). \tag{56}$$

This makes it clear that one does not require a collective strategy in order to recover all of the information bits, but a collective strategy is only required in order to bridge the gap between $I(W_{\mathrm{FC}})$ and $I(W)$.

Observe also that similar reasoning applies to any product measurement, not just the Fuchs-Caves measurements (we focused on the Fuchs-Caves measurement due to its strong analogy with a likelihood ratio and because Arikan's decoding algorithm processes likelihood ratios). With this in mind, we could simply choose the product measurement to be the one that maximizes the accessible information, in order to maximize the number of bits that can be processed efficiently. Let W_{acc} be the classical channel induced by performing the measurement that maximizes the accessible information. One would then process the bits in $\mathcal{G}_N(W_{\mathrm{acc}}, \beta)$ in a way very similar as described above. All of the observations above then justify Claim 1.

The reasoning also leads to a generalization of Lemma 1 that applies when using Fuchs-Caves measurements to distinguish a tensor-product state $\rho_0^{\otimes N}$ from $\rho_1^{\otimes N}$. The test consists of performing product measurements followed by classical post-processing. If one wishes to perform this test in the most delicate way possible, one could perform it as in Figure 1.

▶ **Lemma 3.** *When using product Fuchs-Caves measurements to distinguish $\rho_0^{\otimes N}$ from $\rho_1^{\otimes N}$, the probability of error p_e is bounded from above in terms of the quantum fidelity $F(\rho_0, \rho_1)$:*

$$p_e \leq \frac{1}{2}[F(\rho_0, \rho_1)]^N. \tag{57}$$

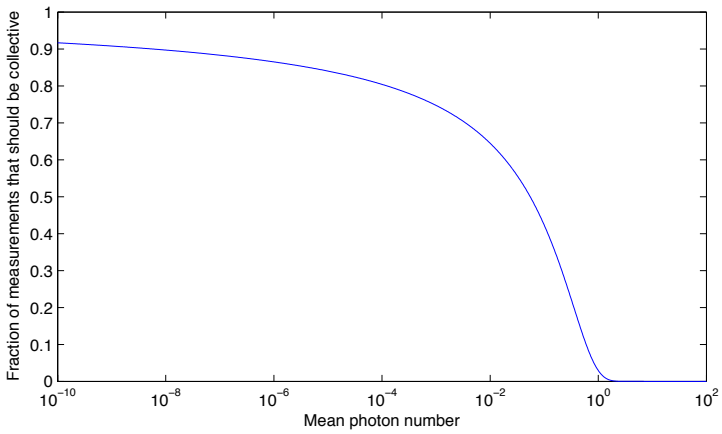


Figure 2 The fraction of collective measurements required for a polar decoder plotted as a function of the mean photon number E at the receiving end, when using a BPSK coding strategy.

Proof. The proof is very similar to the proof of Lemma 1. The test, though, consists of performing individual Fuchs-Caves measurements on the N systems, and these tests result in likelihood ratios $\lambda_{y_1}, \ldots, \lambda_{y_N}$. The decision rule is then as follows:

$$\text{decide } \rho_0^{\otimes N} \text{ if } \lambda_{y_1} \times \cdots \times \lambda_{y_N} \geq 1, \tag{58}$$

$$\text{decide } \rho_1^{\otimes N} \text{ if } \lambda_{y_1} \times \cdots \times \lambda_{y_N} < 1. \tag{59}$$

An analysis proceeding exactly as in (21)-(26) leads to the following bound:

$$\begin{aligned} 2\, p_e(W) &\leq \sum_{y_1,\ldots,y_N} \left[\langle y_1| \cdots \langle y_N|\rho_1^{\otimes N}|y_1\rangle \cdots |y_N\rangle\right]^{1/2} \left[\langle y_1| \cdots \langle y_N|\rho_0^{\otimes N}|y_1\rangle \cdots |y_N\rangle\right]^{1/2} \\ &= \sum_{y_1,\ldots,y_N} \langle y_1|\rho_1|y_1\rangle^{1/2} \cdots \langle y_N|\rho_1|y_N\rangle^{1/2} \langle y_1|\rho_0|y_1\rangle^{1/2} \cdots \langle y_N|\rho_0|y_N\rangle^{1/2} \qquad (60) \\ &= \sum_{y_1} \langle y_1|\rho_1|y_1\rangle^{1/2}\langle y_1|\rho_0|y_1\rangle^{1/2} \cdots \sum_{y_N} \langle y_N|\rho_1|y_N\rangle^{1/2}\langle y_N|\rho_0|y_N\rangle^{1/2} \qquad (61) \\ &= [F(\rho_0,\rho_1)]^N. \qquad (62) \end{aligned}$$

Furthermore, one can implement this test efficiently and non-destructively on a quantum computer as described in Figure 1. The result is to project onto two different subspaces: the one spanned by eigenvectors whose corresponding eigenvalues satisfy (58) and the other. ◀

5 Decoding the Pure-Loss Bosonic Channel

A channel of particular practical interest is the pure-loss bosonic channel. A simple physical model for this channel is a beamsplitter of transmissivity $\eta \in [0,1]$, where the sender has access to one input port, the environment injects the vacuum state into the other input port, the receiver has access to one output port, and the environment obtains the other output port. It is well known that the Holevo capacity of this channel is equal to $g(\eta N_S) \equiv (\eta N_S + 1)\log(\eta N_S + 1) - \eta N_S \log(\eta N_S)$ [6], where N_S is the mean input photon number. In the low-photon number regime, one can come very close to achieving the capacity by employing a binary phase-shift keying (BPSK) strategy (using coherent states

$|\alpha\rangle$ and $|-\alpha\rangle$ as the signaling states) [21]. The BPSK strategy induces a cq channel of the following form: $x \to |(-1)^x\alpha\rangle\langle(-1)^x\alpha|$. The symmetric Holevo rate for this channel is equal to $\chi(E) \equiv h_2\big(\big[1+e^{-2E}\big]/2\big)$, where h_2 is the binary entropy and $E \equiv \eta N_S$. If the receiver performs a Helstrom measurement at every channel output, this induces a classical channel with symmetric mutual information equal to $I_{\text{Hel}}(E) \equiv 1 - h_2([1-\sqrt{1-e^{-4E}}]/2)$. (See Ref. [9], for example, for explicit calculations.) Our results in the previous section demonstrate that the fraction of information bits required to be decoded using a collective strategy is equal to $1 - I_{\text{Hel}}(E)/\chi(E)$. Figure 2 reveals that this fraction is rather small for mean photon number (MPN) larger than one, but then it rises sharply as we enter a quantum regime where the MPN is less than one. Even deep in the quantum regime at a MPN of 10^{-8}, however, roughly 10% of the bits do not require collective decoding.

6 Small Blocklength Polar Decoders

This section briefly discusses how the Helstrom measurements [12, 13] in the quantum successive cancellation decoder from Ref. [23] decompose for very small size polar codes.

6.1 Two-Bit Polar Decoder

We begin by considering the simple two-bit polar code. The channel is of the form $x \to \rho_x$, where $x \in \{0,1\}$ and ρ_x is some conditional density operator. The two-bit polar code performs the simple transformation on the input bits u_1 and u_2:

$$(u_1, u_2) \to (u_1 + u_2, u_2), \tag{63}$$

where addition is modulo 2.

The first step of the successive cancellation decoder is to recover u_1, assuming that bit u_2 is chosen uniformly at random. The optimal measurement is a Helstrom measurement, and in this case, it amounts to distinguishing between the following two states

$$\frac{1}{2}\sum_{u_2} \rho_{u_2} \otimes \rho_{u_2}, \qquad \frac{1}{2}\sum_{u_2} \rho_{u_2+1} \otimes \rho_{u_2}. \tag{64}$$

The Helstrom measurement is given by the projector onto the positive eigenspace of the difference of the two density operators above:

$$\left\{\frac{1}{2}\sum_{u_2} \rho_{u_2} \otimes \rho_{u_2} - \frac{1}{2}\sum_{u_2} \rho_{u_2+1} \otimes \rho_{u_2} \geq 0\right\} = \left\{\sum_{u_2} (\rho_{u_2} - \rho_{u_2+1}) \otimes \rho_{u_2} \geq 0\right\} \tag{65}$$

$$= \left\{\sum_{u_2} (-1)^{u_2} (\rho_0 - \rho_1) \otimes \rho_{u_2} \geq 0\right\} \tag{66}$$

$$= \left\{(\rho_0 - \rho_1) \otimes \sum_{u_2} (-1)^{u_2} \rho_{u_2} \geq 0\right\} \tag{67}$$

$$= \{(\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \geq 0\}. \tag{68}$$

Thus, this test factorizes into the parity of the individual quantum hypothesis tests $\{(\rho_0 - \rho_1) \geq 0\}$. That is, supposing that $\Pi_+ \equiv \{(\rho_0 - \rho_1) \geq 0\}$ and $\Pi_- \equiv \{(\rho_0 - \rho_1) < 0\}$, one can write the two-bit test as the product of two controlled gates

$$U_1 \equiv I_{B_1} \otimes (\Pi_+)_{B_2} \otimes I_A + I_{B_1} \otimes (\Pi_-)_{B_2} \otimes (\sigma_X)_A, \tag{69}$$

$$U_2 \equiv (\Pi_+)_{B_1} \otimes I_{B_2} \otimes I_A + (\Pi_-)_{B_1} \otimes I_{B_2} \otimes (\sigma_X)_A, \tag{70}$$

where B_1 is the first channel output, B_2 is the second channel output, and A is an ancillary system initialized to the state $|0\rangle$. The product of these two unitary gates is equal to

$$U_1 U_2 = ((\Pi_+)_{B_1} \otimes (\Pi_+)_{B_2} + (\Pi_-)_{B_1} \otimes (\Pi_-)_{B_2}) \otimes I_A + ((\Pi_-)_{B_1} \otimes (\Pi_+)_{B_2} + (\Pi_+)_{B_1} \otimes (\Pi_-)_{B_2}) \otimes (\sigma_X)_A. \quad (71)$$

The receiver would then measure the ancillary system A in order to make a decision about u_1.

Next, we determine the decoding of u_2, given that u_1 has already been decoded. By the definition of the polar encoder transformation in (63), the goal is to distinguish between the following two states:

$$\rho_{u_1} \otimes \rho_0, \qquad \rho_{u_1+1} \otimes \rho_1. \quad (72)$$

The optimal quantum hypothesis test is given by the following projector:

$$\{\rho_{u_1} \otimes \rho_0 - \rho_{u_1+1} \otimes \rho_1 \geq 0\}. \quad (73)$$

This optimal quantum hypothesis test is not factorizable into smaller tests, and indeed, it is necessary to perform a collective measurement in order to implement it. Nonetheless, Lemma 3 provides a simple implementation of the Fuchs-Caves measurement for distinguishing these two states.

6.2 Four-Bit Polar Decoder

We now consider the form of Helstrom measurements for a four-bit polar code. Recall that the input transformation for the four-bit polar code is as follows:

$$(u_1, u_2, u_3, u_4) \to (u_1 + u_2 + u_3 + u_4, u_3 + u_4, u_2 + u_4, u_4). \quad (74)$$

It is straightforward to find the form of the four different tests for decoding u_1 through u_4. (See the appendix for derivations.) The test for decoding u_1 is again a parity test:

$$\left\{(\rho_0 - \rho_1)^{\otimes 4} \geq 0\right\}. \quad (75)$$

The test for decoding u_2 given u_1 is

$$\left\{ \left(\sum_{u_3'} \rho_{u_1+u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{u_4} \otimes \rho_{u_4} \right) - \left(\sum_{u_3'} \rho_{u_1+1+u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{1+u_4} \otimes \rho_{u_4} \right) \geq 0 \right\}. \quad (76)$$

It remains unclear to us if there is a simple way to decompose the above test any further into non-collective actions (or even approximately using, e.g., the Fuchs-Caves measurement). The test for decoding u_3 given u_2 and u_1 is

$$\{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes (\rho_{u_2} \otimes \rho_0 - \rho_{u_2+1} \otimes \rho_1) \geq 0\}. \quad (77)$$

One could actually approximate this test "efficiently" by performing a product Fuchs-Caves measurement of the first two systems, a product Fuchs-Caves measurement of the last two, and then take the parity of the results of these two tests (of course implementing these tests coherently). The final Helstrom test for decoding u_4 given u_3, u_2, and u_1 is

$$\{\rho_{u_1+u_2+u_3} \otimes \rho_{u_3} \otimes \rho_{u_2} \otimes \rho_0 - \rho_{u_1+u_2+u_3+1} \otimes \rho_{u_3+1} \otimes \rho_{u_2+1} \otimes \rho_1 \geq 0\}. \quad (78)$$

Clearly, it would be better to perform this last test by processing the likelihood ratios resulting from individual Fuchs-Caves measurements, rather than performing the optimal collective Helstrom measurement.

6.3 Polar Decoder for Larger Blocklengths

One can continue in the above fashion to determine the form of a quantum successive cancellation decoder that recovers each bit of an eight-bit polar code. We again try to simplify each Helstrom measurement and provide an expression for each one in Appendix B. A few tests simplify, in particular those used to recover the first bit u_1 (Eq. (101)), the fifth bit u_5 (Eq. (108), the seventh bit u_7 (Eq. (111)), and the last bit u_8 (Eq. (113)). However, for the other tests, it is unclear if they can be approximated by some combination of Helstrom and Fuchs-Caves measurements, followed by coherent post-processing.

From considering the eight-bit polar decoder, we can make several observations. For any blocklength, it is always possible to recover the first bit efficiently by calculating the parity of individual Helstrom measurements (though, this bit is always the "worst" bit, so the receiver would never actually be decoding it in practice). The receiver can always recover the last bit by performing a Fuchs-Caves measurement (this is always the "best" bit, so this should already be evident from the main observation in this paper). Furthermore, there are many bits that can be recovered by first performing Fuchs-Caves measurements, followed by the parity of these tests. Unfortunately, the fraction of these tests tends to zero in the limit of large blocklength. Thus, there still remains much to understand regarding the structure of a polar decoder.

7 Conclusion

The main result of this paper is an advance over previous schemes for decoding classical information transmitted over channels with classical inputs and quantum outputs. In particular, we have shown that $N \cdot I(W_{\text{acc}})$ of the information bits can be decoded reliably and efficiently on a quantum computer by a "non-collective" coherent decoding strategy, while closing the gap to the Holevo information rate (decoding the other $N(I(W) - I(W_{\text{acc}}))$ bits) should require a collective strategy. For the pure-loss bosonic channel, this implies that the majority of the bits transmitted can be decoded by a product strategy whenever the mean photon number is larger than one, while the fraction of collective measurements required increases sharply as the mean photon number decreases below one, marking the beginning of the quantum regime. Remarkably, even at mean photon numbers as low as 10^{-8}, roughly 10% of the bits do not require collective decoding, however. As another contribution, we have shown that a receiver can also employ collective Fuchs-Caves measurements when decoding a classical-quantum polar code. Finally, we gave the explicit form of the Helstrom measurements of a quantum successive cancellation decoder for two-, four-, and eight-bit polar codes. This should be helpful in determining the explicit form of tests for larger blocklength polar codes.

The main open question is still to determine whether all of the information bits can be efficiently decoded on a quantum computer. To answer this question, one might consider employing the Schur transform [3, 10, 4] and exploiting the structure inherent in polar codes. Unfortunately, it is not clear to us that this approach will lead to a quantum successive cancellation decoder with time complexity $O(N \log N)$ because the complexity of the Schur transform is higher than this.

We acknowledge helpful discussions with Frédéric Dupuis, Saikat Guha, Hari Krovi, David Poulin, and Joseph Renes. MMW acknowledges support from Montreal's Centre de Recherches Mathématiques. OLC aknowledges support from NSERC through a Vanier scholarship. PH acknowledges support from the Canada Research Chairs program, the Perimeter Institute, CIFAR, FQRNT's INTRIQ, NSERC, and ONR through grant N000140811249.

References

1 Erdal Arikan. Channel polarization: A method for constructing capacity-achieving codes for symmetric binary-input memoryless channels. *IEEE Transactions on Information Theory*, 55(7):3051–3073, July 2009.

2 Erdal Arikan and Emre Telatar. On the rate of channel polarization. In *Proceedings of the 2009 International Symposium on Information Theory*, pages 1493–1495, Seoul, Korea, June 2009. arXiv:0807.3806.

3 Robin Blume-Kohout, Sarah Croke, and Michael Zwolak. Ideal state discrimination with an O(1)-qubit quantum computer. arXiv:1201.6625.

4 Matthias Christandl. *The Structure of Bipartite Quantum States - Insights from Group Theory and Cryptography*. PhD thesis, University of Cambridge, April 2006. arXiv:quant-ph/0604183.

5 Christopher A. Fuchs and Carlton M. Caves. Mathematical techniques for quantum communication theory. *Open Systems & Information Dynamics*, 3(3):345–356, 1995. arXiv:quant-ph/9604001.

6 Vittorio Giovannetti, Saikat Guha, Seth Lloyd, Lorenzo Maccone, Jeffrey H. Shapiro, and Horace P. Yuen. Classical capacity of the lossy bosonic channel: The exact solution. *Physical Review Letters*, 92(2):027902, January 2004.

7 Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Achieving the Holevo bound via sequential measurements. *Physical Review A*, 85:012302, January 2012. arXiv:1012.0386.

8 Saikat Guha. Structured optical receivers to attain superadditive capacity and the holevo limit. *Physical Review Letters*, 106:240502, June 2011. arXiv:1101.1550.

9 Saikat Guha and Mark M. Wilde. Polar coding to achieve the holevo capacity of a pure-loss optical channel. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 546–550, Boston, Massachusetts, USA, 2012. arXiv:1202.0533.

10 Aram W. Harrow. *Applications of coherent classical communication and the Schur transform to quantum information theory*. PhD thesis, Massachusetts Institute of Technology, September 2005. arXiv:quant-ph/0512255.

11 Masahito Hayashi. *Quantum Information: An Introduction*. Springer-Verlag, Berlin Heidelberg, 2006.

12 Carl W. Helstrom. Quantum detection and estimation theory. *Journal of Statistical Physics*, 1:231–252, 1969.

13 Carl W. Helstrom. *Quantum Detection and Estimation Theory*. Academic, New York, 1976.

14 Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, 9:177–183, 1973.

15 Alexander S. Holevo. The capacity of the quantum channel with general signal states. *IEEE Transactions on Information Theory*, 44(1):269–273, January 1998. arXiv:quant-ph/9611023.

16 Richard Jozsa. Fidelity for mixed quantum states. *Journal of Modern Optics*, 41(12):2315–2323, 1994.

17 Daniel K. L. Oi, Vaclav Potocek, and John Jeffers. Measuring nothing. July 2012. arXiv:1207.3011.

18 Joseph M. Renes, Frédéric Dupuis, and Renato Renner. Efficient polar coding of quantum information. *Physical Review Letters*, 109:050504, August 2012. arXiv:1109.3195.

19 Benjamin Schumacher and Michael D. Westmoreland. Sending classical information via noisy quantum channels. *Physical Review A*, 56(1):131–138, July 1997.

20 Pranab Sen. Achieving the Han-Kobayashi inner bound for the quantum interference channel by sequential decoding. September 2011. arXiv:1109.0802.

21 Masaki Sohma and Osamu Hirota. Binary discretization for quantum continuous channels. *Physical Review A*, 62:052312, October 2000.

22 Armin Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.

23 Mark M. Wilde and Saikat Guha. Polar codes for classical-quantum channels. *IEEE Transactions on Information Theory*, 59(2):1175–1187, February 2013. arXiv:1109.2591.

24 Mark M. Wilde, Saikat Guha, Si-Hui Tan, and Seth Lloyd. Explicit capacity-achieving receivers for optical communication and quantum reading. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 551–555, Boston, Massachusetts, USA, July 2012. arXiv:1202.0518.

25 Mark M. Wilde and Joseph M. Renes. Polar codes for private classical communication. In *Proceedings of the 2012 International Symposium on Information Theory and its Applications*, Honolulu, Hawaii, USA, October 2012. arXiv:1203.5794.

26 Mark M. Wilde and Joseph M. Renes. Quantum polar codes for arbitrary channels. In *Proceedings of the 2012 International Symposium on Information Theory*, pages 334–338, Boston, Massachusetts, USA, July 2012. arXiv:1201.2906.

A Derivations for the Four-Bit Polar Decoder Measurements

The four-bit polar encoder amounts to the following transformation:

$$(u_1, u_2, u_3, u_4) \to (u_1 + u_2 + u_3 + u_4, u_3 + u_4, u_2 + u_4, u_4). \tag{79}$$

A.1 Recovering u_1

Let us first determine how the quantum successive cancellation decoder (QSCD) recovers the bit u_1, assuming that u_2, u_3, and u_4 are chosen uniformly at random. The test aims to distinguish between the following two states:

$$\frac{1}{2^3} \sum_{u_2,u_3,u_4} \rho_{u_2+u_3+u_4} \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4}, \tag{80}$$

$$\frac{1}{2^3} \sum_{u_2,u_3,u_4} \rho_{u_2+u_3+u_4+1} \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4}, \tag{81}$$

and it performs the following projection:

$$\left\{ \sum_{u_2,u_3,u_4} (\rho_{u_2+u_3+u_4} - \rho_{u_2+u_3+u_4+1}) \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0 \right\}$$

$$= \left\{ \sum_{u_2,u_3,u_4} (-1)^{u_2+u_3+u_4} (\rho_0 - \rho_1) \otimes \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0 \right\} \tag{82}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2,u_3,u_4} (-1)^{u_2+u_3+u_4} \rho_{u_3+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0 \right\} \tag{83}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2,u_3,u_4} (-1)^{u_3+u_4} \rho_{u_3+u_4} \otimes (-1)^{u_2+u_4} \rho_{u_2+u_4} \otimes (-1)^{u_4} \rho_{u_4} \geq 0 \right\} \tag{84}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2', u_3', u_4'} (-1)^{u_2'} \rho_{u_2'} \otimes (-1)^{u_3'} \rho_{u_3'} \otimes (-1)^{u_4'} \rho_{u_4'} \geq 0 \right\} \tag{85}$$

$$= \left\{ (\rho_0 - \rho_1) \otimes \sum_{u_2'} (-1)^{u_2'} \rho_{u_2'} \otimes \sum_{u_3'} (-1)^{u_3'} \rho_{u_3'} \otimes \sum_{u_4'} (-1)^{u_4'} \rho_{u_4'} \geq 0 \right\} \tag{86}$$

$$= \{ (\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \otimes (\rho_0 - \rho_1) \geq 0 \}. \tag{87}$$

Thus, this first test nicely factors as the parity of the four individual tests $\{(\rho_0 - \rho_1) \geq 0\}$.

A.2 Recovering u_2 given u_1

We now determine how the quantum successive cancellation decoder recovers u_2 given u_1, while randomizing over u_3 and u_4. The aim is to distinguish between the following two states:

$$\frac{1}{2^2} \sum_{u_3, u_4} \rho_{u_1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{u_4} \otimes \rho_{u_4}, \tag{88}$$

$$\frac{1}{2^2} \sum_{u_3, u_4} \rho_{u_1 + 1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{1 + u_4} \otimes \rho_{u_4}, \tag{89}$$

which translates to a projection of the following form:

$$\left\{ \sum_{u_3, u_4} \rho_{u_1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{u_4} \otimes \rho_{u_4} - \rho_{u_1 + 1 + u_3 + u_4} \otimes \rho_{u_3 + u_4} \otimes \rho_{1 + u_4} \otimes \rho_{u_4} \geq 0 \right\}. \tag{90}$$

Define $u_3' = u_3 + u_4$ and the above becomes

$$\left\{ \sum_{u_3', u_4} \rho_{u_1 + u_3'} \otimes \rho_{u_3'} \otimes \rho_{u_4} \otimes \rho_{u_4} - \rho_{u_1 + 1 + u_3'} \otimes \rho_{u_3'} \otimes \rho_{1 + u_4} \otimes \rho_{u_4} \geq 0 \right\}$$

$$= \left\{ \left(\sum_{u_3'} \rho_{u_1 + u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{u_4} \otimes \rho_{u_4} \right) \right.$$

$$\left. - \left(\sum_{u_3'} \rho_{u_1 + 1 + u_3'} \otimes \rho_{u_3'} \right) \otimes \left(\sum_{u_4} \rho_{1 + u_4} \otimes \rho_{u_4} \right) \geq 0 \right\}. \tag{91}$$

A.3 Recovering u_3 given u_2 and u_1

Let us determine how the QSCD recovers u_3 given u_2 and u_1, while randomizing over u_4. The test distinguishes between the following two states:

$$\frac{1}{2} \sum_{u_4} \rho_{u_1 + u_2 + u_4} \otimes \rho_{u_4} \otimes \rho_{u_2 + u_4} \otimes \rho_{u_4}, \tag{92}$$

$$\frac{1}{2} \sum_{u_4} \rho_{u_1 + u_2 + 1 + u_4} \otimes \rho_{1 + u_4} \otimes \rho_{u_2 + u_4} \otimes \rho_{u_4}, \tag{93}$$

and amounts to a projector of the following form:

$$\left\{\sum_{u_4} \rho_{u_1+u_2+u_4} \otimes \rho_{u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} - \sum_{u_4} \rho_{u_1+u_2+1+u_4} \otimes \rho_{1+u_4} \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\}$$

$$= \left\{\sum_{u_4} (\rho_{u_1+u_2+u_4} \otimes \rho_{u_4} - \rho_{u_1+u_2+1+u_4} \otimes \rho_{1+u_4}) \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\} \tag{94}$$

$$= \left\{\sum_{u_4} (-1)^{u_4} (\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\} \tag{95}$$

$$= \left\{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes \sum_{u_4} (-1)^{u_4} \rho_{u_2+u_4} \otimes \rho_{u_4} \geq 0\right\} \tag{96}$$

$$= \{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \otimes (\rho_{u_2} \otimes \rho_0 - \rho_{u_2+1} \otimes \rho_1) \geq 0\}. \tag{97}$$

Thus, this test nicely factorizes as the parity of two tests $\{(\rho_{u_1+u_2} \otimes \rho_0 - \rho_{u_1+u_2+1} \otimes \rho_1) \geq 0\}$ and $\{(\rho_{u_2} \otimes \rho_0 - \rho_{u_2+1} \otimes \rho_1) \geq 0\}$.

A.4 Recovering u_4 given u_3, u_2, and u_1

Finally, we determine how the QSCD recovers u_4 given all of the previous bits. The test in this case just aims to distinguish the following states:

$$\rho_{u_1+u_2+u_3} \otimes \rho_{u_3} \otimes \rho_{u_2} \otimes \rho_0, \tag{98}$$

$$\rho_{u_1+u_2+u_3+1} \otimes \rho_{u_3+1} \otimes \rho_{u_2+1} \otimes \rho_1, \tag{99}$$

and amounts to the following projection:

$$\{\rho_{u_1+u_2+u_3} \otimes \rho_{u_3} \otimes \rho_{u_2} \otimes \rho_0 - \rho_{u_1+u_2+u_3+1} \otimes \rho_{u_3+1} \otimes \rho_{u_2+1} \otimes \rho_1 \geq 0\}. \tag{100}$$

B Measurements for the Eight-Bit Polar Decoder

Here, we provide the form of a quantum successive cancellation decoder that recovers each bit of an eight-bit polar code. Full derivations of the results in this section are available from the authors upon request.

B.1 Recovering u_1

The test to recover the first bit u_1 is simply the parity of eight individual Helstrom measurements:

$$\{(\rho_0 - \rho_1)^{\otimes 8} \geq 0\}. \tag{101}$$

B.2 Recovering u_2 given u_1

The test to recover bit u_2 given u_1 projects onto the positive eigenspace of the difference of

$$\left(\sum_{u_3',u_4',u_5'} \rho_{u_1+u_3'+u_4'+u_5'} \otimes \rho_{u_3'} \otimes \rho_{u_4'} \otimes \rho_{u_5'}\right) \otimes \left(\sum_{u_6',u_7',u_8'} \rho_{u_6'+u_7'+u_8'} \otimes \rho_{u_6'} \otimes \rho_{u_7'} \otimes \rho_{u_8'}\right) \tag{102}$$

and

$$\left(\sum_{u'_3,u'_4,u'_5} \rho_{u_1+u'_3+u'_4+u'_5} \otimes \rho_{u'_3} \otimes \rho_{u'_4} \otimes \rho_{u'_5}\right) \otimes \left(\sum_{u'_6,u'_7,u'_8} \rho_{u'_6+u'_7+u'_8} \otimes \rho_{u'_6} \otimes \rho_{u'_7} \otimes \rho_{u'_8}\right). \tag{103}$$

As such, it is not clear to us how one could approximate this test as some combination of Helstrom and Fuchs-Caves tests.

B.3 Recovering u_3 given u_2, and u_1

The test to recover bit u_3 given u_1 and u_2 is equal to the parity of the following two tests:

$$\left\{ \begin{array}{c} \left(\sum_{u'_4} \rho_{u_1+u_2+u'_4} \otimes \rho_{u'_4}\right) \otimes \left(\sum_{u'_5} \rho_{u'_5} \otimes \rho_{u'_5}\right) \\ -\left(\sum_{u'_4} \rho_{u_1+u_2+1+u'_4} \otimes \rho_{u'_4}\right) \otimes \left(\sum_{u'_5} \rho_{1+u'_5} \otimes \rho_{u'_5}\right) \geq 0 \end{array} \right\}, \tag{104}$$

$$\left\{ \begin{array}{c} \left(\sum_{u'_6} \rho_{u_2+u'_6} \otimes \rho_{u'_6}\right) \otimes \left(\sum_{u''_8} \rho_{u''_8} \otimes \rho_{u''_8}\right) \\ -\left(\sum_{u'_6} \rho_{u_2+u'_6+1} \otimes \rho_{u'_6}\right) \otimes \left(\sum_{u''_8} \rho_{1+u''_8} \otimes \rho_{u''_8}\right) \geq 0 \end{array} \right\}. \tag{105}$$

It is again unclear to us how to decompose this measurement further.

B.4 Recovering u_4 given u_3, u_2, and u_1

The test to recover bit u_4 given u_1, u_2, and u_3 projects onto the positive eigenspace of the difference of

$$\left(\sum_{u'_5} \rho_{u_1+u_2+u_3+u'_5} \otimes \rho_{u'_5}\right) \otimes \left(\sum_{u'_6} \rho_{u_3+u'_6} \otimes \rho_{u'_6}\right) \otimes \left(\sum_{u'_7} \rho_{u_2+u'_7} \otimes \rho_{u'_7}\right) \otimes \left(\sum_{u'_8} \rho_{u'_8} \otimes \rho_{u'_8}\right) \tag{106}$$

and

$$\left(\sum_{u'_5} \rho_{u_1+u_2+u_3+1+u'_5} \otimes \rho_{u'_5}\right) \otimes \left(\sum_{u'_6} \rho_{u_3+1+u'_6} \otimes \rho_{u'_6}\right) \otimes \left(\sum_{u'_7} \rho_{u_2+1+u'_7} \otimes \rho_{u'_7}\right) \otimes \left(\sum_{u'_8} \rho_{1+u'_8} \otimes \rho_{u'_8}\right) \tag{107}$$

Again, this one remains unclear how to decompose further.

B.5 Recovering u_5 given $u_4, \ldots, u_1$

The test to recover bit u_5 given u_1 through u_4 is equal to

$$\left\{ \begin{array}{c} (\rho_{u_1+u_2+u_3+u_4} \otimes \rho_0 - \rho_{u_1+u_2+u_3+u_4+1} \otimes \rho_1) \otimes (\rho_{u_3+u_4} \otimes \rho_0 - \rho_{u_3+u_4+1} \otimes \rho_1) \\ \otimes (\rho_{u_2+u_4} \otimes \rho_0 - \rho_{u_2+u_4+1} \otimes \rho_1) \otimes (\rho_{u_4} \otimes \rho_0 - \rho_{u_4+1} \otimes \rho_1) \geq 0 \end{array} \right\}. \tag{108}$$

It is easy to see that one could approximate this test by first performing four Fuchs-Caves measurements on adjacent pairs of channel outputs and taking the parity of these tests.

B.6 Recovering u_6 given $u_5, \ldots, u_1$

The test to recover bit u_6 given u_1 through u_5 is a projection onto the positive eigenspace of the difference of

$$\left(\sum_{u_7'} \rho_{u_1+\cdots+u_5+u_7'} \otimes \rho_{u_5+u_7'} \otimes \rho_{u_3+u_4+u_7'} \otimes \rho_{u_7'} \right) \otimes \left(\sum_{u_8'} \rho_{u_2+u_4+u_8'} \otimes \rho_{u_8'} \otimes \rho_{u_4+u_8'} \otimes \rho_{u_8'} \right) \tag{109}$$

and

$$\left(\sum_{u_7'} \rho_{u_1+\cdots+u_5+1+u_7'} \otimes \rho_{u_5+1+u_7'} \otimes \rho_{u_3+u_4+u_7'} \otimes \rho_{u_7'} \right) \otimes \left(\sum_{u_8'} \rho_{u_2+u_4+1+u_8'} \otimes \rho_{1+u_8'} \otimes \rho_{u_4+u_8'} \otimes \rho_{u_8'} \right). \tag{110}$$

A simple decomposition of this test remains unclear.

B.7 Recovering u_7 given $u_6, \ldots, u_1$

The test for recovering bit u_7 given the previous ones is

$$\left\{ \begin{array}{c} (\rho_{u_1+\cdots+u_6} \otimes \rho_{u_5+u_6} \otimes \rho_{u_3+u_4} \otimes \rho_0 - \rho_{u_1+\cdots+u_6+1} \otimes \rho_{u_5+u_6+1} \otimes \rho_{u_3+u_4+1} \otimes \rho_1) \otimes \\ (\rho_{u_2+u_4+u_6} \otimes \rho_{u_6} \otimes \rho_{u_4} \otimes \rho_0 - \rho_{u_2+u_4+u_6+1} \otimes \rho_{u_6+1} \otimes \rho_{u_4+1} \otimes \rho_1) \geq 0 \end{array} \right\}, \tag{111}$$

which is clearly implementable by performing a Fuchs-Caves measurement on the first four qubits and the last four, and than taking the parity of these two tests.

B.8 Recovering u_8 given $u_7, \ldots, u_1$

The final test for recovering the last bit u_8 given all others is a projection onto the positive eigenspace of the difference of

$$\rho_{u_1+\cdots+u_7} \otimes \rho_{u_5+u_6+u_7} \otimes \rho_{u_3+u_4+u_7} \otimes \rho_{u_7} \otimes \rho_{u_2+u_4+u_6} \otimes \rho_{u_6} \otimes \rho_{u_4} \otimes \rho_0, \tag{112}$$

and

$$\rho_{u_1+\cdots+u_7+1} \otimes \rho_{u_5+u_6+u_7+1} \otimes \rho_{u_3+u_4+u_7+1} \otimes \rho_{u_7+1} \otimes \rho_{u_2+u_4+u_6+1} \otimes \rho_{u_6+1} \otimes \rho_{u_4+1} \otimes \rho_1. \tag{113}$$

It is clear that we can approximate this test with a Fuchs-Caves measurement.

On the Query Complexity of Perfect Gate Discrimination

Giulio Chiribella[1], Giacomo Mauro D'Ariano[2], and Martin Roetteler[3]

1 **Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University**
Beijing, 100084, China
gchiribella@mail.tsinghua.edu.cn
2 **QUIT group, Dipartimento di Fisica "A. Volta", INFN Sezione di Pavia, via Bassi 6, 27100 Pavia, Italy**
3 **NEC Laboratories America Princeton, New Jersey, USA**

Abstract

We investigate the problem of finding the minimum number of queries needed to perfectly identify an unknown quantum gate within a finite set of alternatives, considering both deterministic strategies. For unambiguous gate discrimination, where errors are not tolerated but inconclusive outcomes are allowed, we prove that parallel strategies are sufficient to identify the unknown gate with minimum number of queries and we use this fact to provide upper and lower bounds on the query complexity. In addition, we introduce the notion of generalized t-designs, which includes unitary t-designs and group representations as special cases. For gates forming a generalized t-design we prove that there is no difference between perfect probabilistic and perfect deterministic gate discrimination. Hence, evaluating of the query complexity of perfect discrimination is reduced to the easier problem of evaluating the query complexity of unambiguous discrimination.

1998 ACM Subject Classification J.2 Physical sciences and engineering

Keywords and phrases quantum gate identification, unambiguous discrimination, minimum error discrimination, query complexity

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.178

1 Introduction

Identifying an unknown unitary evolution is a fundamental problem in quantum theory [1, 2, 3, 4, 5, 6, 7, 8, 9], with a wide range of applications in quantum information and computation. In quantum computation, the problem is known as oracle identification [10, 11, 12, 13, 14] and is the core of paradigmatic quantum algorithms such as Grover's [15] and Bernstein-Vazirani's [16]. In addition, identifying an unknown unitary gate has applications in the alignment of reference frames via quantum communication [17, 18, 19, 20, 21, 22], in the design of quantum communication protocols that work in the absence of shared reference frames [23, 24, 25], and in the design of quantum machines that learn to execute a desired operation from a training set of examples [26]. For all these applications, the crucial step is to find efficient strategies that discriminate among a set of unknown gates with minimum number of queries to the black box uses.

A striking feature of gate discrimination is that any two distinct unitaries can be perfectly distinguished from one another in a finite number of queries, either using entanglement [1, 2] or using a sequential strategy where different queries are called at different time steps [5].

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 178–191

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Clearly, this feature implies that an unknown gate in a finite set $(U_x)_{x\in\mathsf{X}}$ can be perfectly identified in a finite number of queries, e.g. by running $|\mathsf{X}|-1$ pairwise tests each of which eliminates one wrong alternative. However, in terms of efficiency the pairwise approach leaves large room for improvements: For example, when the unitaries $(U_x)_{x\in\mathsf{X}}$ are mutually orthogonal, one can identify the black box in a single query using an ancilla, following the lines of the dense coding protocol [27]. In general, finding the minimum number of queries needed for perfect discrimination is a hard problem: for example, solving it would automatically give a general solution for query complexity of oracle identification. One way to approach the problem is to consider the less demanding task of *unambiguous gate discrimination* [3, 4, 6, 7, 28], where the unknown gate is identified without errors but one allows for an inconclusive result. General conditions for unambiguous discrimination were given in Refs. [4, 7, 28] under the assumption that the available queries are used in parallel. However, the case of general strategies and the quantification of the resources required for unambiguous gate discrimination have remained largely unaddressed up to now.

In this paper we prove that parallel strategies are sufficient for unambiguous gate discrimination: if the unambiguous discrimination can be achieved in N queries, then it can be achieved by calling the N queries in parallel (in general, using ancillas). Furthermore, we show that for suitable sets of gates, called *generalized t-designs*, there is no difference between the performances of deterministic strategies using the queries in parallel and the performances of general probabilistic strategies allowing for inconclusive outcomes and sequential queries. Clearly, this implies that, if unambiguous discrimination is possible in N queries, then also perfect discrimination must be possible in N queries. This result reduces the query complexity of perfect discrimination to the query complexity of unambiguous discrimination, which is simpler to evaluate. The reduction to unambiguous discrimination has a fairly large range of applications, including in particular the case when the set of gates is the representation of a finite group. Particular examples are the group of all Boolean oracles [10], the groups of linear [16] and quadratic [29] Boolean functions, the group of permutations [19], and the group of all oracles corresponding to functions from a given finite set to another [7]. Based on the reductions to parallel strategies, we provide lower and upper bounds on the query complexity of perfect/unambiguous discrimination and on the size of the ancilla systems needed by the discrimination strategy. The bounds are general and can often be improved in specific cases. Nevertheless, they suffice to show that unambiguous discrimination of the gates $(U_x)_{x\in\mathsf{X}}$ is always possible with no more than $|\mathsf{X}|-1$ queries. Since $|\mathsf{X}|-1$ is the minimum number of queries that would be needed by the method of pairwise elimination, our result shows that a joint discrimination strategy typically offers an advantage over pairwise elimination. Finally, we discuss the extension of our result to ancilla-unassisted discrimination strategies, where the prohibition to use ancillas implies an overhead in the number of queries needed to achieve perfect/unambiguous discrimination.

2 Results

Unambiguous gate discrimination. We show that unambiguous gate discrimination can be parallelized: if the gates in a given set can be distinguished unambiguously with N queries, then they can be distinguished unambiguously by applying the queries in parallel, possibly using ancillas. Denoting by $N_{\min}$ the minimum number of queries needed to unambiguously identify a gate in $\mathsf{U}:=(U_x)_{x\in\mathsf{X}}$, we prove the bounds

$$|\mathsf{U}|^{\frac{1}{d^2-1}}-1\leq N_{\min}\leq|\mathsf{U}|-\dim(\mathsf{U})+1, \tag{1}$$

where d is the dimension of the Hilbert space where the gates act and $\dim(\mathsf{U})$ is the number of linearly independent operators in U.

In addition, we prove a basic fact about unambiguous state discrimination of pure states, namely that the states in a generic set $\{|\psi_x\rangle\}_{x\in\mathsf{X}}$ can be unambiguously discriminated using N identical input copies whenever N satisfies

$$N > \frac{\log(|\mathsf{X}|-1)}{\log\left(F^{-\frac{1}{2}}\right)} \qquad F := \max_{x,y\in\mathsf{X},x\neq y} |\langle\psi_x|\psi_y\rangle|^2. \tag{2}$$

Applying this result in the case of gate discrimination then gives the upper bound

$$N_{\min} \leq \left\lfloor \frac{\log(|\mathsf{U}|-1)}{\log\left(F_{\mathsf{U}}^{-1/2}\right)} \right\rfloor + 1, \tag{3}$$

where F_{U} is the *minimax fidelity* $F_{\mathsf{U}} := \min_{|\Psi\rangle\in\mathcal{H}\otimes\mathcal{H},\|\Psi\|=1} \max_{x,y\in\mathsf{X},x\neq y} \left|\langle\Psi|(U_x^\dagger U_y\otimes I)|\Psi\rangle\right|^2$. Of course, every upper bound on F_{U} results in a corresponding upper bound for $N_{\min}$. All the upper bounds in Eqs. (1) and (3) are achieved for particular sets of gates. However, in specific cases they can often be improved.

Perfect gate discrimination. We introduce the notion of generalized unitary t-designs, which enables a joint treatment of group representations and unitary t-designs [30, 31, 32, 33]. When the unitary gates form a generalized t-design, we show that probabilistic strategies using $N \leq t$ queres cannot improve the performances of discrimination of parallel deterministic strategies. Precisely, the maximum probability of correct discrimination with $N \leq t$ queries (conditional to the occurrence of conclusive outcomes) is given by

$$p_N = \frac{\dim \mathsf{U}_N}{|\mathsf{U}|} \qquad \mathsf{U}_N := (U_x^{\otimes N})_{x\in\mathsf{X}} \tag{4}$$

and can be achieved by a deterministic strategy that uses the N queries in parallel. As a corollary, for a generalized $|\mathsf{U}|$-design U there is no difference between perfect and unambiguous discrimination: whenever unambiguous discrimination is possible, the probability of the inconclusive result can be reduced to zero. Thanks to this reduction, Eqs. (1) and (3) become bounds on the query complexity of perfect gate discrimination.

3 General gate discrimination strategies

Let $\mathcal{H} \simeq \mathbb{C}^d, d < \infty$ be a finite dimensional Hilbert space, let $\mathsf{Lin}(\mathcal{H})$ be the set of linear operators on $\mathcal{H}$, and let $\mathsf{U} = (U_x)_{x\in\mathsf{X}} \subset \mathsf{Lin}(\mathcal{H})$ be a finite set of unitary matrices. All throughout the paper we will require that two unitaries U_x and U_y corresponding to distinct labels $x \neq y$ be statistically distinguishable, that is

$$\forall x, y \in \mathsf{X}, x \neq y \quad \exists|\psi\rangle \in \mathcal{H} : \quad U_x|\psi\rangle\langle\psi|U_x^\dagger \neq U_y|\psi\rangle\langle\psi|U_y^\dagger. \tag{5}$$

▶ **Definition 1.** If Eq. (5) holds, we say that the mapping $U : x \in \mathsf{X} \mapsto U_x \in \mathsf{Lin}(\mathcal{H})$ is a *projectively faithful representation* of the set X.

Suppose that we are given a black box implementing one of the unitaries in U. In order to identify the action of the black box with N queries, we will consider without loss of generality *pure* strategies: the most general pure strategy consists in

1. preparing a pure input state $|\Psi\rangle \in \mathcal{H}\otimes\mathcal{H}_A$, where A is a suitable ancillary system

2. evolving it through a quantum circuit that uses N queries to the unknown gate U_x, interspersed with known unitary gates $(U_n)_{n=1}^N \subset \mathsf{Lin}(\mathcal{H}\otimes\mathcal{H}_A)$, thus obtaining the output state
$$|\Psi_x\rangle := \left[\prod_{n=1}^{N} U_n(U\otimes I_A)\right] |\Psi\rangle \tag{6}$$
3. performing a measurement on the output state $|\Psi_x\rangle$ with measurement outcomes in the set $\mathsf{Y} = \mathsf{X}\cup\{?\}$. The outcome set Y includes an inconclusive outcome $y =?$ corresponding to the case when the experimenter abstains from producing a guess [3].

Denote by p_x the prior probability of U_x and by $p_N(y|x)$ the conditional probability of the measurement outcome y given that the gate is U_x and that N queries are used. Conditionally to the occurrence of conclusive outcomes, the probability of correct gate identification with N queries is
$$p_N := \frac{\sum_{x\in\mathsf{X}} p_N(x|x)\ p_x}{\sum_{x,y\in\mathsf{X}} p_N(y|x)\ p_x}. \tag{7}$$

We will now spell out three different notions of perfect gate discrimination, in increasing order of strength:

▶ **Definition 2.** Denote by p_N^{max} rhe maximum of p_N over all probabilistic discrimination protocols using N queries (with no constraints on the probability of abstention). A discrimination strategy achieves

- *perfect probabilistic discrimination* iff $p_N^{max} = 1$
- *unambiguous discrimination* iff $p_N^{max} = 1$ and $p(x|x) > 0$ for every $x \in \mathsf{X}$ such that $p_x > 0$
- *perfect deterministic discrimination* iff $p_N^{max} = 1$ and $p_? := \sum_{x\in\mathsf{X}} p(?|x)p_x = 0$.

Clearly, perfect deterministic discrimination implies unambiguous discrimination, which in turn implies perfect probabilistic discrimination. The latter two types of discrimination can be characterized in terms of linear independence:

▶ **Theorem 3.** *The unitaries $(U_x)_{x\in\mathsf{X}}$ can be discriminated in N queries*

1. *in a perfect probabilistic way if and only if there exists $x_0 \in \mathsf{X}$ such that $U_{x_0}^{\otimes N} \notin \mathrm{Span}(U_x^{\otimes N})_{x\in\mathsf{X},x\neq x_0}$*
2. *in an unambiguous way if and only if the unitaries $(U_x^{\otimes N})_{x\in\mathsf{X}}$ are linearly independent.*

Proof. We first prove necessity. The condition for perfect probabilistic discrimination is equivalent to the existence of at least one $x_0 \in \mathsf{X}$ such that $p_N(x_0|x) = 0 \quad \forall x \neq x_0$, which in turn is equivalent to the condition that the output state $|\Psi_{x_0}\rangle$ in Eq. (6) is linearly independent from the states $(|\Psi_x\rangle)_{x\in\mathsf{X},x\neq x_0}$. Since the function $U_x^{\otimes N} \mapsto |\Psi_x\rangle$ is linear, the condition $U_{x_0}^{\otimes N} \notin \mathrm{Span}(U_x^{\otimes N})_{x\in\mathsf{X},x\neq x_0}$ is necessary for perfect probabilistic discrimination. Similarly, the condition for unambiguous discrimination is equivalent to requirement that $p_N(x_0|x) = 0 \quad \forall x, x_0 \in \mathsf{X}, x \neq x_0$, which in turn is equivalent to the requirement that the output states $\{|\Psi_x\rangle\}_{x\in\mathsf{X}}$ are linearly independent. Independence of the states $\{|\Psi_x\rangle\}_{x\in\mathsf{X}}$ implies independence of the unitaries $(U_x^{\otimes N})_{x\in\mathsf{X}}$. Both conditions are also sufficient, because the linear function $U_x^{\otimes N} \mapsto \{|\Phi_x\rangle^{\otimes N}\}_{x\in\mathsf{X}}$ defined by $|\Phi_x\rangle := (U_x\otimes I)|\Phi\rangle$, $|\Phi\rangle := \sum_{n=1}^d |n\rangle|n\rangle/\sqrt{d}$ is invertible, and therefore preserves linear independence. Note that the states $|\Phi_x\rangle$ can be obtained from a parallel strategy where N pairs of systems are prepared in the state $|\Phi\rangle^{\otimes N}$ and the unitary U_x is applied on the first system of each pair. ◀

The equivalence between unambiguous gate discrimination and linear independence of the unitaries was observed in Ref. [7] in the case of a single query (and hence of N parallel queries, which can be treated as a single query to the product box $U_x^{\otimes N}$). Theorem 3 extends the existing characterization to arbitrary discrimination strategies, possibly consisting of multiple time steps. As a consequence of this extension, unambiguous discrimination and perfect probabilistic discrimination can be parallelized:

▶ **Corollary 4.** *If the gates $(U_x)_{x\in\mathsf{X}}$ can be distinguished unambiguously (respectively, in a perfect probabilistic fashion) with N queries, then they can be distinguished unambiguously (respectively, in a perfect probabilistic fashion) using the N queries in parallel.*

We refer to the minimum number N_{min} needed to unambiguously identify a gate in $(U_x)_{x\in\mathsf{X}}$ as the *query complexity of unambiguous gate discrimination* for the gate set $\mathsf{U} := (U_x)_{x\in\mathsf{X}}$. Corollary 4 allows us to conclude that the query complexity of perfect probabilistic/unambiguous discrimination does not change if one restricts to parallel strategies. However, general sequential strategies can help in reducing the probability of the inconclusive result.

4 General bounds on the query complexity of unambiguous gate discrimination

The possibility of parallelizing unambiguous gate discrimination, established by theorem 3, leads immediately to general lower and upper bounds on the query complexity. There bounds do not assume any structure of the set of unitaries U, and can typically be improved when more information about U is available.

4.1 Lower bound

▶ **Theorem 5** (Dimensional bound)**.** *The gates in $\mathsf{U} = (U_x)_{x\in\mathsf{X}}$ can be unambiguously discriminated using N queries only if*

$$|\mathsf{U}| \leq \begin{pmatrix} d^2+N-1 \\ d^2-1 \end{pmatrix}, \tag{8}$$

which implies $N_{\min} > |\mathsf{U}|^{\frac{1}{d^2-1}} - 1$.

Proof. By theorem 3, unambiguous discrimination is possible only if $\dim(U_x^{\otimes N})_{x\in\mathsf{X}} = |\mathsf{U}|$. On the other hand, $\dim(U_x^{\otimes N})_{x\in\mathsf{X}} \leq \dim \mathsf{A}_{N,+}$, where $\mathsf{A}_{N,+} := \mathrm{Span}\left\{A^{\otimes N} \mid A \in \mathsf{Lin}(\mathcal{H})\right\}$. Since $\mathsf{A}_{N,+}$ is the symmetric subspace of the N-fold tensor product of $\mathsf{Lin}(\mathcal{H})$, and the dimension of the latter is d^2, the dimension of $\mathsf{A}_{N,+}$ is $\dim \mathsf{A}_{N,+} = \begin{pmatrix} d^2+N-1 \\ d^2-1 \end{pmatrix}$. ◀

If we do not impose any structure on the set of unitaries $\mathsf{U} = (U_x)_{x\in\mathsf{X}}$, then the bound of Eq. (8) is the best we can hope for. Indeed, for any fixed Hilbert space dimension d and for every number N we can always find a set of unitaries U such that the minimum number of queries needed to unambiguously identify a gate in U is exactly N.

▶ **Example 6.** The bound of Eq. (8) can be saturated choosing $(U_x^{\otimes N})_{x\in\mathsf{X}}$ to be a basis for $A_{N,+}$. This is possible thanks to the Schur-Weyl duality [34], which implies that the unitaries $(U^{\otimes N})_{U\in U(d)}$ are a spanning set for $A_{N,+}$.

4.2 Upper bounds

An upper bound on the query complexity can be obtained by observing that the dimension of $\mathrm{Span}(U_x^{\otimes N})_{x\in\mathsf{X}}$ grows at least linearly with N, a fact that can be proved using an earlier result by Chefles [35]:

▶ **Theorem 7** (Linear bound). *The query complexity of unambiguous discrimination of the gates in* U *is upper bounded by*

$$N_{\min} \leq |\mathsf{U}| + 1 - \dim(\mathsf{U}). \tag{9}$$

Proof. Let $\mathsf{S} = (v_x)_{x\in\mathsf{X}}$ be a finite set of vectors in a vector space V, with the property that every two distinct vectors in S are linearly independent. Under this hypothesis, Chefles proved that $\dim \mathrm{Span}(v_x^{\otimes N+1}) \geq \dim \mathrm{Span}(v_x^{\otimes N}) + 1$ [35]. Applying the result to the set $\mathsf{U}_N := (U_x^{\otimes N})_{x\in\mathsf{X}}$ gives $\dim(\mathsf{U}_N) \geq \dim(\mathsf{U}) + N - 1$. Hence, for the unitaries in U_N are linearly independent for $N = |\mathsf{U}| - \dim(\mathsf{U}) + 1$. ◀

In general, the bound of Eq. (9) can be achieved: for every fixed Hilbert space dimension d and for every fixed cardinality $|\mathsf{U}|$ we can find a set of unitaries such that $N_{\min} = |\mathsf{U}| - \dim(\mathsf{U}) + 1$. This can be seen in the following

▶ **Example 8.** Consider the discrete phase shifts

$$U_x := \omega^x\ |1\rangle\langle 1| + (I - |1\rangle\langle 1|) \qquad \omega := e^{\frac{2\pi i}{|\mathsf{X}|}},$$

with $x = 1, \ldots, |\mathsf{X}|$. In this case the number of linearly independent unitaries in $(U_x^{\otimes N})_{x\in\mathsf{X}}$ is exactly equal to $N + 1$, as it can be seen from the fact that the unitaries $(U_x^{\otimes N})_{x\in\mathsf{X}}$ are in bijective correspondence with the vectors of their eigenvalues, given by $(v_x)_{x\in\mathsf{X}} \subset \mathbb{C}^{N+1}$ where $v_x := (1, \omega, \omega^2, \ldots, \omega^N)^T$. Since the number of linearly independent unitaries in $(U_x^{\otimes N})_{x\in\mathsf{X}}$ is $N + 1$, the minimum number needed for unambiguous discrimination is exactly $N_{\min} = |\mathsf{X}| - 1 = |\mathsf{U}| - \dim(U) + 1$.

Another example where the bound of Eq. (7) gives the exact value is is the example of the so-called "shift-and-multiply" gates:

▶ **Example 9** (Shift-and-multiply gates). Theorem 7 provides a tight bound for the "shift-and-multiply" representation of the group $\mathbf{G} = \mathbb{Z}_d \times \mathbb{Z}_d$, defined by

$$U_{pq} = S^p M^q \qquad (p, q) \in \mathbb{Z}_d \times \mathbb{Z}_d\ , \tag{10}$$

where $S = \sum_{k=1}^d |(k+1) \mathrm{mod}\ d\rangle\langle k|$ and $M = \sum_{k=1}^d e^{(2\pi i k)/d} |k\rangle\langle k|$. In this case, the unitaries $(U_{pq})_{(p,q)\in\mathbb{Z}_d\times\mathbb{Z}_d}$ are linearly independent, and therefore the bound gives $N_{\min} = 1$. Note that, in fact, the unitaries are orthogonal in the Hilbert-Schmidt product, and, therefore, an unknown unitary U_{pq} can be identified perfectly and deterministically, as in the dense coding protocol [27].

Theorem 7 provides an estimate of $N_{\min}$ that is always better than the number of pairwise tests $|\mathsf{U}| - 1$ that would be needed to identify a gate in $(\mathsf{U}_x)_{x\in\mathsf{X}}$ with the method of pairwise eliminations outlined in [1, 2]. Note however that Eq. (9) only ensures *unambiguous* discrimination, while the pairwise elimination method ensures *perfect deterministic* discrimination. In the next Section we will see that the distinction between unambiguous and perfect discrimination disappears when the gates in U form a group representation, or, more generally, a generalized t-design.

Before adding more structure on the set U, we give here a second upper bound that often yields a better estimate than Theorem 7. To state the bound we introduce the *minimax fidelity* of the unitaries U, defined as

$$F_{\mathsf{U}} := \min_{|\Psi\rangle\in\mathcal{H}\otimes\mathcal{H},\||\Psi\|=1} \max_{x,y\in\mathsf{X},x\neq y} \left|\langle\Psi|(U_x^\dagger U_y\otimes I)|\Psi\rangle\right|.$$

The minimax fidelity quantifies the distinguishability of the unitaries $(U_x)_{x\in\mathsf{X}}$ when single-shot ancilla-assisted strategies are used. Clearly, if $F_{\mathsf{U}}=0$, the unitaries can be perfectly distinguished in one shot using a suitable input state. Note also that, under the standing assumption of this paper (projective faithfulness of the mapping $x\mapsto U_x$), F_{U} must be strictly smaller than 1.

▶ **Theorem 10** (Fidelity bound). *The query complexity of unambiguous discrimination of the gates in* U *is upper bounded as*

$$N_{\min} \leq \left\lfloor \frac{\log(|\mathsf{U}|-1)}{\log\left(F_{\mathsf{U}}^{-\frac{1}{2}}\right)} \right\rfloor + 1. \tag{11}$$

The proof is based on a simple observation:

▶ **Lemma 11.** *Let* $(|\psi_x\rangle)_{x\in\mathsf{X}}\in\mathcal{H}$ *be a set of unit vectors such that* $F:=\max_{x,y\in\mathsf{X},x\neq y}|\langle\psi_x|\psi_y\rangle|^2$ *is strictly smaller than one. If* $F^{N/2}<1/(|\mathsf{X}|-1)$*, then the states* $(|\psi_x\rangle^{\otimes N})_{x\in\mathsf{X}}$ *are linearly independent, and, therefore, unambiguously distinguishable.*

Proof. Suppose that $\sum_{y\in\mathsf{X}} c_y|\psi_y\rangle^{\otimes N}=0$. Multiplying by $\langle\psi_x|^{\otimes N}$, taking the modulus, and summing over x we obtain

$$\begin{aligned}\sum_{x\in\mathsf{X}}|c_x| &= \sum_{x\in\mathsf{X}}\left|\sum_{y\in\mathsf{X},y\neq x} c_y\langle\psi_x|\psi_y\rangle^N\right| \\ &\leq \sum_{x\in\mathsf{X}}\sum_{y\in\mathsf{X},y\neq x}|c_y|F^{N/2} \\ &= (|\mathsf{X}|-1)F^{N/2}\left(\sum_{x\in\mathsf{X}}|c_x|\right).\end{aligned}$$

Clearly, if $(|\mathsf{X}|-1)F_{\mathsf{S}}^{N/2}<1$, the only possible solution is $c_x=0\ \forall x\in\mathsf{X}$. Hence, the states $(|\psi_x\rangle^{\otimes N})_{x\in\mathsf{X}}$ are linearly independent. ◀

Proof of theorem 10. Choose the input state $|\Psi\rangle\in\mathcal{H}\otimes\mathcal{H}$ so that $\max_{x,y\in\mathsf{X},x\neq y}|\langle\Psi|(U_x^\dagger U_y\otimes I)|\Psi\rangle|^2=F_{\mathsf{U}}$. For $F_{\mathsf{U}}^{N/2}\leq 1/(|\mathsf{U}|-1)$ the states $(|\Psi_x\rangle^{\otimes N})_{x\in\mathsf{X}}, |\Psi_x\rangle:=(U_x\otimes I)|\Psi\rangle$ are linearly independent. Therefore, also the unitaries $(U_x^{\otimes N})_{x\in\mathsf{X}}$ are linearly independent, i.e. unambiguously distinguishable. ◀

The fidelity bound gives good estimates when F_{U} is close to zero. However, it tends to produce large overheads when F_{U} approaches 1. This phenomenon is illustrated in the following example:

▶ **Example 12** (Permutation gates). Consider the permutations matrices

$$U_\pi=\sum_{k=1}^{d}|\pi(k)\rangle\langle k|, \tag{12}$$

where π is an element of the permutation group S_d. In this case it is clear that the unitary U_π can be perfectly identified with d queries (applying U_π to all the d vectors in the computational basis we can surely identify the permutation $\pi\in S_d$). One the other hand, applying the unitary U_π on a maximally entangled state gives the bound $F_{\mathsf{U}}\geq\left(\frac{d-2}{d}\right)^2$, which inserted in the fidelity bound gives $N_{\min}\leq\log(d!)/\log[d/(d-2)]=O(d^2\log d)$, which is off by a factor $d\log d$ from the actual value.

5 Discrimination of generalized unitary t-designs

Here we impose additional structure on the set of gates $(U_x)_{x\in\mathsf{X}}$. Our analysis includes the case where the set X is a finite group and $x \mapsto U_x$ is a projective representation of X. Also, it will include the case where the unitaries $(U_x)_{x\in\mathsf{X}}$ from a unitary t-design [30, 31, 32, 33]. In order to treat these two cases in a unified way, we introduce the notion of *generalized unitary t-designs*. For the discrimination of generalized unitary t-designs we will show the following properties

1. among all possible discrimination strategies using $N \leq t$ queries, the deterministic strategies using all queries in parallel maximize the probability of correct gate identification
2. for strategies using $N \leq t$ queries, there is no difference between perfect probabilistic, unambiguous, and perfect deterministic discrimination.

5.1 Generalized unitary t-designs: definition and characterization

Let us start from the definition:

▶ **Definition 13** (Generalized unitary t-designs)**.** Let $(U_x)_{x\in\mathsf{X}}$ be a set of unitaries, $(p_x)_{x\in\mathsf{X}}$ be a set of probabilities. We say that the set $(U_x, p_x)_{x\in\mathsf{X}}$ is a *generalized weighted unitary t-design* iff

$$\left(U_y^{\otimes t} \otimes \overline{U}_y^{\otimes t}\right)\left(\sum_{x\in\mathsf{X}} p_x U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t}\right) = \left(\sum_{x\in\mathsf{X}} p_x U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t}\right) \qquad \forall y \in \mathsf{X}. \tag{13}$$

If $p_x = 1/|\mathsf{X}| \quad \forall x \in \mathsf{X}$ we say that $(U_x)_{x\in\mathsf{X}}$ is a *generalized unitary t-design* (or shortly, a *generalized t-design*).

Note that, by definition, every generalized weighted t-design is also a weighted generalized $(t-1)$-design.

▶ **Example 14** (Unitary t-designs)**.** A unitary t-design is a set of unitaries and probabilities $(U_x, p_x)_{x\in\mathsf{X}}$ such that

$$\sum_{x\in\mathsf{X}} p_x U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t} = \int \mathrm{d}U \; U^{\otimes t} \otimes \bar{U}^{\otimes t},$$

where the integral in the l.h.s. runs over the normalized Haar measure of the group $U(d)$. From the definition is clear that any unitary t-design is an example of generalized unitary t-design.

Generalized t-designs can be characterized as follows:

▶ **Proposition 15.** *A set of unitaries $(U_x, p_x)_{x\in\mathsf{X}}$ is a weighted generalized t-design if and only if there exists a compact group $\mathbf{G}$ such that $\mathsf{X} \subseteq \mathsf{G}$ and*

$$\sum_{x\in\mathsf{X}} p_x \; U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t} = \int \mathrm{d}g \; U_g^{\otimes t} \otimes \overline{U}_g^{\otimes t}, \tag{14}$$

where $\int \mathrm{d}g f(g)$ denotes the integral of f with respect to the normalized Haar measure.

Proof. If the condition in proposition 15 is satisfied, clearly $(U_x, p_x)_{x\in\mathsf{X}}$ is a weighted generalized t-design. Conversely, if $(U_x, p_x)_{x\in\mathsf{X}}$ is a generalized weighted t-design, define $\mathbf{G}$ to be the closure of the group generated by the unitaries $(U_x,)_{x\in\mathsf{X}}$. Since we are in

finite dimensions, $\mathbf{G}$ is a compact group. Clearly, $(U_g^{\otimes t} \otimes \overline{U}_g^{\otimes t})(\sum_{x\in\mathsf{X}} p_x\ U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t}) = (\sum_{x\in\mathsf{X}} p_x\ U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t})$ for every $g \in \mathbf{G}$. Hence,

$$\begin{aligned}(\sum_{x\in\mathsf{X}} p_x\ U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t}) &= \int \mathrm{d}g (U_g^{\otimes t} \otimes \overline{U}_g^{\otimes t}) \left(\sum_{x\in\mathsf{X}} p_x\ U_x^{\otimes t} \otimes \overline{U}_x^{\otimes t}\right) \\ &= \int \mathrm{d}g (U_g^{\otimes t} \otimes \overline{U}_g^{\otimes t}).\end{aligned}$$

◀

Thanks to the above characterization, one can easily transfer properties of compact groups to generalized t-designs. In the next sections we will use this trick to prove strong properties of gate discrimination for generalized t-designs.

5.2 Basic group-theoretic facts

Since generalized t-designs have an underlying group-theoretic structure, it is useful to recall here some basic facts about the representation of compact groups. Let $\mathbf{G}$ be a compact group and let $U : \mathbf{G} \to \mathsf{Lin}(\mathcal{H}),\ g \mapsto U_g$ be a a *unitary projective representation (UPR)* of G with multiplier $\omega : \mathbf{G} \times \mathbf{G} \to \mathbb{C}$ [in short, this means that $U_g U_h = \omega(g,h) U_{gh},\ \forall g, h \in \mathbf{G}$]. Unitary representations correspond to the special case UPRs where $\omega(g,h) = 1\ \forall g, h \in \mathbf{G}$.

With a suitable choice of basis, the Hilbert space can be decomposed as a direct sum of tensor product pairs

$$\mathcal{H} = \bigoplus_{\mu\in\mathsf{Irr}(U)} (\mathcal{R}_\mu \otimes \mathcal{M}_\mu)\,, \tag{15}$$

where the sum runs over the set $\mathsf{Irr}(U)$ of all inequivalent irreducible representations (*irreps*) contained in the decomposition of U (known as *isotypic decomposition*), $\mathcal{R}_\mu$ is a *representation space* of dimension d_μ, carrying the irrep U^μ, and $\mathcal{M}_\mu$ is a *multiplicity space* of dimension m_μ, m_μ being the multiplicity of the irrep U^μ in the decomposition of U. Eq. (15) implies that the representation U can be written in the block diagonal form

$$U = \bigoplus_{\mu\in\mathsf{Irr}(U)} \left(U^\mu \otimes I_{\mathcal{M}_\mu}\right), \tag{16}$$

where $I_{\mathcal{M}_\mu}$ denotes the identity matrix on $\mathcal{M}_\mu$. Note that all the irreps $U^\mu \in \mathsf{Irr}(U)$ must have the same multiplier ω.

Using Eq. (16) and the orthogonality of matrix elements, one can prove that the set of unitaries $\mathsf{U} := (U_g)_{g\in\mathbf{G}}$ satisfies

$$\dim(\mathsf{U}) = \sum_{\mu\in\mathsf{Irr}(U)} d_\mu^2. \tag{17}$$

Due to the importance of linear independence in the gate discrimination problem, this equation will become very useful in the following section.

A representation that plays a key role in gate discrimination is the *regular representation*, which for finite groups is a representation of $\mathbf{G}$ on the Hilbert space $\mathcal{H} = \mathbb{C}^{|\mathbf{G}|}$, equipped with the orthonormal basis $\{|g\rangle \mid g \in \mathbf{G}\}$:

▶ **Definition 16.** The *regular representation with multiplier* ω is the projective representation $U^{reg,\omega} : \mathbf{G} \to \mathsf{Lin}(\mathbb{C}^{|\mathbf{G}|})$ defined by

$$U_g^{reg,\omega}|h\rangle = \omega(g,h)\ |gh\rangle\ , \qquad \forall g, h \in \mathbf{G} \tag{18}$$

The regular decomposition is reducible and its isotypic decomposition is

$$U_g^{reg,\omega} = \bigoplus_{\mu\in \mathrm{Irr}(\mathbf{G},\omega)} \left(U_g^\mu \otimes I_{\mathcal{M}_\mu}\right) \qquad \mathcal{M}_\mu \simeq \mathbb{C}^{d_\mu} \tag{19}$$

where $\mathrm{Irr}(\mathbf{G},\omega)$ denotes the set of all the irreps of $\mathbf{G}$ with multiplier ω [in particular, $\mathrm{Irr}(\mathbf{G},1)$ is the set of all *unitary* irreps of $\mathbf{G}$]. Note that every irrep appears with multiplicity $m_\mu = d_\mu$. Choosing $g = e$ (the identity element in the group) and taking the trace on both sides of Eq. (19) one obtains

$$|\mathbf{G}| = \sum_{\mu\in \mathrm{Irr}(\mathbf{G},\omega)} d_\mu^2 \ , \tag{20}$$

which holds for every possible multiplier ω. Finally, combining Eqs. (17) and (20), one gets the following statement:

▶ **Proposition 17.** *Let $\mathbf{G}$ be a finite group and let $U : \mathbf{G} \to \mathsf{Lin}(\mathcal{H})$ be a UPR with multiplier ω. Then, the unitaries $(U_g)_{g\in\mathbf{G}}$ are linearly independent if and only if the isotypic decomposition of U contains all the irreps in $\mathsf{Irr}(\mathbf{G},\omega)$.*

5.3 Optimal discrimination of generalized unitary t-designs

We start from general result about the maximum probability of correct identification, maximized over over all probabilistic strategies consisting of N queries. Precisely, we show that the maximum success probability can be always achieved with a deterministic parallel strategy:

▶ **Theorem 18** (Optimal probabilistic gate discrimination)**.** *Let $(U_x)_{x\in\mathsf{X}}$ be a set of unitary gates and let $(p_x)_{x\in\mathsf{X}}$ the corresponding prior probabilities. Then, the maximum probability of correct gate identification [defined in Eq. (7)] is*

$$p_N^{\max} = \max_{x\in\mathsf{X}} \ p_x \langle\!\langle U_x|^{\otimes N} \ R_N^{-1} \ |U_x\rangle\!\rangle^{\otimes N}, \tag{21}$$

with $|U_x\rangle\!\rangle := (U_x\otimes I)|I\rangle\!\rangle$, $|I\rangle\!\rangle := \sum_{n=1}^d |n\rangle|n\rangle$, $R_N := \sum_{x\in\mathsf{X}} p_x \ (|U_x\rangle\!\rangle\langle\!\langle U_x|)^{\otimes N}$, and R_N^{-1} denotes the Moore-Penrose inverse of R_N. The maximum probability of correct identification can be achieved applying the N queries in parallel on an entangled state.

Proof. Using the formalism of quantum combs [36, 37, 38], we express the probability $p_N(y|x)$ as $p_N(y|x) = \langle\!\langle U_x|^{\otimes N} \ T_y \ |U_x\rangle\!\rangle^{\otimes N}$ where $(T_y)_{y\in\mathsf{Y}}$ is a collection of positive operators satisfying suitable normalization conditions [38, 36] (the actual form of the conditions is irrelevant here). The probability of correct identification can be bounded as

$$\begin{aligned}
p_N &= \frac{\sum_{x\in\mathsf{X}} p_x \ \langle\!\langle U_x|^{\otimes N} R_N^{-\frac{1}{2}} \ \left(R_N^{\frac{1}{2}} T_x R_N^{\frac{1}{2}}\right) \ R_N^{-\frac{1}{2}} |U_x\rangle\!\rangle^{\otimes N}}{\sum_{y\in\mathsf{X}} \mathrm{Tr}[T_y R_N]} \\
&\le \sum_{x\in\mathsf{X}} p_x \ \mathrm{Tr}[\rho_x \ R_N^{-\frac{1}{2}} (|U_x\rangle\!\rangle\langle\!\langle U_x|)^{\otimes N} R_N^{-\frac{1}{2}}] \qquad \rho_x := \frac{R_N^{\frac{1}{2}} T_x R_N^{\frac{1}{2}}}{\sum_{y\in\mathsf{X}} \mathrm{Tr}[R_N^{\frac{1}{2}} T_y R_N^{\frac{1}{2}}]} \\
&\le \sum_{x\in\mathsf{X}} p_x \ \mathrm{Tr}[\rho_x] \ \|R_N^{-\frac{1}{2}} (|U_x\rangle\!\rangle\langle\!\langle U_x|)^{\otimes N} R_N^{-\frac{1}{2}}\|_\infty \\
&\le \max_{x\in\mathsf{X}} p_x \ \langle\!\langle U_x|^{\otimes N} \ R_N^{-1} \ |U_x\rangle\!\rangle^{\otimes N},
\end{aligned}$$

the last inequality coming from the condition $\sum_{x\in\mathsf{X}} \mathrm{Tr}[\rho_x] = 1$. Defining

$$x_{\max} := \mathrm{argmax}\ p_x\ \langle\!\langle U_x|^{\otimes N}\ R^{-1}\ |U_x\rangle\!\rangle^{\otimes N},$$

the bound can be saturated by applying the N queries of U_x in parallel on the maximally entangled state $|\Phi\rangle^{\otimes N}$,$|\Phi\rangle := |I\rangle\!\rangle/\sqrt{d}$, and by performing the POVM $(P_y)_{y\in\mathsf{Y}}$ defined by $P_{x_{\max}} = R^{-1}(|U_{x_{\max}}\rangle\!\rangle\langle\!\langle U_{x_{\max}}|)^{\otimes N}R^{-1}/\langle\!\langle U_{x_{\max}}|)^{\otimes N}R^{-2}|U_{x_{\max}}\rangle\!\rangle$, $P_? = I - P_{x_{\max}}$, $P_y = 0$ for every $y \neq x_{\max}$. ◀

In the case of generalized weighted t-designs, the following strong property holds:

▶ **Theorem 19** (Optimal gate discrimination for generalized N-designs). *Let $(U_x, p_x)_{x\in\mathsf{X}}$ be a generalized weighted N-design. Then, the maximum of the probability of correct discrimination over all probabilistic strategies consisting of N queries is*

$$p_N^{\max} = \dim(\mathsf{U}_N)\ \max_{x\in\mathsf{X}} p_x \qquad \mathsf{U}_N := (U_x^{\otimes N})_{x\in\mathsf{X}}. \tag{22}$$

For uniform prior $p_x = 1/|\mathsf{U}|$, the maximum probability $p_N^{\max} = \dim(\mathsf{U}_N)/|\mathsf{U}|$ can be achieved by a deterministic strategy that uses the N queries in parallel.

Proof. Let $\mathbf{G}$ the compact group such that $\sum_{x\in\mathsf{X}}(U_x\otimes\overline{U}_x)^{\otimes N} = \int \mathrm{d}g\ (U_g\otimes\overline{U}_g)^{\otimes N}$, or equivalently, $\sum_{x\in\mathsf{X}} U_x^{\otimes N} A U_x^{\dagger\otimes N} = \int \mathrm{d}g\ U_g^{\otimes N} A U_g^{\dagger\otimes N}$ for every operator $A \in \mathsf{Lin}(\mathcal{H}^{\otimes N})$. Exploiting the isotypic decomposition of $U^{\otimes N}$, one can write $U_x^{\otimes N} = \bigoplus_{\mu\in\mathsf{Irr}(U)} \left(U_x^\mu \otimes I_{\mathcal{M}_\mu}\right)$ and, therefore, $|U_x\rangle\!\rangle^{\otimes N} = \bigoplus_{\mu\in\mathsf{Irr}(U^{\otimes N})} |U_x^\mu\rangle\!\rangle|I_{\mathcal{M}_\mu}\rangle\!\rangle$. The operator R_N in theorem 18 can be directly computed as

$$\begin{aligned} R_N &= \textstyle\sum_{x\in\mathsf{X}} p_x\ (|U_x\rangle\!\rangle\langle\!\langle U_x|)^{\otimes N} \\ &= \textstyle\int \mathrm{d}g\ (|U_g\rangle\!\rangle\langle\!\langle U_g|)^{\otimes N} \\ &= \textstyle\bigoplus_{\mu\in\mathsf{Irr}(U^{\otimes N})} \frac{m_\mu}{d_\mu}\ \left(I_{\mathcal{R}_\mu}\otimes I_{\mathcal{R}_\mu}\otimes \frac{|I_{\mathcal{M}_\mu}\rangle\!\rangle\langle\!\langle I_{\mathcal{M}_\mu}|}{m_\mu}\right), \end{aligned}$$

so that, computing the inverse, one has $\langle\!\langle U_x|^{\otimes N}\ R_N^{-1}\ |U_x\rangle\!\rangle^{\otimes N} = \sum_{\mu\in\mathsf{Irr}(U^{\otimes N})} d_\mu^2 = \dim(\mathsf{U}^{\otimes N})$ [cf. Eq. (17)]. Inserting this value in Eq. (21) proves Eq. (22). We now prove that for the uniform prior the maximum success probability can be obtained with a deterministic strategy that uses the N queries in parallel. To this purpose, consider the maximum likelihood input state [39, 40]: this is the state in $\mathcal{H}^{\otimes N}\otimes\mathcal{H}_A$ given by

$$|\Phi_{ML}\rangle := \bigoplus_{\mu\in\mathsf{Irr}(U^{\otimes N})} \sqrt{\frac{d_\mu}{\dim(U^{\otimes N})}}\ |I_{\mathcal{R}_\mu}\rangle\!\rangle,$$

where $|I_{\mathcal{R}_\mu}\rangle\!\rangle = \sum_{n=1}^{d_\mu} |\alpha_n^\mu\rangle|\beta_n^\mu\rangle$, $(|\alpha_n^\mu\rangle)_{n=1}^{d_\mu}$ being an orthonormal basis for $\mathcal{R}_\mu$ and $(|\beta_n^\mu\rangle)_{n=1}^{d_\mu}$ being an orthonormal set of vectors in $\mathcal{M}_\mu\otimes\mathcal{H}_A$ [here the dimension of $\mathcal{H}_A$ is chosen in order to satisfy the relation $d_\mu \leq m_\mu d_A, \forall\mu\in\mathsf{Irr}(U^{\otimes N})$]. Applying the N queries in parallel one obtains the output states $|\Phi_{ML,x}\rangle := (U_x^{\otimes N}\otimes I_A)|\Phi_{ML}\rangle$. Optimal discrimination can be achieved deterministically using the square root measurement [41], which in this case has POVM elements $P_x := \frac{\dim(U^{\otimes N})}{|\mathsf{U}|}\ |\Phi_{ML,x}\rangle\langle\Phi_{ML,x}|$. ◀

The general result of theorem 19 is well illustrated by the case of discrete phase shifts:

▶ **Example 20** (Discrete phase shifts). Consider the discrete phase shifts

$$U_k = \sum_{l=0}^{L-1} \omega^{kl}\ P_l \qquad \omega = e^{\frac{2\pi i}{K}}, k \in \{1,\dots K\} \tag{23}$$

where $\{P_l\}_{l=0}^{L-1}$ are orthogonal projectors summing up to the identity in $\mathcal{H}$. The unitaries $\{U_k \mid k = 1, \ldots, K\}$ form a unitary representation of the Abelian group $\mathbf{G} = \mathbb{Z}^K$. Now, the unitary irreps of $\mathbb{Z}^K$ are one-dimensional, and are given by $U_\mu : \mathbb{Z}^d \to \mathbb{C}, k \mapsto \omega^{\mu k}$, with $\mu \in \{0, \ldots, K-1\}$. From Eq. (23) it is then clear that $\mathrm{Irr}(U) = \{0, 1, \ldots L-1\}$ and Eq. (22) gives $p_1^{\max} = L/K$. Similarly, it is clear that $\mathrm{Irr}(U^{\otimes N}) = \{0, 1, \ldots, N(L-1)\}$, and therefore, Eq. (22) gives

$$p_N^{\max} = \frac{NL - N + 1}{K} \qquad N \leq \frac{K-1}{L-1}. \tag{24}$$

The minimum number of queries needed for perfect discrimination is then $N_{\min} = \left\lceil \frac{K-1}{L-1} \right\rceil$.

5.4 Perfect discrimination of generalized unitary t-designs

An immediate consequence of Theorem 19, all possible notions of perfect gate discrimination coincide in the case of generalized unitary t-designs:

▶ **Corollary 21.** *If the unitaries $(U_x)_{x\in\mathsf{X}}$ form a generalized t-design, then the following are equivalent:*

1. *perfect probabilistic discrimination is possible with $N \leq t$ queries*
2. *unambiguous discrimination is possible with $N \leq t$ queries*
3. *perfect deterministic discrimination is possible in $N \leq t$ queries.*

In particular, for a generalized $|\mathsf{U}|$-design there is no difference between the three types of perfect discrimination.

For generalized t-designs the evaluation of the query complexity of perfect discrimination is reduced to the simpler problem of evaluating the query complexity of unambiguous discrimination. In particular, the bounds in Theorems 5, 7, and 10 become automatically bounds on the query complexity of perfect discrimination.

6 Conclusions

We investigated the problem of identifying an unknown unitary gate in a finite set of alternatives, using both deterministic and probabilistic discrimination strategies, and allowing the unknown gate to be queried multiple times and to be be used in parallel or in series in arbitrary quantum circuits. In this scenario, we provided upper and lower bounds on the amount of resources needed to achieve unambiguous and perfect gate identification. Specifically, we gave bounds on the query complexity and the minimum size of the ancillas needed to achieve unambiguous/perfect identification. Most of our results stem from two key observations. The first observation is that unambiguous gate discrimination can be parallelized: if unambiguous discrimination is possible with N queries, then unambiguous gate discrimination must also be possible by applying the N queries in parallel on a suitable entangled state. The second key observation is based on the definition of generalized unitary t-designs, a definition that includes unitary t-designs and group representations as special cases. The remarkable feature of generalized t-designs is that for strategies using $N \leq t$ queries there is no difference between unambiguous and perfect deterministic discrimination. Using this fact, one can reduce the analysis of perfect gate discrimination to the simpler analysis of unambiguous gate discrimination.

Acknowledgments. GC acknowledges support by the National Basic Research Program of China (973) 2011CBA00300 (2011CBA00301), by the 1000 Youth Fellowship Program of China, and by the National Natural Science Foundation of China through Grants 61033001 and 61061130540, and by Perimeter Institute for Theoretical Physics, where part of this work was carried out. Research at Perimeter Institute for Theoretical Physics is supported in part by the Government of Canada through NSERC and by the Province of Ontario through MRI.

References

1 A. Acín, *Statistical Distinguishability between Unitary Operations*, Phys. Rev. Lett. **87**, 177901 (2001).

2 G. M. D'Ariano, P. Lo Presti, and M. G. A. Paris, *Using Entanglement Improves the Precision of Quantum Measurements*, Phys. Rev. Lett. **87**, 270404 (2001).

3 J. A. Bergou, U. Herzog, and M. Hillery, *Quantum Filtering and Discrimination between Sets of Boolean Functions*, Phys. Rev. Lett. **90**, 257901 (2003).

4 A. Chefles and M. Sasaki, *Retrodiction of Generalized Measurement Outcomes* Phys. Rev. A **67**, 032112 (2003).

5 R. Duan, Y. Feng, and M. Ying, *Entanglement is Not Necessary for Perfect Discrimination between Unitary Operations*, Phys. Rev. Lett. **98**, 100503 (2007).

6 J. A. Bergou and M. Hillery, *Quantum State Filtering Applied to the Discrimination of Boolean Functions*, Phys. Rev. A **72**, 012302 (2005).

7 A. Chefles, A. Kitagawa, M. Takeoka, M. Sasaki, and J. Twamley *Unambiguous Discrimination among Oracle Operators*, J. Phys. A: Math. Theor. **40**. 10183 (2007).

8 G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Memory Effects in Quantum Channel Discrimination*, Phys. Rev. Lett. **101** 180501 (2008).

9 R. Duan, Y. Feng, and M. Ying, *Perfect Distinguishability of Quantum Operations*, Phys. Rev. Lett. **103**, 210501 (2009).

10 A. Ambainis, *Quantum Lower Bounds by Quantum Arguments*, Journal of Computer and System Science **64**, 750 (2002).

11 W. van Dam 1998 *Quantum Oracle Interrogation: Getting All Information For Almost Half the Price*, Proceedings of the 39th Annual IEEE Symposium on Foundations of Computer Science (FOCS) **362** (1998).

12 E. Fahri, J. Goldstone, S. Gutman, and M. Sipser, *Bound on the Number of Functions That Can Be Distinguished With k Quantum Queries*, Phys. Rev. A **60**, 4331 (1999).

13 A. Ambainis, K. Iwama, A. Kawachi, H. Masuda, R. H. Putra, and S. Yamashita, *Quantum Identification of Boolean Oracles*, Proceedings of STACS 2004, Lecture Notes in Computer Science **2996**, 105 (2004).

14 A. Ambainis, K. Iwama, A. Kawachi, R. Raymond, and S. Yamashita, *Improved Algorithms for Quantum Identification of Boolean Oracles*, Lecture Notes in Computer Science **4059**, 105 (2006).

15 L. Grover, *Quantum Mechanics Helps in Searching for a Needle in a Haystack*, Phys. Rev. Lett. **79**, 325 (1997).

16 E. Bernstein and U. Vazirani, *Quantum Complexity Theory*, Proceedings of the 25th Annual ACM Symposium on the Theory of Computing, 11 (1993).

17 G. Chiribella, G. M. D'Ariano, P. Perinotti P, and M. F. Sacchi, *Efficient Use of Quantum Resources for the Transmission of a Reference Frame*, Phys. Rev. Lett. **93**, 180503 (2004).

18 E. Bagan, M. Baig, and R. Muñz-Tapia, *Quantum Reverse Engineering and Reference-Frame Alignment Without Nonlocal Correlations*, Phys. Rev. A **70**, 030301(R) (2004).

19 J. von Korff and J. Kempe, *Quantum Advantage in Transmitting a Permutation*, Phys. Rev. Lett. **93** 260502 (2004).

20 D. Collins, L. Diosi, G. Gisin, S. Massar, and S. Popescu, *Quantum Gloves: Quantum States That Encode As Much As Possible Chirality and Nothing Else*, Phys. Rev. A **72**. 022304 (2005).

21 A. Hayashi, T. Hashimoto, and M. Horibe, *Extended Quantum Color Coding*, Phys. Rev. A **71** 012326 (2005).

22 G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Optimal Covariant Quantum Newtorks*, AIP Conf. Proc. **1110**, 47 (2008).

23 S. D. Bartlett, T. Rudolph, R. W. Spekkens, and P. S. Turner, *Quantum Communication Using a Bounded-Size Quantum Reference Frame*, New J. Phys. **11** 063013 (2009).

24 G. Chiribella, V. Giovannetti, L. Maccone, and P. Perinotti, *Teleportation Transfers Only Speakable Quantum Information*, Phys. Rev. A **86** 010304(R) (2012).

25 M. Skotiniotis, B. Kraus, and W. Dür, *Efficient Quantum Communication Under Collective Noise*, Quantum Information and Computation **13**. 0290 (2013).

26 A. Bisio, G. Chiribella, G. M. D'Ariano, S. Facchini, and P. Perinotti, *Optimal Quantum Learning of a Unitary Transformation*, Phys. Rev. A **81** 032324 (2010).

27 C. H. Bennett and S. J. Wiesner, *Communication via One- and Two-Particle Operators on Einstein-Podolsky-Rosen States*, Phys. Rev. Lett. **69** 2881 (1992).

28 G. Wang and M. Ying, *Unambiguous Discrimination Among Quantum Operations*, Phys. Rev. A **73**, 042301 (2006).

29 M. Rötteler, *Quantum Algorithms to Solve the Hidden Shift Problem for Quadratics and for Functions of Large Gowers Norm*, Letc. Notes Comp. Science, **5734,** 993 (2009).

30 C. Dankert, R. Cleve, J. Emerson, and E. Livine, *Exact and Approximate Unitary 2 - Designs and their Application to Fidelity Estimation*, Physical Review A **80**, 012304 (2009).

31 D. Gross, K. Audenaert, and J. Eisert, *Evenly Distributed Unitaries: On the Structure of Unitary Designs*, J. Math. Phys. **48**, 052104 (2007).

32 A. J. Scott, *Optimizing Quantum Process Tomography with Unitary 2-Designs*, J. Phys. A **41**, 055308 (2008).

33 A. Roy and A. J. Scott, *Unitary Designs and Codes*, Des. Codes Cryptogr. **53**, 13 (2009).

34 R. W. Goodman and N. Wallach, *Representations and invariants of the classical groups*, Cambridge University Press (2003).

35 A. Chefles, *Quantum operations, state transformations and probabilities*, Phys. Rev. A **65**, 052314 (2002).

36 G. Chiribella, G. M. D'Ariano, and P. Perinotti, *Quantum Circuits Architecture*, Phys. Rev. Lett. **101**, 060401 (2008).

37 G. Chiribella, G. M. D'Ariano and P. Perinotti, *Theoretical Framework for Quantum Networks*, Phys. Rev. A **80**, 022339 (2009).

38 G. Gutoski and J. Watrous, *Toward a General Theory of Quantum Games*, Proceedings of the 39th ACM Symposium on the Theory of Computation (STOC) **39**, 565 (2007).

39 G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, *Covariant Quantum Measurements That Maximize The Likelihood*, Phys. Rev. A **70**, 062105 (2004).

40 G. Chiribella, G. M. D'Ariano, P. Perinotti, and M. F. Sacchi, *Maximum Likelohood Estimation of an Unknown Group Transformation*, Int. J. Quantum Inf. **4**, 453 (2006).

41 P. Hausladen and W. K. Wootters, *A 'Pretty Good' Measurement for Distinguishing Quantum States*, J. Mod. Opt. **41**, 2385 (1994).

Symmetries of Codeword Stabilized Quantum Codes*

Salman Beigi[1], Jianxin Chen[2,3], Markus Grassl[4], Zhengfeng Ji[3], Qiang Wang[5], and Bei Zeng[2,3]

1 **School of Mathematics, Institute for Research in Fundamental Sciences (IPM)**
Niavaran Square, Tehran, Iran
salman.beigi@gmail.com

2 **Department of Mathematics & Statistics, University of Guelph**
50 Stone Road East, Guelph, Ontario, Canada
{chenkenshin,zengbei}@gmail.com

3 **Institute for Quantum Computing**
200 University Avenue West, Waterloo, Ontario, Canada
jizhengfeng@gmail.com

4 **Centre for Quantum Technologies, National University of Singapore**
3 Science Drive 2, Singapore 117543
Markus.Grassl@nus.edu.sg

5 **School of Mathematics and Statistics, Carleton University**
1125 Colonel By Drive, Ottawa, Ontario, Canada
wang@math.carleton.ca

Abstract

Symmetry is at the heart of coding theory. Codes with symmetry, especially cyclic codes, play an essential role in both theory and practical applications of classical error-correcting codes. Here we examine symmetry properties for codeword stabilized (CWS) quantum codes, which is the most general framework for constructing quantum error-correcting codes known to date. A CWS code $\mathcal{Q}$ can be represented by a self-dual additive code $\mathcal{S}$ and a classical code $\mathcal{C}$, i. e., $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$, however this representation is in general not unique. We show that for any CWS code $\mathcal{Q}$ with certain permutation symmetry, one can always find a self-dual additive code $\mathcal{S}$ with the same permutation symmetry as $\mathcal{Q}$ such that $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$. As many good CWS codes have been found by starting from a chosen $\mathcal{S}$, this ensures that when trying to find CWS codes with certain permutation symmetry, the choice of $\mathcal{S}$ with the same symmetry will suffice. A key step for this result is a new canonical representation for CWS codes, which is given in terms of a unique decomposition as union stabilizer codes. For CWS codes, so far mainly the standard form $(\mathcal{G}, \mathcal{C})$ has been considered, where $\mathcal{G}$ is a graph state. We analyze the symmetry of the corresponding graph of $\mathcal{G}$, which in general cannot possess the same permutation symmetry as $\mathcal{Q}$. We show that it is indeed the case for the toric code on a square lattice with translational symmetry, even if its encoding graph can be chosen to be translational invariant.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases CWS Codes, Union Stabilizer Codes, Permutation Symmetry, Toric Code

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.192

* This work was partially supported by NSERC, CIFAR, and IARPA.

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 192–206

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Coding theory is an important component of information theory having a long history dating back to Shannon's seminal 1948 paper that laid the ground for information theory [21]. Coding theory is at the heart of reliable communication, where codes with symmetry, especially cyclic codes, such as the Reed-Solomon codes, are among the most widely used codes in practice [19].

In recent years, it has become evident that quantum communication and computation offer the possibility of secure and high rate information transmission, fast computational solution of certain important problems, and efficient physical simulation of quantum phenomena. However, quantum information processing depends on the identification of suitable quantum error-correcting codes (QECC) to make such processes and machines robust against faults due to decoherence, ubiquitous in quantum systems. Quantum coding theory has hence been extensively developed during the past 15 years [3, 9, 20].

Codeword stabilized (CWS) quantum codes are by far the most general construction of QECC [6]. A CWS code $\mathcal{Q}$ can be represented by a stabilizer state (i. e. a self-dual additive code) $\mathcal{S}$ and a classical code $\mathcal{C}$, i. e. $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$. When $\mathcal{C}$ is a linear code, the corresponding CWS code $\mathcal{Q}$ is actually a stabilizer code. Also, any CWS code is local Clifford equivalent to a standard form $(\mathcal{G}, \mathcal{C})$, where $\mathcal{G}$ is a graph state [6].

The CWS construction encompasses stabilizer (additive) codes and all the known non-additive codes with good parameters. It also leads to many new codes with good parameters, or good algebraic/combinatorial properties, through both analytical and numerical methods. Alternative perspectives of CWS codes have also been analyzed, including the union stabilizer codes (USt) method [11, 12], and the codes based on graphs [18, 23]. Concatenated codes and their generalizations using CWS codes have been developed [1], and decoding methods for CWS codes have been studied as well [17].

Given all the evidence that the CWS framework is a powerful method to construct and analyze QECC, it remains unclear to what extent the stabilizer state $\mathcal{S}$ and the classical code $\mathcal{C}$ can represent the symmetry of the CWS code $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ in general. Given the vital importance that the code symmetry plays in coding theory, this understanding becomes crucial since if such a correspondence exists, it can provide practical methods for constructing CWS codes with desired symmetry from $\mathcal{S}$ and/or $\mathcal{C}$ with corresponding symmetry.

Unfortunately, there is no immediate clue what answer one can hope for. First of all, the representation $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ is not unique. So for a given CWS code $\mathcal{Q}$, there might be some stabilizer states $\mathcal{S}$ and/or classical codes $\mathcal{C}$ which are more symmetric than others. Perhaps the best known example is the CWS representation for the five-qubit code $\mathcal{Q}_5$, where in the ideal case $\mathcal{S}$ can be chosen as a graph state corresponding to the pentagon graph, and the classical code $\mathcal{C}$ is chosen as the repetition code $\{00000, 11111\}$. In this case, both $\mathcal{S}$ and $\mathcal{C}$ nicely represent the cyclic symmetry of the five-qubit code.

However, there are known 'bad cases', too. One example is the seven-qubit Steane code $\mathcal{Q}_7$, where although the code itself is cyclic, one cannot find any $\mathcal{S}$ corresponding to a cyclic graph, even if local Clifford operations are allowed [10]. Nonetheless, we know that the stabilizer group for this code $\mathcal{Q}_7$ is invariant under cyclic shifts, and the logical Z operator can be chosen as $Z_L = Z^{\otimes 7}$, therefore the logical $|0\rangle_L$ can be chosen as a cyclic stabilizer code. This is to say, there exists a representation for $\mathcal{Q}_7 = (\mathcal{S}, \mathcal{C})$ such that $\mathcal{S}$ is cyclic. In general it remains unclear under which conditions a representation for cyclic CWS code with a cyclic stabilizer state $\mathcal{S}$ exists.

In this work, we address the symmetry properties of CWS codes. We are interested in

the permutation symmetry of CWS codes, which includes the important category of cyclic codes. Our main question is, to which extent can the representation $(\mathcal{S},\mathcal{C})$ and the standard form $(\mathcal{G},\mathcal{C})$ reflect the symmetry of the corresponding CWS code $\mathcal{Q}$. We show that for any CWS code $\mathcal{Q}$ with permutation symmetry, one can always find a stabilizer state $\mathcal{S}$ with the same permutation symmetry as $\mathcal{Q}$ such that $\mathcal{Q} = (\mathcal{S},\mathcal{C})$. As many good CWS codes are found by starting from a chosen $\mathcal{S}$, this ensures that when trying to find CWS codes with certain permutation symmetry, the choice of $\mathcal{S}$ with the same symmetry will suffice. A key step to reach this main result is to obtain a canonical representation for CWS codes, which is in terms of a unique decomposition as union stabilizer codes.

We know that for the standard form of CWS codes using graph states, it is not always possible to find a graph with the same permutation symmetry. This is partially due to the fact that the local Clifford operations transforming the CWS code into the standard form may break the permutation symmetry of the original code. Also, the graphs usually can only represent the symmetry of the stabilizer generators of the stabilizer state, but not the symmetry of the stabilizer state in general. We show that this is indeed the case for the toric code on a two-dimensional square lattice with translational symmetry, even if its encoding graph can be chosen to be translational invariant.

However, we show that the converse always holds, i. e., any graph $\mathcal{G}$ and classical code $\mathcal{C}$ with certain permutation symmetry yields a CWS code $\mathcal{Q} = (\mathcal{G},\mathcal{C})$ with the same symmetry.

2 Preliminaries

The single-qudit (generalized) Pauli group is generated by the operators X and Z acting on the qudit Hilbert space $\mathbb{C}^p$, satisfying $ZX = \omega XZ$, where $\omega = \omega_p = \exp 2i\pi/p$. For simplicity, throughout the paper, we assume that p is a prime, although our results naturally extend to prime powers. Denote the computational basis of $\mathbb{C}^p$ by $\{|j\rangle \colon j = 0, 1, \ldots, p-1\}$. Then, without loss of generality, we can fix the operators X and Z such that $X|j\rangle = |j+1\rangle$ and $Z|j\rangle = \omega^j|j\rangle$, respectively. Let I be the identity operator. The set $\{X^aZ^b \colon a, b = 0, \ldots, p-1\}$ of p^2 operators forms a so-called nice unitary error basis which is a particular basis for the vector space of $p \times p$ matrices [15, 16].

The n-qudit Pauli group $\mathcal{P}_n$ consists of all local operators of the form $\boldsymbol{M} = \alpha_{\boldsymbol{M}} M_1 \otimes \cdots \otimes M_n$, where $\alpha_{\boldsymbol{M}} = \omega^k$ for some integer k is an overall phase factor, and $M_i = X_i^a Z_i^b$ for some $a, b \in \{0, 1, \ldots, p-1\}$, is an element of the single-qudit Pauli group of qudit i. We can write $\boldsymbol{M}$ as $\alpha_{\boldsymbol{M}}(M_1)_1(M_2)_2 \ldots (M_n)_n$ or $\alpha_{\boldsymbol{M}} M_1 M_2 \ldots M_n$ when it is clear what the qudit labels are. The weight of an operator $\boldsymbol{M}$ is the number of tensor factors M_i that differ from identity.

The n-qudit Clifford group $\mathcal{L}_n$ is the group of $p^n \times p^n$ unitary matrices that map $\mathcal{P}_n$ to itself under conjugation. The n-qudit local Clifford group is a subgroup in $\mathcal{L}_n$ containing elements of the form $M_1 \otimes \cdots \otimes M_n$, where each M_i is a single qudit Clifford operation, i. e., $M_i \in \mathcal{L}_1$.

A stabilizer group $\mathcal{S}$ in the Pauli group $\mathcal{P}_n$ is defined as an abelian subgroup of $\mathcal{P}_n$ which does not contain ωI. A stabilizer consists of p^m Pauli operators for some $m \leq n$. As the operators in a stabilizer commute with each other, they can be simultaneously diagonalized. The common eigenspace of eigenvalue 1 is a stabilizer quantum code $\mathcal{Q} = (\!(n, K, d)\!)_p$ with length n, dimension $K = p^{n-m}$, and minimum distance d. The projection $P_{\mathcal{Q}}$ onto the code $\mathcal{Q}$ can be expressed as

$$P_{\mathcal{Q}} = \frac{1}{|\mathcal{S}|} \sum_{\boldsymbol{M} \in \mathcal{S}} \boldsymbol{M}. \tag{1}$$

The centralizer $C(\mathcal{S})$ of the stabilizer $\mathcal{S}$ is given by the elements in $\mathcal{P}_n$ which commute with all elements in $\mathcal{S}$. For $m < n$, the minimum distance d of the code $\mathcal{Q}$ is the minimum weight of all elements in $C(\mathcal{S}) \setminus \mathcal{S}$.

If $m = n$, then there exists a unique n-qudit state $|\psi\rangle$ such that $\boldsymbol{M}|\psi\rangle = |\psi\rangle$ for every $\boldsymbol{M} \in \mathcal{S}$. Such a state $|\psi\rangle$ is called a stabilizer state, and the group $\mathcal{S} = \mathcal{S}(|\psi\rangle)$ is called the stabilizer of $|\psi\rangle$. A stabilizer state can also be viewed as a self-dual code over the finite field $\mathbb{F}_{p^2}$ under the trace inner product [7]. For a stabilizer state, the minimum distance is defined as the minimum weight of the non-trivial elements in $\mathcal{S}(|\psi\rangle)$ [7].

A union stabilizer (USt) code of length n is characterized by a stabilizer code with stabilizer $\mathcal{S} = \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m \rangle$, where $\boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m$ are m independent generators, and a classical code $\mathcal{C}$ over $\mathbb{F}_p$ of length m. Note that for a given $\mathcal{S}$, the choice of the m generators $\boldsymbol{g}_j$ is not unique. Now for a classical code $\mathcal{C}$ of length m with K codewords, for each codeword $\mathbf{c} = (c_1, c_2, \ldots, c_m) \in \mathcal{C}$, the corresponding quantum code is given by the subspace $V_{\boldsymbol{c}}$ stabilized by $\omega^{c_1}\boldsymbol{g}_1$, $\omega^{c_2}\boldsymbol{g}_2$, $\ldots$, $\omega^{c_m}\boldsymbol{g}_m$. Note that for $\boldsymbol{c} \neq \boldsymbol{c}' \in \mathcal{C}$, the subspaces $V_{\boldsymbol{c}}$ and $V_{\boldsymbol{c}'}$ are mutually orthogonal. The corresponding USt code is then given by the subspace $\bigoplus_{\boldsymbol{c}} V_{\boldsymbol{c}}$.

Therefore, the combination of $\mathcal{S}$ (more precisely, the generators of $\mathcal{S}$) and $\mathcal{C}$ gives an $((n, 2^{n-m}K))_p$ USt quantum code $\mathcal{Q}$. Hence we denote a USt code $\mathcal{Q}$ by $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$. The projection onto $\mathcal{Q}$ can be expressed as

$$P_{\mathcal{Q}} = \sum_{\boldsymbol{c} \in \mathcal{C}} \frac{1}{p^m} \sum_{\boldsymbol{y} \in \mathbb{F}_p^m} \omega^{\boldsymbol{c}\cdot\boldsymbol{y}} \boldsymbol{g}_1^{y_1} \cdots \boldsymbol{g}_m^{y_m}, \tag{2}$$

where we identify the elements y_i of the finite field with integers modulo p.

A CWS code $\mathcal{Q}$ of length n is a USt code with $m = n$. That is, it is characterized by a stabilizer state with stabilizer $\mathcal{S}$ and a classical code $\mathcal{C}$ of length n. For a CWS code $\mathcal{Q}$ given by $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$, the stabilizer $\mathcal{S}$ always corresponds to a unique stabilizer state. We will then refer to $\mathcal{S}$ as the stabilizer state when no confusion arises.

For a CWS code, the projection $P_{\mathcal{Q}}$ onto the code space is given by

$$P_{\mathcal{Q}} = \sum_{\boldsymbol{t} \in \mathcal{C}} \frac{1}{p^n} \sum_{\boldsymbol{x} \in \mathbb{F}_p^n} \omega^{\boldsymbol{t}\cdot\boldsymbol{x}} \boldsymbol{g}_1^{x_1} \cdots \boldsymbol{g}_n^{x_n}, \tag{3}$$

where we again identify the elements x_i of the finite field with integers modulo p.

A CWS code has a permutation symmetry σ if

$$P_{\mathcal{Q}}^{\sigma} = P_{\mathcal{Q}}, \tag{4}$$

where $P_{\mathcal{Q}}^{\sigma}$ is the projection onto the space obtained by permuting the qudits of the code $\mathcal{Q}$ according to σ.

3 Canonical form of CWS codes

For a given a CWS code $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$, there might exist another stabilizer state $\mathcal{S}'$ and another classical code $\mathcal{C}'$ such that $\mathcal{Q} = (\mathcal{S}', \mathcal{C}')$. In other words, the representation of a CWS code by the stabilizer state $\mathcal{S}$ and the classical code $\mathcal{C}$ is non-unique.

In order to discuss the relationship between the symmetry of the CWS code $\mathcal{Q}$ and that of the stabilizer state $\mathcal{S}$, we first need to explore the relationship between the different representations of $\mathcal{Q}$ (i. e., the relationship between $\mathcal{S}$ and $\mathcal{S}'$, as well as the relationship between $\mathcal{C}$ and $\mathcal{C}'$).

Let us start by recalling that a stabilizer code can be viewed as a CWS code where the classical code is a linear code [6]. A simple way to see this is that for a given stabilizer code $\mathcal{Q}_s$ with stabilizer generated by $\mathcal{S} = \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m \rangle$, which is a code of dimension p^{n-m}, we can choose the larger stabilizer $\mathcal{S}' = \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \ldots, \boldsymbol{g}_m, \bar{Z}_1, \ldots, \bar{Z}_{n-m} \rangle$, where $\bar{Z}_1, \ldots, \bar{Z}_{n-m} \in C(\mathcal{S})$ mutually commute. Now choose the classical code $\mathcal{C}' = \big\{(0, \ldots, 0, x_{m+1}, \ldots, x_n) \colon x_j \in \{0, \ldots, p-1\}\big\}$ of length n with p^{n-m} codewords, where the first m coordinates of each codeword are zero. Then we have $\mathcal{Q}_s = (\mathcal{S}', \mathcal{C}')$, i. e., the stabilizer code $\mathcal{Q}_s$ can then be viewed as a CWS code with stabilizer state $\mathcal{S}'$ and classical code $\mathcal{C}'$. However, note that the choice of $\mathcal{S}'$ (and hence $\mathcal{C}'$) is non-unique, as in particular the choice of $\bar{Z}_1, \ldots, \bar{Z}_{n-s} \in C(\mathcal{S})$ is non-unique.

▶ **Example 1.** As an example, consider the five-qubit code with stabilizer

$$\boldsymbol{g}_1 = XZZXI, \quad \boldsymbol{g}_2 = IXZZX, \quad \boldsymbol{g}_3 = XIXZZ, \quad \boldsymbol{g}_4 = ZXIXZ. \tag{5}$$

In the CWS picture, the stabilizer state can be chosen as

$$\mathcal{S} = \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3, \boldsymbol{g}_4, \boldsymbol{Z}_L \rangle, \tag{6}$$

where $\boldsymbol{Z}_L = Z^{\otimes 5}$ is the logical Z operator. Alternatively, one can choose the stabilizer state

$$\mathcal{S}' = \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3, \boldsymbol{g}_4, \boldsymbol{X}_L \rangle, \tag{7}$$

where $\boldsymbol{X}_L = X^{\otimes 5}$ is the logical X operator. For both S and S', the classical code can be chosen as $\mathcal{C} = \{00000, 00001\}$.

Similarly, a USt code $(\mathcal{S}, \mathcal{C})$ can be viewed as a CWS code $(\mathcal{S}', \mathcal{C}')$ with the classical code $\mathcal{C}'$ of length n possessing some coset structure, i. e., $\mathcal{C}' = \bigcup_{\boldsymbol{t}_i \in \tilde{\mathcal{C}}} \mathcal{C}_0 + \boldsymbol{t}_i$, where $\mathcal{C}_0$ is a linear code. This linear code $\mathcal{C}_0$ of length n can be readily chosen as the classical code for the CWS representation of the stabilizer code $\mathcal{S}$. The code $\tilde{\mathcal{C}}$ of length n can be derived from $\mathcal{C}$ of length m by appending $n - m$ zero coordinates. However, again, the choices of $\mathcal{S}'$ and $\mathcal{C}'$ are non-unique.

In the general situation, we have some freedom in choosing the stabilizer state when representing a stabilizer code or a USt code in the CWS framework. Consequently, for a given CWS code $\mathcal{Q}$, there are also many different ways to write it in terms of a USt code in general. We will show, however, that we can always obtain a unique stabilizer $\mathcal{S}$, when expressing a given CWS code as a USt code. The following theorem gives a canonical form for any CWS code.

▶ **Theorem 2.** *Every CWS code has a unique representation as a union stabilizer code.*

Proof. To prove this theorem, we will need some lemmas.

▶ **Lemma 3** (translational invariant codes)**.** *Let $\mathcal{C} \subset \mathbb{F}_p^n$ be a code over $\mathbb{F}_p$ with $|\mathcal{C}| = M$ and assume that for some non-zero $\boldsymbol{s} \in \mathbb{F}_p^n$ we have $\mathcal{C} = \mathcal{C} + \boldsymbol{s}$, i. e., the code is invariant with respect to translation by $\boldsymbol{s}$. Then $\mathcal{C}$ can be written as a disjoint union of cosets of the one-dimensional space $\mathcal{C}_0 = \langle \boldsymbol{s} \rangle$ generated by $\boldsymbol{s}$, i. e.,*

$$\mathcal{C} = \bigcup_{\boldsymbol{t}_i \in \mathcal{C}'} \mathcal{C}_0 + \boldsymbol{t}_i,$$

where $\mathcal{C}' \subset \mathbb{F}_p^n$ with $|\mathcal{C}'| = M/p$.

Proof. By assumption, for every $\boldsymbol{x} \in \mathcal{C}$, the vector $\boldsymbol{x} + \boldsymbol{s}$ is in the code as well. Hence we can arrange the elements of $\mathcal{C}$ as follows:

$\mathcal{C}'$	$\boldsymbol{t}_1$	$\boldsymbol{t}_2$	$\dots$	$\boldsymbol{t}_{M/p}$
$\mathcal{C}' + \boldsymbol{s}$	$\boldsymbol{t}_1 + \boldsymbol{s}$	$\boldsymbol{t}_2 + \boldsymbol{s}$	$\dots$	$\boldsymbol{t}_{M/p} + \boldsymbol{s}$
$\mathcal{C}' + 2\boldsymbol{s}$	$\boldsymbol{t}_1 + 2\boldsymbol{s}$	$\boldsymbol{t}_2 + 2\boldsymbol{s}$	$\dots$	$\boldsymbol{t}_{M/p} + 2\boldsymbol{s}$
$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
$\mathcal{C}' + (p-1)\boldsymbol{s}$	$\boldsymbol{t}_1 + (p-1)\boldsymbol{s}$	$\boldsymbol{t}_2 + (p-1)\boldsymbol{s}$	$\dots$	$\boldsymbol{t}_{M/p} + (p-1)\boldsymbol{s}$

Every column in this arrangements is a coset $\mathcal{C}_0 + \boldsymbol{t}_i$. ◀

▶ **Lemma 4** (vanishing character sum). *Let $\mathcal{C} \subset \mathbb{F}_q^n$ be an arbitrary code of length n. Assume that the function*

$$f\colon \mathbb{F}_p^n \to \mathbb{C}; \quad f(\boldsymbol{y}) = \sum_{\boldsymbol{c}\in\mathcal{C}} \omega^{\boldsymbol{c}\cdot\boldsymbol{y}},$$

where $\omega = \exp(2\pi i/p)$, vanishes outside a proper subspace $V_0 < \mathbb{F}_p^n$. Then there exists a non-zero vector $\boldsymbol{s} \in \mathbb{F}_p^n$ such that $\mathcal{C} = \mathcal{C} + \boldsymbol{s}$. What is more, the code $\mathcal{C}$ can be written as a union of cosets of the linear code $\mathcal{C}_0 = V_0^\perp$, i. e.,

$$\mathcal{C} = \bigcup_{\boldsymbol{t}\in\mathcal{C}'} \mathcal{C}_0 + \boldsymbol{t}. \tag{8}$$

Proof. Let $\chi_C(\boldsymbol{y})$ denote the characteristic function of the code $\mathcal{C}$, i. e., $\chi_{\mathcal{C}}(\boldsymbol{y}) \in \{0,1\}$, and $\chi_{\mathcal{C}}(\boldsymbol{y}) = 1$ if and only if $\boldsymbol{y} \in \mathcal{C}$. Define $g(\boldsymbol{y}) = 1 - (1-\omega)\chi_{\mathcal{C}}(\boldsymbol{y})$. Then $g(\boldsymbol{y}) = \omega^{\chi_{\mathcal{C}}(\boldsymbol{y})}$.

The Fourier transform of $g(\boldsymbol{y})$ over $\mathbb{F}_p^n$ reads

$$\begin{aligned}
\hat{g}(\boldsymbol{y}) &= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \omega^{\boldsymbol{x}\cdot\boldsymbol{y}} g(\boldsymbol{x}) \\
&= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \omega^{\boldsymbol{x}\cdot\boldsymbol{y}} (1-(1-\omega)\chi_{\mathcal{C}}(\boldsymbol{x})) = \sqrt{p^n}\delta_{\boldsymbol{y},\boldsymbol{0}} - \frac{1-\omega}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \omega^{\boldsymbol{x}\cdot\boldsymbol{y}} \chi_{\mathcal{C}}(\boldsymbol{x}) \\
&= \sqrt{p^n}\delta_{\boldsymbol{y},\boldsymbol{0}} - \frac{1-\omega}{\sqrt{p^n}} \sum_{\boldsymbol{c}\in C} \omega^{\boldsymbol{c}\cdot\boldsymbol{y}} = \sqrt{p^n}\delta_{\boldsymbol{y},\boldsymbol{0}} - \frac{1-\omega}{\sqrt{p^n}} f(\boldsymbol{y}),
\end{aligned}$$

where $\delta_{\boldsymbol{y},\boldsymbol{0}} = 1$ if $\boldsymbol{y} = \boldsymbol{0}$, and $\delta_{\boldsymbol{y},\boldsymbol{0}} = 0$ otherwise.

This shows that for $\boldsymbol{y} \neq \boldsymbol{0}$, the Fourier transform $\hat{g}(\boldsymbol{y})$ is proportional to the function $f(\boldsymbol{y})$, and hence $\hat{g}$ vanishes outside of V_0 as well. Recall that that $\dim V_0 \leq n-1$, as V_0 is a proper subspace by assumption. Let $\boldsymbol{s} \in V_0^\perp$ be a non-zero vector that is orthogonal to all vectors in V_0. Furthermore, let $V_0^c = \mathbb{F}_p^n \setminus V_0$ denote the set-complement of V_0 in the full vector space.

We want to show that the code $\mathcal{C}$ is invariant with respect to translations by $\boldsymbol{s}$, i. e., $\mathcal{C} = \mathcal{C} + \boldsymbol{s}$ or equivalently, $\chi_{\mathcal{C}}(\boldsymbol{y} + \boldsymbol{s}) = \chi_{\mathcal{C}}(\boldsymbol{y})$. This is in turn equivalent to showing that $g(\boldsymbol{y}) = g(\boldsymbol{y} + \boldsymbol{s})$. In the following, $\mathcal{F}^{-1}$ denotes the inverse Fourier transform:

$$\begin{aligned}
g(\boldsymbol{y}+\boldsymbol{s}) = (\mathcal{F}^{-1}\hat{g})(\boldsymbol{y}+\boldsymbol{s}) &= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \omega^{-\boldsymbol{x}\cdot(\boldsymbol{y}+\boldsymbol{s})} \hat{g}(\boldsymbol{x}) \\
&= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in V_0} \omega^{-\boldsymbol{x}\cdot(\boldsymbol{y}+\boldsymbol{s})} \hat{g}(\boldsymbol{x}) + \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in V_0^c} \omega^{-\boldsymbol{x}\cdot(\boldsymbol{y}+\boldsymbol{s})} \hat{g}(\boldsymbol{x}) \\
&= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in V_0} \omega^{-\boldsymbol{x}\cdot\boldsymbol{s}} \omega^{-\boldsymbol{x}\cdot\boldsymbol{y}} \hat{g}(\boldsymbol{x})
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in V_0} \omega^{-\boldsymbol{x}\cdot\boldsymbol{y}} \hat{g}(\boldsymbol{x}) \\
&= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in V_0} \omega^{-\boldsymbol{x}\cdot\boldsymbol{y}} \hat{g}(\boldsymbol{x}) + \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in V_0^c} \omega^{-\boldsymbol{x}\cdot\boldsymbol{y}} \hat{g}(\boldsymbol{x}) \\
&= \frac{1}{\sqrt{p^n}} \sum_{\boldsymbol{x}\in \mathbb{F}_p^n} \omega^{-\boldsymbol{x}\cdot\boldsymbol{y}} \hat{g}(\boldsymbol{x}) \\
&= (\mathcal{F}^{-1}\hat{g})(\boldsymbol{y}) = g(\boldsymbol{y}).
\end{aligned}
$$

Here we have used the fact that $\hat{g}(\boldsymbol{x})$ vanishes outside of V_0 and that $\boldsymbol{s}$ is orthogonal to all vectors in V_0.

From Lemma 3, it follows that the code $\mathcal{C}$ can be written as a union of cosets of the code $\mathcal{C}_0 = V_0^\perp$ generated by all vectors $\boldsymbol{s}$ that are orthogonal to V_0. ◀

Now we are ready to prove Theorem 2. Let $P_\mathcal{Q}$ denote the projection operator onto a CWS code $\mathcal{Q} = (\!(n, K, d)\!)_p$, i. e.

$$
\begin{aligned}
P_\mathcal{Q} = \sum_{\boldsymbol{t}\in\mathcal{C}} \frac{1}{p^n} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \omega^{\boldsymbol{t}\cdot\boldsymbol{x}} \boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} &= \frac{1}{p^n} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \left(\sum_{\boldsymbol{t}\in\mathcal{C}} \omega^{\boldsymbol{t}\cdot\boldsymbol{x}} \right) \boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} \\
&= \frac{1}{p^n} \sum_{\boldsymbol{x}\in\mathbb{F}_p^n} \alpha_{\boldsymbol{x}} \boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} \qquad (9)
\end{aligned}
$$

where $\boldsymbol{g}_1, \dots, \boldsymbol{g}_n$ are the generators of the stabilizer, and $\mathcal{C} = (n, K)_p$ is a classical code.

First note that the coefficients $\alpha_{\boldsymbol{x}}$ in (9) are uniquely determined since the p^n operators $\{\boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} : \boldsymbol{x} \in \mathbb{F}_p^n\}$ are a subset of the error-basis of linear operators on the space $\mathbb{C}^{p^n}$. The coefficient $\alpha_{\boldsymbol{x}}$ is proportional to $\mathrm{tr}(\boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} \cdot P_\mathcal{Q})$. On the other hand, $\alpha_{\boldsymbol{x}} = \sum_{\boldsymbol{t}\in\mathcal{C}} \omega^{\boldsymbol{t}\cdot\boldsymbol{x}} = f(\boldsymbol{x})$, where $f(\boldsymbol{x})$ is the function appearing in Lemma 4. So if the coefficients $\alpha_{\boldsymbol{x}} = f(\boldsymbol{x})$ vanish outside of a proper subspace $V_0 < \mathbb{F}_p^n$, the classical code $\mathcal{C}$ can be decomposed as union of cosets of $\mathcal{C}_0 = V_0^\perp$. Then (9) can be re-written as follows:

$$
\begin{aligned}
P_\mathcal{Q} &= \frac{1}{p^n} \sum_{\boldsymbol{x}\in V_0} \left(\sum_{\boldsymbol{t'}\in\mathcal{C}'} \sum_{\boldsymbol{c}\in\mathcal{C}_0} \omega^{(\boldsymbol{t'}+\boldsymbol{c})\cdot\boldsymbol{x}} \right) \boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} \\
&= \frac{1}{p^n} \sum_{\boldsymbol{x}\in V_0} \left(\sum_{\boldsymbol{c}\in\mathcal{C}_0} \omega^{\boldsymbol{c}\cdot\boldsymbol{x}} \sum_{\boldsymbol{t'}\in\mathcal{C}'} \omega^{\boldsymbol{t'}\cdot\boldsymbol{x}} \right) \boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} \\
&= \frac{|\mathcal{C}_0|}{p^n} \sum_{\boldsymbol{x}\in V_0} \left(\sum_{\boldsymbol{t'}\in\mathcal{C}'} \omega^{\boldsymbol{t'}\cdot\boldsymbol{x}} \right) \boldsymbol{g}_1^{x_1} \dots \boldsymbol{g}_n^{x_n} \qquad (10)
\end{aligned}
$$

In the last step we have used the fact that the spaces V_0 and $\mathcal{C}_0$ are orthogonal to each other, i. e., the inner product $\boldsymbol{c}\cdot\boldsymbol{x}$ vanishes. Now assume that the space V_0 has dimension m and that $\{\boldsymbol{b}_1, \dots, \boldsymbol{b}_m\} \subset \mathbb{F}_p^n$ is a basis of V_0. Then every vector $\boldsymbol{x} \in V_0$ can be expressed as $\boldsymbol{x} = \sum_{j=1}^m y_j \boldsymbol{b}_j$. For every $\boldsymbol{t'} \in \mathcal{C}'$ we define the vectors $\boldsymbol{s} \in \mathbb{F}_p^m$ with $s_j = \sum_{i=1}^n t_i b_{ji}$, forming another classical code $\mathcal{D} \subset \mathbb{F}_p^m$. Further, we define the m operators $\tilde{\boldsymbol{g}}_j = \prod_{i=1}^n \boldsymbol{g}_i^{b_{ji}}$. This allows us to express (10) as

$$
P_\mathcal{Q} \;=\; \frac{1}{p^m} \sum_{\boldsymbol{y}\in\mathbb{F}_p^m} \left(\sum_{\boldsymbol{s}\in\mathcal{D}} \omega^{\boldsymbol{s}\cdot\boldsymbol{y}} \right) \tilde{\boldsymbol{g}}_1^{y_1} \dots \tilde{\boldsymbol{g}}_m^{y_m}. \qquad (11)
$$

Hence, whenever the classical code associated to a CWS code has some non-trivial shift invariance, the projection onto a CWS code can be expressed as a projection onto a USt code (cf. (2)), thereby increasing the dimension of the underlying stabilizer code and reducing the size of the classical code. In order to obtain a unique representation, we may assume that the stabilizer code is of maximal dimension, and hence the classical code is "without any linear structure."

In order to show uniqueness, consider the coefficients $\operatorname{tr}(\boldsymbol{M} \cdot P_{\mathcal{Q}})$ of the expansion of the projection $P_{\mathcal{Q}}$ in terms of the operator basis formed by the n-qudit Pauli matrices $\boldsymbol{M}$. Clearly, we have $\{\boldsymbol{M} \colon \operatorname{tr}(\boldsymbol{M} \cdot P_{\mathcal{Q}}) \neq 0\} \subset \mathcal{S} = \langle \tilde{\boldsymbol{g}}_1, \ldots, \tilde{\boldsymbol{g}}_m \rangle$. If the group $\mathcal{S}' = \langle \boldsymbol{M} \colon \operatorname{tr}(\boldsymbol{M} \cdot P_{\mathcal{Q}}) \neq 0 \rangle$ was a proper subgroup of $\mathcal{S}$, the coefficients $\sum_{\boldsymbol{s} \in \mathcal{D}} \omega^{\boldsymbol{s} \cdot \boldsymbol{y}}$ would vanish for $\boldsymbol{y}$ outside a proper subspace $V_0 < \mathbb{F}_p^m$, contradicting the assumption the classical code $\mathcal{D}$ has no linear structure.

Note that the stabilizer $\mathcal{S}$ is only unique up to the choice of some phase factors of the error basis. For example, replacing $\tilde{\boldsymbol{g}}_1$ by $\omega \tilde{\boldsymbol{g}}_1$ will introduce some phase factor which has to be compensated by changing the first coordinate s_1 of the codewords $\boldsymbol{s}$ of the classical code $\mathcal{D}$. To finally fix these degrees of freedom, we can enforce $\boldsymbol{g}_i = M_1 \otimes \cdots \otimes M_n$, with $M_j = X_j^a Z_j^b$ for $j = 1, 2, \ldots, n$ and $a, b \in \{0, 1, \ldots p-1\}$. ◀

4 Symmetries of the stabilizer state of a CWS code

We are now ready to discuss the relationship between the symmetries of the CWS code $\mathcal{Q}$ and that of the corresponding stabilizer state $\mathcal{S}$.

▶ **Theorem 5.** *For any CWS code $\mathcal{Q}$ with permutation symmetry σ, there exists a stabilizer state $\mathcal{S}$ with the same permutation symmetry σ such that $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$.*

Proof. To prove this theorem, we will need some lemmas.

▶ **Lemma 6.** *If the projection operator $P_{\mathcal{C}}$ given in Eq. (9) is invariant under a permutation σ of the qudits, then the stabilizer code related to expressing $P_{\mathcal{C}}$ in terms of a USt code as in Eq. (11) is invariant with respect to the permutation as well.*

Proof. The statement follows directly from the uniqueness of the stabilizer group $\mathcal{S} = \langle \tilde{\boldsymbol{g}}_1, \ldots, \tilde{\boldsymbol{g}}_m \rangle$ generated by the operators in Eq. (11). ◀

We now prove a lemma for a special case of Theorem 5, when the CWS code is a Calderbank–Shor–Steane (CSS) code [4, 22].

▶ **Lemma 7.** *For a CSS code $\mathcal{Q}$ with permutation symmetry σ, there exists a stabilizer state $|\psi\rangle \in \mathcal{Q}$ such that $|\psi\rangle$ has the same permutation symmetry as $\mathcal{Q}$.*

Proof. For a CSS code $\mathcal{Q}$, the stabilizer generators can always be chosen such that every generator is either a tensor product of powers of X (denoted by $\mathcal{S}_X$) or a tensor product of powers of Z (denoted by $\mathcal{S}_Z$). We can use the following matrix form:

$$\left[\begin{array}{c|c} \mathcal{S}_X & 0 \\ 0 & \mathcal{S}_Z \end{array}\right]$$

As the permutation symmetry σ of $\mathcal{Q}$ does not change the type of an operator, both $\mathcal{S}_X$ and $\mathcal{S}_Z$ have necessarily the same symmetry σ. Furthermore, the logical operators can also be chosen as either tensor products of powers of X or tensor products of powers of Z, which correspond to the dual of the classical codes associated to either the Z stabilizers or the X stabilizers, respectively. Without loss of generality let us choose a set $\mathcal{L}_Z$ of commuting

logical operators which are all of Z type. Then the group generated by the set $\mathcal{S}_X \cup \mathcal{S}_Z \cup \mathcal{L}_Z$ of mutually commuting operators is again invariant under the permutation σ. As the stabilizer group is maximal, it stabilizes a unique state $|\psi\rangle$. Hence $|\psi\rangle$ is the stabilizer state with the desired symmetry, and the CSS code can be expressed as CWS code in terms of $|\psi\rangle$ and some classical code $\mathcal{C}$. ◀

We now prove a lemma for the stabilizer code case of Theorem 5, which improves the result of Lemma 7.

▶ **Lemma 8.** *For a stabilizer code $\mathcal{Q}$ with permutation symmetry σ, there exists a stabilizer state $|\psi\rangle \in \mathcal{Q}$ such that $|\psi\rangle$ has the same permutation symmetry as $\mathcal{Q}$.*

Proof. To prove this lemma, we shall use a standard form for stabilizers (see [20, Section 10.5.7]):

$$\left[\begin{array}{ccc|ccc} I & A_1 & A_2 & B & 0 & C \\ 0 & 0 & 0 & D & I & E \end{array}\right] = \left[\begin{array}{c|c} \mathcal{S}_X & \mathcal{S}_Z \\ 0 & \mathcal{S}'_Z \end{array}\right] = \left[\begin{array}{c} \mathcal{S} \\ \mathcal{S}' \end{array}\right]$$

where A_1 is an $r \times (n-k-r)$ matrix, A_2 is an $r \times k$ matrix, B is an $r \times r$ matrix, C is an $r \times k$ matrix, D is an $(n-k-r) \times r$ matrix, and E is an $(n-k-r) \times k$ matrix. Similar as in the CSS case, we can choose a set $\mathcal{L}_Z$ of commuting logical operators which are all of Z type. In matrix form, they are given by $[0\ 0\ 0|\ -A_2^t\ 0\ I]$. Then the group generated by the mutually commuting operators in $\mathcal{S} \cup \mathcal{S}' \cup \mathcal{L}_X$ stabilizes a unique state $|\psi\rangle$ which is invariant with respect to the permutation σ. Hence $|\psi\rangle$ is the stabilizer state with the desired symmetry that can be used to express $\mathcal{Q}$ as CWS code with some classical code $\mathcal{C}$. ◀

To prove Theorem 5, given a CWS code $\mathcal{Q}$, we first find its unique decomposition as a USt code $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$, based on Theorem 2. Here $\mathcal{S}$ is in general a stabilizer code with $m = n - k$ generators. If $\mathcal{Q}$ has a permutation symmetry σ, then according to Lemma 6, the stabilizer code $\mathcal{S}$ must also have the symmetry σ. Now according to Lemma 8, there exists a quantum state $|\psi\rangle$ in the stabilizer code $\mathcal{S}$ which also has the symmetry σ. Hence $|\psi\rangle$ is the stabilizer state with the desired symmetry. Note that the stabilizer $\mathcal{S}'$ of the state $|\psi\rangle$ contains the original stabilizer $\mathcal{S}$. Therefore, common eigenspaces of $\mathcal{S}$ are further decomposed into one-dimensional joint eigenspaces of $\mathcal{S}'$, and we can rewrite the projection $P_{\mathcal{Q}}$ onto the USt code in the form corresponding to a CWS code. ◀

5 Symmetries of the Classical Code

Theorem 5 does not make any statement about the symmetry of the classical code. In general, if we insist to use the canonical form of the CWS code as given by Theorem 2, we cannot expect that the (non-linear) classical code $\mathcal{C}$ associated with the CWS code $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ has the same symmetry as $\mathcal{Q}$. That is, in this case, even if the stabilizer $\mathcal{S}$ has the same permutation symmetry σ as the quantum code $\mathcal{Q}$, one might not be able to find a classical code $\mathcal{C}$ with the same symmetry σ in general. Let us look at an example.

▶ **Example 9.** Consider the stabilizer state $1/\sqrt{2}(|00\ldots0\rangle - |11\ldots1\rangle)$ (hence a CWS code, denoted by $\mathcal{Q}$), which is invariant under all permutations. Using the canonical form of $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ as given by Theorem 2, the group $\mathcal{S}$ is generated by $XX\ldots X$ and all pairs of Z, which is permutation invariant. However, the classical code $\mathcal{C}$ consists of the vector which is one in the first coordinate and zero elsewhere, i. e., $\mathcal{C}$ is a code with a single codeword $10\ldots0$, which has a smaller symmetry group than that of $\mathcal{Q}$.

On the other hand, if we choose the group $\mathcal{S}'$ generated by $-XX\ldots X$ and all pairs of Z, the corresponding classical code $\mathcal{C}'$ consists just of the zero vector. So in the representation $\mathcal{Q}=(\mathcal{S}',\mathcal{C}')$, both $\mathcal{S}'$ and $\mathcal{C}'$ have the same permutation symmetries as $\mathcal{Q}$.

This example indicates that exploiting the phase factor freedom in the USt code decomposition of a CWS code, and thereby deviating slightly from the canonical form, there is some chance to find both a stabilizer and a classical code with the same permutation symmetry as the CWS code.

To study the properties of the classical code $\mathcal{C}$ associated with a CWS code $\mathcal{Q}=(\mathcal{S},\mathcal{C})$, consider the case where the stabilizer state $\mathcal{S}$ has some permutation symmetry σ. Then for given generators $\{\boldsymbol{g}_1,\boldsymbol{g}_2,\ldots,\boldsymbol{g}_n\}$ of the stabilizer $\mathcal{S}$, the permuted operators $\{\boldsymbol{g}_1^\sigma,\boldsymbol{g}_2^\sigma,\ldots,\boldsymbol{g}_n^\sigma\}$ generate the same stabilizer $\mathcal{S}$. The transformation $\boldsymbol{g}_i\mapsto\boldsymbol{g}_i^\sigma$ can be characterized by a $\mathbb{Z}_p$-valued, invertible $n\times n$ matrix R given by

$$\boldsymbol{g}_i^\sigma=\prod_{j=1}^n\boldsymbol{g}_j^{R_{ji}}. \tag{12}$$

Let us write the K classical codewords in $\mathcal{C}$ as an $K\times n$ matrix with entries c_{ij}. We are now ready to present the following theorem, which gives a sufficient condition for $\mathcal{C}$ to guarantee that $\mathcal{Q}$ has the same permutation symmetry as $\mathcal{S}$

▶ **Theorem 10.** *Let $\mathcal{Q}=(\mathcal{S},\mathcal{C})$ be a CWS code, and let $\{\boldsymbol{g}_1,\boldsymbol{g}_2,\ldots,\boldsymbol{g}_n\}$ be generators of $\mathcal{S}$. If $\mathcal{S}$ has permutation symmetry σ, where $\boldsymbol{g}_i^\sigma=\prod_{j=1}^n\boldsymbol{g}_j^{R_{ji}}$, and $\mathcal{C}R\cong\mathcal{C}$, then $\mathcal{Q}$ has the same permutation symmetry σ as $\mathcal{S}$. Here by $\mathcal{C}R\cong\mathcal{C}$ we mean that the set of rows of $\mathcal{C}R$, corresponding to the transformed code, equals the code $\mathcal{C}$ (not as a matrix).*

Proof. We start by applying the permutation σ to the projection $P_\mathcal{Q}$ onto the code space given by Eq. (3):

$$\begin{aligned}P_\mathcal{Q}^\sigma&=\sum_{\boldsymbol{t}\in\mathcal{C}}\frac{1}{p^n}\sum_{\boldsymbol{x}\in\mathbb{F}_p^n}\omega^{\boldsymbol{t}\cdot\boldsymbol{x}}(\boldsymbol{g}_1^\sigma)^{x_1}\ldots(\boldsymbol{g}_n^\sigma)^{x_n}=\sum_{\boldsymbol{t}\in\mathcal{C}}\frac{1}{p^n}\sum_{\boldsymbol{x}\in\mathbb{F}_p^n}\omega^{\boldsymbol{t}\cdot\boldsymbol{x}}(\prod_j\boldsymbol{g}_j^{R_{j1}})^{x_1}\ldots(\prod_j\boldsymbol{g}_j^{R_{jn}})^{x_n}\\&=\sum_{\boldsymbol{t}\in\mathcal{C}}\frac{1}{p^n}\sum_{\boldsymbol{x}\in\mathbb{F}_p^n}\omega^{\boldsymbol{t}\cdot\boldsymbol{x}}\boldsymbol{g}_1^{\sum_j R_{1j}x_j}\ldots\boldsymbol{g}_n^{\sum_j R_{nj}x_j}\end{aligned} \tag{13}$$

Let $x'_j=\sum_i R_{ji}x_i$ and $t_i=\sum_j R_{ji}t'_j$. Then for $\boldsymbol{t}\in\mathcal{C}$, we have $\boldsymbol{t}'\in\mathcal{C}'$, where the transformed code $\mathcal{C}'$, considered as a $K\times n$ matrix, is given by

$$\mathcal{C}=\mathcal{C}'R. \tag{14}$$

Then Eq. (13) becomes

$$\begin{aligned}P_\mathcal{Q}^\sigma&=\sum_{\boldsymbol{t}'\in\mathcal{C}'}\frac{1}{p^n}\sum_{\boldsymbol{x}\in\mathbb{F}_p^n}\omega^{\sum_i\sum_j R_{ji}t'_jx_i}\boldsymbol{g}_1^{\sum_j R_{1j}x_j}\ldots\boldsymbol{g}_n^{\sum_j R_{nj}x_j}\\&=\sum_{\boldsymbol{t}'\in\mathcal{C}'}\frac{1}{p^n}\sum_{\boldsymbol{x}\in\mathbb{F}_p^n}\omega^{\sum_j t'_j(\sum_i R_{ji}x_i)}\boldsymbol{g}_1^{\sum_j R_{1j}x_j}\ldots\boldsymbol{g}_n^{\sum_j R_{nj}x_j}\\&=\sum_{\boldsymbol{t}'\in\mathcal{C}'}\frac{1}{p^n}\sum_{\boldsymbol{x}'\in\mathbb{F}_p^n}\omega^{\boldsymbol{t}'\cdot\boldsymbol{x}'}\boldsymbol{g}_1^{x'_1}\ldots\boldsymbol{g}_n^{x'_n}.\end{aligned} \tag{15}$$

Now because of $\mathcal{C}R \cong \mathcal{C}$, the rows of $\mathcal{C}R$ are a permutation of the rows of $\mathcal{C}$. Hence there exists a permutation matrix P such that $P\mathcal{C}R = \mathcal{C}$, which gives

$$P\mathcal{C} = \mathcal{C}R^{-1} = \mathcal{C}'. \tag{16}$$

The second equality follows from Eq. (14). Hence the rows of $\mathcal{C}'$ are a permutation of the rows of $\mathcal{C}$, i. e., $\mathcal{C}$ and $\mathcal{C}'$ are the same code. Therefore Eq. (15) becomes

$$P_{\mathcal{Q}}^{\sigma} = \sum_{\boldsymbol{t}' \in \mathcal{C}} \frac{1}{p^n} \sum_{\boldsymbol{x}' \in \mathbb{F}_p^n} \omega^{\boldsymbol{t}' \cdot \boldsymbol{x}'} \boldsymbol{g}_1^{x_1'} \dots \boldsymbol{g}_n^{x_n'} = P_{\mathcal{Q}}, \tag{17}$$

which proves the theorem. ◀

Note that although Theorem 10 is stated in terms of a set of generator $\boldsymbol{g}_i$ of $\mathcal{S}$, it is actually independent of the choice of the generators. That is to say, if $\boldsymbol{g}_i^{\sigma} = \prod_{j=1}^{n} \boldsymbol{g}_j^{R_{ji}}$, and $\mathcal{C}R \cong \mathcal{C}$ holds, then for some other generators $\boldsymbol{g}_i'$ of $\mathcal{S}$, where $\mathcal{Q} = (\mathcal{S}_{\boldsymbol{g}'}, \mathcal{C}')$ and $(\boldsymbol{g}_i')^{\sigma} = \prod_{j=1}^{n} (\boldsymbol{g}_j')^{R'_{ji}}$, one would then have $\mathcal{C}'R' \cong \mathcal{C}'$.

Theorem 10 gives a sufficient condition that the CWS code $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ may have the same permutation symmetry as $\mathcal{S}$. Note that [8, Proposition 5.2] considers a special case of Proposition 10, where the permutation σ is the cyclic shift. However, it turns out that the argument in [8] is false; the cyclic symmetry of the stabilizer $\mathcal{S}$ is not sufficient to guarantee the cyclic symmetry of the resulting quantum code $\mathcal{Q}$; the classical code $\mathcal{C}$ must also have a cyclic symmetry, as discussed in Corollary 12.

It remains unclear whether the condition given in Theorem 10 is also necessary, at least in the case when both the CWS code $\mathcal{Q}$ and the stabilizer $\mathcal{S}$ have a permutation symmetry σ. We expect that in this case the condition $\mathcal{C}R \cong \mathcal{C}$ would be necessary. However, while the condition might be violated for a particular choice of $\mathcal{S}$, it might hold for a different representation $\mathcal{Q} = (\mathcal{S}', \mathcal{C}')$.

6 The Standard Form $\mathcal{Q} = (\mathcal{G}, \mathcal{C})$

Starting with the unique representation of a CWS code as a USt code, we can derive a standard form of a CWS code. We know that up to local Clifford (LC) operations, any CWS code $\mathcal{Q}$ can be represented by a graph $\mathcal{G}$ and a binary classical code $\mathcal{C}$ [5, 6]. Starting with a given CWS code $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$, one can transform the stabilizer $\mathcal{S}$ into a graph state using LC operations, and then $\mathcal{C}$ will be transformed accordingly [5]. Our concern is that if $\mathcal{Q}$ has some permutation symmetry σ, whether it can be kept during this LC operations, in other words, whether one can always obtain a graph with the same permutation symmetry σ as $\mathcal{Q}$ has.

Indeed, even if one can always find a stabilizer state $\mathcal{S}$ with the same symmetry as $\mathcal{Q}$ has, we are asking too much here for the graph $\mathcal{G}$. In general, one cannot find a graph with the same permutation symmetry as $\mathcal{Q}$ has. Let us look at an example.

▶ **Example 11.** The stabilizer $\mathcal{S}$ for the 7-qubit Steane code is generated by

$$\boldsymbol{g}_1 = XIXXXII, \quad \boldsymbol{g}_2 = IXIXXXI, \quad \boldsymbol{g}_3 = IIXIXXX, \tag{18}$$

which are the three X-type generators, and the three Z-type generators

$$\boldsymbol{g}_4 = ZIZZZII, \quad \boldsymbol{g}_5 = IZIZZZI, \quad \boldsymbol{g}_6 = IIZIZZZ. \tag{19}$$

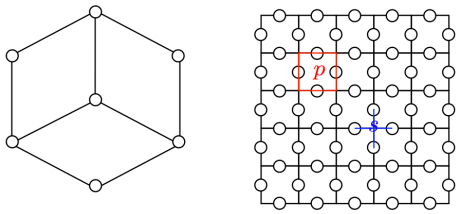


Figure 1 (a) The graph for the Steane code with three-fold cyclic symmetry. (b) The toric code on a square lattice. Qubits are sitting on edges of the lattice. p denotes a plaquette, which contains 4 qubits as shown across the red lines. s denotes a star, which contains 4 qubits as shown across the blue lines.

This code is cyclic, and for its CWS representation, one can choose, e. g., the stabilizer state $|\psi\rangle$ with stabilizer $\mathcal{S}'$ generated by $\mathcal{S}' = \langle \boldsymbol{g}_1, \boldsymbol{g}_2, \boldsymbol{g}_3, \boldsymbol{g}_4, \boldsymbol{g}_5, \boldsymbol{g}_6, Z^{\otimes 7} \rangle$. Then $|\psi\rangle$ is cyclic as well. However, when transforming the Steane code into the standard form of its CWS representation, one cannot find it a cyclic graph [10]. In fact, the best symmetric graph one can find is with a three-fold cyclic symmetry instead of a 7-fold cyclic symmetry, as shown in Fig. 1(a). The three-fold symmetry is in fact the symmetry of the generators of $\mathcal{S}'$ instead of the symmetry of the entire stabilizer group $\mathcal{S}'$. This is related to the fact that the graph $\mathcal{G}$ in some sense represents only the stabilizer generators of its corresponding graph state.

The toric code turns out to provide another example, as shown in Fig. 1(b), which is in some sense even worse than the Steane code example. Despite the fact that the generators of the stabilizer group for the toric code have a translational symmetry, we will show in Theorem 13 that one cannot find a graph with translational symmetry. However, both the Steane code and the toric code do not provide counterexamples to Theorem 5, as the logical zero has the desired symmetry in both cases.

Nevertheless, there might still be some interesting relationship between the permutation symmetries of $\mathcal{Q}$ and the symmetries of $\mathcal{G}$ and $\mathcal{C}$. Let us start with a simple case:

▶ **Corollary 12.** *For a CWS code $\mathcal{Q} = (\mathcal{G}, \mathcal{C})$, if both $\mathcal{G}$ and $\mathcal{C}$ have a permutation symmetry σ, then the code $\mathcal{Q}$ has the permutation symmetry σ as well.*

Proof. This is actually a direct implication of Theorem 10; in this case the matrix R is nothing but a permutation matrix corresponding to the permutation σ. ◀

This turns out to be good luck, as due to the structure of the stabilizer generators of graph states, a permutation of the qubits corresponds to the same permutation of the generators $\boldsymbol{g}_i$, and hence also corresponds to a permutation of the coordinates in the classical code $\mathcal{C}$. Prominent examples are the $((5, 2, 3))$ code and the $((5, 6, 2))$ code, whose corresponding graph is a pentagon in both cases, and the corresponding classical codes are cyclic (see [6, Sec. IIIA,B]).

Finally, let us examine the graph symmetry for the toric code. The toric code was first proposed by Kitaev in 1997 as an example demonstrating topologically ordered quantum systems [13, 14]. The setting is a two-dimensional square lattice with periodic boundary conditions and with a qubit sitting on each edge of the lattice. There are two types of stabilizer generators:

1. $+$ (star) type, indicated in Fig. 1(b) as s:

$$A_s^X = \prod_{j \in \mathrm{star}(s)} X_j \tag{20}$$

2. $\square$ (plaquette) type, indicated in Fig. 1(b) as p:

$$A_p^Z = \prod_{j \in \text{plaquette}(p)} Z_j \tag{21}$$

It is straightforward to check that A_s^X and A_p^Z commute for any pair s and p.

These stabilizer generators are by definition translational invariant, for the translation along each direction of the two-dimensional square lattice. What is more, one can even find an encoding graph which is also translational invariant [2]. We will show that unfortunately one cannot find a translational invariant graph to represent the toric code as a CWS code.

▶ **Theorem 13.** *A graph corresponding to the toric code cannot have the same translational symmetry as the code.*

Proof. Let $\mathcal{T}$ be the toric code stabilizer generated by the star and plaquette operators as given by Eq. (20) and Eq. (21). Suppose that $\mathcal{Q} = (\mathcal{G}, \mathcal{C})$ is a code where $\mathcal{G}$ is as symmetric as the toric code stabilizer (i. e., translational invariant) and is local Clifford equivalent to $\mathcal{T}$. This means that if we let $\mathcal{S}$ be the stabilizer of $\mathcal{Q}$, then there are local Clifford elements $C_1, C_2, \ldots$ such that $\mathcal{S} = \widehat{C}\mathcal{T}\widehat{C}^\dagger$, where $\widehat{C} = C_1 \otimes C_2 \otimes \cdots$ (here we choose an arbitrary indexing of qubits).

Let σ be a permutation symmetry of the toric code and define $\widehat{C}_\sigma = C_{\sigma(1)} \otimes C_{\sigma(2)} \otimes \cdots$. Since σ is assumed to be a symmetry of $\mathcal{S}$ as well, we have

$$\mathcal{S} = \widehat{C}\mathcal{T}\widehat{C}^\dagger = \widehat{C}_\sigma \mathcal{T} \widehat{C}_\sigma^\dagger.$$

Then for $D_i = C_i^\dagger C_{\sigma(i)}$, we have $\widehat{D}\mathcal{T}\widehat{D}^\dagger = \mathcal{T}$, where $\widehat{D} = D_1 \otimes D_2 \otimes \cdots$.

Let $XXXX$ be the element of this stabilizer group $\mathcal{T}$ corresponding to some star $+$. Since $\widehat{D}$ is local, and $XXXX$ is the only element of $\mathcal{T}$ that acts on edges corresponding to $+$, we must have $\mathcal{D}XXXX\mathcal{D}^\dagger = XXXX$. The same argument applies to the Z-terms corresponding to a plaquette $\square$. As a result, conjugation by D_i maps X to $\pm X$ and Z to $\pm Z$. Hence D_i is an element of the Pauli group.

Now we know that $\widehat{D}$ is in the Pauli group, and it holds for every permutation σ. On the other hand, the symmetry group of the toric code is transitive. Therefore, for every i, j, the product $C_i^\dagger C_j$ is in the Pauli group, and furthermore

$$C_1 \otimes C_2 \otimes \cdots = (H \otimes H \otimes \cdots)(P_1 \otimes P_2 \otimes \cdots),$$

where the factors P_i are in the Pauli group and H is some Clifford element acting on a single qubit.

$\widehat{C}\mathcal{T}\widehat{C}^\dagger$ is supposed to correspond to a graph state, but $(P_1 \otimes P_2 \otimes \cdots)$ just changes some signs in the stabilizer group, and $(H \otimes H \otimes \cdots)$ cannot turn the stabilizer group of the toric code into a graph-type stabilizer group. ◀

7 Summary and Discussion

In this work we have investigated the symmetry properties of CWS codes. Our main result shows that for a given CWS code $\mathcal{Q}$ with some permutation symmetry σ, there always exits a stabilizer state $\mathcal{S}$ with the same symmetry σ such that $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ for some classical code $\mathcal{C}$. As many good CWS codes are found by starting from a chosen $\mathcal{S}$, this ensures that when trying to find CWS codes with certain permutation symmetry, the choice of $\mathcal{S}$ with the same symmetry will suffice. A key point to reach our main result is to obtain a canonical representation for CWS codes, i. e., a unique decomposition as USt codes.

One natural question is whether there is any chance to find a classical code $\mathcal{C}$ with the same symmetry σ as that of $\mathcal{Q}$, which, together with some $\mathcal{S}$ with symmetry σ, gives $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$. We do not know the answer in general, but we know that one can no longer restrict $\mathcal{S}$ to the stabilizer used in the canonical form, but might have to introduce some phase factors. We have developed a sufficient condition that $\mathcal{C}$ has to satisfy in order to ensure that in combination with some $\mathcal{S}$ with symmetry σ, one will have $\mathcal{Q} = (\mathcal{S}, \mathcal{C})$ with the same symmetry σ. Observing the fact that the permutation on the code $\mathcal{Q}$ does not directly translate into a permutation of the classical $\mathcal{C}$ (but a linear transformation given by the matrix R), in general one cannot expect to find a classical code $\mathcal{C}$ with the same symmetry as that of $\mathcal{Q}$.

One interesting case are cyclic codes. If there exists a graph $\mathcal{G}$ which has the same symmetry σ as the CWS code $\mathcal{Q} = (\mathcal{G}, \mathcal{C})$, then the permutation of the code $\mathcal{Q}$ translates directly into a permutation of the classical code $\mathcal{C}$. Hence, combining a graph $\mathcal{G}$ whose symmetry group contains the cyclic group of order n, with a cyclic classical code $\mathcal{C}$ of length n, gives a cyclic CWS code $\mathcal{Q} = (\mathcal{G}, \mathcal{C})$. It would be nice to see whether the converse is true as well, i. e., given a cyclic CWS code $\mathcal{Q}$ which corresponds to a graph $\mathcal{G}$ whose symmetry group contains the cyclic group of order n, can we always find a cyclic classical code $\mathcal{C}$ of length n, such that $\mathcal{Q} = (\mathcal{G}, \mathcal{C})$. We leave this for future investigation.

In general, although every CWS code $\mathcal{Q}$ is local Clifford equivalent to a standard form $(\mathcal{G}, \mathcal{C})$, the local Clifford operation may destroy the permutation symmetry of the original code. In other words, one cannot expect to always find a graph $\mathcal{G}$ which has the same symmetry as that of $\mathcal{Q}$. The seven-qubit Steane code is such an example where the graph can only possess a three-fold cyclic symmetry which is the symmetry of the stabilizer generators, instead of the seven-fold cyclic symmetry of the code. For the toric code, despite the stabilizer generators being translational invariant, we show that there does not exist any associated translational invariant graph. A general understanding of the conditions that graphs can possess the same symmetry as the CWS code is worth further investigation.

Acknowledgements. SB was in part supported by National Elites Foundation and by a grant from IPM (No. 91810409). JC is supported by NSERC and NSF of China (Grant No. 61179030). The CQT is funded by the Singapore MoE and the NRF as part of the Research Centres of Excellence programme. ZJ acknowledges support from NSERC, ARO and NSF of China (Grant Nos. 60736011 and 60721061). QW is supported by NSERC. BZ is supported by NSERC and CIFAR. MG acknowledges support by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract No. D11PC20166. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

The authors would like to thank Martin Roetteler for his suggestion to use the Fourier transformation to prove Lemma 4.

References

1 S. Beigi, I. Chuang, M. Grassl, P. Shor, and B. Zeng. Graph concatenation for quantum codes. *Journal of Mathematical Physics*, 52(2):022201, February 2011.

2 S. Bravyi and R. Raussendorf. Measurement-based quantum computation with the toric code states. *Physical Review A*, 76(2):022304, August 2007.

3 A. R. Calderbank, E. Rains, P. W. Shor, and N. J. A. Sloane. Quantum error correction via codes over $GF(4)$. *IEEE Transactions on Information Theory*, 44(4):1369–1387, 1998.

4 A. R. Calderbank and P. W. Shor. Good quantum error-correcting codes exist. *Physcial Review A*, 54:1098–1105, August 1996.

5 I. Chuang, A. Cross, G. Smith, J. Smolin, and B. Zeng. Codeword stabilized quantum codes: Algorithm and structure. *Journal of Mathematical Physics*, 50(4):042109, April 2009.

6 A. Cross, G. Smith, J. A. Smolin, and B. Zeng. Codeword stabilized quantum codes. *IEEE Transactions on Information Theory*, 55(1):433–438, 2009.

7 L. E. Danielsen. On self-dual quantum codes, graphs, and Boolean functions. Master's thesis, University of Bergen, 2005. `http://arxiv.org/abs/quant-ph/0503236`.

8 S. Dutta and P. P Kurur. Quantum Cyclic Code. Preprint arXiv:1007.1697 [cs.IT], June 2010.

9 D. Gottesman. *Stabilizer Codes and Quantum Error Correction*. PhD thesis, California Institute of Technology, Pasadena, CA, 1997.

10 M. Grassl, A. Klappenecker, and M. Rötteler. Graphs, quadratic forms, and quantum codes. In *Proceedings 2002 IEEE International Symposium on Information Theory (ISIT 2002)*, page 45, Lausanne, Switzerland, June/July 2002. `http://arxiv.org/abs/quant-ph/0703112`.

11 M. Grassl and M. Rötteler. Non-additive quantum codes from Goethals and Preparata codes. *Proceedings of 2008 IEEE Information Theory Workshop*, pages 396–400, 2008.

12 M. Grassl and M. Rötteler. Quantum Goethals-Preparata codes. *Proceedings of 2008 IEEE International Symposium on Information Theory*, pages 300–304, 2008.

13 A. Yu. Kitaev. Quantum computations: algorithms and error correction. *Russian Math. Surveys*, 52:1191–1249, 1997.

14 A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2002.

15 E. Knill. Group Representations, Error Bases and Quantum Codes. Technical Report LAUR-96-2807, LANL, 1996. Preprint `http://arxiv.org/quant-ph/9608049`.

16 E. Knill. Non-binary Unitary Error Bases and Quantum Codes. Technical Report LAUR-96-2717, LANL, 1996. Preprint `http://arxiv.org/quant-ph/9608048`.

17 Y. Li, I. Dumer, and L. P. Pryadko. Clustered Error Correction of Codeword-Stabilized Quantum Codes. *Physical Review Letters*, 104(19):190501, May 2010.

18 S. Y. Looi, L. Yu, V. Gheorghiu, and R. B. Griffiths. Quantum error correcting codes using qudit graph states. *Physical Review A*, 78(4):042303, 2008.

19 F. J. MacWilliams and N. J. A. Sloane. *The Theory of Error-Correcting Codes*. North-Holland Publishing Company, Amsterdam, 1977.

20 M. Nielsen and I. Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, England, 2000.

21 C. E. Shannon. A mathematical theory of communication. *Bell Labs Technical Journal*, 27:379–423, 1948.

22 A. Steane. Multiple particle interference and quantum error correction. *Proceedings of the Royal Society of London, Series A*, 452:2551–2577, 1996.

23 S. Yu, Q. Chen, and C. H. Oh. Graphical quantum error-correcting codes. Preprint arXiv:0709.1780 [quant-ph], September 2007.

Certifying the Absence of Apparent Randomness under Minimal Assumptions

Gonzalo de la Torre, Chirag Dhara, and Antonio Acin

ICFO-Institut de Ciencies Fotoniques
Av. Carl Friedrich Gauss, 3, 08860 Castelldefels, Barcelona, Spain

Abstract

Contrary to classical physics, the predictions of quantum theory for measurement outcomes are of a probabilistic nature. Questions about the completeness of such predictions lie at the core of quantum physics and can be traced back to the foundations of the field. Recently, the completeness of quantum probabilistic predictions could be established based on the assumption of freedom of choice. Here we ask when can events be established to be as unpredictable as we observe them to be relying only on minimal assumptions, ie. distrusting even the free choice assumption but assuming the existence of an arbitrarily weak (but non-zero) source of randomness. We answer the latter by identifying a sufficient condition weaker than the monogamy of correlations which allow us to provide a family of finite scenarios based on GHZ paradoxes where quantum probabilistic predictions are as accurate as they can possibly be. Our results can be used for a protocol of full randomness amplification, without the need of privacy amplification, in which the final bit approaches a perfect random bit exponentially fast on the number of parties.

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

Keywords and phrases randomness, Bell nonlocality, free choice

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.207

1 Introduction

Physical theories aim at providing the best possible predictions for both natural phenomena taking place in the universe and on the controlled environment of our laboratories. Interestingly, the type of predictions theories can make has been changing over time, depending heavily on the specific physical theory considered. Moreover, the latter happens not only at the less surprising quantitative level, ie. general relativistic predictions about the perihelion precession shift of mercury are more accurate than those of newtonian theory of gravity, but more strikingly, also at a qualitative level ie. the uncertainty on predictions hold fundamentally different statuses in classical and quantum theory.

Classical mechanics, the theory governing our physical understanding until the XIX century, is a deterministic theory by construction. The latter neither does imply that probabilistic predictions do not play any role nor that we cannot observe physical phenomena behaving as random and yet being governed by classical mechanical laws. Instead, it means that all uncertainty in the predictions of the theory can be traced back to a lack of knowledge about all the relevant degrees of freedom of the physical phenomena considered. As an example, accurate knowledge of the applied force and torque, viscosity and gravitational potential would make the outcome prediction of a coin flip fully predictable. Thus, no room for intrinsic unpredictability is available within classical theory and the best possible predictions are deterministic.

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 207–219


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

With the advent of quantum mechanics, the former intuitions had to change dramatically. Contrary to classical mechanics, quantum theory is a probabilistic theory as dictated by its axioms. This means that, in general, the predictions of the theory for measurement outcomes have an inevitable amount of uncertainty, even when full knowledge and control over all relevant degrees of freedom is assumed.

Such a striking change on the role of predictability greatly shaked the foundations of physics. The completeness of quantum predictions has indeed been widely debated by some of the most eminent physicists that contributed to its development [1, 2]. However, John Bell was the first to derive consequences on the issue of completeness from experimentally fasifiable predictions under rigorous assumptions [3]. He proved that according to the predictions of quantum theory and under the assumption of locality and freedom of choice, the outcomes of some quantum experiments would be incompatible with an underlying deterministic theory. Very recently, completeness of quantum theory could be established under the assumptions of locality, the correctness of quantum theory and the crucial assumption of freedom of choice [4]. However, one may consider cases where the freedom of choice assumption cannot be fully trusted and ask whether it is truly a necessary requirement in order to exclude all apparent randomness. In the present letter we pose such question, that is, under what conditions and minimal assumptions can we certify that an event is as intrinsically unpredictable as it is observed to be. In other words, when can we exclude all possible apparent randomness of an event. Interestingly, in a recent work the full unpredictability of an event could be certified under minimal assumptions [5]. Nevertheless, this result required a complex scenario on the infinite number of parties limit. The proof was based on the monogamy of correlations in such limit. The main result of this letter is to identify a sufficient condition weaker than the monogamy of correlations [6] that certify events without any apparent randomness under the assumptions of locality and the existence of a source of arbitrarily deterministic bits. Using this condition, we construct a family of finite scenarios based on GHZ paradoxes [7] where events are indeed as intrinsically random as they appear to be. Moreover, our results imply a perfect free random bit can be approached exponentially fast in the number of parties and is therefore suitable for a full randomness amplification protocol without privacy amplification[5].

1.1 Geometric interpretation of the problem

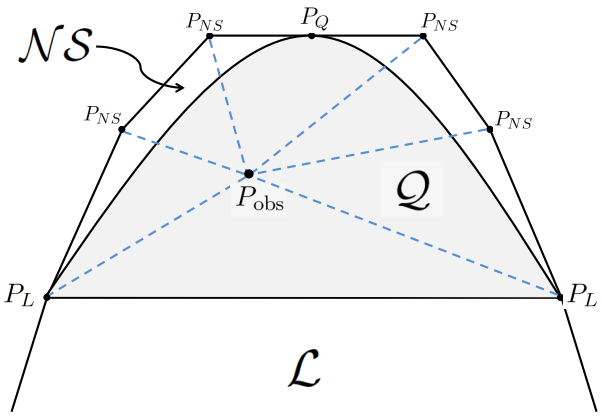


Figure 1 Qualitative picture of local, quantum and no-sigalling sets.

Fig. 1 is a useful qualitative geometric picture which serves to clarify the general idea and to explain the scenario we work with. Given some non-local distribution P_{obs}, its intrinsic randomness content is quantitatively dependent on whether we use the quantum or non-signalling framework. For example, the Tsirelson correlations [8] in the $(2, 2, 2)$ scenario considered strictly within the quantum set yields 1.23 bits of randomness [9]. However, its randomness in the larger non-signalling set is a much smaller 0.34 bits. Another example is the GHZ correlations [10] which contain (considering the tripartite states in particular) 3 bits of randomness within the quantum set. However, in the non-signalling set it reduces to just 1 bit since the extremal points are fully characterized in [11]. In fact, it is generally the case that the intrinsic randomness of a point considered to be embedded in the non-signalling set is lower than its intrinsic randomness within the quantum set. The reason is simply that there are more general decompositions possible within the non-signalling set which increases our ignorance about its underlying preparation. It is in this context that we can finally pose the question that is the theme of this work. *Is it possible to guarantee that the observed correlations do not contain any classical randomness for some correlations* P_{obs} *even allowing the largest possible ignorance by embedding it in the non-signalling set?*

The challenge to answering this question in full generality is that the definition of intrinsic randomness in such scenarios is defined as the optimization over all possible preparations of P_{obs}. However, this computation requires a complete characterization of the corresponding no-signalling polytope. This is known only for the smallest dimension and thus the computation is infeasible for anything but the smallest systems. What we show here is that despite the infeasibility of calculating the intrinsic randomness in full generality it is possible to choose scenarios carefully in which the computation is rendered feasible. We not only demonstrate one such case but also certify that the observed randomness is fully intrinsic in our chosen scenario. What make the result counter-intuitive is that our results are valid for a whole class of non-extremal distributions.

There is a further layer of subtlety which we additionally address in our work. This is related to a paradox of randomness certification using Bell inequalities, which is the freedom of choice assumption. The assumption of freedom of choice may be regarded a reasonable assumption in many cases but it is particularly problematic for randomness certification. Recently there has been a significant body of work in deriving Bell inequalities with relaxations of this assumption [12, 13, 14, 15, 5]. A significant feature of our results are that they are valid even under a complete (non-zero) relaxation of the measurement assumption. For this reason, these results may also be interpreted as an alternative approach for full randomness amplification with the benefit of significantly easier techniques.

2 Preliminaries

Suppose that a Bell test is performed repeatedly among N parties and the resulting statistics is given by $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$, where $\mathbf{a} = (a_1, \ldots, a_N)$ and $\mathbf{x} = (x_1, \ldots, x_N)$ are the string of outcomes and measurement inputs of the parties involved. Let g be a function acting on the measurement results $\mathbf{a}$. As previously explained, there are different physically relevant notions of randomness.

First, the *observed randomness* of g for measurements $\mathbf{x}$ is the randomness computed directly from the statistics. Operationally, this may be defined as the optimal probability of guessing the outcome of g for input $\mathbf{x}$,

$$G_{\text{obs}}(g, \mathbf{x}, P_{\text{obs}}) = \max_{k \in \text{Im}(g)} P_{\text{obs}}(g(\mathbf{a}) = k|\mathbf{x}). \tag{1}$$

where $\mathrm{Im}(g)$ is the image of function g.

Moving to the definition of the *intrinsic randomness*, one should consider all possible preparations of the observed statistics in terms of no-signalling probability distributions. In our context, a particular preparation reads

$$P_{\mathrm{obs}}(\mathbf{a}|\mathbf{x}) = \sum_e p(e|\mathbf{x}) P_e^{\mathrm{ex}}(\mathbf{a}|\mathbf{x}) \tag{2}$$

where the P_e^{ex} are extremal points of the no-signaling set [16]. The terms $p(e|\mathbf{x})$ may depend on $\mathbf{x}$, which accounts for possible correlations between the preparation e and the measurement settings $\mathbf{x}$, given that the choice of measurements are not assumed to be free. Hence, we define the intrinsic randomness of a function g by optimizing over all possible non-signalling preparations of P_{obs} so as to minimize the randomness of g. In other words,

$$G_{\mathrm{int}}(g, \mathbf{x}, P_{\mathrm{obs}}) = \max_{\{p(e|\mathbf{x}), P_e^{\mathrm{ex}}\}} \sum_e p(e|\mathbf{x}) G_{\mathrm{obs}}(g, \mathbf{x}, P_e^{\mathrm{ex}})$$

subject to:

$$\sum_e p(e|\mathbf{x}) P_e^{\mathrm{ex}}(\mathbf{a}|\mathbf{x}) = P_{\mathrm{obs}}(\mathbf{a}|\mathbf{x}) \tag{3}$$

$$p(\mathbf{x}|e) \geq \delta \quad \text{with} \quad \delta > 0; \ \forall \ \mathbf{x}, e \tag{4}$$

where $G_{\mathrm{obs}}(g, \mathbf{x}, P_e^{\mathrm{ex}}) = \max_k \ P_e^{\mathrm{ex}}(g(\mathbf{a}) = k|\mathbf{x})$ is also the intrinsic randomness of P_e^{ex}, since intrinsic and observed randomness must coincide for extremal points of the non-signalling set. Note that condition $p(\mathbf{x}|e) \geq \delta > 0$ allows for an arbitrary (but not absolute) relaxation of the freedom of choice assumption by allowing for arbitrary (yet not complete) correlations between the preparation and the measurement settings. Physically, this condition ensures that all measurement combinations appear for all possible preparations e. An example of a source of randomness fulfilling this condition is a Santha-Vazirani source [17]. Note however that our definition allows sources more general than the Santha-Vazirani sources.

From a cryptographic point of view, the observed randomness is the one perceived by the parties performing the Bell test, whereas the intrinsic randomness is that perceived by a non-signalling eavesdropper possessing knowledge of the preparation of the observed correlations and with the ability to arbitrarily (yet not fully) bias the choice of the measurement settings.

In general, G_{obs} is strictly larger than G_{intr}, as the set of non-signalling correlations is larger than the quantum. The results in [6, 18] provide a Bell test in which G_{intr} approaches G_{obs} (and to 1/2) in the limit of an infinite number of measurements and assuming free choices, that is, $p(\mathbf{x}|e)$ in (2) is independent of e. The results in [19] allow some relaxation of this last condition. The results in [5] arbitrarily relaxed the free-choice condition and give a Bell test in which G_{intr} tends to G_{obs} (and both tend to 1/2) in the limit of an infinite number of parties. Here, we provide a significantly stronger proof, as we allow the same level of relaxation on free choices and provide Bell tests in which $G_{\mathrm{intr}} = G_{\mathrm{obs}}$ for any number of parties. Moreover, a perfect random bit is obtained in the limit of an infinite number of parties.

3 Scenario

Our scenario consists of N parties where each performs two measurements of two outcomes. In what follows, we adopt a spin-like notation and label the outputs by ± 1. Then, any non-signalling probability distribution can be written as (for simplicity we give the expression

for three parties, but it easily generalizes to an arbitrary number)

$$\begin{aligned} &P(a_1,a_2,a_3|x_1,x_2,x_3) = \\ &\frac{1}{8}\Big(1 + a_1\langle A_1^{(x_1)}\rangle + a_2\langle A_2^{(x_2)}\rangle + a_3\langle A_3^{(x_3)}\rangle + \\ &a_1a_2\langle A_1^{(x_1)}A_2^{(x_2)}\rangle + a_1a_3\langle A_1^{(x_1)}A_3^{(x_3)}\rangle + \\ &a_2a_3\langle A_2^{(x_2)}A_3^{(x_3)}\rangle + a_1a_2a_3\langle A_1^{(x_1)}A_2^{(x_2)}A_3^{(x_3)}\rangle\Big), \end{aligned} \tag{5}$$

where $A_i^{(x_i)}$ denotes the outputs of measurement x_i by each party i. In this scenario, we consider Mermin Bell inequalities, whose Bell operator reads

$$M_N = \frac{1}{2}M_{N-1}(A_N^{(0)} + A_N^{(1)}) + \frac{1}{2}M'_{N-1}(A_N^{(0)} - A_N^{(1)}), \tag{6}$$

where M_2 is the Clauser-Horne-Shimony-Holt operator and M'_{N-1} is obtained from M_{N-1} after swapping $A_i^{(0)} \leftrightarrow A_i^{(1)}$. We study probability distributions that give the maximal non-signalling violation of the Mermin inequalities and focus our analysis on a function f that maps the N measurement results into one bit as follows:

$$f(\mathbf{a}) = \begin{cases} +1 & n_-(\mathbf{a}) = (4j+2); \quad \text{with } j \in \{0,1,2,\ldots\} \\ \\ -1 & \text{otherwise} \end{cases} \tag{7}$$

where $n_-(\mathbf{a})$ denotes the number of results in $\mathbf{a}$ that are equal to -1.

4 Results

Our goal in what follows is to quantify the intrinsic randomness of the bit defined by $f(\mathbf{a})$ for those distributions maximally violating the Mermin inequality for odd N. We first prove the following

▶ **Lemma 1.** *Let $P_{\mathrm{M}}(\mathbf{a}|\mathbf{x})$ be an N-partite (odd N) non-signalling probability distribution maximally violating the corresponding Mermin inequality. Then, for any input $\mathbf{x}$ appearing in the inequality*

$$P_{\mathrm{M}}(f(\mathbf{a}) = h_N|\mathbf{x}) \geq 1/2, \text{ with } h_N = \sqrt{2}\cos\left(\frac{\pi(N+4)}{4}\right). \tag{8}$$

Note that, as N is odd, $h_N = \pm 1$. Operationally, the Lemma implies that, for all points maximally violating the Mermin inequality, the bit defined by f is biased towards the same value h_N. Since the proof of the Lemma for arbitrary odd N is convoluted, we give the explicit proof for $N = 3$ here, which already conveys the main ingredients of the general proof, and relegate the generalization to the Supplementary Information.

Proof for three parties. With some abuse of notation, the tripartite Mermin inequality may be expressed as,

$$M_3 = \langle 001\rangle + \langle 010\rangle + \langle 100\rangle - \langle 111\rangle \leq 2, \tag{9}$$

where $\langle x_1x_2x_3\rangle = \langle A_1^{(x1)}A_2^{(x2)}A_3^{(x3)}\rangle$ and similar for the other terms. The maximal non-signalling violation assigns $M_3 = 4$ which can only occur when the first three correlators in (9) take their maximum value of $+1$ and the last takes its minimum of -1.

Take any input combination appearing in the inequality (9), say, $\mathbf{x_m} = (0,0,1)$. Maximal violation of M_3 imposes the following conditions:

1. $\langle 001\rangle = 1$. This further implies $\langle 0\rangle_1 = \langle 01\rangle_{23}$, $\langle 0\rangle_2 = \langle 01\rangle_{13}$ and $\langle 1\rangle_3 = \langle 00\rangle_{12}$.
2. $\langle 010\rangle = 1$ implying $\langle 0\rangle_1 = \langle 10\rangle_{23}$, $\langle 1\rangle_2 = \langle 00\rangle_{13}$ and $\langle 0\rangle_3 = \langle 01\rangle_{12}$.
3. $\langle 100\rangle = 1$ implying $\langle 1\rangle_1 = \langle 00\rangle_{23}$, $\langle 0\rangle_2 = \langle 10\rangle_{13}$ and $\langle 0\rangle_3 = \langle 10\rangle_{12}$.
4. $\langle 111\rangle = -1$ implying $\langle 1\rangle_1 = -\langle 11\rangle_{23}$, $\langle 1\rangle_2 = -\langle 11\rangle_{13}$ and $\langle 1\rangle_3 = -\langle 11\rangle_{12}$

Imposing these relations on (5) for input $\mathbf{x_m} = (0, 0, 1)$ one gets

$$\begin{aligned} P_{\mathrm{M}}(a_1, a_2, a_3|0, 0, 1) = \frac{1}{8}\,(1 + a_1a_2a_3 + (a_1 + a_2a_3)\langle 0\rangle_1 + \\ (a_2 + a_1a_3)\langle 0\rangle_2 + (a_3 + a_1a_2)\langle 1\rangle_3) \end{aligned} \tag{10}$$

Using all these constraints and the definition of the function (20), Eq. (8) can be expressed as

$$\begin{aligned} &P_M(f(\mathbf{a}) = +1|\mathbf{x_m}) \\ &= P_{\mathrm{M}}(1, -1, -1|\mathbf{x_m}) + P_{\mathrm{M}}(-1, 1, -1|\mathbf{x_m}) \\ &+ P_{\mathrm{M}}(-1, -1, 1|\mathbf{x_m}) \\ &= \frac{1}{4}(3 - \langle 0\rangle_1 - \langle 0\rangle_2 - \langle 1\rangle_3) \end{aligned} \tag{11}$$

Proving that $P(f(\mathbf{a}) = +1|\mathbf{x_m}) \geq 1/2$ then amounts to showing that $\langle 0\rangle_1 + \langle 0\rangle_2 + \langle 1\rangle_3 \leq 1$. This form is very convenient since it reminds one of a positivity condition of probabilities.

We then consider the input combination $\bar{\mathbf{x}}_\mathbf{m}$ such that all the bits in $\bar{\mathbf{x}}_\mathbf{m}$ are different from those in $\mathbf{x}_m$. We call this the swapped input, which in the previous case is $\bar{\mathbf{x}}_\mathbf{m} = (1, 1, 0)$. Note that this is *not* an input appearing in the Mermin inequality. However, using the previous constraints derived for distributions P_{M} maximally violating the inequality, one has

$$\begin{aligned} &P_{\mathrm{M}}(a_1, a_2, a_3|1, 1, 0) \\ =&\frac{1}{8}(1 + a_1\langle 1\rangle_1 + a_2\langle 1\rangle_2 + a_3\langle 0\rangle_3 + a_1a_2\langle 11\rangle_{12} \\ &+ a_1a_3\langle 10\rangle_{13} + a_2a_3\langle 10\rangle_{23} + a_1a_2a_3\langle 110\rangle_{123}) \\ =&\frac{1}{8}(1 + a_1\langle 1\rangle_1 + a_2\langle 1\rangle_2 + a_3\langle 0\rangle_3 - a_1a_2\langle 1\rangle_3 \\ &+ a_1a_3\langle 0\rangle_2 + a_2a_3\langle 0\rangle_1 + a_1a_2a_3\langle 110\rangle_{123}), \end{aligned} \tag{12}$$

where the second equality results from the relations $\langle 11\rangle_{12} = -\langle 1\rangle_3$, $\langle 10\rangle_{13} = \langle 0\rangle_2$ and $\langle 10\rangle_{23} = \langle 0\rangle_1$.

It can be easily verified that summing the two positivity conditions $P_{\mathrm{M}}(1, 1, -1|\bar{\mathbf{x}}_\mathbf{m}) \geq 0$ and $P_{\mathrm{M}}(-1, -1, 1|\bar{\mathbf{x}}_\mathbf{m}) \geq 0$ gives the result we seek, namely $1 - \langle 0\rangle_1 - \langle 0\rangle_2 - \langle 1\rangle_3 \geq 0$, which completes the proof. ◀

Using the previous Lemma, it is rather easy to prove the following

▶ **Theorem 2.** *Let $P_{\mathrm{obs}}(\mathbf{a}|\mathbf{x})$ be an N-partite (odd N) non-signalling probability distribution maximally violating the corresponding Mermin inequality. Then the intrinsic and the observed randomness of the function f are equal for any input $\mathbf{x}$ appearing in the Mermin inequality:*

$$G_{\mathrm{int}}(f, \mathbf{x}, P_{\mathrm{obs}}) = G_{\mathrm{obs}}(f, \mathbf{x}, P_{\mathrm{obs}})$$

where

$$G_{\mathrm{obs}}(f, \mathbf{x}, P_{\mathrm{obs}}) \quad = \quad \max_{k\in\{+1,-1\}} P_{\mathrm{obs}}(f(\mathbf{a}) = k|\mathbf{x})$$

Proof of Theorem 1. Since P_{obs} maximally and algebraically violates the Mermin inequality, all the extremal distributions P_e^{ex} appearing in its decomposition must also necessarily lead to the maximal violation of the Mermin inequality (see Supplementary Information for details). Hence, the randomness of f in these distributions as well satisfies Eqn. (8) of Lemma 1. Using this, we find,

$$\begin{aligned} G_{\text{obs}}(f,\mathbf{x},P_e^{\text{ex}}) &= \max_{k\in\{+1,-1\}} P_e^{\text{ex}}(f(\mathbf{a})=k|\mathbf{x}) \\ &= |P_e^{\text{ex}}(f(\mathbf{a})=h_N|\mathbf{x})-1/2|+1/2 \\ &= P_e^{\text{ex}}(f(\mathbf{a})=h_N|\mathbf{x}), \end{aligned} \tag{13}$$

for every e. Therefore,

$$\begin{aligned} G_{\text{int}}(f,\mathbf{x},P_{\text{obs}}) &= \max_{\{p(e|\mathbf{x}),P_e^{\text{ex}}\}} \sum_e p(e|\mathbf{x}) G_{\text{obs}}(f,\mathbf{x},P_e^{\text{ex}}) \\ &= \max_{\{p(e|\mathbf{x}),P_e^{\text{ex}}\}} \sum_e p(e|\mathbf{x}) P_e^{\text{ex}}(f(\mathbf{a})=h_N|\mathbf{x}) \\ &= P_{\text{obs}}(f(\mathbf{a})=h_N|\mathbf{x}), \end{aligned} \tag{14}$$

where the last equality follows from the constraint $\sum_e p(e|\mathbf{x})P_e(\mathbf{a}|\mathbf{x}) = P_{\text{obs}}(\mathbf{a}|\mathbf{x})$. On the other hand the observed randomness for f is, $G_{\text{obs}}(f,\mathbf{x},P_{\text{obs}}) = P_{\text{obs}}(f(\mathbf{a})=h_N|\mathbf{x})$. ◀

The previous technical results are valid for any non-signalling distribution maximally violating the Mermin inequality. For odd N this maximal violation can be attained by a unique quantum distribution, denoted by $P_{\text{ghz}}(\mathbf{a}|\mathbf{x})$, resulting from measurements on a Greenberger-Horne-Zeilinger (GHZ) state. When applying Theorem 2 to this distribution, one gets

Main result: Let $P_{\text{ghz}}(\mathbf{a}|\mathbf{x})$ be the N-partite (odd N) quantum probability probability distribution attaining the maximal violation of the Mermin inequality. The intrinsic and observed randomness of f for a Mermin input satisfy

$$G_{\text{int/obs}}(f,\mathbf{x},P_{\text{ghz}}) = \frac{1}{2} + \frac{1}{2^{(N+1)/2}} \tag{15}$$

This follows straightforwardly from Theorem 2, since $P_{\text{ghz}}(\mathbf{a}|\mathbf{x}) = 1/2^{N-1}$ for outcomes $\mathbf{a}$ with an even number of results equal to -1 and for those measurements appearing in the Mermin inequality.

It is important to remark that $f(\mathbf{a}|\mathbf{x}_m)$ approaches a perfect random bit exponentially with the number of parties. In fact, this bit defines a process in which full randomness amplification takes place. Yet, it is not a complete protocol as, contrary to the existing proposal in [5], no estimation part is provided.

5 Discussion

We have seen that for the choice of our function, the observed randomness in distributions maximally violating the Mermin inequality is wholly intrinsic. This includes the physically realizable GHZ correlations. For the latter, the randomness of the function approaches that of a perfect bit exponentially fast in the size of the system. In adversarial terms, this implies that no non-signalling adversary has additional knowledge or can predict the outcome of f better than the parties performing the Bell test.

In the context of the GHZ correlations (being the only correlations in the class we have defined that may be attained by quantum systems), our result bears a resemblance to those in [4, 18] where the completeness of quantum theory was discussed. These results show that the predictive power of quantum theory is maximal. However, our scenario departs significantly from the one considered there. For one thing, we do not assume quantum theory is correct at the level of the dynamics, i.e we do not assume the unitarity of the dynamics, but only at the level of correlations. Besides, we consider a function of the outcomes. Most important of all, our setup allows us to relax the critical free choice assumption arbitrarily, as long as it is not absolute. This was not possible in [4, 18], except perhaps in a very limited sense due to the results of [19].

Furthermore, our results bear a deep relationship with full randomness amplification [20]. Since the free choice can be relaxed and we find our function approaching a perfect random bit with increasing system size, this is precisely the task set out to full randomness amplification. The missing link is the full protocol including estimation, which we do not provide here.

Future directions of work include exploiting such relations to upper bound the classical randomness where exact relations are not possible. Moreover, an interesting line of work is to extend these techniques for distributions non-maximally violating Bell inequalities. These could perhaps lead to experimentally viable tests of fully general device independence randomness certification.

References

1 Niels Bohr. Can quantum-mechanical description of physical reality be considered complete? *Phys. Rev.*, 48:696–702, 1935.

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

3 John Bell. On the einstein podolsky rosen paradox. *Physics*, 1:195–200, 1964.

4 Roger Colbeck and Renato Renner. No extension of quantum theory can have improved predictive power. *Nat Commun*, 2:411–, 2011.

5 Rodrigo Gallego, Lluis Masanes, Gonzalo de la Torre, Chirag Dhara, Leandro Aolita, and Antonio Acín. Full randomness from arbitrarily deterministic events. *arXiv*, arXiv:1210.6514 [quant-ph], 2012.

6 Jonathan Barrett, Adrian Kent, and Stefano Pironio. Maximally nonlocal and monogamous quantum correlations. *Phys. Rev. Lett.*, 97(17):170409–4, 2006.

7 N. David Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Phys. Rev. Lett.*, 65(15):1838–1840, 1990.

8 B.S. Tsirelson. Quantum analogues of the bell inequalities. the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36:557, 1987.

9 Antonio Acín, Serge Massar, and Stefano Pironio. Randomness versus nonlocality and entanglement. *Phys. Rev. Lett.*, 108(10):100402–, 2012.

10 D. M. Greenberger, M. A. Horne, and A. Zeilinger. *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*. Kluwer, 1989.

11 Stefano Pironio, Jean-Daniel Bancal, and Valerio Scarani. Extremal correlations of the tripartite no-signaling polytope. *Journal of Physics A: Mathematical and Theoretical*, 44(6):065303–, 2011.

12 Johannes Kofler, Tomasz Paterek, and Caslav Brukner. Experimenter'ss freedom in bell's theorem and quantum cryptography. *Phys. Rev. A*, 73(2):022104–, 2006.

13 J. Barrett and N. Gisin. How much measurement independence is needed in order to demonstrate nonlocality? *Arxiv*, arXiv:1008.3612, 2010.

14 Nicolas Gisin. Is realism compatible with true randomness? *arXiv*, arXiv:1012.2536 [quant-ph], 2010.

15 Michael J. W. Hall. Relaxed bell inequalities and kochen-specker theorems. *Phys. Rev. A*, 84(2):022102–, 2011.

16 J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts. Nonlocal correlations as an information-theoretic resource. *Phys. Rev. A*, 71:022101, 2005.

17 Miklos Santha and Umesh V. Vazirani. Generating quasi-random sequences from semi-random sources. *J. Comput. Syst. Sci.*, 33(1):75–87, 1986.

18 R. Colbeck and R. Renner. The completeness of quantum theory for predicting measurement outcomes. *arXiv*, arXiv:1208.4123 [quant-ph], 2012.

19 Roger Colbeck and Renato Renner. Free randomness can be amplified. *Nat Phys*, 8(6):450–454, 2012.

20 Rodrigo Gallego, Lars Erik Würflinger, Antonio Acín, and Miguel Navascues. Operational framework for nonlocality. *Phys. Rev. Lett.*, 109(7):070401–, 2012.

21 Arthur T. Benjamin, Bob Chen, and Kimberly Kindred. Sums of evenly spaced binomial coefficients. *Mathematics Magazine*, 83:370–373, 2010.

A Proof of the main theorem

Here we prove the principal theorem of the main text. It is basically a generalization of the the proof for $N = 3$. We would like to prove that the function f defined in the main text, satisfies the property:

$$P(f(\mathbf{a}) = h_N|\mathbf{x_m}) \geq 1/2 \tag{16}$$

for any N-partite distribution (odd N) that maximally violates the Mermin inequality. As in the tripartite case, in order to prove the result we (I) express condition (16) in terms of some correlators and (II) use positivity conditions from the swapped input to prove the inequality.

An N-partite no-signalling probability distribution $P(\mathbf{a}|\mathbf{x})$ with inputs $\mathbf{x} \in \{0,1\}^N$ and outputs $\mathbf{a} \in \{+1,-1\}^N$ can be parameterized in terms of correlators as,

$$P(\mathbf{a}|\mathbf{x}) = \frac{1}{2^N}\left(1 + \sum_{i=1}^{N} a_i\langle x_i\rangle + \sum_{i<j} a_i a_j \langle x_i x_j\rangle \sum_{i<j<k} a_i a_j a_k \langle x_i x_j x_k\rangle + \cdots + a_1 a_2 \ldots a_N \langle x_1 x_2 \ldots x_N\rangle\right) \tag{17}$$

Restricting $P(\mathbf{a}|\mathbf{x})$ to those maximally violating the N-partite Mermin inequality is equivalent to requiring all correlators of input strings of odd parity to take their extremal values. Namely, we have,

$$\langle x_1 x_2 \ldots x_N\rangle = (-1)^{(-1+\sum_{i=1}^{N} x_i)/2}, \tag{18}$$

for all N-point correlators satisfying $\sum_{i=0}^{N} x_i = 1 \bmod 2$. For instance, $\langle 0,0,\ldots,1\rangle = 1$ and similarly for all permutations. Also, $\langle 0,0,\ldots,0,1,1,1\rangle = -1$ as well as for for all permutations, etc. In the following we will use the notation $\langle . \rangle_k$ to denote a k-point correlator. The input combination used to extract randomness is a generalization of the tripartite case and denoted by $\mathbf{x_m} = (0,0,\ldots,0,1)$. The corresponding N-point correlator satisfies $\langle 0,0,\ldots,0,1\rangle = 1$ for all N. The latter implies two useful relations:

1. Half the total outcomes vanish. In particular these are the terms for which the product of outcomes is -1 *i.e.* $P(\prod_{i=1}^{N} a_i = -1|\mathbf{x_m}) = 0$.
2. $\langle . \rangle_{N-k} = \langle . \rangle_k$ for all $1 \leq k \leq (N-1)/2$ where the correlators $\langle . \rangle_{N-k}$ and $\langle . \rangle_k$ are complementary in the input $\mathbf{x_m}$.

One can use these in Eqn. (17) to express $P(\mathbf{a}|\mathbf{x_m})$ in terms of only the first $(N-1)/2$-point correlators as,

$$P(\mathbf{a}|\mathbf{x_m}) = \frac{1}{2^{N-1}} \left(1 + \sum a_i \langle x_i \rangle + \sum a_i a_j \langle x_i x_j \rangle + \cdots + \sum a_i a_j \cdots a_p \langle x_i x_j \cdots x_p \rangle_{(N-1)/2}\right). \tag{19}$$

where $a_1 \cdot a_2 \cdot a_3 \ldots a_N = +1$ since $P(\mathbf{a}|\mathbf{x_m}) = 0$ when $a_1 \cdot a_2 \cdot a_3 \ldots a_N = -1$.

B Expressing the inequality in terms of correlators

As mentioned, our first goal is to express Eq.(16) as a function of some correlators. Let us recall the function we use in our main theorem,

$$f(\mathbf{a}) = \begin{cases} +1 & n_-(\mathbf{a}) = (4j+2); \quad \text{with } j \in \{0, 1, 2, \ldots\} \\ \\ -1 & \text{otherwise} \end{cases} \tag{20}$$

where $n_-(\mathbf{a})$ denotes the number of results in $\mathbf{a}$ that are equal to -1.

It turns out that the quantity (Eq. (16)) we would like to calculate, namely, $P(f(\mathbf{a}) = h_N|\mathbf{x}_m) - 1/2$ can be equivalently expressed as $h_N \cdot (P(f(\mathbf{a}) = +1|\mathbf{x}_m) - 1/2)$. The latter form is convenient since the function only takes value $+1$ for all N.

We proceed to express the latter in terms of correlators (as in the proof for three parties in the main text),

$$(h_N \cdot P(f(\mathbf{a}) = +1|\mathbf{x}_m) - 1/2) = 2^{-(N-1)} \boldsymbol{\alpha}' \cdot \mathbf{c}, \tag{21}$$

where

$$\begin{aligned} \boldsymbol{\alpha}' &= h_N \cdot (\alpha_0 - 2^{N-2}, \alpha_1, \alpha_2, \ldots, \alpha_{(N-1)/2}) \\ \mathbf{c} &= \left(1, \sum_{\mathcal{S}^1} \langle . \rangle_1, \sum_{\mathcal{S}^2} \langle . \rangle_2, \ldots, \sum_{\mathcal{S}^{(N-1)/2}} \langle . \rangle_{(N-1)/2}\right) \end{aligned} \tag{22}$$

Note that, since the function f symmetric under permutations, the vector $\mathbf{c}$ consists of the different sums of all k-point correlators, denoted by $\mathcal{S}^k$, where k ranges from 0 to $(N-1)/2$ because of Eq. (19). The vector $\boldsymbol{\alpha}'$ is the vector of coefficients for each sum of correlators. Our next goal is to compute this vector.

Recall that function f is such that $f(\mathbf{a}) = +1$ if $n_-(\mathbf{a}) = 4j+2$ for any $j \in \mathbb{N} \cup \{0\}$. By inspection, the explicit values of α_i can be written as

$$\alpha_i = \sum_{r=0}^{i} (-1)^r \binom{i}{r} \sum_{j \geq 0} \binom{n-i}{4j+2-r}. \tag{23}$$

For example, $\alpha_0 = \sum_{j \geq 0} \binom{n}{4j+2}$ as one would expect since α_0 simply counts the total number of terms $P(\mathbf{a}|\mathbf{x}_m)$ being summed to obtain $P(f(\mathbf{a}) = +1|\mathbf{x}_m)$.

Making use of the closed formula $\sum_{j \geq 0} \binom{n}{rj+a} = \frac{1}{r} \sum_{k=0}^{r-1} \omega^{-ka} (1 + \omega^k)^n$ [21], where $\omega = e^{i2\pi/r}$ is the r^{th} root of unity, we can simplify the second sum appearing in Eq. 23. Finally

we recall that the phase h_N was defined (in the main text) to be $h_N = \sqrt{2}\cos(N+4)\pi/4$. Putting all this together and performing the first sum in Eq. (23) gives us,

$$\alpha'_i = 2^{\frac{N-3}{2}} \left(-2\cos\frac{(N-2i)\pi}{4}\cos\frac{(N+4)\pi}{4} \right) \tag{24}$$

Notice that the term in the parenthesis is a phase taking values in the set $\{+1,-1\}$ since N is odd while the amplitude is independent of N. Thus, we can simplify Eqn. (24) for even and odd values of i as,

$$\alpha'_i = \begin{cases} 2^{(N-3)/2}(-1)^{\frac{N-i}{2}} & i \text{ odd} \\ 2^{(N-3)/2}(-1)^{\frac{i}{2}} & i \text{ even} \end{cases} \tag{25}$$

Thus, to prove that f possesses the property $h_N \cdot (P(f(\mathbf{a}) = +1|\mathbf{x_m}) - 1/2) \geq 0$ necessary to proving the main theorem is equivalent to proving

$$\boldsymbol{\alpha}' \cdot \mathbf{c} \geq 0, \tag{26}$$

for $\mathbf{c}$ as defined in Eqn. (22) and for the values of $\boldsymbol{\alpha}'$ given by Eqn. (25). This is the task of the following section, where we show that it follows from positivity constraints on $P(\mathbf{a}|\mathbf{x})$.

C Proving the inequality from positivity constraints

We show that positivity conditions derived from the swapped input $\bar{\mathbf{x}}_\mathbf{m} = (1,1,\ldots,1,0)$ may be used to show $\boldsymbol{\alpha}' \cdot \mathbf{c} \geq 0$. Notice that the components of $\bar{\mathbf{x}}_\mathbf{m}$ and $\mathbf{x_m}$ are opposite, ie. $\{\bar{\mathbf{x}}_\mathbf{m}\}_i = \{\mathbf{x_m}\}_i \oplus 1$ for all i. In the following we will repeatedly use the Mermin conditions of Eqn. (18).

We start by summing the positivity conditions $P(++\cdots+-|\bar{\mathbf{x}}_\mathbf{m}) \geq 0$ and $P(--\cdots-+|\bar{\mathbf{x}}_\mathbf{m}) \geq 0$. Using Eqn. (17), one can easily see that upon summing, all k-point correlators for *odd* k are cancelled out since these are multiplied by coefficients (products of a_is) that appear with opposite signs in the two positivity expressions. In contrast, k-point correlators for *even* k add up since they are multiplied by coefficients that appear with the same sign in the two expressions. For example, N being odd, the full correlator always cancels out while the $(N-1)$-point correlators always appear.

This leaves us with an expression containing only the even-body correlators,

$$1 + \sum_{i<j} a_i a_j \langle x_i x_j \rangle + \sum_{i<j<k<l} a_i a_j a_k a_l \langle x_i x_j x_k x_l \rangle + \cdots + \sum a_i \ldots a_p \underbrace{\langle x_i \ldots x_p \rangle}_{(N-1)\text{-pt. corr}} \geq 0. \tag{27}$$

Note once again, that this inequality is derived from the so-called swapped input $\bar{\mathbf{x}}_\mathbf{m}$. We aim to cast it in a form that can be compared directly with Eqn. (22), which comes from the chosen Mermin input $\mathbf{x_m}$. To this end, we need to convert Eqn. (27) to an expression of the form,

$$(\beta_0, \beta_1, \ldots \beta_{(N-1)/2}). \qquad \left(1, \sum\langle . \rangle_1, \ldots, \sum\langle . \rangle_{(N-1)/2}\right) \geq 0 \tag{28}$$

We first highlight the similarities and differences between the two preceding expressions, namely, the one we have *i.e.* Eqn. (27) and the one we want, *i.e.* Eqn. (28). Each contains $(N-1)/2$ distinct classes of terms. However the former contains only even k-point correlators

for $k = 2$ to $(N-1)$ while the latter contains all terms from $k = 1$ to $(N-1)/2$. Thus, terms of Eq. (27) must be mapped to ones in Eqn. (28). Moreover, since the point of making this mapping is to finally compare with Eqn. (22), we also note that the correlators appearing in Eqn. (27) are locally swapped relative to those appearing in Eqn. (22). Thus, our mapping must also convert correlators of the swapped input into those corresponding to the chosen input.

We demonstrate next that one may indeed transform the inequality (27) into the inequality (28) satisfying both the demands above. To this end, all the *even* k-point correlators (for $k \geq \frac{N-1}{2}$) appearing in Eqn. (27) are mapped to odd $(N-k)$-point correlators in Eqn. (28). Likewise, all the even k-point correlators (for $k < \frac{N-1}{2}$) of the swapped input appearing in Eqn. (27) are mapped to the corresponding k-point correlators of the chosen input in Eqn. (28).

These mappings make systematic use of the Mermin conditions Eqn. (18) and are made explicit in the following section.

C.1 Even-point correlators

Consider a $2k$-point correlator where $2k \leq (N-1)/2$. The correlators are of two forms and we show how they are transformed in each case:

- $\langle 11\ldots1\rangle_{2k}$. We would like to map this to the correlator $\langle 00\ldots0\rangle_{2k}$ appearing in $\mathbf{x_m}$. We achieve the mapping by completing each to the corresponding Mermin full-correlators $\langle \underbrace{11\ldots1}_{2k}\underbrace{100\ldots0}_{(N-2k)}\rangle_N = (-1)^k$ and $\langle \underbrace{00\ldots0}_{2k}\underbrace{100\ldots0}_{(N-2k)}\rangle_N = (-1)^0 = 1$. From the signs, we have the relation, $\langle 11\ldots1\rangle_{2k} = (-1)^k\langle 00\ldots0\rangle_{2k}$
- $\langle 11\ldots10\rangle_{2k}$, which we would like to map to $\langle 00\ldots01\rangle_{2k}$. Using the same ideas we get $\langle \underbrace{11\ldots10}_{2k}\underbrace{110\ldots0}_{(N-2k)}\rangle_N = (-1)^k$ and $\langle \underbrace{00\ldots01}_{2k}\underbrace{110\ldots0}_{(N-2k)}\rangle_N = (-1)^1 = -1$. Thus, giving us the relation $\langle 11\ldots10\rangle_{2k} = (-1)^{k+1}\langle 00\ldots01\rangle_{2k}$.

By inspection one can write the relationship

$$\underbrace{a_1a_2\ldots a_{2k}}_{\text{even}}\underbrace{\langle x_1x_2\ldots x_{2k}\rangle}_{\text{cor in }\bar{\mathbf{x}}_\mathbf{m}} = (-1)^k \underbrace{\langle x_1x_2\ldots x_{2k}\rangle}_{\text{cor in }\mathbf{x_m}(desired)}$$

for correlators of either form discussed above on multiplying with their corresponding coefficients. Since we have finally converted to the desired correlators of the chosen input $\bar{\mathbf{x}}$, we can read off β_i as the corresponding phase. Thus, $\beta_i = (-1)^{i/2}$ for even i.

C.2 Odd-point correlators

Consider now a $2k$-point correlator where $2k \geq (N-1)/2$. The correlators are again of two forms and may be transformed to the required $(N-2k)$-point correlators in each case. The only difference from before is that the two correlators are now complementary to each other in the swapped input. Since the details are similar, we simply state the final result $\beta_i = (-1)^{(N-i)/2}$ for odd i.

The final expression thus reads,

$$\beta_i = \begin{cases} (-1)^{\frac{N-i}{2}} & i \text{ odd} \\ \\ (-1)^{\frac{i}{2}} & i \text{ even} \end{cases} \tag{29}$$

Thus, the values of β given in Eqs. (29) exactly match the ones for α'_i (up to the constant factor) given in Eqn. 25. Together with the correlators matching those in $\mathbf{c}$, it proves that f satisfies the required $\boldsymbol{\alpha}' \cdot \mathbf{c} \geq 0$ and hence the full result.

D Proof that all distributions in decomposition maximally violate the Mermin inequality

We end by proving the claim made in the main text that if an observed probability distribution $P_{\text{obs}}(\mathbf{a}|\mathbf{x})$ violates maximally and algebraically the corresponding Mermin inequality, all the no-signaling components $P_e^{\text{ex}}(\mathbf{a}|\mathbf{x})$ present in its preparation must also algebraically violate the inequality.

We recall that the decomposition appears in the definition of intrinsic randomness given by,

$$G_{\text{int}}(g, \mathbf{x}, P_{\text{obs}}) = \max_{\{p(e|\mathbf{x}), P_e^{\text{ex}}\}} \sum_e p(e|\mathbf{x}) G_{\text{obs}}(g, \mathbf{x}, P_e^{\text{ex}})$$

subject to:

$$\sum_e p(e|\mathbf{x}) P_e^{\text{ex}}(\mathbf{a}|\mathbf{x}) = P_{\text{obs}}(\mathbf{a}|\mathbf{x}) \tag{30}$$

$$p(\mathbf{x}|e) \geq \delta \;\text{ with } \delta > 0 \;\; \forall\, \mathbf{x}, e \tag{31}$$

Since P_{obs} algebraically violates the Mermin inequality, this definition imposes stringent conditions on the correlators of P_{obs} satisfying the Mermin condition (18), namely that,

$$\langle x_1 \dots x_N \rangle_{P_{\text{obs}}} = \pm 1 = \sum_e p(e|x_1, \dots, x_N)\, \langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}} \tag{32}$$

where by normalization $\sum_e p(e|x_1, \dots, x_N) = +1$ and $-1 \leq \langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}} \leq +1$. Note that condition $p(\mathbf{x}|e) \geq \delta$ for all $\mathbf{x}, e$ for $\delta > 0$ can be inverted using the Bayes' rule to obtain $p(e|\mathbf{x}) > 0$ for all $\mathbf{x}, e$. Now is clear by convexity that the condition $p(\mathbf{x}|e) \geq \delta$ (denying *absolute* relaxation of freedom of choice) implies that all the correlator $\langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}}$ appearing in the Mermin inequality must also necessarily satisfy $\langle x_1 \dots x_N \rangle_{P_e^{\text{ex}}} = \pm 1$ for all e thus maximally violating the Mermin inequality. In fact it is also clear that this constraint on $p(\mathbf{x}|e)$ is strictly necessary to ensure that the decomposition correlations satisfy maximal Mermin violation. To see this, suppose $p(\mathbf{x}|e_0) = 0$, then the corresponding $\langle x_1 \dots x_N \rangle_{P_{e_0}^{\text{ex}}}$ is fully unconstrained while satisfying Eq. (32).

Is Global Asymptotic Cloning State Estimation?

Yuxiang Yang[1] **and Giulio Chiribella**[2]

1 **Department of Physics, Tsinghua University**
Beijing, 100084, China
yangyx09@mails.tsinghua.edu.cn

2 **Center for Quantum Information, Institute for Interdisciplinary Information Sciences, Tsinghua University**
Beijing, 100084, China
gchiribella@mail.tsinghua.edu.cn

Abstract

We investigate the asymptotic relationship between quantum cloning and quantum estimation from the global point of view where all the copies produced by the cloner are considered jointly. For an N-to-M cloner, we consider the overall fidelity between the state of the M output systems and the state of M ideal copies, and we ask whether the optimal fidelity is attained by a measure-and-prepare protocol in the limit $M \to \infty$. In order to gain intuition into the general problem, we analyze two concrete examples: (i) cloning qubit states on the equator of the Bloch sphere and (ii) cloning two-qubit maximally entangled states. In the first case, we show that the optimal measure-and-prepare fidelity converges to the fidelity of the optimal cloner in the limit $M \to \infty$. In the second case, we restrict our attention to economical covariant cloners, and again, we exhibit a measure-and-prepare protocol that achieves asymptotically the optimal fidelity. Quite counterintuitively, in both cases the optimal states that have to be prepared in order to maximize the overall fidelity are not product states corresponding to M identical copies, but instead suitable M-partite entangled states.

1998 ACM Subject Classification J.2 Physical sciences and engineering

Keywords and phrases quantum cloning, quantum estimation

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.220

1 Introduction

It is well known that every quantum machine producing a large number of indistinguishable clones—referred to as *asymptotic cloning machine*—is "equivalent" to a machine that measures the input states and re-prepares many identical copies of a state depending on the outcome [1, 2, 3, 4]. Here, "equivalent" has to be understood in the following sense: when one restricts the attention to a few clones, their state will be almost indistinguishable from the state that can be produced by a measure-and-prepare protocol. Precisely, the trace distance between the state of k clones produced by machine and the state of k clones produced by the measure-and-prepare protocol goes to zero as k/M, where M is the number of output copies [3, 4]. For $k = 1$, the fact that the state of each individual clone is asymptotically equal to the state produced by a measure-and-prepare protocol implies that the single-copy fidelity of quantum cloning is asymptotically equal to the fidelity of state estimation, a fact that is commonly known as "equivalence between asymptotic cloning and state estimation" [5].

In this paper we raise the question whether the equivalence between asymptotic cloning and state estimation continues to hold when one considers all the M clones together, rather than restricting the attention to a single clone or a small subset of k clones. We refer to this

new form of equivalence as *global asymptotic equivalence between quantum cloning and state estimation* and we conjecture that the equivalence holds. A few observations supporting the conjecture are the following: First of all, in the known cases—cloning of arbitrary pure states [6, 7] and the cloning of coherent states [8, 9, 10, 11, 12]—the equivalence holds, and in a fairly strong sense: the joint state of all the output clones converges in trace distance to the output state of a measure-and-prepare protocol. A more general argument supporting our conjecture comes from the intuition that producing a large number of identical copies means "classicalizing" the information contained in the input states, and therefore it is natural to expect that the optimal way to classicalize quantum information is to perform a measurement.

In order to discuss the question of the global equivalence one needs first to fix the rules of the game, by defining a suitable figure of merit. Here we consider the *global fidelity*, namely the overlap between the output state of all clones and the desired quantum state of M identical copies. In this setting, proving the equivalence means proving that the global fidelity of the optimal N-to-M cloner can be achieved by a measure-and-prepare protocol in the asymptotic limit $M \to \infty$. In order to gain intuition into the problem, we consider two concrete examples: the cloning of qubit states on the equator of the Bloch sphere and the cloning of two-qubit maximally entangled states. In the first case it is known that the optimal cloner, derived in Ref. [13], is *economical* [14, 15, 16], that is, it can be implemented by a unitary interaction between the N input copies and $M - N$ blank copies. An economical cloner is far from being implementable by a measure-and-prepare protocol, and observing an asymptotic equality of fidelities becomes here a quite non-trivial matter. In the second case (cloning of maximally entangled states), we will deliberately restrict ourselves to economical cloning machines, asking the question whether the global fidelity of the optimal economical cloner can be achieved by measurement and re-preparation. In both cases we will give an affirmative answer, showing that the difference between the global fidelity of the optimal economical N-to-M cloner and the global fidelity of the optimal measure-and-prepare protocol becomes negligible in the asymptotic limit $M \to \infty$, for every fixed value of N. Quite counterintuitively, we observe that the obvious protocols consisting in estimation of the unknown state and re-preparation of M identical copies do not reach the maximum fidelity, *even in the asymptotic limit.* This feature is in stark contrast with the intuition coming from the single-copy scenario, where re-preparing identical copies of the same state is asymptotically the best strategy.

2 Preliminaries

In this section we formalize the problem of the joint asymptotic equivalence and give an overview of the methods used in the rest of the paper.

2.1 The problem of the global asymptotic equivalence

Consider a set of states $\{|\psi_x\rangle\}_{x\in\mathsf{X}}$ in a finite dimensional Hilbert space $\mathcal{H}$. The task of optimal quantum cloning is to convert N perfect copies of an unknown state $|\psi_x\rangle$, given with probability p_x, into M approximate copies that are as accurate as possible. Examples of this problem are the universal cloning of pure states [17, 18, 19, 6, 7, 20] and the phase-covariant cloning [21, 13, 22, 23, 15, 16].

The most general cloning process will be described by a quantum channel (completely positive trace-preserving map) $\mathcal{C}$ transforming density matrices on $\mathcal{H}^{\otimes N}$ to density matrices

on $\mathcal{H}^{\otimes M}$. As a figure of merit for the quality of the copies we will consider the *global fidelity*,

$$F[N \to M] = \sum_{x \in \mathsf{X}} \mathrm{Tr}\left[\psi_x^{\otimes M} \mathcal{C}(\psi_x^{\otimes N})\right] \qquad \psi_x := |\psi_x\rangle\langle\psi_x|. \tag{1}$$

When the set $\{|\psi_x\rangle\}_{x\in\mathsf{X}}$ is continuous, it is understood that the sum over the possible input states has to replaced with an integral with a suitable probability distribution $p(x)\, dx$. The optimal cloner will be the quantum channel that maximizes $F[N \to M]$. The fidelity of the optimal cloner will be denoted by by $F_{clon}[N \to M]$.

In addition to the maximum over all channels, it is important to consider the maximum of $F[N \to M]$ over the set of *measure-and-prepare channels*. Operationally, a measure-and-prepare channel can be realized by measuring the input copies with a POVM $(P_y)_{y\in\mathsf{Y}}$ and, when the measurement gives outcome y, by re-preparing a state ρ_y. Averaging over the measurement outcomes, the action of the measure-and-prepare channel on the density matrices is given by $\mathcal{C}(\rho) = \sum_{y\in\mathsf{Y}} \mathrm{Tr}[P_y\rho]\ \rho_y$. We will denote by $F_{est}[N \to M]$ the maximum of the fidelity over the set of measure-and-prepare channels. Such a maximum is known in the literature as *classical fidelity threshold* [24, 25, 26, 27, 12] and can be used as a benchmark for the experimental demonstration of quantum advantages.

In the following we will ask the question whether the difference between $F_{clon}[N \to M]$ and $F_{est}[N \to M]$ becomes negligible in the asymptotic limit $M \to \infty$, while keeping N fixed. An affirmative answer to this question would mean that the quantum way to process information and the classical way fare equally well in the asymptotic limit. In the formalization of the problem there is a catch, because both fidelities converge to zero in many interesting cases when the family of states to be cloned is continuous: a non-vanishing fidelity would indeed violate the Heisenberg limit of quantum metrology [28]. In order not to trivialize the question, it is then important to consider the *relative* difference between the two fidelities, given by

$$\Delta[N \to M] := \frac{F_{clon}[N \to M] - F_{est}[N \to M]}{F_{clon}[N \to M]}. \tag{2}$$

Our conjecture is that, for every fixed N, the relative difference vanishes in the limit $M \to \infty$. In formula:

$$\lim_{M\to\infty} \frac{F_{clon}[N \to M] - F_{est}[N \to M]}{F_{clon}[N \to M]} = 0 \qquad \forall N \in \mathbb{N}. \tag{3}$$

We refer to the conjectured equality as *global asymptotic equivalence between quantum cloning and quantum state estimation*. Of course, here the word "global" refers to the fact that we are considering the global fidelity as the performance measur, as opposed to the single-copy fidelity considered in the previous literature. [19, 6, 20]

From previous results on optimal cloning we know that the relation is satisfied in the case of universal quantum cloning [6, 7] (see [4] for the proof that the optimal channel converges to a measure-and-prepare channel) and in the case of the coherent-state quantum cloning [8, 9, 10, 11, 12] (see [12] for the proof that $F_{clon}[N \to M]$ becomes asymptotically equal to $F_{est}[N \to M]$, up to a negligible error). In the following we will exhibit two new examples supporting the conjecture that joint cloning is asymptotically equivalent to state estimation.

In the first example, we consider the optimal cloning of qubit states on the equator of the Bloch sphere. In this case, the optimal N-to-M cloner is known [13]) and has a very interesting feature: it can be realized through a unitary interaction between the N input copies and only $M - N$ blank copies of the input system. In formula, the optimal quantum

channel has the form

$$\mathcal{C}(\rho) = U\left[\rho \otimes |0\rangle\langle 0|^{\otimes(M-N)}\right] U^\dagger, \tag{4}$$

where $U : \mathcal{H}^{\otimes M} \to \mathcal{H}^{\otimes M}$ is a unitary operator and $|0\rangle$ is a fixed state in $\mathcal{H}$. Cloning channels of this form are usually referred to as *economical* [14, 15, 16]. For the optimal cloner of qubit states on the equator, we will show that our conjecture holds, by explicitly constructing a family of measure-and-prepare channels that attains the maximum fidelity $F_{clon}[N \to M]$. In a sense, this example is more intriguing than the previous ones, because the economical cloner considered here is far from being achieved by measure-and-prepare protocols: the asymptotic equivalence is then a non-trivial relation between the optimal joint fidelities.

In the second example, we consider the cloning of two-qubit maximally entangled states. For simplicity, here we restrict our attention to economical quantum cloners satisfying a natural symmetry requirement, and we denote by $F_{clon,eco}[N \to M]$ the maximum fidelity achieved by these channels. In this case, we show that the maximum value $F_{clon,eco}[N \to M]$ can be achieved by a suitable family of measure-and-prepare channels, in the limit $M \to N$. Again, this example supports the validity of our conjecture.

2.2 General methods

Here we make some general considerations that apply to the two specific examples considered in the paper.

2.2.1 Covariant economical channels

In many relevant cases, the unknown state to be cloned is of the form $|\psi_g\rangle := U_g|\psi\rangle$, where $|\psi\rangle \in \mathcal{H}$ is unit vector and $U : \mathsf{G} \to \mathsf{Lin}(\mathcal{H}), g \mapsto U_g$ is a unitary representation of a compact group G on the set $\mathsf{Lin}(\mathcal{H})$ of linear operators on $\mathcal{H}$. Examples of this problem are the universal cloning of pure states [7] and the phase-covariant cloning [21, 13, 22, 23, 15, 16]. Due to the symmetry of the states, the maximum of the fidelity can be achieved by choosing a *covariant channel*, namely a quantum channel satisfying the property

$$\mathcal{C} \circ \mathcal{U}_g^{\otimes N} = \mathcal{U}_g^{\otimes M} \circ \mathcal{C} \qquad g \in \mathsf{G}, \tag{5}$$

where $\mathcal{U}_g$ is the unitary channel defined by $\mathcal{U}_g(\rho) = U_g \rho U_g^\dagger$, for every density matrix ρ. For covariant channels, the expression of the fidelity is reduced to

$$F[N \to M] = \mathrm{Tr}\left[\psi^{\otimes M} \mathcal{C}(\psi^{\otimes N})\right]. \tag{6}$$

A further simplification arises if we assume that the covariant channel $\mathcal{C}$ is economical, namely $\mathcal{C}(\rho) = V\rho V^\dagger$ for a suitable isometry V: the fidelity takes the simple form

$$F[N \to M] = \left|\langle\psi|^{\otimes M} V |\psi\rangle^{\otimes N}\right|^2 \tag{7}$$

and the covariance condition becomes

$$U_g^{\otimes M} V \left(U_g^{\otimes N}\right)^\dagger = \omega_g \, V \qquad g \in \mathsf{G}, \tag{8}$$

where $\omega : \mathsf{G} \to \mathbb{C}$ is a one-dimensional representation of the group G. For the cloning of maximally entangled states of qudits, where the group is $SU(d)$, Eq. (8) is simply

$$U_g^{\otimes M} V \left(U_g^{\otimes N}\right)^\dagger = \; V \qquad g \in \mathsf{G}, \tag{9}$$

because the only one-dimensional representation of $SU(d)$ is the trivial one ($\omega_g = 1, \; \forall g$).

2.2.2 Covariant measure-and-prepare protocols

In order to prove the global asymptotic equivalence, our goal is to construct a family of measure-and-prepare protocols that attains the fidelity of the best quantum cloners in the limit $M \to \infty$. To achieve this goal, we will make a series of assumptions motivated by physical intuition. A posteriori, the fact that our protocols attain the desired fidelity will provide a confirmation that the intuition was sound.

First of all, for an input state of the form $|\psi_g\rangle = U_g|\psi\rangle$ we will consider measure-and-prepare strategies that are based on state estimation, namely strategies where the set of measurement outcomes coincides with the set parametrizing the input states, namely $\mathsf{X} \equiv \mathsf{G}$. Hence, the measurement is described by a POVM $P_{\hat{g}}\ d\hat{g}$ with in the group G, normalized as $\int \mathrm{d}\,\hat{g}\ P_{\hat{g}} = I^{\otimes N}$.

For the re-preparation stage , we will require that the states that are re-prepared have the form $|\Phi_{\hat{g}}\rangle = U_{\hat{g}}^{\otimes M}|\Phi\rangle$, for a given unit vector $|\Phi\rangle \in \mathcal{H}^{\otimes M}$. With this particular choice, the optimization of the measure-and-prepare protocol is equivalent to the optimization of a state estimation protocol that is designed to maximize the average of the function

$$f(\hat{g}, g) := \mathrm{Tr}\left[\Phi_{\hat{g}}\ \psi_g^{\otimes M}\right]. \tag{10}$$

In this case, is known that the optimal POVM can be chosen to be *covariant* [29], that is, $P_{\hat{g}} = U_{\hat{g}}^{\otimes N} \eta U_{\hat{g}}^{\dagger \otimes N}$ where $\eta \in \mathsf{Lin}\left(\mathcal{H}^{\otimes N}\right)$ is a suitable positive operator, called the seed of POVM. For a covariant POVM, the probability density $p(\hat{g}|g) = \mathrm{Tr}[P_{\hat{g}}\psi_g^{\otimes N}]$ satisfies the relation

$$p(h\hat{g}|hg) = p(\hat{g}|g) \qquad \forall h, \hat{g}, g \in \mathsf{G}. \tag{11}$$

Hence, the fidelity of the corresponding measure-and-prepare protocol becomes

$$F[N \to M] = \int \mathrm{d}\, g\ \ \mathrm{Tr}\left[\eta\ \psi_g^{\otimes N}\right]\ \ \mathrm{Tr}\left[\Phi\ \psi_g^{\otimes M}\right]. \tag{12}$$

Finding the optimal measure-and-prepare protocol is then reduced to finding the optimal operator η and the optimal state $|\Phi\rangle$. To this purpose, in the two examples considered in this paper we will make a suitable ansatz on the form of the state $|\Phi\rangle$, which guarantees that $\mathrm{Tr}\left[\eta\ \psi_g^{\otimes N}\right]$, as a function of g, varies slowly with respect to $\mathrm{Tr}\left[\Phi\ \psi_g^{\otimes M}\right]$, which is concentrated around its maximum at $g = e$, the identity element of the group. Under this ansatz, the fidelity can be approximated as

$$F[N \to M] \approx \left\{\int \mathrm{d}\, g\ \ \mathrm{Tr}\left[\Phi\ \psi_g^{\otimes M}\right]\right\}\ \ \mathrm{Tr}\left[\eta\ \psi^{\otimes N}\right] = \langle\Phi|\rho_{aver}^{(M)}|\Phi\rangle\ p_{true}^{(N)}. \tag{13}$$

where $\rho_{aver}^{(M)} := \int dg\ \psi_g^{\otimes M}$ is the average state of M ideal copies and $p_{true}^{(N)} := \mathrm{Tr}\left[\eta\ \psi^{\otimes N}\right]$ is the probability density that the estimated value $\hat{g}$ coincides with true values g.

Thanks to Eq. (13), optimizing the measure-and-prepare protocol is reduced to two independent optimization problems: the maximization of the fidelity between the state $|\Phi\rangle$ and the average state $\rho_{aver}^{(M)}$ (under the restriction that $|\Phi\rangle$ must be compatible with the ansatz) and the maximization of the probability density $p_{true}^{(N)}$. In the specific cases considered in this paper, we will show that the ansatz can be done without loss of generality: indeed, the fidelity achieved by measure-and-prepare protocols satisfying the ansatz approaches the fidelity of the optimal quantum channel.

3 Cloning equatorial qubit states

Here we consider the optimal N-to-M cloning of pure qubit states on the equator of the Bloch sphere, evaluating the asymptotic expression of the optimal quantum fidelity and showing that it can be achieved via a suitable measure-and-prepare protocol.

3.1 The performance of the optimal quantum cloner

Consider the qubit states on the equator of the Bloch sphere, defined as

$$\begin{aligned} |\psi_\theta\rangle &= \frac{|0\rangle + e^{i\theta}|1\rangle}{\sqrt{2}} \qquad \theta \in [-\pi, \pi) \\ &= U_\theta |\psi\rangle \\ U_\theta &:= \exp\left[\frac{i\theta(\sigma_z + I)}{2}\right], \quad |\psi\rangle := \frac{|0\rangle + |1\rangle}{\sqrt{2}}. \end{aligned}$$

The state of the N input copies can be represented as

$$|\psi\rangle^{\otimes N} = \sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}} |N/2, n\rangle \qquad b_{N,n} := \frac{1}{2^N} \binom{N}{N/2+n}$$

where $\{|N/2, n\rangle \mid n = -N/2, \dots, N/2\}$ are the Dicke states and $b_{N,n}$ is the binomial distribution.

The optimal cloning channel was derived in Ref. [13]. When $M - N$ is even, the optimal channel is covariant with respect to the action of the phase shifts U_θ and economical, i.e. of the form $\mathcal{C}(\rho) = V\rho V^\dagger$ where $V : \mathcal{H}^{\otimes N} \to \mathcal{H}^{\otimes M}$ is an isometry(i.e. $V^\dagger V = I$). Specifically, the isometry of the optimal cloner is $V = \sum_{n=-N/2}^{N/2} |M/2, n\rangle\langle N/2, n|$ and produces the output state

$$V|\psi\rangle^{\otimes N} = \sum_{m=-N/2}^{N/2} \sqrt{b_{N,m}} |M/2, m\rangle. \tag{14}$$

Inserting this expression in Eq. (7) one gets the maximum fidelity [13]

$$F_{clon}[N \to M] = \left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n} b_{M,n}} \right)^2. \tag{15}$$

When M is large compared to N, the fidelity becomes:

$$F_{clon}[N \to M] \approx b_{M,0} \left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}} \right)^2 \qquad M \gg N. \tag{16}$$

In the next subsection we will construct a measure-and-prepare channel that achieves this asymptotic value for every $N \in \mathbb{N}$.

Note that the optimal quantum fidelity in Eq. (15) has a simple expression when N is large ($N \gg 1$). In this case, the probability distributions $b_{N,n}$ and $b_{M,n}$ are well approximated by

the Gaussian distributions $g_N(x) = \sqrt{\frac{2}{\pi N}} e^{-2x^2/N}$ and $g_M(x) = \sqrt{\frac{2}{\pi M}} e^{-2x^2/M}$, respectively. Replacing the summation in Eq. (15) with a Gaussian integral, one gets

$$F_{clon}[N \to M] \approx \frac{\sqrt{4MN}}{M+N} \qquad N \gg 1. \tag{17}$$

Incidentally, it is interesting to observe that in this regime the fidelity is close to 1 whenever the number of extra-copies $M - N$ is negligible compared to N, whereas it is close to 0 whenever N is negligible compared to M. This fact is an illustration of the standard quantum limit for cloning introduced in Ref. [28].

3.2 A family of measure-and-prepare protocols achieving asymptotically the optimal fidelity

Here we consider the maximization of the cloning fidelity over measure-and-prepare protocols based on state estimation (cf. subsection 2.2.2). For equatorial qubit states, the measure-and-prepare protocol consists in the estimation of the parameter $\theta \in [0, 2\pi)$ from the N input copies and in the re-preparation of an M-qubit output state $|\Phi_{\hat{\theta}}\rangle$ conditional to the estimate $\hat{\theta}$. In order to maximize the global fidelity, the states $|\Phi_{\hat{\theta}}\rangle = U_{\hat{\theta}}^{\otimes N}|\Phi\rangle$ should be contained in the symmetric space spanned by the Dicke states $\{|M/2, m\rangle \mid m = -M/2, \ldots, M/2\}$, i.e. $|\Phi_{\hat{\theta}}\rangle = U_{\hat{\theta}}^{\otimes M}|\Phi\rangle$ with

$$|\Phi\rangle = \sum_{m=-M/2}^{M/2} \sqrt{p_{M,m}} |M/2, m\rangle, \tag{18}$$

for some suitable coefficients $\{p_{M,m}\}$ that can be chosen to be positive without loss of generality. For states of this form, the optimal covariant POVM is known [29] and is given by $P_{\hat{\theta}} = U_{\hat{\theta}}^{\otimes N} \eta U_{\hat{\theta}}^{\dagger \otimes N}$ where the seed η is the rank-one operator $\eta = |\eta\rangle\langle\eta|$ with

$$|\eta\rangle := \sum_{n=-N/2}^{N/2} |N/2, n\rangle. \tag{19}$$

Now, the expression for the fidelity is given by

$$F[N \to M] = \int \frac{\mathrm{d}\,\theta}{2\pi} \ \mathrm{Tr}\left[\eta \ \psi_\theta^{\otimes N}\right] \ \mathrm{Tr}\left[\Phi \ \psi_\theta^{\otimes M}\right]. \tag{20}$$

and the goal is to maximize it over all possible choices for the coefficients in Eq. (18). The optimization can be carried out for given values of N and M. However, the full optimization is not needed if one just wants to discuss the large M asymptotics. To this purpose, we make a variational ansatz for the coefficients $\{p_{M,m}\}$ and later we will prove that asymptotically the ansatz is not too restrictive, because it allows one to achieve the fidelity of the optimal cloner. Our variational ansatz is the following:

$$p_{M,m}(\lambda) = \begin{cases} b_{\lceil M/\lambda \rceil, m} & -\frac{\lceil M/\lambda \rceil}{2} \leqslant m \leqslant \frac{\lceil M/\lambda \rceil}{2} \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

for some $\lambda \geqslant 1$. We denote by $|\Phi(\lambda)\rangle$ the state in Eq. (18) with the above choice of coefficients.

With our variational choice, the expression for the fidelity in Eq. (20) can be simplified in the regime

$$\frac{M}{1+\lambda} \gg N. \tag{22}$$

Indeed, under this condition the function $\mathrm{Tr}\left[\eta\ \psi_\theta^{\otimes N}\right]$ varies slowly with respect to $\mathrm{Tr}\left[\Phi(\lambda)\ \psi_\theta^{\otimes M}\right]$ (see the proof in the Appendix) and therefore we can approximate Eq. (20) with

$$\begin{aligned} F[N\to M] &\approx \mathrm{Tr}\left[\eta\ \psi^{\otimes N}\right]\ \left(\int \frac{\mathrm{d}\,\theta}{2\pi}\ \mathrm{Tr}\left[\Phi(\lambda)\ \psi_\theta^{\otimes M}\right]\right) \\ &= p^{(N)}_{true}\ \langle\Phi(\lambda)|\rho^{(M)}_{aver}|\Phi(\lambda)\rangle \end{aligned} \tag{23}$$

where $\rho^{(M)}_{aver} = \sum_{m=-M/2}^{M/2} b_{M,m}|M/2,m\rangle\langle M/2,m|$ and $p^{(N)}_{true} = \left(\sum_n \sqrt{b_{N,n}}\right)^2$.

If there were no constraint on $|\Phi(\lambda)\rangle$, the optimal choice that maximizes the expectation value $\langle\Phi(\lambda)|\rho^{(M)}_{aver}|\Phi(\lambda)\rangle$ would be $|\Phi(\lambda)\rangle = |M/2,0\rangle$, the eigenvector corresponding to the maximum eigenvalue of $\rho^{(M)}_{aver}$. However, from Eq. (21) it is clear that this would require $M/\lambda < 1$, in contradiction with the condition $M/(1+\lambda) \gg N$, under which Eq. (23) was derived. What can be done instead is to choose λ in such a way that both conditions $\lambda \gg 1$ and $M/(1+\lambda) \gg N$ are satisfied. With this choice, the expectation value $\langle\Phi(\lambda)|\rho^{(M)}_{aver}|\Phi(\lambda)\rangle$ is still close to the maximum eigenvalue:

$$\langle\Phi(\lambda)|\rho^{(M)}_{aver}|\Phi(\lambda)\rangle = \sum_{m=-\lceil M/\lambda\rceil/2}^{\lceil M/\lambda\rceil/2} b_{M,m}b_{\lceil M/\lambda\rceil,m} \approx b_{M,0} \qquad \lambda \gg 1$$

Hence, the fidelity of our variational measure-and-prepare protocol, denoted by $F_\lambda[N\to M]$, becomes

$$F_\lambda[N\to M] \approx \langle\Phi(\lambda)|\rho^{(M)}_{aver}|\Phi(\lambda)\rangle\ p^{(N)}_{true} \approx b_{M,0}\left(\sum_{n=-N/2}^{N/2}\sqrt{b_{N,n}}\right)^2 \approx F_{clon}[N\to M], \tag{24}$$

where the last approximate equality comes from Eq. (16). Since by definition the maximum fidelity $F_{est}[N\to M]$ over all measure-and-prepare channels is lower bounded by $F_\lambda[N\to M]$ and upper bounded by $F_{clon}[N\to M]$, we conclude that

$$F_{est}[N\to M] \approx F_\lambda[N\to M] \approx F_{clon}[N\to M] \qquad \frac{M}{1+\lambda} \gg N, \lambda \gg 1.$$

This shows that asymptotically, there is no loss of generality in our ansatz: the protocols satisfying the ansatz have a fidelity that is arbitrarily close to the fidelity of the best measure-and-prepare protocol, which in turn is asymptotically equal to the fidelity of the best quantum cloner.

Let us consider now the fidelity of the naive measure-and-prepare protocol that consists in estimating the phase θ and re-preparing M identical copies of the estimated state. In this case, we have $|\Phi\rangle = |\psi\rangle^{\otimes M} \equiv |\Phi(\lambda=1)\rangle$ [cf. Eqs. (18) and (21)], and, therefore, $\langle\Phi(\lambda=1)|\rho^{(M)}_{aver}|\Phi(\lambda=1)\rangle = \sum_{m=-M/2}^{M/2} b^2_{M,m}$.

For large M, the Gaussian approximation gives $\langle\Phi(\lambda=1)|\rho^{(M)}_{aver}|\Phi(\lambda=1)\rangle \approx \sqrt{1/(\pi M)} \approx b_{M,0}/\sqrt{2}$, and the fidelity becomes

$$F_{\lambda=1}[N\to M] \approx \frac{b_{M,0}}{\sqrt{2}}\left(\sum_{n=-N/2}^{N/2}\sqrt{b_{N,n}}\right)^2 \approx \frac{F_{clon}[N\to M]}{\sqrt{2}} \qquad \forall N\in\mathbb{N}.$$

This proves that re-preparing M identical copies is a strictly suboptimal strategy, which cannot reach the global fidelity of the optimal cloner.

In summary, in this section we showed that the fidelity of the optimal quantum cloner $F_{clon}[N \to M]$ is asymptotically equal to the fidelity of the optimal measure-and-prepare protocol $F_{est}[N \to M]$ in the limit $M \to \infty$. Hence, the conjectured equality in Eq. (3) is verified. However, achieving the optimal fidelity requires one to prepare suitable M-partite entangled states: the simple strategy consisting in re-preparing M identical copies of estimated state does not give the maximal fidelity, *even in the asymptotic limit.*

4 Cloning two-qubit maximally entangled states

In this section we consider the N-to-M cloning of two-qubit maximally entangled states, computing the fidelity of the optimal economical covariant cloner and showing that it can be asymptotically attained via a suitable measure-and-prepare protocol.

Consider a general two-qubit maximally entangled state $|\psi_g\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$, $\mathcal{H}_A \simeq \mathcal{H}_B \simeq \mathbb{C}^2$, which can be parametrized as

$$|\psi_g\rangle = \frac{(U_g \otimes I)|I\rangle\!\rangle}{\sqrt{2}}, \qquad g \in SU(2). \tag{25}$$

Here we are using the "double-ket notation" $|A\rangle\!\rangle := \sum_{m,n} \langle m|A|n\rangle \ |m\rangle|n\rangle$ for a generic operator $A \in \mathsf{Lin}(\mathcal{H})$ [30].

We now give a convenient decomposition of the input state $|\psi_g\rangle^{\otimes N} = (\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N} \simeq \mathcal{H}_A^{\otimes N} \otimes \mathcal{H}_B^{\otimes N}$. With a suitable choice of basis, the Hilbert space $\mathcal{H}_A^{\otimes N}$, can be decomposed as a direct sum of tensor product pairs

$$\mathcal{H}_A^{\otimes N} = \bigoplus_{j=j_{min}^{(N)}}^{N/2} \left(\mathcal{R}_A^{(j,N)} \otimes \mathcal{M}_A^{(j,N)} \right), \tag{26}$$

where j is the quantum number of the total angular momentum and $j_{min}^{(N)} = 0$ for even N while $j_{min}^{(N)} = \frac{1}{2}$ for odd N, $\mathcal{R}_A^{(j,N)}$ is a representation space, of dimension $d_j = 2j+1$, and $\mathcal{M}_A^{(j,N)}$ is a multiplicity space, of dimension $m_j^{(N)} = \frac{2j+1}{N/2+j+1}\binom{N}{N/2+j}$ (see e.g. Ref. [31]). Relative to this decomposition, we can express $U_g^{\otimes N}$ as a block diagonal matrix, where each block corresponds to an irreducible representation of $SU(2)$, namely

$$U_g^{\otimes N} = \bigoplus_{j=j_{min}^{(N)}}^{N/2} \left[U_g^{(j,N)} \otimes I_{m_j}^{(N)} \right]. \tag{27}$$

where $U_g^{(j,N)} \in \mathsf{Lin}(\mathcal{R}_A^{(j,N)})$ is the unitary operator representing the action of the element $g \in SU(2)$ and $I_{m_j}^{(N)}$ denotes the identity on $\mathcal{M}_A^{(j,N)}$.

Using Eq. (27), the input state $|\psi_g\rangle^{\otimes N}$ can be cast in the form

$$|\psi_g\rangle^{\otimes N} = 2^{-N/2}(U_g \otimes I)^{\otimes N}|I\rangle\!\rangle^{\otimes N} = 2^{-N/2} \bigoplus_{j=j_{min}^{(N)}}^{N/2} \left(|U_g^{(j,N)}\rangle\!\rangle \otimes |I_{m_j}^{(N)}\rangle\!\rangle \right) \tag{28}$$

with $|U_g^{(j,N)}\rangle\!\rangle \in \mathcal{R}_A^{(j,N)} \otimes \mathcal{R}_B^{(j,N)}$ and $|I_{m_j}^{(N)}\rangle\!\rangle \in \mathcal{M}_A^{(j,N)} \otimes \mathcal{M}_B^{(j,N)}$. Hence, we obtained the decomposition

$$|\psi_g\rangle^{\otimes N} = \bigoplus_{j=j_{min}^{(N)}}^{N/2} \sqrt{c_j^{(N)}} |\psi_g^{(j,N)}\rangle \qquad |\psi_g^{(j,N)}\rangle := \frac{|U_g^{(j,N)}\rangle\!\rangle}{\sqrt{d_j}} \otimes \frac{|I_{m_j}^{(N)}\rangle\!\rangle}{\sqrt{m_j^{(N)}}} \tag{29}$$

and $c_j^{(N)} := \frac{d_j m_j^{(N)}}{2^N} = \frac{(2j+1)^2}{(N/2+j+1)}\, b_{N,j}$, $b_{N,j}$ being the binomial distribution $b_{N,j} = \binom{N}{N/2+j}/2^N$. Note that every state $|\psi_g\rangle^{\otimes N}$ in Eq. (29) belongs to the subspace

$$\mathcal{H}_{ent}^{(N)} := \bigoplus_{j=j_{min}^{(N)}}^{N/2} \left(\mathcal{R}_A^{(j,N)} \otimes \mathcal{R}_B^{(j,N)} \otimes \mathcal{M}_A^{(j,N)} \otimes \mathcal{M}_B^{(j,N)}\right) \subset (\mathcal{H}_A \otimes \mathcal{H}_B)^{\otimes N}. \tag{30}$$

Hence, for the optimization of the fidelity we can restrict our attention to this subspace and consider quantum channels that map states on $\mathcal{H}_{ent}^{(N)}$ to states on $\mathcal{H}_{ent}^{(M)}$.

4.1 The performance of the optimal economical covariant cloner

Here we focus here on a special type of cloning machines, namely economical covariant cloning machines [14, 15, 16]. An economical covariant cloner is described by an isometric channel $\mathcal{C}(\rho) = V\rho V^\dagger$, where $V : \mathcal{H}_{ent}^{(N)} \to \mathcal{H}_{ent}^{(M)}$ is an isometry satisfying the covariance requirement, which in our case is expressed by the relation

$$(U_g \otimes U_h)^{\otimes M} V \left(U_g^\dagger \otimes U_h^\dagger\right)^{\otimes N} = V \qquad \forall g, h \in \mathsf{G}. \tag{31}$$

Note that the action of $(U_g \otimes U_h)^{\otimes N}$, restricted to the subspace $\mathcal{H}_{ent}^{(N)}$ is

$$(U_g \otimes U_h)^{\otimes N}\big|_{\mathcal{H}_{ent}^{(N)}} = \bigoplus_{j=j_{min}^{(N)}}^{N/2} \left(U_g^{(j,N)} \otimes U_h^{(j,N)} \otimes I_{m_j,A}^{(N)} \otimes I_{m_j,B}^{(N)}\right),$$

where $I_{m_j,A}^{(N)}$ ($I_{m_j,B}^{(N)}$) denotes the identity on the multiplicity space $\mathcal{M}_A^{(j,N)}$ ($\mathcal{M}_B^{(j,N)}$). A similar decomposition holds for the action of $(U_g \otimes U_h)^{\otimes M}$ restricted to the subspace $\mathcal{H}_{ent}^{(M)}$.

Now, using the Schur's lemma, Eq. (31) is reduced to

$$V = \bigoplus_{j=j_{min}^{(N)}}^{N/2} (R_j \otimes M_j), \tag{32}$$

where $R_j : \mathcal{R}_A^{(j,N)} \otimes \mathcal{R}_B^{(j,N)} \to \mathcal{R}_A^{(j,M)} \otimes \mathcal{R}_B^{(j,M)}$ is an isometry acting on the representation spaces and satisfying $\left(U_g^{(j,M)} \otimes U_h^{(j,M)}\right) R_j \left(U_g^{(j,N)} \otimes U_h^{(j,N)}\right) = V \quad \forall g, h \in \mathsf{G}$, and $M_j : \mathcal{M}_A^{(j,N)} \otimes \mathcal{M}_B^{(j,N)} \to \mathcal{M}_A^{(j,M)} \otimes \mathcal{M}_B^{(j,M)}$ is an isometry acting on the multiplicity spaces.

The fidelity of the economical cloner with isometry V in Eq. (32) is given by

$$F[N \to M] = \left|\langle\psi|^{\otimes M}\ V|\psi\rangle^{\otimes N}\right|^2 = \left|\sum_{j=j_{min}^{(N)}}^{N/2} \sqrt{c_j^{(N)} c_j^{(M)}}\, \langle\psi^{(j,N)}|V|\psi^{(j,M)}\rangle\right|^2$$
$$\leq \left(\sum_{j=j_{min}^{(N)}}^{N/2} \sqrt{c_j^{(N)} c_j^{(M)}}\right)^2.$$

The equality holds when $V|\psi^{(j,N)}\rangle = |\psi^{(j,M)}\rangle$ for every $j = j_{min}^{(N)}, \ldots, N/2$, or, equivalently,

$$R_j \frac{|I^{(j,N)}\rangle\!\rangle}{\sqrt{d_j}} = \frac{|I^{(j,N)}\rangle\!\rangle}{\sqrt{d_j}}, \qquad M_j \frac{|I_{m_j}^{(N)}\rangle\!\rangle}{\sqrt{m_j^{(N)}}} = \frac{|I_{m_j}^{(M)}\rangle\!\rangle}{\sqrt{m_j^{(M)}}}$$

Interestingly, the optimal covariant economical cloner can be achieved using local operations, because the above isometries only represent an embedding of the state of N systems on A's and B's sides into Hilbert space of M systems, and these embedding operations can be carried out locally.

The maximal fidelity among all possible economical covariant cloner is then

$$F_{eco,clon}[N \to M] = \left(\sum_{j=j_{min}^{(N)}}^{N/2} \sqrt{c_j^{(N)} c_j^{(M)}} \right)^2 , \tag{33}$$

and, when M is large compared to N, becomes

$$F_{eco,clon}[N \to M] \approx \frac{2 b_{M,0}}{M} \left(\sum_{j=j_{min}^{(N)}}^{N/2} \sqrt{\frac{b_{N,j}(2j+1)^4}{N/2+j+1}} \right)^2 \quad M \gg N. \tag{34}$$

In the next subsection we will construct a measure-and-prepare channel that achieves this asymptotic value, despite the fact that the fact that the cloning machine considered here is economical, and, therefore, far from a measure-and-prepare channel.

Before concluding, it is worth noting that the expression for the optimal quantum fidelity becomes simpler when N is large ($N \gg 1$). Approximating the summation in Eq. (33) with a Gaussian integral, one obtains the fidelity

$$F_{eco,clon}[N \to M] \approx \left(\int_0^{N/2} 8x^2 \sqrt{\frac{g_M(x) g_N(x)}{MN}}\, \mathrm{d}\, x \right)^2 \approx \left(\frac{4N}{M} \right)^{3/2} \quad N \gg 1. \tag{35}$$

Also in this case, it is also interesting to observe that the fidelity is close to 1 whenever the number of extra-copies $M - N$ is negligible compared to N, whereas it is close to 0 whenever N is negligible compared to M, in agreement with the standard quantum limit for cloning [28].

4.2 A family of measure-and-prepare protocols achieving asymptotically the optimal fidelity

Here we show how to reach the fidelity $F_{clon,eco}[N \to M]$ with a suitable measure-and-prepare protocol. Also in this case, we will first make a series of assumptions on the protocol, and we will eventually show that asymptotically our choice achieves the desired fidelity.

To start with, we consider strategies where the states re-prepared are of the form

$$|\Phi_{\hat{g}}\rangle = \bigoplus_{j=j_{min}^{(M)}}^{M/2} \sqrt{p_j^{(M)}}\ |\psi_g^{(j,M)}\rangle \tag{36}$$

where $|\psi_g^{(j,M)}\rangle$ is the vector defined in Eq. (29) and $\{p_j^{(M)}\}$ are some non-negative coefficients. Our choice is quite natural, as it is motivated by the form of the desired states $|\psi_g\rangle^{\otimes M}$ [cf. Eq. (29)].

Once we assume states of this form for the re-preparation, the optimal POVM for the measurement is known [31] and is given by the square-root measurement [32], which in this case has the expression $P_{\hat{g}} = U_{\hat{g}}^{\otimes N} \eta U_{\hat{g}}^{\dagger \otimes N}$, where $\eta = |\eta\rangle\langle\eta|$ and $|\eta\rangle = \bigoplus_{j=j_{min}^{(N)}}^{N/2} d_j\ |\psi^{(j,N)}\rangle$

Then, we make a variational anstaz on the form of the coefficients $\{p_j^{(M)}\}$ in Eq. (36), similar to the ansatz made in subsection 3.2: we assume

$$p_j^{(M)}(\lambda) = \begin{cases} c_j^{(\lceil M/\lambda \rceil)} & j \leqslant \lceil \frac{M}{2\lambda} \rceil \\ 0 & j > \lceil \frac{M}{2\lambda} \rceil, \end{cases} \tag{37}$$

for a parameter $\lambda \geq 0$ to be optimized. Denoting by $|\Phi(\lambda)\rangle$ the state of Eq. (36) with the variational choice of coefficients, one can argue that asymptotically $|\langle\eta|\psi_g\rangle^{\otimes N}|^2$ varies slowly with respect to $|\langle\Phi(\lambda)|\psi_g\rangle^{\otimes M}|^2$ provided that $M/(1+\lambda) \gg N$, following the same lines illustrated in the Appendix for the case of equatorial qubits. Hence, the fidelity can be turned into Eq. (13) with:

$$\rho_{aver}^{(M)} = \sum_{j=j_{min}^{(M)}}^{M/2} c_j^{(M)} \left[\frac{I_A^{(j,M)}}{d_j} \otimes \frac{I_B^{(j,M)}}{d_j} \otimes \frac{|I_{m_j}^{(M)}\rangle\!\rangle\langle\!\langle I_{m_j}^{(M)}|}{m_j^{(M)}} \right] \tag{38}$$

where $I_A^{(j,M)}$ ($I_B^{(j,M)}$) denotes the identity on the representation space $\mathcal{R}_A^{(j,M)}$ ($\mathcal{R}_B^{(j,M)}$), and

$$p_{true}^{(N)} = \left(\sum_{j=j_{min}^{(N)}}^{N/2} \sqrt{c_j^{(N)}} d_j \right)^2 . \tag{39}$$

With similar observation as in subsection 3.2, λ should be chosen in such a way that both conditions $\lambda \gg 1$ and $M/\lambda \gg N$ are satisfied. With this choice, the expectation value $\langle\Phi(\lambda)|\rho_{aver}^{(M)}|\Phi(\lambda)\rangle$ is

$$\langle\Phi(\lambda)|\rho_{aver}^{(M)}|\Phi(\lambda)\rangle = \sum_{j=j_{min}^{(\lceil M/\lambda \rceil)}}^{\lceil M/\lambda \rceil/2} \frac{c_j^{(\lceil M/\lambda \rceil)} c_j^{(M)}}{d_j^2} \approx \frac{2b_{M,0}}{M} \qquad \lambda \gg 1$$

Hence, the fidelity of our measure-and-prepare protocol, denoted by $F_\lambda[N \to M]$, becomes

$$\begin{aligned} F_\lambda[N \to M] &\approx \langle\Phi(\lambda)|\rho_{aver}^{(M)}|\Phi(\lambda)\rangle \; p_{true}^{(N)} \approx \frac{2b_{M,0}}{M} \left(\sum_{j=j_{min}^{(N)}}^{N/2} (2j+1)^2 \sqrt{\frac{b_{N,j}}{N/2+j+1}} \right)^2 \\ &\approx F_{eco,clon}[N \to M], \end{aligned} \tag{40}$$

the last approximate equality coming from Eq. (34). This shows that asymptotically, the fidelity of the protocols satisfying the assumptions gets arbitrarily close to the fidelity of the best economical covariant cloner.

Also in this case, we can compare the fidelity of our measure-and-prepare protocol with the naive protocol that consists in estimating the state and re-preparing M identical copies according to the estimate. In this case, we have $|\Phi\rangle = |\psi\rangle^{\otimes M} \equiv |\Phi(\lambda = 1)\rangle$ [cf. Eqs. (36) and (37)], and, therefore,

$$\langle\Phi(\lambda=1)|\rho_{aver}^{(M)}|\Phi(\lambda=1)\rangle = \sum_{j=j_{min}^{(M)}}^{M/2} \left(\frac{c_j^{(M)}}{d_j} \right)^2 . \tag{41}$$

For large M, this gives $\langle\Phi(\lambda=1)|\rho_{aver}^{(M)}|\Phi(\lambda=1)\rangle \approx \sqrt{1/(\pi M^3)} \approx b_{M,0}/(\sqrt{2}M)$, and the fidelity becomes

$$F_{\lambda=1}[N\to M] \approx \frac{b_{M,0}}{\sqrt{2}M}\left(\sum_{j=j_{min}^{(N)}}^{N/2}\sqrt{c_j^{(N)}}d_j\right)^2 \approx \frac{F_{eco,clon}[N\to M]}{2^{3/2}}.$$

This proves that re-preparing M identical copies is a strictly suboptimal strategy, which cannot reach the fidelity of the optimal economical covariant cloner.

In summary, in this section we showed that for the case of two-qubit maximally entangled states, the fidelity of the optimal economical covariant cloner $F_{eco,clon}[N\to M]$ can be achieved by a measure-and-prepare protocol $F_\lambda[N\to M]$ in the asymptotic limit $M\to\infty$. However, achieving the optimal fidelity requires one to prepare suitable M-partite entangled states: the simple strategy consisting in re-preparing M identical copies of estimated state does not give the desired fidelity, even in the asymptotic limit.

5 Discussion and conclusions

In this paper we posed the question whether the asymptotic cloning is equivalent to state estimation in terms of the global fidelity between the output state of all clones and the desired state of M perfect copies. To gain insight into the problem, we provided two examples (cloning of equatorial qubit states and cloning of two-qubit maximally entangled states) where the equivalence between cloning and estimation is satisfied in a rather non-trivial way, despite the cloning machines under consideration are economical. Our results suggest the existence of a general mechanism that guarantees the equality of fidelities in Eq. (3). Finding a general proof, or finding a counterexample to the conjectured equivalence between global asymptotic cloning and state estimation is the most pressing open question raised by our work.

Acknowledgements. This work is supported by the National Basic Research Program of China (973) 2011CBA00300 (2011CBA00301), by the National Natural Science Foundation of China through Grants 61033001, 61061130540, and 11350110207, and by the 1000 Youth Fellowship Program of China.

References

1 Dagmar Bruß, Artur Ekert, and Chiara Macchiavello. Optimal universal quantum cloning and state estimation. *Physical Review Letters*, 81(12):2598–2601, 1998.

2 Joonwoo Bae and Antonio Acín. Asymptotic quantum cloning is state estimation. *Physical Review Letters*, 97(3):030402, 2006.

3 Giulio Chiribella and Giacomo Mauro D'Ariano. Quantum information becomes classical when distributed to many users. *Physical Review Letters*, 97(25):250503, 2006.

4 Giulio Chiribella. On quantum estimation, quantum cloning and finite quantum de finetti theorems. *Proc. TQC 2010, Lecture Notes in Computer Science*, 6519/2011:9–25, 2011.

5 M. Keyl. *Asymptotic cloning is state estimation.* `http://qig.itp.uni-hannover.de/qiproblems/22`.

6 Dagmar Bruß, David P. DiVincenzo, Artur Ekert, Christopher A. Fuchs, Chiara Macchiavello, and John A. Smolin. Optimal universal and state-dependent quantum cloning. *Physical Review A*, 57:2368–2378, Apr 1998.

7 R. F. Werner. Optimal cloning of pure states. *Physical Review A*, 58:1827–1832, 1998.

8 N. J. Cerf, A. Ipe, and X. Rottenberg. Cloning of continuous quantum variables. *Physical Review Letters*, 85:1754–1757, Aug 2000.

9 Göran Lindblad. Cloning the quantum oscillator. *Journal of Physics A: Mathematical and General*, 33(28):5059, 2000.

10 N. J. Cerf and S. Iblisdir. Optimal n-to-m cloning of conjugate quantum variables. *Physical Review A*, 62(4):040301, Sep 2000.

11 P. T. Cochrane, T. C. Ralph, and A. Dolińska. Optimal cloning for finite distributions of coherent states. *Physical Review A*, 69:042313, Apr 2004.

12 Giulio Chiribella and Jinyu Xie. Optimal design and quantum benchmarks for coherent state amplifiers. *Physical Review Letters*, 110:213602, May 2013.

13 Giacomo Mauro D'Ariano and Chiara Macchiavello. Optimal phase-covariant cloning for qubits and qutrits. *Physical Review A*, 67(4):042306, 2003.

14 Chi-Sheng Niu and Robert B Griffiths. Two-qubit copying machine for economical quantum eavesdropping. *Physical Review A*, 60(4):2764, 1999.

15 Francesco Buscemi, Giacomo Mauro D'Ariano, and Chiara Macchiavello. Economical phase-covariant cloning of qudits. *Physical Review A*, 71(4):042327, 2005.

16 Thomas Durt, Jaromír Fiurášek, and Nicolas J Cerf. Economical quantum cloning in any dimension. *Physical Review A*, 72(5):052322, 2005.

17 Vladimir Bužek and Mark Hillery. Quantum copying: Beyond the no-cloning theorem. *Physical Review A*, 54(3):1844, 1996.

18 Vladimìr Bužek, Samuel L Braunstein, Mark Hillery, and Dagmar Bruß. Quantum copying: A network. *Physical Review A*, 56(5):3446, 1997.

19 Nicolas Gisin and Serge Massar. Optimal quantum cloning machines. *Physical Review Letters*, 79(11):2153–2156, 1997.

20 Michael Keyl and Reinhard F. Werner. Optimal cloning of pure states, testing single clones. *Journal of Mathematical Physics*, 40:3283, 1999.

21 Dagmar Bruß, Mirko Cinchetti, Giacomo Mauro D'Ariano, and Chiara Macchiavello. Phase-covariant quantum cloning. *Physical Review A*, 62:012302, Jun 2000.

22 Heng Fan, Keiji Matsumoto, Xiang-Bin Wang, and Miki Wadati. Quantum cloning machines for equatorial qubits. *Physical Review A*, 65:012304, Dec 2001.

23 Heng Fan, Keiji Matsumoto, Xiang-Bin Wang, and Hiroshi Imai. Phase-covariant quantum cloning. *Journal of Physics A: Mathematical and General*, 35(34):7415, 2002.

24 K Hammerer, MM Wolf, Eugene Simon Polzik, and JI Cirac. Quantum benchmark for storage and transmission of coherent states. *Physical Review Letters*, 94(15):150503, 2005.

25 Gerardo Adesso and Giulio Chiribella. Quantum benchmark for teleportation and storage of squeezed states. *Physical Review Letters*, 100(17):170503, 2008.

26 Masaki Owari, Martin B Plenio, Eugene Simon Polzik, Alessio Serafini, and Michael Marc Wolf. Squeezing the limit: quantum benchmarks for the teleportation and storage of squeezed states. *New Journal of Physics*, 10(11):113014, 2008.

27 J. Calsamiglia, M. Aspachs, R. Muñoz Tapia, and E. Bagan. Phase-covariant quantum benchmarks. *Physical Review A*, 79:050301, May 2009.

28 Giulio Chiribella, Yuxiang Yang, and Andrew Chi-Chih Yao. Reliable quantum replication at the heisenberg limit. *arXiv:1304.2910*, 2013.

29 Alexander Holevo. *Probabilistic and statistical aspects of quantum theory*, volume 1. Birkhauser, 2011.

30 G. M. D'Ariano, P. Lo Presti, and M. F. Sacchi. Bell measurements and observables. *Physics Letters A*, 272(1):32–38, 2000.

31 G. Chiribella, G. M. D'Ariano, and M. F. Sacchi. Optimal estimation of group transformations using entanglement. *Physical Review A*, 72:042338, Oct 2005.

32 Paul Hausladen and William K Wootters. A 'pretty good' measurement for distinguishing quantum states. *Journal of Modern Optics*, 41(12):2385–2390, 1994.

A Justification of the asymptotic approximation for the fidelity

In order to study the asymptotic behavior of the fidelity for measure-and-prepare protocols, we can use a Taylor expansion of $\mathrm{Tr}\left[\eta\ \psi_g^{\otimes N}\right]$ up to the second order term:

$$\begin{aligned}\mathrm{Tr}\left[\eta\ \psi_g^{\otimes N}\right] &= \sum_{n,m=-N/2}^{N/2} \sqrt{b_{N,n} b_{M,m}}\, e^{i(n-m)\theta} \\ &\approx \left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}}\right)^2 - \frac{\theta^2}{2}\left[\sum_{n,m=-N/2}^{N/2} \sqrt{b_{N,n} b_{N,m}}\,(n-m)^2\right] \\ &= \left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}}\right)^2 - \theta^2 \left(\sum_{n=-N/2}^{N/2} n^2 \sqrt{b_{N,n}}\right)\left(\sum_{m=-N/2}^{N/2} \sqrt{b_{N,m}}\right). \end{aligned} \tag{42}$$

In the asymptotic limit of large amplification, i.e. $M \gg N$, $b_{M,m}$ can be approximated by the Gaussian $g_M(x) := \sqrt{\frac{2}{\pi M}} e^{-\frac{2x^2}{M}}$, thus we have:

$$\begin{aligned}\mathrm{Tr}\left[\Phi\ \psi_g^{\otimes M}\right] &= \left| \sum_{m=-\lceil M/2\lambda\rceil}^{\lceil M/2\lambda\rceil} \sqrt{b_{\lceil M/\lambda\rceil, m} b_{M,m}}\, e^{im\theta} \right|^2 \approx \left| \int_{-\lceil M/2\lambda\rceil}^{\lceil M/2\lambda\rceil} \mathrm{d}\, x\ \sqrt{\frac{2}{\pi M}}\ \lambda^{\frac{1}{4}}\ e^{-\frac{(1+\lambda)x^2}{M} + i\theta x} \right|^2 \\ &= \frac{2\sqrt{\lambda}}{1+\lambda}\ e^{-\frac{M\theta^2}{2(1+\lambda)}}. \end{aligned}$$

Taking these into Eq. (12) we get the expression of measure-and-prepare fidelity for large amplification:

$$\begin{aligned} F[N \to M] &= \int \frac{\mathrm{d}\,\theta}{2\pi}\ \mathrm{Tr}\left[\eta\ \psi_\theta^{\otimes N}\right]\ \mathrm{Tr}\left[\Phi\ \psi_\theta^{\otimes M}\right] \\ &\approx \frac{2\sqrt{\lambda}}{1+\lambda} \left(\left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}}\right)\left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}} \int_{-\pi}^{\pi} \frac{\mathrm{d}\,\theta}{2\pi}\ e^{-\frac{M\theta^2}{2(1+\lambda)}}\right) \right. \\ &\quad \left. - \sum_{n=-N/2}^{N/2} n^2 \sqrt{b_{N,n}} \int_{-\pi}^{\pi} \frac{\mathrm{d}\,\theta}{2\pi}\ \theta^2\ e^{-\frac{M\theta^2}{2(1+\lambda)}} \right) \\ &= \sqrt{\frac{2\lambda}{\pi M(1+\lambda)}} \left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}}\right) \\ &\qquad \left(\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}} - \frac{1+\lambda}{2\pi M} \sum_{n=-N/2}^{N/2} n^2 \sqrt{b_{N,n}}\right). \end{aligned} \tag{43}$$

It is clear that the contribution resulting from the second order term of the expansion is negligible whenever $M/(1+\lambda) \gg N$. In particular, for large N the binomial can be approximated by a Gaussian, giving

$$\sum_{n=-N/2}^{N/2} \sqrt{b_{N,n}} \approx (2\pi N)^{1/4} \qquad \sum_{n=-N/2}^{N/2} n^2\ \sqrt{b_{N,n}} \approx (2\pi N)^{1/4} \frac{N}{2}. \tag{44}$$

so that the ratio between the second and first order term in Eq. (43) is $N(1+\lambda)/M$.

Distillation of Non-Stabilizer States for Universal Quantum Computation

Guillaume Duclos-Cianci[1] and Krysta M. Svore[2]

1 **Départment de Physique, Université de Sherbrooke**
Sherbrooke, Québec, J1K 2R1 (Canada)
Guillaume.Duclos-Cianci@USherbrooke.ca

2 **Quantum Architectures and Computation Group, Microsoft Research**
Redmond, WA 98052 (USA)
ksvore@microsoft.com

Abstract

Magic state distillation is a fundamental technique for realizing fault-tolerant universal quantum computing, and produces high-fidelity Clifford eigenstates, called magic states, which can be used to implement the non-Clifford $\pi/8$ gate. We propose an efficient protocol for distilling other *non-stabilizer* states that requires only Clifford operations, measurement, and magic states. One critical application of our protocol is efficiently and fault tolerantly implementing arbitrary, non-Clifford, single-qubit rotations in average *constant* online circuit depth and polylogarithmic (in precision) offline resource cost, resulting in significant improvements over state-of-the-art decomposition techniques. Finally, we show that our protocol is robust to noise in the resource states.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases quantum computing, resource estimation, magic state distillation

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.235

1 Introduction

Given recent progress in quantum algorithms, quantum error correction, and quantum hardware, a *scalable* quantum computer is becoming closer and closer to reality. For many proposed quantum computer architectures, e.g., the surface-code model based on code deformation [1], Clifford operations, stabilizer-state preparations, and measurements can be implemented efficiently. However, these operations alone are not sufficient for quantum universality and can be simulated classically [2]. Magic state distillation [3, 4, 5, 6] produces Clifford eigenstates, which in turn can be used to realize a non-Clifford operation, e.g., the single-qubit $\pi/8$ gate, T.

In this paper, we present an efficient protocol for distilling other *non-stabilizer* states. Our protocol uses only $|H\rangle$-type magic resource states, Clifford operations, and measurements, and is robust to noise in the resource states. One notable application of our protocol is producing an arbitrary single-qubit, fault-tolerant unitary operation. Most commonly, a single-qubit unitary U is decomposed into a discrete set of gates, typically $\{H, T\}$, using Solovay-Kitaev decomposition [7, 8], which efficiently produces an approximate fault-tolerant implementation of U with circuit depth $\Theta(\log^c(1/\epsilon))$, where ϵ is the precision and c is around 3.97 [9, 8]. Remarkably, efficient decomposition algorithms have recently been proposed which lower c to 1 [10, 11]. Each T gate in the decomposed sequence requires a number of copies of a quantum magic state $|H\rangle$, dependent on the specific state distillation protocol and

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 235–243


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

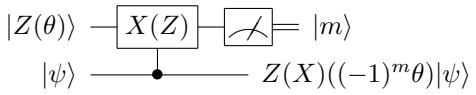


Figure 1 Circuit to rotate by angle $\pm\theta$ around the $Z(X)$-axis.

purity of the state [3, 4, 5, 6]. We show that our protocol requires only *constant* online circuit depth and fewer resources than state-of-the-art decomposition techniques. Our protocols may be useful for other applications as well.

2 Distilling Magic States and Implementing Rotations

We first review how to perform an arbitrary rotation about the Z-axis using a resource state. A state $|\psi\rangle$ is *magic* if we can "distill" a purer $|\psi\rangle$ state from a Clifford circuit applied to n noisy copies of $|\psi\rangle$. We focus on the $+1$ eigenstate of the Hadamard operation H, $|H\rangle = \cos\frac{\pi}{8}|0\rangle + \sin\frac{\pi}{8}|1\rangle$. We assume throughout that Clifford operations are perfect and resource states are arbitrarily pure. We can arbitrarily purify these states by applying a distillation protocol recursively [3, 4, 5, 6]. We concentrate on single-qubit states found in either the XZ- or XY-plane of the Bloch sphere; note that a state can be rotated from one plane to the other through application of the Clifford $HSHX$ operation.

Suppose we have states $|Z(\theta)\rangle = |0\rangle + \mathrm{e}^{i\theta}|1\rangle$ and $|\psi\rangle = a|0\rangle + b|1\rangle$. The circuit to implement a rotation around the Z-axis using $|Z(\theta)\rangle$ as a resource state is presented in Fig. 1. Upon measurement of the first qubit in the computational basis, we obtain either

$$\xrightarrow{m=0} \quad a|0\rangle + b\mathrm{e}^{i\theta}|1\rangle\,, \text{ or}$$
$$\xrightarrow{m=1} \quad a\mathrm{e}^{i\theta}|0\rangle + b|1\rangle = a|0\rangle + b\mathrm{e}^{-i\theta}|1\rangle\,,$$

each with probability $1/2$. Thus, the rotation angle is randomly either θ or $-\theta$, up to global phase. An analogous circuit performs a rotation about the X-axis [1].

As an example, consider the XY-plane version of $|H\rangle$:

$$|Z(\pi/4)\rangle = HSHX|H\rangle = |0\rangle + \mathrm{e}^{i\pi/4}|1\rangle\,.$$

Using the circuit in Fig. 1, we can implement a Z-rotation of angle $\pm\pi/4$, producing at random either the T gate or its adjoint, $T^\dagger$. We can deterministically *correct* the angle by applying the phase gate S: $ST^\dagger|\psi\rangle = T|\psi\rangle$. For general rotations, deterministic correction is not possible.

3 Distilling Other Non-Stabilizer States

We now present our protocol for producing other non-stabilizer states using a very simple two-qubit Clifford circuit and $|H\rangle$ states as an initial resource.

Consider the circuit of Fig. 2. One can easily verify that it measures the parity of the two input qubits and decodes the resulting state into the second qubit. Consider the two inputs to be $|H\rangle$ states and define $\theta_0 = \frac{\pi}{8}$ and $|H\rangle = |H_0\rangle = \cos\theta_0|0\rangle + \sin\theta_0|1\rangle$. Then upon application of the controlled-`NOT` gate $\Lambda(X)$,

$$|H_0\rangle|H_0\rangle \quad \xrightarrow{\Lambda(X)} \quad \cos^2\theta_0|00\rangle + \sin^2\theta_0|01\rangle + \cos\theta_0\sin\theta_0(|11\rangle + |10\rangle).$$

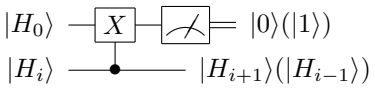


Figure 2 Two-qubit circuit used to obtain new $|H_i\rangle$ states from initial resource states $|H_0\rangle$. Upon measuring the 0 (1) outcome, the output state is $|H_{i+1}\rangle$ ($|H_{i-1}\rangle$).

Upon measurement m of the first qubit, we have

$$\xrightarrow{m=0} \frac{\cos^2\theta_0\,|0\rangle + \sin^2\theta_0\,|1\rangle}{\cos^4\theta_0 + \sin^4\theta_0}, \text{ or } \xrightarrow{m=1} \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle).$$

We define θ_1 such that

$$\cos\theta_1\,|0\rangle + \sin\theta_1\,|1\rangle \quad = \quad \frac{\cos\theta_0\,|0\rangle + \sin\theta_0\,|1\rangle}{\cos^4\theta_0 + \sin^4\theta_0},$$

from which we deduce $\cot\theta_1 = \cot^2\theta_0$. We define $|H_1\rangle = \cos\theta_1\,|0\rangle + \sin\theta_1\,|1\rangle$, a non-stabilizer state obtained from $|H\rangle$ states, Clifford operations, and measurements. If the measurement outcome is 1, then we obtain a stabilizer state and discard the output (see Fig. 2). The measurement outcomes occur with respective probabilities $p_0 = \cos^4\theta_0 + \sin^4\theta_0 = \frac{3}{4}$ and $p_1 = 1 - p_0 = \frac{1}{4}$.

We now recurse on this protocol using the non-stabilizer states produced by the previous round of the protocol as input to the circuit in Fig. 2. We define $|H_i\rangle = \cos\theta_i\,|0\rangle + \sin\theta_i\,|1\rangle$, where $\cot\theta_i = \cot^{i+1}\theta_0$. Using as input the previously produced $|H_i\rangle$ state and a new $|H_0\rangle$ state, we have

$$|H_0\rangle\,|H_i\rangle \xrightarrow{\Lambda(X)} \cos\theta_0\cos\theta_i\,|00\rangle + \sin\theta_0\sin\theta_i\,|01\rangle + \sin\theta_0\cos\theta_i\,|10\rangle + \cos\theta_0\sin\theta_i\,|11\rangle\,.$$

Upon measurement of the first qubit, we have

$$\begin{aligned}
&\xrightarrow{m=0} && (\cos\theta'\,|0\rangle + \sin\theta'\,|1\rangle),\\
&\xrightarrow{m=1} && (\cos\theta''\,|0\rangle + \sin\theta''\,|1\rangle), \text{ where}\\
\cot\theta' \quad &= && \cot\theta_i\cot\theta_0 = \cot^{i+2}\theta_0 = \cot\theta_{i+1},\\
\cot\theta'' \quad &= && \cot\theta_i\tan\theta_0 = \cot^{i}\theta_0 = \cot\theta_{i-1}.
\end{aligned}$$

Thus, if we measure $m = 0$, we obtain the state $|H_{i+1}\rangle$ and if we measure $m = 1$, we obtain $|H_{i-1}\rangle$. The probability of measuring 0 is given by $p_{0,i} = \cos^2\theta_i\cos^2\theta_0 + \sin^2\theta_i\sin^2\theta_0$. Note that $\frac{3}{4} \le p_{0,i} < \cos^2\frac{\pi}{8} = 0.853\ldots$

We can view this recursive process as a semi-infinite random walk with biased non-homogeneous probabilities, as Fig. 3 illustrates. Every time a step is taken along this "ladder" of states, one $|H\rangle \equiv |H_0\rangle$ is consumed, except at the first step of the ladder when we require two $|H\rangle$ states; if $m = 1$ at the first node, we discard the output and start with two new $|H\rangle$ states.

4 A Denser Ladder

We can produce a denser ladder of states by using additional resource states $\left|\psi_0^{0,1,2}\right\rangle$. Consider the Clifford circuit of Fig. 4(a) that takes as input four $|H\rangle$ states. With probability $3(2+\sqrt{2})/32 \approx 0.320$, the measurement outcome is 000 and the resulting state is $\left|\psi_0^0\right\rangle =$

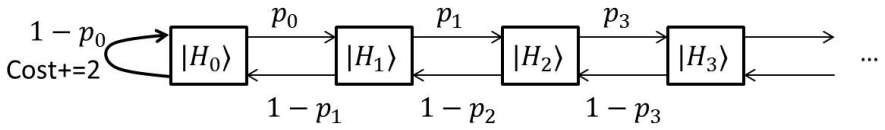


Figure 3 Obtaining non-stabilizer states from initial $|H\rangle$ states. Using $|H_i\rangle$ and $|H_0\rangle$ states probabilistically yields a $|H_{i-1}\rangle$ or $|H_{i+1}\rangle$ using the circuit of Fig. 2. Each ladder step costs one $|H_0\rangle$ state, except the first one which costs two.

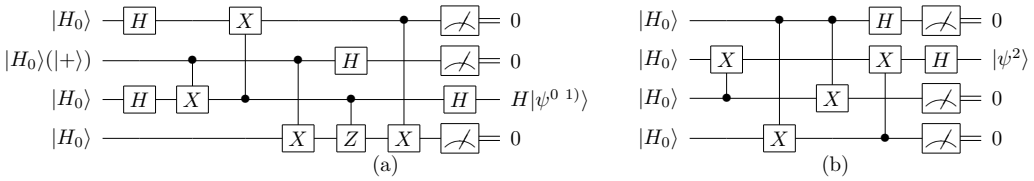


Figure 4 (a) Circuit to produce $\left|\psi_0^0\right\rangle$ ($\left|\psi_0^1\right\rangle$) states. (b) Circuit to produce $\left|\psi_0^2\right\rangle$ states.

$\cos\phi_0^0\,|0\rangle+\sin\phi_0^0\,|1\rangle$ with $\phi_0^0=\frac{\pi}{2}-\cot^{-1}\left(\frac{2+3\sqrt{2}}{6+5\sqrt{2}}\right)\approx 0.446$. Otherwise the output is discarded. Since the probability of success is 0.320 and every trial consumes four copies of $|H_0\rangle$, the average cost to produce $\left|\psi_0^0\right\rangle$ is 12.50 $|H_0\rangle$ states.

Another interesting state is obtained using the same circuit with one input state replaced with a $|+\rangle$ state. Measurement 000 is obtained with probability $(6+\sqrt{2})/32\approx 0.232$, resulting in the state $\left|\psi_0^1\right\rangle=\cos\phi_0^1\,|0\rangle+\sin\phi_0^1\,|1\rangle$ with $\phi_0^1=\frac{\pi}{2}-\cot^{-1}\left(\frac{2\sqrt{2}}{3+\sqrt{2}}\right)\approx 0.570$. Since the probability of success is 0.232 and every trial consumes three $|H_0\rangle$ states, the average cost to produce $\left|\psi_0^1\right\rangle$ is 12.95 $|H_0\rangle$ states. Fig. 4(b) shows a circuit which produces the output state $\left|\psi_0^2\right\rangle=\cos\phi_0^2\,|0\rangle+\sin\phi_0^2\,|1\rangle$ with $\phi_0^2=\frac{\pi}{2}-\cot^{-1}\left(\frac{7}{6\sqrt{2}}\right)\approx 0.690$, when measurement 000 is obtained (with probability $11/32\approx 0.344$). The probability of success is 0.344 and the average cost to produce $\left|\psi_0^2\right\rangle$ is 11.64 $|H_0\rangle$ states.

Now we can use one of these non-stabilizer states as input to the circuit in Fig. 2 in place of the top $|H_0\rangle$ state. Begin with states $\left|\psi_0^i\right\rangle$ and $|H_0\rangle$. If $m=1$, the state is discarded. Otherwise, we obtain $\left|\psi_1^i\right\rangle=\cos\phi_1^i\,|0\rangle+\sin\phi_1^i\,|1\rangle$, where $\cot\phi_1^i=\cot\phi_0^i\cot\theta_0$. As before, we define $\left|\psi_i^j\right\rangle=\cos\phi_i^j\,|0\rangle+\sin\phi_i^j\,|1\rangle$, where $\cot\phi_i^j=\cot\phi_0^j\cot^i\theta_0$. If we input states $\left|\psi_i^j\right\rangle$ and $|H_0\rangle$, we obtain

$$|H_0\rangle\left|\psi_i^j\right\rangle\xrightarrow{\Lambda(X)}\cos\theta_0\cos\phi_i^j\,|00\rangle+\sin\theta_0\sin\phi_i^j\,|01\rangle+\sin\theta_0\cos\phi_i^j\,|10\rangle+\cos\theta_0\sin\phi_i^j\,|11\rangle\,,$$

such that the output state is, depending on the measurement outcome,

$$\xrightarrow{m=0}\left|\psi_{i+1}^j\right\rangle,\qquad\text{or}\quad\xrightarrow{m=1}\left|\psi_{i-1}^j\right\rangle.$$

Denser "ladders" of states can be obtained using $\left|\psi_0^{0,1,2}\right\rangle$ as inputs in place of the top $|H_0\rangle$ state.

5 Noisy states

A priori, noise in the $|H_0\rangle$ resource states could be amplified by the circuit in Fig. 2 and affect the purity of the $|H_i\rangle$ states. However, we show this is not the case. We measure the accuracy of the imperfect $|H_i\rangle$ states using the trace distance on states ρ and

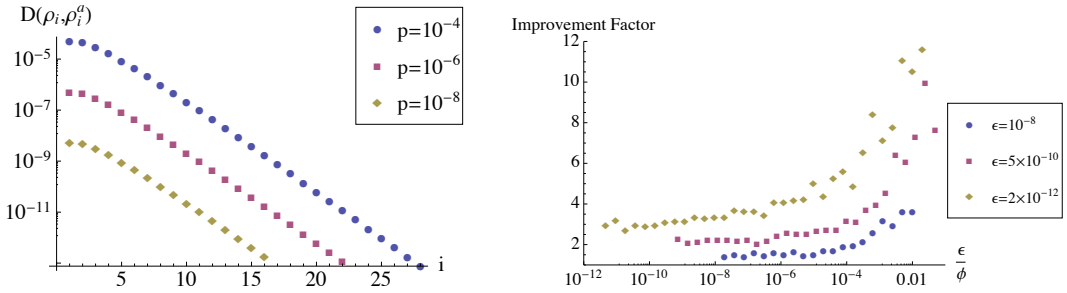


Figure 5 (a) Evolution of the trace distance between imperfect ρ_i^a and perfect $|H_i\rangle$ states with noise p. Exponential decay fits give $(2.08*10^{-3})\times 2.31^{-i}$, $(1.63*10^{-5})\times 2.28^{-i}$ and $(1.26*10^{-7})\times 2.24^{-i}$ for the circle, square and diamond data set, respectively. (b) Improvement factor of the total offline cost using the noisiest $|H_0\rangle$ states to distill $|H_i\rangle$ states of precision ϵ as a function of the relative precision of the rotation ϵ/ϕ.

σ: $D(\rho,\sigma)=\mathrm{tr}(\sqrt{(\rho-\sigma)^\dagger(\rho-\sigma)})/2$. We assume errors only occur on the $|H_0\rangle$ states. We numerically study three types of errors. For the first error, we assume that the mixed state, ρ_0^a, is on the line joining the center of the Bloch sphere and the the perfect state, i.e.,

$$\rho_0^a(p) \quad = \quad (1-p)|H_0\rangle\langle H_0| + p|-H_0\rangle\langle -H_0|,$$

where $|-H_0\rangle = \sin\frac{\pi}{8}|0\rangle - \cos\frac{\pi}{8}|1\rangle$ is the state orthogonal to $|H_0\rangle$. We denote the imperfect version of $|H_i\rangle$ obtained from ρ_0^a states as ρ_i^a. We can always bring any mixed state into this form using twirling [4]. For the protocol to be practical, we require it to remain stable under the two following types of errors, where we assume the state is pure and the rotation is off of the desired axis by δ:

$$\begin{aligned}\rho_0^b(\delta) &= \frac{1}{2}\left(I+\sin\left(\frac{\pi}{4}+\delta\right)X+\cos\left(\frac{\pi}{4}+\delta\right)Z\right),\\ \rho_0^c(\delta) &= \frac{1}{2}\left(I+\sin\frac{\pi}{4}\cos\delta X+\sin\frac{\pi}{4}\sin\delta Y+\cos\frac{\pi}{4}Z\right).\end{aligned}$$

We numerically generated 1000 pseudo-random instances of the protocol to produce $|H_i\rangle$ states for different values of i for each error type and for noise strengths $10^{-4}, 10^{-6}$, and 10^{-8}. Figure 5(a) shows an exponential decay of the distance between erroneous and ideal states; if we start with a $|H_0\rangle$ state distilled to our target accuracy, all subsequent derived $|H_i\rangle$ states will also be distilled to at least that accuracy. This further suggests that for larger values of i, noisier $|H_0\rangle$ states could be used to still achieve the desired accuracy, and in turn decrease the number of distillation recursions (and resources) necessary to prepare the $|H_0\rangle$ states.

Extrapolating from Fig.5(a), one could for example prepare ρ_{12} states with accuracy 10^{-9} using only input $|H_0\rangle$ states of accuracy 10^{-6}, saving at least one round of distillation prior to our protocol, reducing the total offline cost (including magic state distillation). Using states as noisy as possible and using the costs and accuracies presented in Table I of [4], we were able to estimate, via numerical simulations, the improvement factor to be gained in offline cost for different rotations and precisions. The results are presented in Fig. 5(b). Two important behaviors are noted. First, for any given realtive precision ϵ/ϕ, the improvement factor increases as the absolute precision ϵ goes down. Second, and more importantly, there is as much as an order of magnitude to be gained for rotation angles that are comparable to

the desired accuracy ϵ, e.g., for $\epsilon = 5 \times 10^{-10}$ and $\phi \sim 100\epsilon$, there is a factor ~ 11 reduction in resource offline cost.

6 Application to Single-qubit Rotations

We now show how to use the ladders of states to enable the fault-tolerant approximation of any single-qubit rotation. Results do not include the improvements in offline cost discussed in previous section, so an additional gain factor between 2 and 10, depending on ϵ, is expected. Recall the circuit given in Fig. 1. If we input either $HSHX\,|H_i\rangle$ or $HSHX\left|\psi_i^j\right\rangle$ in place of the top qubit, we obtain rotation $Z(\pm 2\theta_i)$ on $|\psi\rangle$. Note that there is a factor of two difference between the angle θ_i involved in the description of the state and the rotation applied, e.g., the $|H_0\rangle$ state is over $\theta_0 = \frac{\pi}{8}$, and can be used to implement a $\frac{\pi}{4}$ rotation. Also, since $0 < \theta_i < \frac{\pi}{4}$ $(\forall i)$, the discontinuity of cotangent is not a problem.

Although the circuit in Fig. 1 randomly applies $\pm\theta$, our protocols still result in efficient application of the desired Z-rotation.

We propose the following protocol to approximate a Z-rotation $Z(\phi)$:

1. Set desired accuracy ϵ.
2. Pick a target rotation angle $0 < \phi < 2\pi$.
3. Find the state $|H_i\rangle$ (or denser state $\left|\psi_i^j\right\rangle$) such that $2\theta_i$ is close to ϕ.
4. Simulate an instance of the ladder to obtain that state and add its cost to the offline cost.
5. Apply a rotation using $|H_i\rangle$ (or denser state $\left|\psi_i^j\right\rangle$) as input to the circuit of Fig. 1 and add one to the online cost.
6. Recurse on steps 3 through 5 until the desired accuracy is reached.

Thus, one has to implement a sequence of j rotations $\{Z(2\theta_{i_j})\}$ on $|\psi\rangle$ using the sequence of states $\{\left|H_{i_j}\right\rangle\}$, such that $Z(\phi) \approx \prod_j Z(2\theta_{i_j})$. The online cost is also given by $\left|\{\left|H_{i_j}\right\rangle\}\right|$.

We define the accuracy of the applied rotation V compared to the target rotation $U = Z(\phi)$ as

$$\max_{|\psi\rangle} D(U\,|\psi\rangle\,\langle\psi|\,U^\dagger, V\,|\psi\rangle\,\langle\psi|\,V^\dagger),$$

where $D(\rho,\sigma)$ is the trace distance between states ρ and σ. If U and V are rotations about the same axis, one can show that in our case, for small angles of rotation, this reduces to the difference of rotation angles: $\epsilon = \Delta\phi$. In [9], the distance measure used is $D(U,V) = \sqrt{(2 - |\mathrm{tr}(UV^\dagger)|)/2}$. In the case of rotations about the same axis, it can be reduced to $\sqrt{1 - |\cos(\Delta\phi)|} \approx \Delta\phi/\sqrt{2}$ for small $\Delta\phi$.

We define an online and offline cost to apply a unitary gate. The *online cost*, C_{on}, is the expected number of $|H_i\rangle$ states required to implement the unitary. The *offline cost*, C_{off}, is the total number of distilled $|H_0\rangle$ states required to obtain all of the intermediate $|H_i\rangle$ states used to perform the given unitary. For Solovay-Kitaev decomposition, the offline cost equals the online cost and is given by the total number of T and $T^\dagger$ gates in the decomposition. In both cases, we do not count the cost of initially distilling $|H_0\rangle$ states.

We simulated $\sim 1.8 \times 10^4$ instances of our $|H\rangle$ protocol, each for a random angle ϕ and target accuracy between $10^{-12} < \epsilon < 10^{-4}$. We assume that $C_{\mathrm{on}} \sim \ln_{\mathrm{on}}^{c}(\frac{1}{\epsilon})$, and $C_{\mathrm{off}} \sim \ln^{c_{\mathrm{off}}}(\frac{1}{\epsilon})$, where C_{on} and C_{off} are the online and offline costs, respectively, such that $\ln C_{\mathrm{on}} \sim c_{\mathrm{on}} \ln\ln(\frac{1}{\epsilon})$, and $\ln C_{\mathrm{off}} \sim c_{\mathrm{off}} \ln\ln(\frac{1}{\epsilon})$. From linear fits to the data, we find $\ln(C_{\mathrm{on}}) = -0.21 + 1.23\ln(\ln(1/\epsilon))$ with a standard deviation around the mean of $\ln(\Delta C_{\mathrm{on}}) = -0.30 + 0.83\ln(\ln(1/\epsilon))$, and $\ln(C_{\mathrm{on}}) = -0.44 + 2.22\ln(\ln(1/\epsilon))$ with a standard

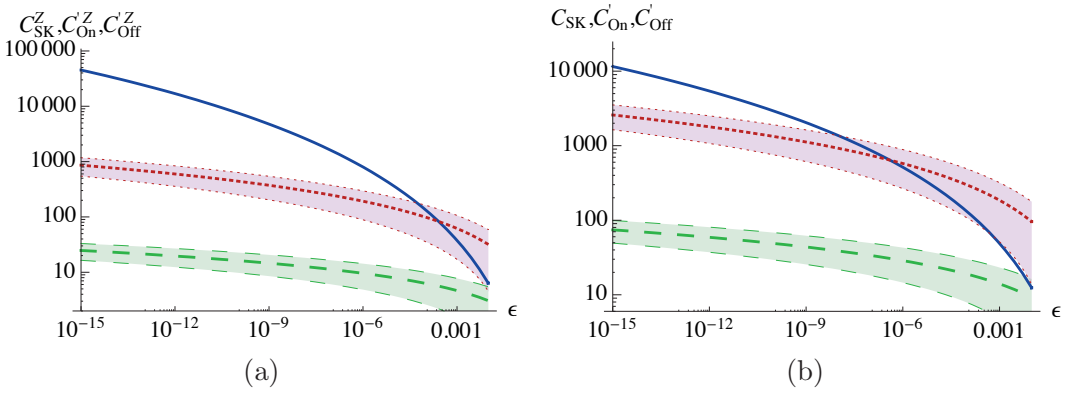


Figure 6 Cost of (a) random Z-rotations and (b) random unitaries as a function of precision ϵ. Solid line: SK decomposition [9]. Dotted line: Offline cost using $|H\rangle$, or $\{|\psi^0\rangle, |\psi^1\rangle, |\psi^2\rangle\}$ as initial resources. Dashed line: Online cost using $|H\rangle$, or $|\psi^0\rangle, |\psi^1\rangle, |\psi^2\rangle\}$ as initial resources. The shaded regions around the dashed and dotted lines represent the standard deviation around the mean. (a) $\ln(C'^Z_{SK}) = -4.88 + 4.41\ln(\ln(1/\epsilon))$; $\ln(C'^Z_{On}) = -0.46 + 1.04\ln(\ln(1/\epsilon))$; $\ln(C'^Z_{Off}) = 0.96 + 1.64\ln(\ln(1/\epsilon))$. (b) $\ln(C'_{SK}) = -2.67 + 3.40\ln(\ln(1/\epsilon))$; $\ln(C'_{On}) = -0.46 + 1.04\ln(\ln(1/\epsilon)) + \ln 3$; $\ln(C'_{Off}) = 0.96 + 1.64\ln(\ln(1/\epsilon)) + \ln 3$.

deviation around the mean of $\ln(\Delta C_{on}) = 0.02 + 1.87\ln(\ln(1/\epsilon))$. We deduce that $c_{on} \sim 1.23$ and $c_{off} \sim 2.22$ for our protocol.

For the denser protocol, the offline costs are 12.50, 12.95, and 11.64 for $|\psi_0^0\rangle$, $|\psi_0^1\rangle$, and $|\psi_0^2\rangle$, respectively. The denser set of states results in improved scalings for both the online and offline costs: $c'_{on} \sim 1.04$ and $c'_{off} \sim 1.64$, where $'$ denotes the denser protocol. However, the offline costs of our new states $|\psi_0^i\rangle$ are improved only when precisions are smaller then $\epsilon \approx 1.28 \times 10^{-5}$.

Figure. 6 shows the behavior of the protocols on Z rotations and arbitrary rotations. For an arbitrary rotation, recall that a single-qubit unitary U is composed of three rotations around the X- and Z-axes [12]: $U \propto X(\alpha)Z(\beta)X(\gamma)$, for some angles α, β, γ. We can use our protocol to implement both Z and X rotations as previously outlined. Fig. 6(a) plots the fit for Solovay-Kitaev decomposition [9] (solid line), the online cost (dashed), and offline cost (dotted). For all practical precisions, the online cost of our proposed scheme is consistently smallest. The offline cost is advantageous when $\epsilon \leq 4.41 \times 10^{-4}$ for Z-rotations and $\epsilon < 1.03 \times 10^{-6}$ for random unitaries.

7 Minimizing Online Cost

We can further minimize the online cost by considering instead the following protocol to implement a Z rotation by angle ϕ: Prepare *offline* the state $|Z(\phi)\rangle$ using the protocol described to apply $|Z(\phi)\rangle$ to a $|0\rangle$ ancilla. Then, use $|Z(\phi)\rangle$ *online* to apply the rotation to the desired qubit. With probability $\frac{1}{2}$, the rotation $Z(\phi)$ is applied and the online cost is 1. If it fails, prepare *offline* $|Z(2\phi)\rangle$; with probability $\frac{1}{2}$, $Z(\phi)$ is applied online and the online cost is 2. If it fails, prepare *offline* $|Z(4\phi)\rangle$, and so on. The probability of success after n iterations decreases exponentially with n; the process is a negative binomial of parameter $p = \frac{1}{2}$ and the expected number of online rotations for success is $\sim \frac{1}{p} = 2$. We simulated this process for random angles $0 < \phi < 2\pi$ and accuracies $10^{-12} < \epsilon < 10^{-4}$ and found the expected number of online rotations is $\langle C''_{on}\rangle = 1.99$ and the offline cost is $c''_{off} \sim 1.75$. Note

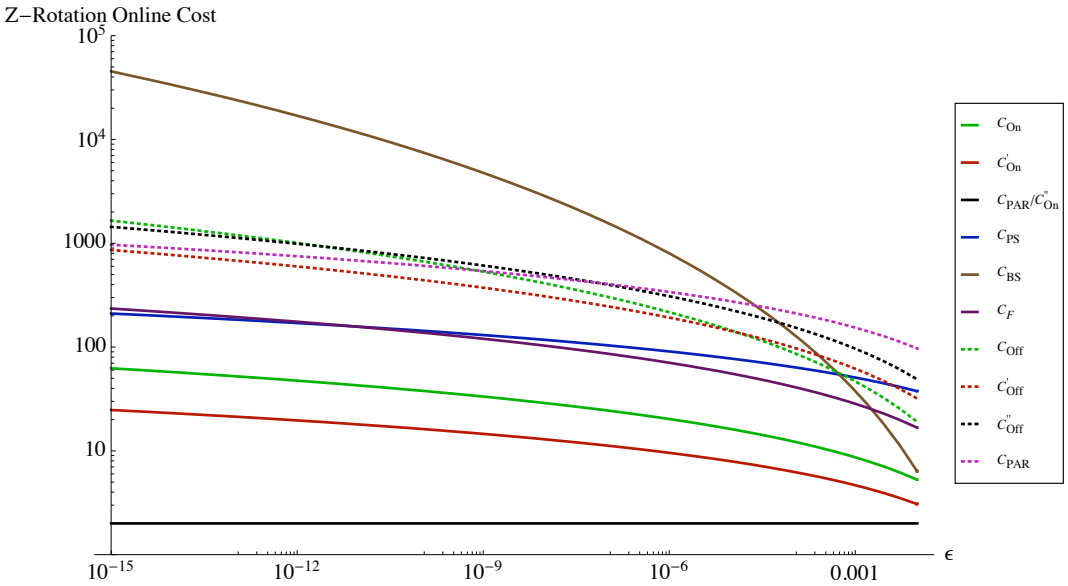


Figure 7 Comparison of online (solid) and offline (dashed) costs to decompose Z rotations vs. accuracy ϵ. Methods plotted include C_{PAR} [13], C_{PS} [10], C_{BS} [9], C_{F} [14]. C_{On}, C_{Off}, C'_{On}, C'_{Off}, C''_{On}, C''_{Off} represent our $|H\rangle$ ladder, dense ladder, and minimal online cost with the dense ladder, respectively. The offline costs for $C_{\{\mathrm{BS,PS,F}\}}$ are equal to their online costs.

that any method can be used to prepare the ancilla state offline, and here we use our protocol for preparation. We discovered after writing that a similar technique was described in [13].

Figure 7 compares the cost of state-of-the-art decomposition techniques with our protocols. The plot highlights the tradeoffs between the various methods. Note that we only plotted two methods, our protocol C'' and C_{PAR} (which uses C_{F} to prepare the state), using the minimal online framework, but the other techniques could also be used to prepare the state offline, yielding an expected online cost of 2 and a roughly doubled offline cost. Our protocols C, C', and C'' (red, green, black) exhibit a very clear tradeoff between online circuit depth and offline cost. For example, if operations on logical qubits must be minimized (due to noise), then trading offline resources for low online circuit depth is desirable, making C, C', and C'' advantageous compared to $C_{\{\mathrm{BS,F,PS}\}}$. C'' is competitive with the minimal-online versions of C_{F} (plotted as C_{PAR}) and C_{PS} (not plotted). In practice, several decomposition techniques will be used throughout the compilation of a quantum algorithm.

Finally, our protocol can be used to fault-tolerantly implement elements of the V basis, which consists of $V_{\{1,2,3\}} = (I + 2i\{X, Y, Z\})/\sqrt{5}$ and their inverses. The V basis was shown to be *efficiently universal*, guaranteeing decompositions to be of depth $O(\log(1/\epsilon))$ [15]. It was previously dismissed as a candidate basis for decomposition due to the inability to implement the gates fault-tolerantly. However, our protocol enables fault-tolerant implementation: $V = Z(\pi/4)Z(2\theta_2)$, which is a T gate followed by a rotation using the $|H_2\rangle$ resource state. On average, it requires an offline cost of 10 $|H_0\rangle$ states. This has prompted the development of decomposition algorithms targeted to the V basis that may outperform those for the $\{H, T\}$ basis [16].

8 Conclusions

We have proposed a protocol to distill non-stabilizer states efficiently using magic states, Clifford operations, and measurements. One application of our protocol is implementing arbitrary single-qubit rotations with lower resource cost than state-of-the-art decomposition methods and *constant* online circuit depth. An extension of our work is to study other stabilizer circuits as "ladders" of states, or to use SH eigenstates distilled using the protocols of [3, 5]. Finally optimizing the sequence of angles required to implement the desired rotation, or determining when to use a given decomposition technique, will be a necessary component of any quantum compiler. We thank Alex Bocharov and Cody Jones for many useful discussions.

References

1. A. G. Fowler, A. M. Stephens, and P. Groszkowski, Phys. Rev. A **80**, 052312 (2009).
2. S. Aaronson and D. Gottesman, Physical Review A **70**, 052328 (2004), http://arxiv.org/abs/arXiv:quant-ph/0406196 .
3. S. Bravyi and A. Kitaev, Phys. Rev. A **71**, 022316 (2005).
4. A. M. Meier, B. Eastin, and E. Knill, "Magic state distillation with the four-qubit code," (2012).
5. S. Bravyi and J. Haah, "Magic state distillation with low overhead," (2012), 1209.2426 .
6. N. C. Jones, "Multilevel distillation of magic states for quantum computing," (2012), 1210.3388 .
7. A. Kitaev et al., *Classical and Quantum Computation* (American Mathematical Society, Providence, RI, 2002).
8. C. M. Dawson and M. A. Nielsen, "The Solovay-Kitaev algorithm," (2005), arxiv:quant-ph/0505030 .
9. A. Bocharov and K. M. Svore, Phys. Rev. Lett. **109**, 190501 (2012).
10. P. Selinger, "Efficient clifford+T approximation of single-qubit operators," (2012), 1212.6253 .
11. V. Kliuchnikov, D. Maslov, and M. Mosca, "Practical approximation of single-qubit unitaries by single-qubit quantum clifford and T circuits," (2012), 1212.6964 .
12. M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, UK, 2000).
13. N. C. Jones, J. D. Whitfield, P. L. McMahon, M.-H. Yung, R. V. Meter, A. Aspuru-Guzik, and Y. Yamamoto, "Simulating chemistry efficiently on fault-tolerant quantum computers," (2012), 1204.0567 .
14. A. Fowler, Quantum Information and Computation **11**, 867 (2011), quant-ph/0411206 .
15. A. W. Harrow, B. Recht, and I. L. Chuang, J. Math. Phys. **43** (2002).
16. A. Bocharov, Y. Gurevich, and K. M. Svore, in *Quantum Information Processing (QIP) 2012, Poster* (2012).

Realistic Cost for the Model of Coherent Computing

Akira SaiToh

National Institute of Informatics
2-1-2 Hitotsubashi, Chiyoda, Tokyo 101-8430, Japan
akirasaitoh@nii.ac.jp

Abstract

For the model of so-called coherent computing recently proposed by Yamamoto *et al.* [Y. Yamamoto *et al.*, New Gen. Comput. 30 (2012) 327–355], a theoretical analysis of the success probability is given. Although it was claimed as their prospect that the Ising spin configuration problem would be efficiently solvable in the model, here it is shown that the probability of finding a desired spin configuration decreases exponentially in the number of spins for certain hard instances. The model is thus physically unfeasible for solving the problem within a polynomial cost.

1998 ACM Subject Classification C.4 Performance of Systems

Keywords and phrases Reliability, Laser-network computing, Computational complexity

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.244

1 Introduction

It has been of long-standing interest to study the ability of analog computing systems to solve computationally difficult problems [1, 2]. It is recently of growing interest to investigate the power of quantum adiabatic time evolution in this direction [3]. Nevertheless, it has been commonly believed, with strong theoretical and numerical evidences, that a desired solution should not be obtained with a sufficiently large probability within polynomial time owing to the exponential decrease in the energy gap between desired and undesired eigenstates during an adiabatic change of Hamiltonians [4, 5, 6, 7, 8, 9].

Recently, Yamamoto *et al.* wrote a series of papers [10, 11, 12] on their model—so called the coherence computing model—of an injection-locked slave laser network, which uses quantum states to some extent in contrast to conventional classical optical computing models [14, 15]. It was claimed to be promising in solving the Ising spin configuration problem [16] and those polynomial-time reducible to this problem faster than known conventional models.

The Ising spin configuration problem has been well-known as a typical NP-hard problem described by an Ising-type Hamiltonian [16]. A typical description is as follows.

> *Ising spin configuration problem:* Given a graph $G = (V, E)$ with set V of vertices and set E of edges, and weighting functions $J : E \to \{0, \pm 1\}$ and $B : V \to \{0, \pm 1\}$, find the minimum eigenvalue $\lambda_{\rm g}$ of the Hamiltonian $H = \sum_{(ij)\in E} J_{ij}\sigma_{z,i}\sigma_{z,j} + \sum_{i\in V} B_i\sigma_{z,i}$. Here, $\sigma_{z,i}$ is the Pauli Z operator acting on the space of the ith spin (there are $n = |V|$ spin-1/2's).

In an intuitive point of view, the problem is difficult in the sense that the number of given parameters grows quadratically while the number of eigenvalues including multiplicity grows exponentially. Although the Hamiltonian is diagonal in the Z basis, writing it in the matrix form itself takes exponential time. Hereafter, we employ n for representing the input length

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 244–253


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

of an instance although, precisely speaking, the bit length of an encoded instance is $O(n^2)$. We do not go into the controversy on the definition of the input length [17]. As for known results on the complexity of the problem, it becomes P in case the graph is a planer graph and $B_i = 0 \ \ \forall i$ (see Ref. [18]); for nonplaner graphs, it is in general NP-hard, and it is so under many different conditions [18]. In addition, a planer graph together with nonzero B_i's also makes the problem NP-hard [16]. It is also worthwhile to mention that the typical value of $\lambda_{\rm g}$ is $c_{\rm g}n$ with coefficient $c_{\rm g}$ (so-called the ground-state energy density) typically between -2 and $-1/2$ when the values of J_{ij} are chosen in a certain random manner and B_i are set to zero [19, 20, 21, 22, 23, 24, 25, 26] ($c_{\rm g}$ is between -1.5 and -1 when the graph is a ladder and J_{ij} and B_i are randomly chosen from $\{\pm 1\}$ [27]). Furthermore, it should be mentioned that the distribution of eigenenergies of H (namely, the envelope of the multiplicity of eigenenergies with a normalization) is a normal distribution with mean zero and standard deviation proportional to $\sqrt{n}$ in the random energy model [22, 33, 25]. Here, the important observation is that the standard deviation increases with n in spite of the exponentially increasing number of spin configurations.

Let us also introduce the NP-complete variant of the Ising spin configuration problem as follows.

> *NPC Ising spin configuration problem:*
> *Instance:* Positive integer n, integer K, and parameters $J_{ij} \in \{0, \pm 1\}$ $(i < j)$ and $B_i \in \{0, \pm 1\}$ for integers $0 \leq i, j \leq n-1$.
> *Question:* Is there an eigenvalue λ of the Hamiltonian $H = \sum_{i,j=0;i<j}^{n-1} J_{ij}\sigma_{z,i}\sigma_{z,j} + \sum_{i=0}^{n-1} B_i\sigma_{z,i}$ such that $\lambda < K$?

This is the problem we are going to investigate in this contribution as for its computational difficulty under the coherent computing model.

Let us now briefly look into Yamamoto *et al.*'s coherent computing model [10, 11, 12] which is schematically depicted as Fig. 1. It has one master laser whose output is split into n paths and injected to n slave lasers. Each slave laser is initially locked to the superposed state $(|R\rangle_i + |L\rangle_i)$ where $|R\rangle$ and $|L\rangle$ are the right and left circular polarized states (see, *e.g.*, Refs. [28, 29] for physics of the injection-locked laser system). The initial state of the n slave lasers is therefore $\bigotimes_{i=0}^{n-1}(|R\rangle_i + |L\rangle_i)$. The laser network is a macroscopic system; thus initially it holds many photons in this same state. The computational basis is set to $\{|R\rangle, |L\rangle\}^n$ and σ_z is written as $|R\rangle\langle R| - |L\rangle\langle L|$. The ith slave laser and the jth slave laser are connected for nonzero J_{ij}. At time $t = 0$, they mutually inject a small amount of horizontally polarized signal via an attenuator, a phase shifter, and a horizontal linear polarizer, which determine the amplitude attenuation coefficient that is regarded as J_{ij}. Among the three instruments, the attenuator's transmission coefficient controls $|J_{ij}|$ and the other instruments controls $\mathrm{sgn}J_{ij}$. In addition, a small amount of injection of horizontally polarized signal is also made from the master laser to each slave laser at $t = 0$. This amount corresponds to B_i for the ith slave laser. It is controlled by the combination of a half-wave plate and a quarter wave plate. For more details of implementation of the coefficients, see section 7 of Utsunomiya *et al.* [10].

Then one waits for a small time duration $t_{\rm st}$ to let the system evolve. Laser modes satisfying the matching condition with the above-mentioned setting grow rapidly and other modes are suppressed. For $t > t_{\rm st}$, the system is thought to be in a steady state. Then for each slave laser its output is guided to a polarization beam splitter and the right and the left polarization components are separately detected by photodetectors. By a majority vote of photon number counting, the computational result of each slave laser, $|a\rangle_i \in \{|R\rangle, |L\rangle\}$, is

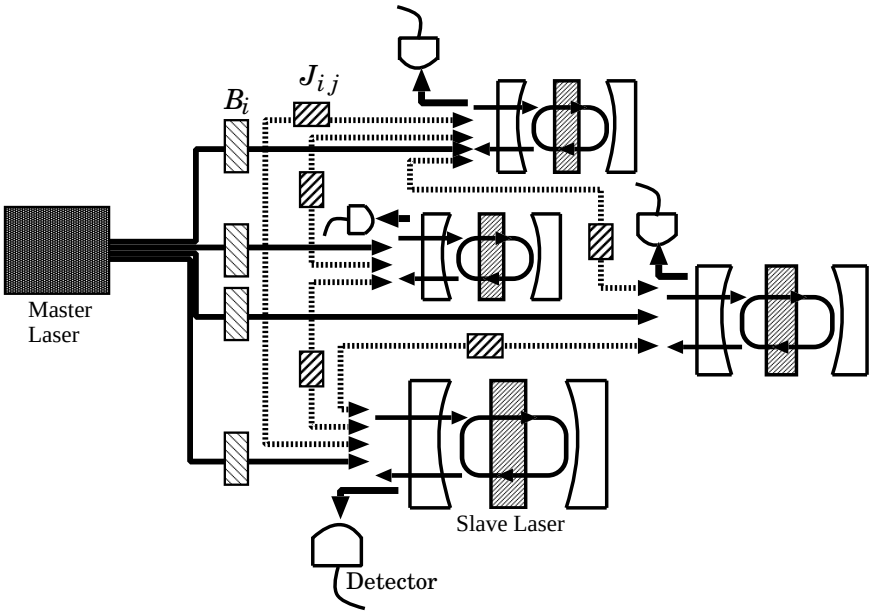


Figure 1 Schematic description of the coherent computing model. See the text for how J_{ij} and B_i are realized by optical instruments.

retrieved. The steady state $|a\rangle_0 \cdots |a\rangle_{n-1}$ is thus determined. Once this is determined, it takes only polynomial time to calculate the corresponding eigenvalue since there are only $O(n^2)$ terms in the Hamiltonian (here, we do not use its matrix form).

Thus, in short, the state starts from $(|R\rangle + |L\rangle)^{\otimes n}$ and eventually reaches a steady state representing a configuration that corresponds to the minimum energy of the given Hamiltonian. Yamamoto *et al.* [10, 11, 12] employed rate equations involving several factors characterizing each oscillator and connections with other oscillators to analyze photon numbers of the right and left polarization components for each slave laser; they concluded that the system reaches a steady state within 10 nano seconds without obvious dependence on n.

It has been unknown so far if the coherent computing model is a valid computer model in view of a rigid and fair description of computational costs. Conventional analog computing models do not solve NP-hard problems within a polynomial cost; they require either exponentially long convergence time or exponentially fine accuracy [13]. Thus it should be natural to be skeptical against the power of the coherent computing model. In this contribution, we investigate the signal per noise ratio in the output of the coherent computer when the NPC Ising spin configuration problem is handled. We will reach the fact that for certain hard instances, the relative signal intensity corresponding to solutions is bounded above by a function decreasing exponentially in n. This is because the number of modes that are possibly locked in the laser network increases rapidly in n owing to the fact that the locking range of the laser network does not shrink as n grows considering imperfectness of optical instruments.

The analysis of computational difficulty is described in Sec. 2. The result is discussed in Sec. 3 and summarized in Sec. 4.

2 Computational difficulty in the coherent computing model

The coherent computing model illustrated in Fig. 1 was so far analyzed by Utsunomiya *et al.* [10, 11, 12] on the basis of the assumption that given coefficients J_{ij} and B_i are exactly implemented by optical instruments although fluctuations and quantum noise in the system were considered in their analyses of time evolutions using rate equations, which led to a quite ideal convergence taking only 10 nano seconds.

Here, we assume that individual optical instruments are imperfect[1] so that there are errors in J_{ij} and B_i, which are due to calibration errors and/or thermal fluctuations. Then the following proposition is achieved.

▶ **Proposition 1.** *Consider the NPC Ising spin configuration problem. Suppose calibration errors and/or thermal fluctuations of optical instruments cause nonzero physical deviations,*[1] $\varepsilon_{ij} \in \mathbf{R}$ *for nonzero* J_{ij} *and* $\kappa_i \in \mathbf{R}$ *for nonzero* B_i*. We assume that* ε_{ij} *are* i.i.d. *random variables with mean zero and a certain standard deviation* σ_ε *and* κ_i *are* i.i.d. *random variables with mean zero and a certain standard deviation* σ_κ*. Then, for large* n*, there exist YES instances such that the probability to obtain a spin configuration corresponding to one of* λ*'s* $< K$ *using the coherent computer is* $\leq \mathrm{poly}(n)2^{-n}$*.*

Proof. Here we consider instances generated in the way that J_{ij}'s and B_i's are independent uniformly distributed random variables with values in $\{0, \pm 1\}$. Since a problem instance is a given data set, the standard deviation for J_{ij} and that for B_i intrinsic to the problem instance itself are not of our concern. We only consider physical deviations as errors.

As the model is a sort of a bulk model (there are many photons), it is convenient to consider populations of individual configurations. Let $P_{\lambda,l_\lambda}(t)$ be the population of eigenstate $|\varphi_{\lambda,l_\lambda}\rangle$ ($l_\lambda \in \{0, \ldots, d_\lambda - 1\}$) corresponding to eigenenergy λ of the Hamiltonian (specified by the problem instance), where t stands for time and d_λ is the multiplicity of λ. We also introduce $P_\lambda(t) = \sum_{l_\lambda=0}^{d_\lambda-1} P_{\lambda,l_\lambda}(t)$. It should be kept in mind that we do not start from the thermal distribution; for the initial state, we have identical copies of $\sum_\lambda \sum_{l_\lambda} |\varphi_{\lambda,l_\lambda}\rangle = (|R\rangle + |L\rangle)^{\otimes n}$. In the present setting, the random-energy model [22, 33] is valid[2] and hence, for large n, with an appropriate scaling factor M, one can write $P_\lambda(0) = M\mathcal{N}(0, \sigma_\lambda^2)$ with $\sigma_\lambda = \Theta(\sqrt{n})$ where $\mathcal{N}(\mu, \sigma^2)$ is the density function of the normal distribution with mean μ and standard deviation σ. Here, we have $M = 2^n P_{\lambda_\mathrm{g},0}(0)$ with λ_g the ground state energy because the initial population is same for all the configurations.

Let us denote the set of solution states (spin configurations corresponding to λ's $< K$) as Y. The total population of solution states at t is given by $P_Y(t) = \sum_{\lambda<K} P_\lambda(t)$. Similarly, the total population of nonsolution states is given by $P_X(t) = \sum_{\lambda\geq K} P_\lambda(t)$; here, $X = \{|\varphi_{\lambda,l_\lambda}\rangle \mid \lambda \geq K\}$. Ideally, only $|\varphi_{\lambda,l_\lambda}\rangle$'s $\in Y$ will enjoy population enhancement by mode selections. However, there exists $v \geq K$ such that $P_\lambda(t > t_\mathrm{st}) \gg 0$ for $\lambda \leq v$. This is because

[1] It is a common case that each optical instrument has a few permil uncertainty in the calibration of each property (see Ref. [30]). In addition, there is a quantum limit in any classical instrument [31, 32] so that a manufacturing error and a manipulation error cannot be made arbitrarily small.

[2] Let us pick up a certain configuration $|\varphi\rangle$. Suppose, by applying m bit flips, its energy changes by $\Delta E(\varphi \overset{m}{\mapsto} \varphi')$ with $|\varphi'\rangle$ a resultant configuration. This process should obey the random energy change and hence for large m, $\Delta E(\varphi \overset{m}{\mapsto} \varphi')$ should obey the normal distribution with mean zero and a standard deviation proportional to $\sqrt{m}$ by the central limit theorem (in regard with a sum of random variables). In addition, the most typical number of bit flips is $n/2$ when we generate all other configurations from $|\varphi\rangle$. Typical bit flips generate a dominant number of configurations. Thus the distribution of energies is approximated by the normal distribution with mean zero and a standard deviation proportional to $\sqrt{n}$. In this way, we have just obtained the energy distribution function in the random-energy model.

the matching condition is imperfect in reality; the locking range is broader than the ideal range considering errors in optical instruments.[3] Let us write $P_Z(t) = \sum_{K \le \lambda \le v} P_\lambda(t)$; here, $Z = \{|\varphi_{\lambda,l_\lambda}\rangle \mid K \le \lambda \le v\}$.

By assumption, we are considering physical deviations (including calibration errors and thermal fluctuations), ε_{ij} for nonzero J_{ij} and κ_i for nonzero B_i. The Hamiltonian implemented on the laser network is written as $\widetilde{H} = \sum_{i<j|J_{ij}\neq 0}(J_{ij}+\varepsilon_{ij})\sigma_{z,i}\sigma_{z,j} + \sum_{i|B_i\neq 0}(B_i+\kappa_i)\sigma_{z,i}$. This suggests that $v = K + K'(n)$ with $K'(n) \simeq \sigma_\varepsilon\sqrt{n(n-1)/3} + \sigma_\kappa\sqrt{2n/3}$ by the central limit theorem in regard with a sum of random variables (see, *e.g.*, Refs. [35, 36]), considering the expected number of nonzero J_{ij}'s and that of nonzero B_i's. Therefore, $P_Z(0) = M\int_K^{K+K'(n)} \mathcal{N}(0,\sigma_\lambda^2)\mathrm{d}\lambda$.

Let us write $H = H_J + H_B$ with $H_J = \sum_{i<j} J_{ij}\sigma_{z,i}\sigma_{z,j}$ and $H_B = \sum_i B_i\sigma_{z,i}$. As we have mentioned, it is known [19, 20, 21, 22, 23, 24, 25, 26] that the ground state energy of H_J is typically $c_{\mathrm{g}}n$ with $-2 < c_{\mathrm{g}} < -1/2$. Therefore, for any normalized vector $|v\rangle$ in the Hilbert space of the system of our concern, $\langle v|H|v\rangle$ is typically bounded below by $-3n$. Thus, for typical instances we can choose $K = K(n)$ with $-K(n) = O(n)$. Recall that $K'(n) = \Theta(n)$ and $\sigma_\lambda = \Theta(\sqrt{n})$. We find that $\int_K^{K+K'(n)} \mathcal{N}(0,\sigma_\lambda^2)\mathrm{d}\lambda = \left[\frac{1}{2}\mathrm{erf}(\frac{\lambda}{\sqrt{2}\sigma_\lambda})\right]_K^{K+K'(n)}$ is a monotonically increasing function of n. Hence, for a certain constant $b > 0$, $P_Z(0) \ge b2^n P_{\lambda_{\mathrm{g}},0}(0)$.

Let us assume that locked modes have equally enhanced intensities for $t > t_{\mathrm{st}}$. This leads to the signal per noise ratio for $t > t_{\mathrm{st}}$: $P_Y(t > t_{\mathrm{st}})/P_Z(t > t_{\mathrm{st}}) = P_Y(0)/P_Z(0)$. (In case one can assume that only one of $|\varphi_{\lambda,l_\lambda}\rangle$'s in $Y \cup Z$ survives, the ratio of the probability of finding $|\varphi_{\lambda,l_\lambda}\rangle$ originated from Y and that of finding $|\varphi_{\lambda,l_\lambda}\rangle$ originated from Z at $t > t_{\mathrm{st}}$ is given by the same equation.)

Consider some typical instances for which d_{g} is small and is not clearly dependent on n (d_{g} is the multiplicity in the ground level). This is a typical situation because the multiplicity of λ obeys the distribution $\mathcal{N}(0,\sigma_\lambda^2)$ with $\sigma_\lambda = \Theta(\sqrt{n})$ in the present setting, as we have explained. It is always possible to choose[4] the value of K such that all $|\varphi_{\lambda,l_\lambda}\rangle \in Y$ are configurations with at most a constant number of bits different from one of the ground states. In this case, $P_Y(0) = \mathrm{poly}(n)P_{\lambda_{\mathrm{g}},0}(0)$ and thus, for large n, $P_Y(t > t_{\mathrm{st}})/P_Z(t > t_{\mathrm{st}}) \le \mathrm{poly}(n)2^{-n}$. ◀

▶ Remark. It is trivial to find a similar proof for the existence of hard instances of the Ising spin configuration problem for finding a ground level in the coherent computing model.

By Proposition 1, it is now easy to prove the following theorem.

▶ **Theorem 1.** *There exists an instance of the NPC Ising spin configuration problem such that a decision takes* $\Omega(2^n/\mathrm{poly}(n))$ *time in the coherent computing model when nonzero physical deviations,*[1] $\varepsilon_{ij} \in \mathbf{R}$ *for nonzero* J_{ij} *and* $\kappa_i \in \mathbf{R}$ *for nonzero* B_i*, are considered. Here,* ε_{ij} *(*κ_i*) are assumed to be* i.i.d. *random variables with zero mean and a certain standard deviation* σ_ε *(*σ_κ*).*

Proof. By Proposition 1, there exists an YES instance such that the probability p_s for a single trial of coherent computing to find $\lambda < K$ is $\le \mathrm{poly}(n)2^{-n}$. The success probability after τ trials is given by $1-(1-p_s)^\tau$. In order to make this probability larger than a certain constant $c > 0$, we need $\tau > \log(1-c)/\log(1-p_s) = (\log\frac{1}{1-c})/[p_s + \mathcal{O}(p_s^2)] = \Omega(2^n/\mathrm{poly}(n))$. ◀

[3] See, *e.g.*, Ref. [34] for an experimental gain curve.
[4] Recall that we are proving the *existence* of hard instances.

3 Discussion

We have theoretically shown a weakness of the coherent computing model for the problem to examine the existence of a suitably small (large negative) eigenvalue of an Ising spin glass Hamiltonian. As the number n of spins grows, the desired signal decreases exponentially for certain hard instances because exponentially many undesired configurations obtain gains in a realistic setting.

Indeed, Yamamoto *et al.* made numerical simulations [10, 11, 12] to examine their prospect that a desired configuration would be found efficiently in the coherent computing model. But, in general, the following points should be taken into account whenever a computer simulation of a physical system is performed.

First, in classical computing, exponentially fine accuracy is achievable by linearly increasing the register size of a variable or an array size of combined variables. Nevertheless, in physical systems, noise decreases as $\propto 1/\sqrt{T}$ with T the number of trials or the number of identical systems according to the well-known central limit theorem. In the field of quantum computing, this has been well-studied in the context of NMR bulk-ensemble computation at room temperature which suffers from exponential decrease of signal intensity corresponding to the computation result as the input size grows (see, *e.g.*, [37, 38]). In the coherent computing model, the ratio of the population of correct configurations and that of wrong configurations at the steady state should not decrease in a super-polynomial manner if the model were physically feasible for solving the problem efficiently. So far, Yamamoto *et al.* reported [10, 11, 12] that each slave laser maintains a sufficiently large discrepancy between the populations of $|R\rangle$ and $|L\rangle$ at the steady state for some instances with a small number of spins ($n \leq 10$), using a simulation based on rate equations. They also showed their simulation results for $n = 1000$ for a very restricted type of instances such that J_{ij}'s take the same value and B_i's for odd i take the same value and so do for even i. Nevertheless, the populations (in other words, the joint probabilities) of correct and wrong configurations and how they scale for large n were not reported. Recently, Wen [39] showed his simulation results for the case where the graph was a two-layer lattice for n up to 800. Although it was reported that his simulations of the coherent computer found eigenvalues lower than those found by a certain semidefinite programming method, the populations of correct and wrong configurations were not shown. Thus, it is difficult to discuss the power of the coherent computing model on the basis of presently known simulation results.

Second, the coefficients of a problem Hamiltonian cannot be implemented as they are, in reality. Seemingly negligible errors in the coefficients might be crucial in complexity analyses for a large input size. This point has not been considered in conventional simulation studies [11, 12, 39] of the coherent computing model. In the coherent computing model, nonzero J_{ij}'s and nonzero B_i's in the Ising spin glass Hamiltonian should accompany calibration errors and/or thermal fluctuations. In particular, optical instruments usually have nonnegligible calibration errors [30]. As we have written in the proof of Proposition 1, a well-known application of the central limit theorem for the sum of random variables [35, 36] indicates the important observation that the sum of such physical deviations is an increasing function of the number of spins. This fact has led to our conclusion that the relative population of desired configurations decreases exponentially in n for certain hard instances.

The second point is also usually overlooked in computer simulations [3] of adiabatic quantum computing. Discussions on the complexity of adiabatic time evolution are usually made as to how long time should be spent in light of a minimum energy gap between the ground state and the nearest excited state during adiabatically changing the Hamiltonian

toward its final form. The coefficients in the starting and the final Hamiltonians are quite often considered to be given accurate numbers [9]. Nevertheless, they should have certain errors due to imperfect calibrations [30] and/or fluctuations in reality, as we have discussed. The target state will not appear as a stable state if a nontarget state of the final Hamiltonian becomes a ground state of the Hamiltonian owing to the errors. A real physical setup for adiabatic quantum computing should suffer from the demand of considerably fine tuning of individual apparatus to implement desired coupling for large n. So far, n has not been very large in physical implementations [40, 41, 42] so that this problem has not been significant. (In addition, even under the setting without error in Hamiltonian coefficients, adiabatic quantum computing tends to suffer from exponentially decreasing energy gap when random instances of certain NP-hard problems are tried, according to the numerical analysis by Farhi *et al.*[9])

A possible way to avoid very fine tuning is to use error correction schemes similar to those for standard circuit-model quantum computing. There have been several studies on error correction codes [43] and dynamical decoupling [44, 45] in the context of adiabatic quantum computing. It is of interest if similar schemes apply to the coherent computing model. As for error correction codes, each Pauli operator in an original Hamiltonian should be encoded to a certain multi-partite coupling term in an encoded Hamiltonian. Thus one needs to find a scheme to implement such a term in the coherent computing model. It is highly nontrivial to introduce, *e.g.*, a four-partite coupling among slave lasers. Further investigation is needed for the usability of error correction codes. Another scheme is dynamical decoupling. This scheme looks effective for suppressing thermal fluctuations at a glance. Consider the minimum gap between two distinct eigenvalues of a problem Hamiltonian and normalize it with the maximum gap. This decreases only polynomially in n for any instance of the Ising spin configuration problem by the definition of the problem. Thus the minimum operation interval of dynamical decoupling required for an effective noise suppression decreases only polynomially in n according to Eq. (52) of Ref. [46]. One problem is how to use this scheme for cancelling calibration errors. In addition, we need to find an implementation of the scheme such that the scheme itself does not introduce an uncontrollable noise. This will be difficult for large n because imperfections in decoupling operations probably lead to a similar argument as Proposition 1.

As we have proved, there are hard instances of the NPC Ising spin configuration problem for which one cannot efficiently achieve a correct decision in the coherent computing model (Theorem 1). This is a reasonable result in light of the fact that no known conventional computer model could solve an NP-complete problem within a polynomial cost. It is still an open problem if an unreasonable computational power is achievable by combining error protection schemes with the coherent computing model.

4 Conclusion

The model of coherent computing has been theoretically investigated in view of computational cost under a realistic setting. It has been proved that there exist hard instances of the NPC Ising spin configuration problem, which require exponential time for a correct decision in the model.

Acknowledgements. The author would like to thank William J. Munro, Kae Nemoto, and Yoshihisa Yamamoto for helpful discussions. This work is supported by the Grant-in-Aid for Scientific Research from JSPS (Grant No. 25871052).

References

1 T. Roska, L. O. Chua, The CNN Universal Machine: An Analogic Array Computer, IEEE Trans. Circuits Sys. II 40 (1993) 163-173.

2 M. Ercsey-Ravasz, T. Roska, and Z. Néda, Cellular neural networks for NP-hard optimization, in the 11th International Workshop on Cellular Neural Networks and their Applications (CNNA2008), Santiago de Compostela, Spain, 14-16 July 2008, pp.52-56; *ibid*, EURASIP J. Adv. Signal Process. 2009 (2009) 646975.

3 E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, A Quantum Adiabatic Evolution Algorithm Applied to Random Instances of an NP-Complete Problem, Science 292 (2001) 472-475.

4 W. van Dam, M. Mosca, and U. Vazirani, How Powerful is Adiabatic Quantum Computation? in: Proceedings of the 42nd IEEE Symposium on Foundations of Computer Science (FOCS 2001), Las Vegas, NV, 14-17 October 2001 (IEEE, New York, 2001) pp.279-287.

5 M. Žnidarič, Scaling of the running time of the quantum adiabatic algorithm for propositional satisfiability, Phys. Rev. A 71 (2005) 062305.

6 M. Žnidarič and M. Horvat, Exponential complexity of an adiabatic algorithm for an NP-complete problem, Phys. Rev. A 73 (2006) 022329.

7 M. H. S. Amin, Effect of Local Minima on Adiabatic Quantum Optimization, Phys. Rev. Lett. 100 (2008) 130503.

8 I. Hen and A. P. Young, Exponential complexity of the quantum adiabatic algorithm for certain satisfiability problems, Phys. Rev. E 84 (2011) 061152.

9 E. Farhi, D. Gosset, I. Hen, A. W. Sandvik, P. Shor, A. P. Young, and F. Zamponi, Performance of the quantum adiabatic algorithm on random instances of two optimization problems on regular hypergraphs, Phys. Rev. A 86 (2012) 052334.

10 S. Utsunomiya, K. Takata, and Y. Yamamoto, Mapping of Ising models onto injection-locked laser systems, Opt. Exp. 19 (2011) 18091.

11 K. Takata, S. Utsunomiya, and Y. Yamamoto, Transient time of an Ising machine based on injection-locked laser network, New J. Phys. 14 (2012) 013052.

12 Y. Yamamoto, K. Takata, and S. Utsunomiya, Quantum Computing v.s. Coherent Computing, New Gen. Comput. 30 (2012) 327-355.

13 S. Aaronson, Guest Column: NP-complete problems and physical reality, ACM SIGACT News, 36(1) (2005) 30-52.

14 N. T. Shaked, S. Messika, S. Dolev, and J. Rosen, Optical solution for bounded NP-complete problems, Appl. Opt. 46 (2007) 711-724.

15 S. Dolev, T. Haist, and M. Olteans, Eds., Optical Supercomputing, 1st Int. Workshop, Vienna, Austria, 26 August 2008, Proceedings, LNCS 5172 (Springer, Berlin, 2008).

16 F. Barahona, On the computational complexity of Ising spin glass models, J. Phys. A: Math. Gen. 15 (1982) 3241-3253.

17 T. E. O'Neil, The Importance of Symmetric Representation, in: Proceedings of the 2009 International Conference on Foundations of Computer Science (FCS'09), Las Vegas, NV, 13-16 July 2009 (CSREA Press, USA, 2009) pp.115-119.

18 S. Istrail, Statistical mechanics, three-dimensionality and NP-completeness: I. Universality of intractability for the partition function of the Ising model across non-planar surfaces, in: Proceedings of the Thirty-Second Annual ACM Symposium on Theory of Computing (STOC'00), Portland, OR, 21-23 May 2000 (ACM, New York, 2000) pp.87-96.

19 J. Vannimenus and G. Toulouse, Theory of the frustration effect: II. Ising spins on a square lattice, J. Phys. C: Solid State Phys. 10 (1977) L537-L542.

20 S. Kirkpatrick, Frustration and ground-state degeneracy in spin glasses, Phys. Rev. B 16 (1977) 4630-4641.

21 I. Morgenstern and K. Binder, Magnetic correlations in two-dimensional spin-glasses, Phys. Rev. B 22 (1980) 288-303.

22 B. Derrida, Random-energy model: Limit of a family of disordered models, Phys. Rev. Lett. 45 (1980) 79-82.

23 B. Derrida, Random-energy model: An exactly solvable model of disordered systems, Phys. Rev. B 24 (1981) 2613-2626.

24 C. De Simone, M. Diehl, M. Jünger, P. Mutzel, G. Reinelt, and G. Rinaldi, Exact ground states of Ising spin glasses: New experimental results with a branch-and-cut algorithm, J. Stat. Phys. 80 (1995) 487-496.

25 A. Andreanov, F. Barbieri, and O.C. Martin, Large deviations in spin-glass ground-state energies, Eur. Phys. J. B 41 (2004) 365-375.

26 S. Boettcher, Simulations of ground state fluctuations in mean-field Ising spin glasses, J. Stat. Mech. 2010 (2010) P07002.

27 T. Kadowaki, Y. Nonomura, and H. Nishimori, Exact ground-state energy of the Ising spin glass on strips, J. Phys. Soc. Jpn. 65 (1996) 1609-1616.

28 H. Haken, H. Sauermann, Ch. Schmid, and H. D. Vollmer, Theory of Laser Noise in the Phase Locking Region, Z. Phys. 206 (1967) 369-393.

29 H. A. Haus and Y. Yamamoto, Quantum noise of an injection-locked laser oscillator, Phys. Rev. A 29 (1984) 1261-1274.

30 The SP 250 Series on NIST Measurement Services, `http://www.nist.gov/calibrations/sp250_series.cfm`, see *e.g.,* SP 250-64: R. W. Leonhardt, Calibration Service for Low-level Pulsed-Laser Radiometers at 1.06 μm: Pulse Energy and Peak Power.

31 A. A. Clerk, M. H. Devoret, S. M. Girvin, F. Marquardt, and R. J. Schoelkopf, Introduction to quantum noise, measurement, and amplification, Rev. Mod. Phys. 82 (2010) 1155-1208.

32 M. D. LaHaye, O. Buu, B. Camarota, and K. C. Schwab, Approaching the Quantum Limit of a Nanomechanical Resonator, Science 304 (2004) 74-77.

33 B. Derrida, A generalization of the Random Energy Model which includes correlations between energies, J. Physique Lett. 46 (1985) L-401-L-407.

34 S. Kobayashi and T. Kimura, Injection Locking in AlGaAs Semiconductor Laser, IEEE J. Quant. Ele. QE-17 (1981) 681-689.

35 A. N. Shiryaev, Probability (2nd ed., translated by R. P. Boas, Springer-Verlag, New York, 1996).

36 A. Klenke, Probability Theory: A Comprehensive Course (Springer-Verlag, London, 2008).

37 E. Knill, I. Chuang, and R. Laflamme, Effective pure states for bulk quantum computation, Phys. Rev. A 57 (1998) 3348-3363.

38 A. SaiToh and M. Kitagawa, Matrix-product-state simulation of an extended Brüschweiler bulk-ensemble database search, Phys. Rev. A 73 (2006) 062332.

39 K. Wen, Injection-locked laser network for solving NP-complete problems, PhD Thesis, Stanford University (2012), `http://purl.stanford.edu/xp446hc0861`.

40 M. Steffen, W. van Dam, T. Hogg, G. Breyta, and I. Chuang, Experimental Implementation of an Adiabatic Quantum Optimization Algorithm, Phys. Rev. Lett. 90 (2003) 067903.

41 X. Peng, Z. Liao, N. Xu, G. Qin, X. Zhou, D. Suter, and J. Du, Quantum Adiabatic Algorithm for Factorization and Its Experimental Implementation, Phys. Rev. Lett. 101 (2008) 220405.

42 M. W. Johnson *et al.*, Quantum annealing with manufactured spins, Nature 473 (2011) 194-198.

43 S. P. Jordan, E. Farhi, and P. W. Shor, Error-correcting codes for adiabatic quantum computation, Phys. Rev. A 74 (2006) 052322.

44 D. A. Lidar, Towards Fault Tolerant Adiabatic Quantum Computation, Phys. Rev. Lett. 100 (2008) 160506.
45 G. Quiroz and D. A. Lidar, High-fidelity adiabatic quantum computation via dynamical decoupling, Phys. Rev. A 86 (2012) 042333.
46 H. K. Ng, D. A. Lidar, and J. Preskill, Combining dynamical decoupling with fault-tolerant quantum computation, Phys. Rev. A 84 (2011) 012305.

Optimal Robust Self-Testing by Binary Nonlocal XOR Games

Carl A. Miller[1] and Yaoyun Shi[2]

1 **Electrical Engineering and Computer Science Department**
University of Michigan
2260 Hayward St. Ann Arbor, MI 48109 USA carlmi@umich.edu
2 **Electrical Engineering and Computer Science Department**
University of Michigan
2260 Hayward St. Ann Arbor, MI 48109 USA shiyy@umich.edu

Abstract

Self-testing a quantum apparatus means verifying the existence of a certain quantum state as well as the effect of the associated measuring devices based only on the statistics of the measurement outcomes. Robust (i.e., error-tolerant) self-testing quantum apparatuses are critical building blocks for quantum cryptographic protocols that rely on imperfect or untrusted devices. We devise a general scheme for proving optimal robust self-testing properties for tests based on nonlocal binary XOR games. We offer some simplified proofs of known results on self-testing, and also prove some new results.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases self-testing, quantum cryptography, random number generation, nonlocal games

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.254

1 Introduction

Consider a quantum apparatus with a classical input/output interface, and suppose that the internal behavior of the apparatus — the quantum state inside and the measurements selected by the classical input — cannot be trusted to conform to a desired specification. The apparatus is said to be *self-testing* [8], if there exists a self-test, i.e., a set of constraints on the input-output correlations, that once satisfied will guarantee the accuracy to the specification.

The notion of quantum self-testing was explicitly formulated by Mayers and Yao [8], who pointed out its importance for quantum cryptography: self-testing enables quantum cryptographic protocols that rely on imperfect or untrusted quantum devices. Such protocols were advanced in the recent thrust of research on *device-independent* quantum cryptography [1, 15, 9, 14, 6, 5, 18].

Multiple self-testing results are known. Such results are often based on nonlocal games. Popescu and Rohrlich [16] proved that any state that achieves a maximal violation of the CHSH inequality [3] must be equivalent to a direct sum of singlets. A self-testing result was proved for the GHZ paradox by Colbeck [4].

In order for self-testing to be practically useful, it must tolerate error. That is, an apparatus close to passing the self-test must be close to the specification. Robust self-testing results have been proved in [7, 10, 12, 11, 17]. These papers include two recent results which prove robust self-testing for the CHSH game [11, 17].

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 254–262

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Existing proofs of self-testing are fairly lengthy and technical, and appear specific to the underlying (class of) quantum states. Also, there is some variation in the error terms afforded by these results. Some of the results on nonlocal games show that if the score achieved is within ϵ of a passing score, the deviation of the apparatus from perfect behavior is no more than $C\sqrt{\epsilon}$. For other results (e.g., in [10, 11]) the error term is $C\epsilon^{1/4}$. It is natural to ask whether these error bounds can be tightened.

Most existing self-tests are based on binary nonlocal XOR games. In this paper, working within this class, we provide a simple criterion which determines whether a particular game is a robust self-test. The criterion guarantees an error term of $C\sqrt{\epsilon}$, which is easily seen to be the best possible (up to the constant C). The criterion is fairly simple to check, it encompasses known results on the CHSH game and the GHZ paradox, and it allows the proof of new results.

The starting point of our theory is the idea, first observed by Werner and Wolf [19], that the optimal score for a binary nonlocal XOR game can be expressed as the maximum of a certain multivariable sinusoidal function. In the present paper, we take the idea a step further and show that the robust self-testing property can be checked using the local and global properties of this function.

We will begin with some definitions and then state our main results. The results are stated initially for multiqubit states only, and a higher-dimensional generalization is given at the end of the paper. The proofs are sketched here—full proofs can be found in [13]. We offer some examples. We give a simple proof that the CHSH game is a robust self-test. (This result improves on the error term in [11], and it matches that of the independent work [17].) We also augment a recent paper [2] on randomness and quantum correlations by showing that a certain one-parameter family of games satisfies the robust self-testing condition.

2 Definitions

For our purposes, a *binary nonlocal XOR game* is simply a function $f\colon \{0,1\}^n \to \mathbb{R}$. The function f describes a scoring rule for the game: if the input sequence is $(i_1, i_2, \ldots, i_n)$, and the output sequence satisfies $\oplus_k o_k = 0$, then the score is $f(i_1, i_2, \ldots, i_n)$; if the input sequence is $(i_1, i_2, \ldots, i_n)$ and the output sequence satisfies $\oplus_k o_k = 1$, then the score is $-f(i_1, i_2, \ldots, i_n)$.

To any nonlocal game f, let us associate a polynomial $P_f\colon \mathbb{C}^n \to \mathbb{C}$ like so: for any n-tuple $(\lambda_1, \ldots, \lambda_n)$ of complex numbers, let $P_f(\lambda_1, \ldots, \lambda_n)$ be equal to

$$\sum_{(i_1,\ldots,i_n)\in\{0,1\}^n} f(i_1, \ldots, i_n)\lambda_1^{i_1}\lambda_2^{i_2}\cdots\lambda_n^{i_n}. \tag{1}$$

For example, if g is the CHSH game ($g(1,1) = -1$, $g(0,0) = g(0,1) = g(1,0) = 1$) then

$$P_g = 1 + \lambda_1 + \lambda_2 - \lambda_1\lambda_2. \tag{2}$$

Additionally, for any binary nonlocal XOR game $f\colon \{0,1\}^n \to \mathbb{R}$, and any real numbers $\theta_0, \theta_1, \ldots, \theta_n$, let $Z_f(\theta_0, \ldots, \theta_n)$ denote the quantity

$$\sum_{(i_k)\in\{0,1\}^n} f(i_1, \ldots, i_n)\cos\left(\theta_0 + \sum_k i_k\theta_k\right). \tag{3}$$

Thus,

$$Z_g(\theta_0, \theta_1, \theta_2) = \cos(\theta_0) + \cos(\theta_0 + \theta_1) + \cos(\theta_0 + \theta_2) - \cos(\theta_0 + \theta_1 + \theta_2). \tag{4}$$

Note that the function Z_f is 2π-periodic in every variable.

The two quantities P_f and Z_f are related by the following identity.

$$Z_f(\theta_0, \ldots, \theta_n) \quad = \quad \mathrm{Re}[e^{i\theta_0} P_f(e^{i\theta_1}, \ldots, e^{i\theta_n})]. \tag{5}$$

Note also that

$$\left|P_f\left(e^{i\theta_1}, \ldots, e^{i\theta_n}\right)\right| \quad = \quad \max_{t\in[-\pi,\pi]} Z_f(t, \theta_1, \ldots, \theta_n). \tag{6}$$

3 Quantum strategies

For our purposes, a *quantum strategy* for a binary n-player nonlocal game is a pure state

$$|\psi\rangle \in \mathcal{Q}_1 \otimes \mathcal{Q}_2 \otimes \ldots \otimes \mathcal{Q}_n, \tag{7}$$

where each $\mathcal{Q}_j$ is a finite-dimensional Hilbert space, together with two projective measurements

$$\left\{P_j^{(0,+)}, P_j^{(0,-)}\right\}, \left\{P_j^{(1,+)}, P_j^{(1,-)}\right\} \tag{8}$$

on the space $\mathcal{Q}_j$. These measurements can be more compactly expressed as Hermitian operators:

$$M_j^{(0)} \quad := \quad P_j^{(0,+)} - P_j^{(0,-)} \tag{9}$$

$$M_j^{(1)} \quad := \quad P_j^{(1,+)} - P_j^{(1,-)} \tag{10}$$

The *score* for such a strategy is

$$\langle\psi| \sum_{(i_k)} f(i_1, \ldots, i_n) M_1^{(i_1)} \otimes \ldots \otimes M_n^{(i_n)} |\psi\rangle . \tag{11}$$

Let us use the term *qubit strategy* to refer to a strategy whose Hilbert spaces $\mathcal{Q}_j$ are all copies of $\mathbb{C}^2$ and whose projection operators $P_j^{(i,*)}$ are all one-dimensional projectors.

For any nonlocal game f, let q_f denote the highest possible score for f that can be achieved by a qubit strategy. This quantity has a relationship to the functions Z_f and P_f which was proved in [19]. For the benefit of our exposition, we include a proof here.

▶ Proposition 1. Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a nonlocal binary XOR game. Then,

$$q_f = \max_{|\lambda_1|=\ldots=|\lambda_n|=1} |P_f(\lambda_1, \ldots, \lambda_n)| \tag{12}$$

and

$$q_f = \max_{\theta_0,\ldots,\theta_n\in[-\pi,\pi]} Z_f(\theta_0, \ldots \theta_n). \tag{13}$$

Proof. Let $\left(\psi, \left\{\{M_j^{(0)}, M_j^{(1)}\}\right\}_j\right)$ be a qubit strategy for f. Each of the operators $M_j^{(i)}$ is a Hermitian operator on a 2-dimensional space that has eigenvalues in the set $\{-1, +1\}$. After an appropriate change of basis, we may make the assumption that

$$M_j^{(0)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad M_j^{(1)} = \begin{bmatrix} 0 & e^{i\theta_j} \\ e^{-i\theta_j} & 0 \end{bmatrix} \tag{14}$$

for some $\theta_0, \ldots, \theta_n \in [-\pi, \pi]$.

The score for this quantum strategy is clearly bounded by the operator norm of the operator

$$\mathbf{M} := \sum_{(i_k)} f(i_1, \ldots, i_n) M_1^{(i_1)} \otimes \ldots \otimes M_n^{(i_n)} \tag{15}$$

The operator $\mathbf{M}$ is on a Hilbert space which has basis $\{|a_1 a_2 \ldots a_n\rangle \mid a_i \in \{0,1\}\}$. If we take the elements of this basis in lexicographical order, the resulting matrix expression is a reverse-diagonal matrix:

$$\begin{bmatrix} 0 & 0 & \ldots & 0 & * \\ 0 & 0 & \ldots & * & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & * & \ldots & 0 & 0 \\ * & 0 & \ldots & 0 & 0 \end{bmatrix} \tag{16}$$

The entries along the reverse diagonal are given by the expressions

$$P_f\left(e^{i(-1)^{a_1}\theta_1}, \ldots, e^{i(-1)^{a_n}\theta_n}\right) \tag{17}$$

for $(a_k) \in \{0,1\}^n$.

Using the simple observation that the eigenvalues of any matrix of the form

$$\begin{bmatrix} & & & & & & z_1 \\ & & & & & z_2 & \\ & & & & \iddots & & \\ & & & z_n & & & \\ & & \overline{z_n} & & & & \\ & \iddots & & & & & \\ \overline{z_2} & & & & & & \\ \overline{z_1} & & & & & & \end{bmatrix}, \tag{18}$$

are $\pm|z_1|, \pm|z_2|, \ldots, \pm|z_n|$, we find that the operator norm of $\mathbf{M}$ is

$$\max_{(a_i)\in\{0,1\}^n} \left| P_f\left(e^{i(-1)^{a_1}\theta_1}, \ldots e^{i(-1)^{a_n}\theta_n}\right)\right|. \tag{19}$$

Formula (12) follows. Formula (13) follows also via equality (5). ◀

4 Self-testing

Let f be a binary nonlocal XOR game. Let us say that f is a *self-test* if the following condition holds:

(*) There is a single optimal qubit strategy $(\phi, \{M_j^{(0)}, M_j^{(1)}\}_j)$ such that for any other optimal qubit strategy $(\psi, \{N_j^{(0)}, N_j^{(1)}\}_j)$, there exist unitary matrices $U_j \colon \mathbb{C}^2 \to \mathbb{C}^2$ such that

$$(U_1 \otimes U_2 \otimes \ldots \otimes U_n)\psi = \phi \tag{20}$$

and

$$U_j N_j^{(i)} U_j^\dagger = M_j^{(i)} \tag{21}$$

for all $i \in \{0,1\}$, $j \in \{1, 2, \ldots, n\}$.

▶ Proposition 2. Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a nonlocal binary XOR game. Then f is a self-test if and only if the following two conditions hold:

1. There exists a maximum $(\alpha_0, \ldots, \alpha_n)$ of f such that none of $\alpha_1, \ldots, \alpha_n$ is a multiple of π.
2. Every other maximum of f is congruent modulo 2π to either $(\alpha_0, \ldots, \alpha_n)$ or $(-\alpha_0, \ldots, -\alpha_n)$.

Proof. Suppose that f satisfies both of these conditions. Let

$$\phi = \frac{1}{\sqrt{2}} \left(|00\ldots0\rangle + \frac{P_f(\alpha_1, \ldots, \alpha_n)}{|P_f(\alpha_1, \ldots, \alpha_n)|} |11\ldots1\rangle \right).$$

Suppose that $(\psi, \{\{M_j^{(0)}, M_j^{(1)}\}\}_j)$ is an optimal qubit strategy for f. After a unitary change of basis, we may assume that the operators $M_j^{(i)}$ have the form

$$M_j^{(0)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad M_j^{(1)} = \begin{bmatrix} 0 & e^{i\theta_j} \\ e^{-i\theta_j} & 0 \end{bmatrix}, \tag{22}$$

with $(\theta_j) \in [-\pi, \pi]^n$, and we may make the additional assumption that the vectors $(\alpha_1, \ldots, \alpha_n)$ and $(\theta_1, \ldots, \theta_n)$ lie in the same quadrant. (That is, for every $j \in \{1, 2, \ldots, n\}$, $\theta_j \alpha_j \geq 0$.)

Again we let

$$\mathbf{M} = \sum_{(i_k)} f(i_1, \ldots, i_n) M_1^{(i_1)} \otimes \ldots \otimes M_n^{(i_n)}. \tag{23}$$

Since the chosen strategy is optimal, by formula (19) we must have $(\theta_1, \ldots, \theta_n) = (\alpha_1, \ldots, \alpha_n)$. Moreover, the vector ψ must lie in the eigenspace corresponding to the largest eigenvalue of $\mathbf{M}$. This eigenspace is spanned by ϕ. We conclude that f is a self-test.

It is easy to show that if f fails to satisfy either of the two conditions of the theorem, then there exist multiple optimal strategies for f which are inequivalent. ◀

The reader may note one consequence of this proof: if a binary XOR game f is a self-test, then all optimal qubit-strategies for f use states that are equivalent to the GHZ state $\frac{1}{\sqrt{2}}(|00\ldots0\rangle + |11\ldots1\rangle)$.

5 Robustness

Let us say that two qubit strategies $(\psi, \{\{N_j^{(0)}, N_j^{(1)}\}\}_j)$ and $(\gamma, \{\{S_j^{(0)}, S_j^{(1)}\}\}_j)$ are δ-*close* if

$$\|\psi - \gamma\| \leq \delta \tag{24}$$

and

$$\left\| N_j^{(i)} - S_j^{(i)} \right\| \leq \delta \tag{25}$$

for all $j \in \{1, 2, \ldots, n\}$ and $i \in \{0, 1\}$. Let us say that a binary nonlocal XOR game $f\colon \{0,1\}^n \to \mathbb{R}$ is a *second-order robust self-test* if both condition (*) and the following condition hold:

(**) There exists a constant $C > 0$ such that any qubit strategy whose score is within ϵ of the optimal score is $(C\sqrt{\epsilon})$-close to an optimal qubit strategy.

The next proposition uses the concept of a Hessian matrix. For any twice-differentiable function $G\colon \mathbb{R}^m \to \mathbb{R}$ and any element $\mathbf{c} = (c_1, \ldots, c_m) \in \mathbb{R}^m$, let

$$\mathrm{Hess}_{\mathbf{c}}(G) = \left[\frac{\partial^2 G}{\partial x_i \partial x_j}(\mathbf{c}) \right]_{i,j}. \tag{26}$$

The Hessian matrix can be used to calculate the second derivatives of the function G in any direction. When the function G is such that the Hessians at all of its maxima are nonsingular (meaning that all second-derivatives at maxima are negative) the function has the property that near-maxima tend to lie close to true maxima. This fact is the basis for the following proposition, which is proved in full detail in the supplementary information of [13].

▶ Proposition 3. Let $f\colon \{0,1\}^n \to \mathbb{R}$ be a binary nonlocal XOR game. Then f is a second-order robust self-test if and only if the conditions of Proposition 2 are satisfied and the Hessian matrix of Z_f at $(\alpha_0, \ldots, \alpha_n)$ is nonsingular. ◀

Proof sketch. Let $\mathbb{T}$ be the set of all n-qubit strategies $\left(\psi, \left\{\{M_j^{(0)}, M_j^{(1)}\}\right\}_j\right)$ which are such that the operators $M_j^{(i)}$ have the form

$$M_j^{(0)} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad M_j^{(1)} = \begin{bmatrix} 0 & e^{i\theta_j} \\ e^{-i\theta_j} & 0 \end{bmatrix} \tag{27}$$

$(j = 1, \ldots, n)$ and the state ψ has the form

$$\psi \quad = \quad \frac{1}{\sqrt{2}} \left(|00\ldots0\rangle + e^{i\theta_0} |11\ldots1\rangle \right) \tag{28}$$

with $\theta_j \in [-\pi, \pi]$. Direct calculation shows that the score for such a strategy is given by $Z_f(\theta_0, \ldots, \theta_n)$. The Hessian assumption implies that f is a second-order robust self-test within the class $\mathbb{T}$.

Let $\mathbb{S}$ be the set of all n-qubit strategies $\left(\phi, \left\{\{M_j^{(0)}, M_j^{(1)}\}\right\}_j\right)$ such that the operators $M_j^{(i)}$ have the form (27) and the state ϕ is permitted to be any n-qubit state satisfying $\langle \phi \,|\, 00\ldots0\rangle \geq 0$. Then, it can be shown that there exists a constant $K > 0$ such that any n-qubit strategy in $\mathbb{S}$ which achieves a score of $q_f - \epsilon$ must be $(K\sqrt{\epsilon})$-close to some n-qubit strategy in $\mathbb{T}$ which achieves an equal or higher score. As a consequence, robust self-testing holds within the class $\mathbb{S}$ as well. The proof is then completed by the observation that any qubit strategy is equivalent under local unitary transformations to a strategy in $\mathbb{S}$. ◀

6 Examples

It is easy to show that the function Z_g (4) corresponding to the CHSH game has two maxima: $(-\frac{\pi}{4}, \frac{\pi}{2}, \frac{\pi}{2})$ and $(\frac{\pi}{4}, -\frac{\pi}{2}, -\frac{\pi}{2})$. The Hessian matrices at these maxima are both equal to

$$\left(-\frac{1}{\sqrt{2}}\right) \begin{bmatrix} 4 & 2 & 2 \\ 2 & 2 & 1 \\ 2 & 1 & 2 \end{bmatrix}, \tag{29}$$

which is a nonsingular matrix. Therefore, the CHSH game is a second-order robust self-test.

Let d be the 3-player GHZ game:

$$\begin{aligned} Z_d(\theta_0, \theta_1, \theta_2, \theta_3) \quad = \quad & \cos(\theta_0) - \cos(\theta_0 + \theta_1 + \theta_2) \\ & - \cos(\theta_0 + \theta_2 + \theta_3) - \cos(\theta_0 + \theta_1 + \theta_3). \end{aligned}$$

It easy to show that the maxima of this function are $(0, \pm\frac{\pi}{2}, \pm\frac{\pi}{2}, \pm\frac{\pi}{2})$, and that the Hessian matrices at these maxima are nonsingular. Therefore the GHZ game is also a self-test that satisfies second-order robustness. (This fact can also be proved using the results on self-testing graph states in [10].)

Let us see how Proposition 3 can be used to prove new results. The recent paper [2] by Acin *et al.* considers a family of nonlocal games $\{h_\alpha \colon \{0,1\}^2 \to \mathbb{R}\}_{\alpha>1}$ defined by

$$\begin{aligned} h_\alpha(0,0) &= \alpha \quad & h_\alpha(0,1) &= \alpha \\ h_\alpha(1,0) &= 1 \quad & h_\alpha(1,1) &= -1. \end{aligned} \tag{30}$$

The authors characterize the qubit-devices that achieve an optimal score at these games, and show that these devices achieve more randomness than optimal devices for the standard CHSH inequality. The games h_α may therefore be suitable for randomness expansion. However in randomness expansion protocols, it is only possible to approximately determine the expected score of a device. Thus it is important to ask whether the games from this family satisfy robust self-testing.

With the aid of the theory in [2], one can show that the function $Z_{h_\alpha}(\theta_0, \theta_1, \theta_2)$ has two maxima in $[-\pi, \pi]^3$, and the Hessian matrices at these maxima are

$$-(1+\alpha^2)^{-1/2} \begin{bmatrix} 2\alpha^2+2 & \alpha^2+1 & 2 \\ \alpha^2+1 & \alpha^2+1 & 1 \\ 2 & 1 & 2 \end{bmatrix} \tag{31}$$

which is nonsingular for any $\alpha > 1$. Therefore, each of the games in the family $\{h_\alpha\}_{\alpha>1}$ is a second-order robust self-test.

7 General quantum strategies

Now suppose that we consider quantum strategies of arbitrary finite dimension. Whenever there are two Hermitian operators $M^{(0)}, M^{(1)}$ on a single finite-dimensional Hilbert space $\mathcal{Q}$, each having eigenvalues in the set $\{-1, 1\}$, there exists a decomposition

$$\mathcal{Q} = \bigoplus_{\ell=1}^{m} \mathcal{Q}_\ell \tag{32}$$

which is respected by both of the operators $M^{(0)}, M^{(1)}$, with $\dim \mathcal{Q}_\ell \leq 2$. This allows us to reduce general quantum strategies to n-qubit strategies. In particular, this implies that for any binary nonlocal XOR game f, the maximum score achievable by qubit strategies (q_f) is the maximum score achievable by any quantum strategy.

The following generalization of Proposition 3 follows from the above decomposition. (See [13].)

▶ Proposition 4. Let $f \colon \{0,1\}^n \to \mathbb{R}$ be a binary nonlocal XOR game which satisfies the conditions of Proposition 2 and, additionally, satisfies the condition that the Hessian matrices of the maxima of Z_f are all nonsingular. Then, there exists a constant $K > 0$ and an n-qubit state $\chi \in \left(\mathbb{C}^2\right)^{\otimes n}$ such that the following holds: for any quantum strategy

$$\Phi \in \mathcal{Q}_1 \otimes \ldots \otimes \mathcal{Q}_n \tag{33}$$

$$M_j^{(i)} \colon \mathcal{Q}_j \to \mathcal{Q}_j \tag{34}$$

achieving a score of $q_f - \epsilon$, there exist unitary embeddings $U_j \colon \mathcal{Q}_j \to \mathbb{C}^2 \otimes \mathcal{Q}'_j$ and a vector $\Gamma \in \mathcal{Q}'_1 \otimes \ldots \otimes \mathcal{Q}'_n$ such that

$$\left\| (U_1 \otimes \cdots \otimes U_n)\, \Phi - \chi \otimes \Gamma \right\| \leq K\sqrt{\epsilon}. \quad \blacktriangleleft \tag{35}$$

As in the 2-dimensional case, we can take the state χ to be the n-qubit GHZ state.

8 Conclusion

We have provided some general results which allow for easy proofs of robust self-testing in the context of nonlocal binary XOR games. A natural question is whether our results could be generalized to a larger class of games. A possible next step would be to consider games in which the score is based on the XOR of a subset of the outputs (as in the tests used [10]). It would also be interesting to explore further applications to randomness expansion.

Acknowledgements. The authors thank Ryan Landay and Evan Noon for discussions that helped with the development of this material. This research was supported in part by the National Basic Research Program of China under Awards 2011CBA00300 and 2011CBA00301, and the NSF of the United States under Awards 1017335 and 1216729.

References

1 A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani. Device-independent security of quantum cryptography against collective attacks. *Phys. Rev. Lett.*, 98:230501, Jun 2007.

2 A. Acín, S. Massar, and S. Pironio. Randomness versus nonlocality and entanglement. *Physical Review Letters*, 108:100402, 2012.

3 J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Phys. Rev. Lett.*, 23:880–884, October 1969.

4 R. Colbeck. *Quantum And Relativistic Protocols For Secure Multi-Party Computation*. PhD thesis, University of Cambridge, 2006. arXiv:0911.3814.

5 R. Colbeck and A. Kent. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical*, 44(9):095305, 2011.

6 E. Hänggi. *Device-independent quantum key distribution*. PhD thesis, ETH Zurich, December 17 2010. arXiv:1012.3878.

7 F. Magniez, D. Mayers, M. Mosca, and H. Ollivier. Self-testing of quantum circuits. In Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener, editors, *Automata, Languages and Programming, 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part I*, volume 4051 of *Lecture Notes in Computer Science*, pages 72–83. Springer, 2006.

8 D. Mayers and A. Yao. Quantum cryptography with imperfect apparatus. *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, pages 503–509, 1998.

9 M. McKague. Device independent quantum key distribution secure against coherent attacks with memoryless measurement devices. *New Journal of Physics*, 11(10):103037, 2009.

10 M. McKague. Self-testing graph states. arXiv:1010.1989, 2010.

11 M. McKague, T. Z. Yang, and V. Scarani. Robust self-testing of the singlet. arXiv:1203.2976v1, 2012.

12 C. Miller and Y. Shi. Randomness expansion from the Greenberger-Horne-Zeilinger paradox, 2011. Attached as an appendix to arXiv:1207.1819v4.

13 C. Miller and Y. Shi. Optimal robust quantum self-testing by binary nonlocal XOR games. arXiv:1207.1819v4, 2012.

14 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, and C. Monroe. Random numbers certified by Bell's theorem. *Nature*, 464:1021–1024, 2010.

15 S. Pironio, A. Acín, N. Brunner, N. Gisin, S. Massar, and V. Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009.

16 S. Popescu and D. Rohrlich. Which states violate Bell's inequality maximally? *Physics Letters A*, 169(6):411–414, 1992.

17 B. Reichardt, F. Unger, and U. Vazirani. A classical leash for a quantum system: Command of quantum systems via rigidity of CHSH games. arXiv:1209.0448, 2012.

18 U. Vazirani and T. Vidick. Certifiable quantum dice. *Phil. Trans. R. Soc. A*, 370:3432, 7 2012.

19 R. Werner and M. Wolf. All-multipartite Bell-correlation inequalities for two dichotomic observables per site. *Physical Review A*, 64(3), 2001.

Exact Quantum Query Complexity of EXACT and THRESHOLD*

Andris Ambainis, Jānis Iraids, and Juris Smotrovs

Faculty of Computing, University of Latvia,
Raiņa bulvāris 19, Rīga, LV-1586, Latvia
andris.ambainis@lu.lv, janis.iraids@gmail.com, juris.smotrovs@lu.lv

Abstract

A quantum algorithm is *exact* if it always produces the correct answer, on any input. Coming up with exact quantum algorithms that substantially outperform the best classical algorithm has been a quite challenging task.

In this paper, we present two new exact quantum algorithms for natural problems:

- for the problem EXACT_k^n in which we have to determine whether the sequence of input bits $x_1, \ldots, x_n$ contains exactly k values $x_i = 1$;
- for the problem THRESHOLD_k^n in which we have to determine if at least k of n input bits are equal to 1.

1998 ACM Subject Classification F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Quantum query algorithms, Complexity of Boolean functions

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.263

1 Introduction

We consider quantum algorithms in the query model. The algorithm needs to compute a given Boolean function $f : \{0,1\}^n \to \{0,1\}$ by querying its input bits until it is able to produce the value of the function, either with certainty, or with some error probability. The complexity of the algorithm is measured as the number of queries it makes (other kinds of computation needed to produce the answer are disregarded).

In the *bounded error* setting where the algorithm is allowed to give an incorrect answer with probability not exceeding a given constant ϵ, $0 < \epsilon < \frac{1}{2}$, many efficient quantum algorithms are known, with either a polynomial speed-up over classical algorithms (e.g., [12, 1, 9, 16, 4]), or, in the case of partial functions, even an exponential speed-up (e.g., [18, 17]).

Less studied is the *exact* setting where the algorithm must give the correct answer with certainty. Though for partial functions quantum algorithms with exponential speed-up are known (for instance, [8, 5]), the results for total functions up to recently have been much less spectacular: the best known quantum speed-up was just by a factor of 2.

Even more, as remarked in [13], all the known algorithms achieved this speed-up by the same trick: exploiting the fact that XOR of two bits can be computed quantumly with one query, while a classical algorithm needs two queries [8, 7, 10].

A step forward was made by [13] which presented a new algorithm achieving the speed-up by a factor of 2, without using the "XOR trick". The algorithm is for the Boolean function

* This research was supported by EU FP7 projects QCS, QALGO and MQC. The presentation of the paper at the conference was supported by ERDF Project No. 2010/0202/2DP/2.1.1.2.0/10/APIA/VIAA/013.

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 263–269

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

EXACT_2^4 which is true iff exactly 2 of its 4 input bits are equal to 1. It computes this function with 2 queries, while a classical (deterministic) algorithm needs 4 queries.

This function can be generalized to EXACT_k^n in the obvious way. Its deterministic complexity is n (due to its sensitivity being n, see [15]). [13] conjectured that its quantum query complexity is $\max\{k, n-k\}$.

In this paper we prove the conjecture. We also solve the problem for a similar function, THRESHOLD_k^n which is true iff *at least* k of the input bits are equal to 1. When $n = 2k-1$, this function is well-known as the MAJORITY function. The quantum query complexity of THRESHOLD_k^n turns out to be $\max\{k, n-k+1\}$, as conjectured in [13].

In a recent work [2], a function $f(x_1, \ldots, x_n)$ with the deterministic query complexity n and the exact quantum query complexity $O(n^{.8675\ldots})$ was constructed. The quantum advantage that is achieved by our algorithms is smaller but we think that our results are still interesting, for several reasons.

First, we present quantum algorithms for computational problems that are natural and simple to describe. Second, our algorithms contain new ideas which may be useful for designing other exact algorithms. Currently, the toolbox of ideas for designing exact quantum algorithms is still quite small. Expanding it is an interesting research topic.

2 Technical Preliminaries

We denote $[m] = \{1, 2, \ldots, m\}$. We assume familiarity with basics of quantum computation [14]. We now briefly describe the quantum query algorithm model.

Let $f : \{0,1\}^n \to \{0,1\}$ be the Boolean function to compute, with the input bit string $x = x_1 x_2 \ldots x_n$. The quantum query algorithm works in a Hilbert space with some fixed basis states. It starts in a fixed starting state, then performs on it a sequence of unitary transformations U_1, Q, U_2, Q, $\ldots$, U_t, Q, U_{t+1}. The unitary transformations U_i do not depend on the input bits, while Q, called the *query transformation*, does, in the following way. Each of the basis states corresponds to either one or none of the input bits. If the basis state $|\psi\rangle$ corresponds to the i-th input bit, then $Q|\psi\rangle = (-1)^{x_i}|\psi\rangle$. If it does not correspond to any input bit, then Q leaves it unchanged: $Q|\psi\rangle = |\psi\rangle$. For convenience in computations, we denote $\hat{x}_i = (-1)^{x_i}$.

Finally, the algorithm performs a full measurement in the standard basis. Depending on the result of the measurement, it outputs either 0 or 1 which must be equal to $f(x)$.

By the principle of delayed measurement, sometimes a measurement performed in the middle of computation is equivalent to it being performed at the end of computation [14]. We will use that in our algorithms, because they are most easily described as recursive algorithms with the following structure: perform unitary U_1, query Q, unitary U_2, then measure; depending on the result of measurement, call a smaller (by 2 input bits) instance of the algorithm. The principle of delayed measurement ensures that such recursive algorithm can be transformed by routine techniques into the commonly used query algorithm model described above.

The minimum number of queries made by any quantum algorithm computing f is denoted by $Q_E(f)$. We use $D(f)$ to denote the minimum number of queries used by a deterministic algorithm that computes f.

3 Algorithm for EXACT

▶ **Definition 1.** The function $EXACT_k^n$ is a Boolean function of n variables being true iff *exactly* k of the variables are equal to 1.

▶ **Theorem 2.**

$$Q_E(EXACT_k^{2k}) \leq k$$

Proof. We present a recursive algorithm. When $k = 0$ the algorithm returns 1 without making any queries. Suppose $k = m$. For the recursive step we will use basis states $|0\rangle$, $|1\rangle$, $\ldots$, $|n\rangle$ and $|i,j\rangle$ with $i,j \in [2m]$, $i < j$. The i-th input bit will be queried from the state $|i\rangle$. We begin in the state $|0\rangle$ and perform a unitary transformation U_1:

$$U_1 |0\rangle \to \sum_{i=1}^{2m} \frac{1}{\sqrt{2m}} |i\rangle.$$

Next we perform a query:

$$\sum_{i=1}^{2m} \frac{1}{\sqrt{2m}} |i\rangle \xrightarrow{Q} \sum_{i=1}^{2m} \frac{\hat{x}_i}{\sqrt{2m}} |i\rangle.$$

Finally, we perform a unitary transformation U_2, such that

$$U_2 |i\rangle = \sum_{j>i} \frac{1}{\sqrt{2m}} |i,j\rangle - \sum_{j<i} \frac{1}{\sqrt{2m}} |j,i\rangle + \frac{1}{\sqrt{2m}} |0\rangle$$

One can verify that such a unitary transformation exists by checking the inner products:

1) for any $i \in [2m]$,

$$\langle i| U_2^\dagger U_2 |i\rangle = \sum_{j>i} \frac{1}{2m} + \sum_{j<i} \frac{1}{2m} + \frac{1}{2m} = 1.$$

2) for any $i,j \in [2m]$, $i \neq j$,

$$\langle j| U_2^\dagger U_2 |i\rangle = \left(\sum_{l>j} \frac{1}{\sqrt{2m}} \langle j,l| - \sum_{l<j} \frac{1}{\sqrt{2m}} \langle l,j| + \frac{1}{\sqrt{2m}} \langle 0| \right) \cdot \left(\sum_{l>i} \frac{1}{\sqrt{2m}} |i,l\rangle - \sum_{l<i} \frac{1}{\sqrt{2m}} |l,i\rangle + \frac{1}{\sqrt{2m}} |0\rangle \right) = 0$$

The resulting quantum state is

$$\sum_{i=1}^{2m} \frac{\hat{x}_i}{\sqrt{2m}} |i\rangle \xrightarrow{U_2} \sum_{i=1}^{2m} \frac{\hat{x}_i}{2m} |0\rangle + \sum_{i<j} \frac{\hat{x}_i - \hat{x}_j}{2m} |i,j\rangle.$$

If we measure the state and get $|0\rangle$, then $EXACT_m^{2m}(x) = 0$. If on the other hand we get $|i,j\rangle$, then $x_i \neq x_j$ and $EXACT_m^{2m}(x) = EXACT_{m-1}^{2m-2}(x \setminus \{x_i, x_j\})$, therefore we can use our algorithm for $EXACT_{m-1}^{2m-2}$.

◀

Note that we can delay the measurements by using $|i,j\rangle$ as a starting state for the recursive call of the algorithm.

For the sake of completeness, we include the following corollary already given in [13]:

▶ **Corollary 3.** *[13]*

$$Q_E(EXACT_k^n) \leq \max\{k, n-k\}$$

Proof. Assume that $k < \frac{n}{2}$. The other case is symmetric. Then we append the input x with $n-2k$ ones producing x' and call $EXACT_{n-k}^{2n-2k}(x')$. Then concluding that there are $n-k$ ones in x' is equivalent to there being $(n-k)-(n-2k) = k$ ones in the original input x. ◀

The lower bound can be established by the following fact:

▶ **Proposition 4.** *If g is a partial function such that $g(x) = f(x)$ whenever g is defined on x, then $Q_E(g) \leq Q_E(f)$.*

▶ **Proposition 5.**

$$Q_E(EXACT_k^n) \geq \max\{k, n-k\}$$

Proof. Assume that $k \leq \frac{n}{2}$. The other case is symmetric. Define

$$g(x_{k+1}, \ldots, x_n) = EXACT_k^n(1, \ldots, 1, x_{k+1}, \ldots, x_n).$$

Observe that g is in fact negation of the OR function on $n-k$ bits which we know [3] to take $n-k$ queries to compute. Therefore by virtue of Proposition 4 no algorithm for $EXACT_k^n$ may use less than $n-k$ queries. ◀

4 Algorithm for THRESHOLD

We will abbreviate THRESHOLD as Th.

▶ **Definition 6.** The function Th_k^n is a Boolean function of n variables being true iff *at least* k of the variables are equal to 1.

The function Th_{k+1}^{2k+1} is commonly referred to as MAJ_{2k+1} or $MAJORITY_{2k+1}$ because it is equal to the majority of values of input variables.

Remarkably an approach similar to the one used for $EXACT$ works in this case as well.

▶ **Theorem 7.**

$$Q_E(MAJ_{2k+1}) \leq k+1.$$

Proof. Again, a recursive solution is constructed as follows. The base case $k = 0$ is trivial to perform with one query, because the function returns the value of the single variable. The recursive step $k = m$ shares the states, unitary transformation U_1 and the query with our algorithm for $EXACT$, but the unitary U_2 is slightly different:

$$U_1|0\rangle \to \sum_{i=1}^{2m+1} \frac{1}{\sqrt{2m+1}}|i\rangle.$$

$$\sum_{i=1}^{2m+1} \frac{1}{\sqrt{2m+1}}|i\rangle \xrightarrow{Q} \sum_{i=1}^{2m+1} \frac{\hat{x}_i}{\sqrt{2m+1}}|i\rangle.$$

$$U_2 |i\rangle = \sum_{j>i} \frac{\sqrt{2m-1}}{2m} |i,j\rangle - \sum_{j<i} \frac{\sqrt{2m-1}}{2m} |j,i\rangle + \sum_{j\neq i} \frac{1}{2m} |j\rangle.$$

The resulting state is

$$\sum_{i=1}^{2m+1} \frac{\hat{x}_i}{\sqrt{2m+1}} |i\rangle \xrightarrow{U_2} \sum_{i=1}^{2m+1} \sum_{j\neq i} \frac{\hat{x}_j}{2m\sqrt{2m+1}} |i\rangle + \sum_{i<j} \frac{(\hat{x}_i - \hat{x}_j)\sqrt{2m-1}}{2m\sqrt{2m+1}} |i,j\rangle.$$

We perform a complete measurement. There are two kinds of outcomes:

1) If we get state $|i\rangle$, then either
   a) x_i is the value in the majority which according to the polynomial $\sum_{j\neq i} \hat{x}_j$ not being zero implies that in $x \setminus \{x_i\}$ the number of ones is greater than the number of zeroes by at least 2; or
   b) x_i is a value in the minority.

   In both of these cases, for all $j : j \neq i$ it is true that $MAJ_{2m+1}(x) = MAJ_{2m-1}(x \setminus \{x_i, x_j\})$. Therefore, we can solve both cases by removing x_i and one other arbitrary input value and calculating majority from the remaining values.
2) If we get state $|i,j\rangle$, then it is even better: we know that $x_i \neq x_j$ and therefore $MAJ_{2m+1}(x) = MAJ_{2m-1}(x \setminus \{x_i, x_j\})$.

◀

▶ **Corollary 8.** *If $0 < k < n$, then*

$$Q_E(Th_k^n) \leq \max\{k, n-k+1\}.$$

Proof. Assume that $k \leq \frac{n}{2}$. The other case is symmetric. Then we append the input x with $n-2k+1$ ones producing x' and call $MAJ_{2n-2k+1}(x')$. Then x' containing at least $n-k+1$ ones is equivalent to x containing at least $(n-k+1)-(n-2k+1) = k$ ones. ◀

▶ **Proposition 9.**

$$Q_E(Th_k^n) \geq \max\{k, n-k+1\}$$

Proof. Assume that $k \leq \frac{n}{2}$. The other case is symmetric. Define

$$g(x_k, x_{k+1}, \ldots, x_n) = Th_k^n(1, \ldots, 1, x_k, x_{k+1}, \ldots, x_n).$$

Observe that g is in fact the OR function on $n-k+1$ bits which we know [3] takes $n-k+1$ queries to compute. Therefore by virtue of Proposition 4 no algorithm for Th_k^n may use less than $n-k+1$ queries. ◀

5 Conclusion

Coming up with exact quantum algorithms that are substantially better than any classical algorithm has been a difficult open problem. Until a few months ago, no example of total Boolean function with $Q_E(f) < D(f)/2$ was known and the examples of functions with $Q_E(f) = D(f)/2$ were almost all based on one idea: applying 1-query quantum algorithm for $x_1 \oplus x_2$ as a subroutine.

The first exact quantum algorithm with $Q_E(f) < D(f)/2$ (for a total f) was constructed in [2]. However, no symmetric function with $Q_E(f) < D(f)/2$ is known. It has been proven

that if $f(x)$ is a symmetric, non-constant function of n variables, then $Q_E(f) \geq n/2 - o(n)$ [11, 6].

In this paper, we construct exact quantum algorithms for two symmetric functions: *EXACT* and *THRESHOLD*. Both of those algorithms achieve $Q_E(f) = D(f)/2$ (exactly or in the limit) and use new ideas. At the same time, our algorithms are quite simple and easy to understand.

The main open problem is to come with more algorithmic techniques for constructing exact quantum algorithms. Computer experiments via semidefinite optimization [13] show that there are many functions for which exact quantum algorithms are better than deterministic algorithms. Yet, in many of these cases, the only way to construct these algorithms is by searching the space of all quantum algorithms, using semidefinite optimization as the search tool.

For example, from the calculations in [13] (based on semidefinite optimization) it is apparent that there are 3 symmetric functions of 6 variables for which $Q_E(f) = 3$: $PARITY$, $EXACT_3^6$ and $EXACT_{2,4}^6$ (exactly 2 or 4 of 6 variables are equal to 1).

Unlike for the first two functions, we are not aware of any simple quantum algorithm or lower bounds for $EXACT_{2,4}^6$. Based on the evidence from semidefinite optimization, we conjecture that if n is even and $2k < n$ then the quantum query complexity of $EXACT_{k,n-k}^n$ is $n-k-1$. In particular, this would mean that the complexity of $EXACT_{n/2-1,n/2+1}^n$ is $\frac{n}{2}$ and this function also achieves a gap of $Q_E(f) = D(f)/2$.

At the moment, we know that this conjecture is true for $k = 0$ and $k = 1$. Actually, both of those cases can be solved by a classical algorithm which uses the 1-query algorithm for $x_1 \oplus x_2$ as a quantum subroutine. This approach fails for $k \geq 1$ and it seems that the approach in the current paper is also not sufficient — without a substantial new component.

References

1 A. Ambainis. Quantum walk algorithm for element distinctness. *SIAM Journal on Computing*, 37(1): 210-239, 2007. Also FOCS'04 and quant-ph/0311001.

2 A. Ambainis: Superlinear advantage for exact quantum algorithms. In *Proceedings of 45th ACM STOC*, pages 891–900, 2013. Also arXiv:1211.0721.

3 R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. Earlier version in FOCS'98. Also arXiv:9802049.

4 A. Belovs. Span programs for functions with constant-sized 1-certificates. In *Proceedings of 44th ACM STOC*, pages 77–84, 2012. Also arXiv:1105.4024.

5 G. Brassard and P. Høyer. An exact quantum polynomial-time algorithm for Simon's problem. *Proceedings of the Israeli Symposium on Theory of Computing and Systems (ISTCS)*, pages 12–23, 1997. Also arXiv:9704027.

6 H. Buhrman and R. de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002.

7 R. Cleve, A. Eckert, C. Macchiavello, and M. Mosca. Quantum algorithms revisited. *Proceedings of the Royal Society of London*, volume A454, pages 339–354, 1998. Also arXiv:9708016.

8 D. Deutsch and R. Jozsa. Rapid solution of problems by quantum computation. In *Proceedings of the Royal Society of London*, volume A439, pages 553–558, 1992.

9 E. Farhi, J. Goldstone, S. Gutman, A Quantum Algorithm for the Hamiltonian NAND Tree. *Theory of Computing*, 4:169-190, 2008. Also quant-ph/0702144.

10 E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser. A limit on the speed of quantum computation in determining parity. *Physical Review Letters*, 81(5):5442–5444, 1998. Also arXiv:9802045.

11 J. von zur Gathen and J. R. Roche. Polynomials with two values. *Combinatorica*, 17(3):345–362, 1997.

12 L. K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of 28th ACM STOC*, pages 212–219, 1996. Also arXiv:9605043.

13 A. Montanaro, R. Jozsa, G. Mitchison. On exact quantum query complexity. arXiv preprint arXiv:1111.0475 (2011)

14 M. Nielsen and I. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

15 N. Nisan and M. Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994. Earlier version in STOC'92.

16 B. Reichardt, R. Špalek. Span-program-based quantum algorithm for evaluating formulas. *Proceedings of STOC'08*, pp. 103-112. Also arXiv:0710.2630.

17 P. W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. Earlier version in FOCS'94. Also arXiv:9508027.

18 D. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. Earlier version in FOCS'94.

The Quantum Entropy Cone of Stabiliser States

Noah Linden[1], František Matúš[2], Mary Beth Ruskai[3,4], and Andreas Winter[5,1,6]

1 **School of Mathematics, University of Bristol**
Bristol BS8 1TW, United Kingdom
n.linden@bristol.ac.uk

2 **Institute of Information Theory and Automation**
Academy of Sciences of the Czech Republic
Prague, Czech Republic
matus@utia.cas.cz

3 **Institute for Quantum Computing, University of Waterloo**
Waterloo, Ontario, Canada
mbruskai@gmail.com

4 **Tufts University, Medford, MA 02155 USA**
marybeth.ruskai@tufts.edu

5 **ICREA & Física Teòrica: Informació i Fenomens Quàntics**
Universitat Autònoma de Barcelona
ES-08193 Bellaterra (Barcelona), Spain
andreas.winter@uab.cat

6 **Centre for Quantum Technologies, National University of Singapore**
2 Science Drive 3, Singapore 117542, Singapore

Abstract

We investigate the universal linear inequalities that hold for the von Neumann entropies in a multi-party system, prepared in a stabiliser state. We demonstrate here that entropy vectors for stabiliser states satisfy, in addition to the classic inequalities, a type of linear rank inequalities associated with the combinatorial structure of normal subgroups of certain matrix groups.

In the 4-party case, there is only one such inequality, the so-called Ingleton inequality. For these systems we show that strong subadditivity, weak monotonicity and Ingleton inequality exactly characterize the entropy cone for stabiliser states.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases Entropy inequalities, Stabiliser states, Ingleton inequality

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.270

1 Introduction

1.1 Background

Undoubtedly, the single most important quantity in (classical) information theory is the Shannon entropy, and its properties play a key role: for a discrete probability distribution p on $\mathcal{T}$

$$H(p) = -\sum_{t\in\mathcal{T}} p(t)\log p(t)\ . \tag{1}$$

The quantum (von Neumann) entropy is understood to be of equal importance to quantum information: for a quantum state (density operator) $\rho \geq 0$, $\operatorname{Tr}\rho = 1$

$$S(\rho) = -\operatorname{Tr}\rho\log\rho \tag{2}$$

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 270–284


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

which reduces to (1) when ρ is diagonal.

For N-party systems, one can apply these definitions to obtain the entropy of all marginal probability distributions (in the classical case) and reduced density operators (aka quantum marginals) in the quantum case. The collection of these entropies can be regarded as a vector in $\mathbf{R}_{2^N}$, and the collection of all such vectors forms a set whose closure is a convex cone. It is an interesting open question to determine the inequalities which characterize this cone. As discussed in Section 1.3, it is now known that in the classical setting the Shannon inequalities given below do not suffice; they describe a strictly larger cone.

This work has motivated us to consider analogous questions for the von Neumann entropy in N-party quantum systems. Although we are unable to answer this question, we can fully characterize the cone associated with a subset of quantum states known as stabiliser states in the 4-party case. Moreover, we can show that for any number of parties, entropy vectors for stabiliser states satisfy additional inequalities in the class known as linear rank inequalities discussed in Section 3. In the classical setting, distributions whose entropies satisfy this subclass of stronger inequalities, suffice to achieve maximum throughput in certain network coding problems [28].

1.2 Classic inequalities and Definitions

It is well-known that the classical Shannon entropy for an N-party classical probability distribution p on a discrete space $\mathcal{T}_1 \times \cdots \times \mathcal{T}_N$, has the following properties, commonly known as the *Shannon inequalities*:

1. It is non-negative, *i.e.* $H(A) \geq 0$; $H(\emptyset) = 0$. (+)
2. It is strongly subadditive (aka submodular), *i.e.*

$$H(A) + H(B) - H(A \cap B) - H(A \cup B) \geq 0. \tag{SSA}$$

3. It is monotone non-decreasing, *i.e.*

$$A \subset B \quad \Longrightarrow \quad H(A) \leq H(B). \tag{MO}$$

where $H(A)$ denotes the entropy $H(p_A)$ of the marginal distribution p_A on $\mathcal{T}_A = \bigotimes_{j \in A} \mathcal{T}_j$.

The monotonicity property (MO) implies that if $H(A) = 0$ then $H(B) = 0$ for all $B \subset A$ and, thus, $p_A = \bigotimes_{j \in A} \delta_{t_j}$ is a product of point masses. Some of the most remarkable features of quantum systems arise when (MO) is violated. Indeed, for a pure entangled state $\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$ for which $S(\rho_{AB}) = 0$, but the entropy of the reduced states $\rho_A = \mathrm{Tr}_B\, \rho_{AB}$ and $\rho_B = \mathrm{Tr}_A\, \rho_{AB}$ can be (and usually is) strictly positive. In fact, $S(\rho_{AB}) - S(\rho_A)$ can be as large as $-\log d$, where d is the Hilbert space dimension of the smaller of A and B.

For multi-party quantum systems, (+) and (SSA) are still valid [29], but (MO) has to be replaced by the third property below – in analogy to the classical case, we call them Shannon inequalities:

1. Non-negativity: $S(A) \geq 0$; $S(\emptyset) = 0$. (+)
2. Strong subadditivity:

$$S(A) + S(B) - S(A \cap B) - S(A \cup B) \geq 0. \tag{SSA}$$

3. Weak monotonicity:

$$S(A) + S(B) - S(A \setminus B) - S(B \setminus A) \geq 0. \tag{WMO}$$

However, in contrast to the classical setting, this weaker version of monotonicity is not completely independent of strong subadditivity (SSA). In fact, it can be obtained from the latter by the (non-linear) process known as *purification* described in Section 2.2. Using a slight abuse of notation, we use $I(A:B)$ and $I(A:B|C)$ to denote, respectively, the mutual information and conditional mutual information for both classical and quantum systems, defined explicitly in the latter case as

$$\begin{aligned} I(A:B) &= S(A) + S(B) - S(AB), \\ I(A:B|C) &= S(AC) + S(BC) - S(ABC) - S(C), \end{aligned}$$

for pairwise disjoint sets A, B, C. Note that SSA can then be written as $I(A:B|C) \geq 0$.

1.3 Entropy cones and non-Shannon inequalities

The first non-Shannon entropy inequality was obtained in 1997-98 by Yeung and Zhang [43, 44, 45] for 4-party systems. Their work established that the classical entropy cone is strictly smaller than the polyhedral cone defined by the Shannon inequalities. This was the only non-Shannon inequality known until 2006, when Dougherty, Freiling and Zeger [12, 13] used a computer search to generate new inequalities. Then Matúš [34] found two infinite families, one of which can be written as

$$t\,\mathrm{Ing}(AB:CD) + I(A:B|D) + \frac{t(t+1)}{2}\big[I(B:D|C) + I(C:D|B))\big] \quad \geq \quad 0 \tag{3}$$

where t is a non-negative integer, and $\mathrm{Ing}(AB:CD)$ is defined in (ING) below. The case $t = 1$ in (3) yields the inequality in [45]. Moreover, either of the Matúš families can be used to show that the 4-party entropy cone is not polyhedral. In [15] additional non-Shannon inequalities were found.

In the quantum setting, Lieb [30] considered the question of additional inequalities in a form that could be regarded as extending SSA to more parties, but found none. Much later Pippenger [39] rediscovered one of Lieb's results and used it to show constructively that there are no additional inequalities for 3-party systems. He also explicitly raised the question of whether or not additional inequalities hold for more parties. Despite the fact that (SSA) is still the only known inequality, it has been shown that for 4-party systems there are constrained inequalities [4, 31] that do not follow from SSA. (Numerical evidence for additional inequalities is given in the thesis of Ibinson [21].)

1.4 Structure of the paper

This paper is organized as follows. In Section 2 we give some basic notation and review some well-known facts. In Section 3 we discuss what is known about linear rank inequalities beginning with the Ingleton inequality in Section 3.1 and concluding with a discussion of their connection to the notion of common information in Section 3.3. In Section 4 we discuss stabiliser states, beginning with some basic definitions in Section 4.1. In Section 4.2 we consider the entropies of stabiliser states, showing half of our main result that pure stabiliser states generate entropy vectors which satisfy the Ingleton inequality and a large class of other linear rank inequalities. In Section 5 we prove the other half, *i.e.*, that all extremal rays of the 4-party Ingleton cone can be achieved using 5-party stabiliser states. We conclude with some open questions and challenges.

2 Preliminaries

2.1 Notation

We now introduce some notation needed to make precise the notion of entropy vectors and entropy cones. We will let $\mathcal{X} = \{A, B, C, \ldots\}$ denote an index set of finite size $|\mathcal{X}| = N$ so that in many cases we could just assume that $\mathcal{X} = \{1, 2, \ldots N\}$. However, it will occasionally be useful to consider the partition of some the index set into smaller groups, e.g, by grouping A and B as well as D and E, $\mathcal{X}_5 = \{A, B, C, D, E\}$ gives rise to a 3-element $\mathcal{X}_3 = \{AB, C, DE\}$. When the size of $\mathcal{X}$ is important, we write $\mathcal{X}_N$.

An arbitrary N-partite quantum system is associated with a Hilbert space $\mathcal{H} = \bigotimes_{x\in\mathcal{X}} \mathcal{H}_x$ (with no restrictions on the dimension of the Hilbert spaces $\mathcal{H}_x$) with $|\mathcal{X}| = N$. The reduced states (properly called reduced density operators, but more often referred to as reduced density matrices (RDM) and also known as quantum marginals) are given by $\rho_J = \mathrm{Tr}_{J^c}\rho$, where $J^c = \mathcal{X} \setminus J$. This gives rise to a function $S : J \longmapsto S(J) = S(\rho_J)$ on the subsets $J \subset \mathcal{X}$. An element of the output of S can be viewed as a vector in $\mathbf{R}^{2^\mathcal{X}}$, whose coordinates are indexed by the power set $2^\mathcal{X}$ of $\mathcal{X}$. We study the question of which such vectors arise from classical or quantum states, *i.e.*, when their elements are given by the entropies $S(\rho_J)$ of the reduced states of some N-party quantum state.

Classical probability distributions can be embedded into the quantum framework by restricting density matrices to those which are diagonal in a fixed product basis. A function $H : 2^\mathcal{X} \to \mathbf{R}$, associating real numbers to the subsets of a finite set $\mathcal{X}$, which satisfies the Shannon inequalities, eqs. (+), (SSA) and (MO), is called *poly-matroid*. By analogy with poly-matroids, we propose to call a function $S : 2^\mathcal{X} \to \mathbf{R}$ a *poly-quantoid*, if it satisfies (+), (SSA) and (WMO) [36].

We will let $\Gamma^C_\mathcal{X}$ and $\Gamma^Q_\mathcal{X}$ denote, respectively, the convex cone of vectors in a poly-matroid or poly-quantoid. The existence of non-Shannon entropy inequalities implies that there are vectors in $\Gamma^C_\mathcal{X}$ which can not be achieved by any classical state. Neither the classical nor quantum set of true entropy vectors is convex, because their boundaries have a complicated structure [4, 31, 35, 39]. However, the closure of the set of classical or quantum entropy vectors, which we denote $\overline{\Sigma}^C_\mathcal{X}$ or $\overline{\Sigma}^Q_\mathcal{X}$, respectively, is a closed convex cone. The inclusion $\overline{\Sigma}^C_\mathcal{X} \subset \Gamma^C_\mathcal{X}$ is strict for $N \geq 4$ [45]. It is an important open question whether or not this also holds in the quantum setting, *i.e.*, is the inclusion $\overline{\Sigma}^Q_\mathcal{X} \subseteq \Gamma^Q_\mathcal{X}$ also strict?

In this paper, we consider entropy vectors which satisfy additional inequalities known as linear rank inequalities, i.e. those satisfied by the dimensions of subspaces of a vector space and their intersections. A poly-matroid H is called *linearly represented* if $H(J) = \dim \sum_{j\in J} V_j$ for subspaces V_j of a common vector space V.

The simplest linear rank inequality is the 4-party Ingleton inequality (see section 3 below). Poly-matroids and poly-quantoids which also satisfy these additional inequalities will be denoted $\Lambda^C_\mathcal{X}$ and $\Lambda^Q_\mathcal{X}$ respectively.

2.2 Purification and complementarity

For statements about J and $J^c = \mathcal{X} \setminus J$, it suffices to consider a bipartite quantum system with Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$. It is well-known that any pure state $|\psi_{AB}\rangle$ can be written in the form

$$|\psi_{AB}\rangle = \sum_k \mu_k |\phi^A_k\rangle \otimes |\phi^B_k\rangle \tag{4}$$

with $\mu_k > 0$ and $\{\phi_k^A\}$ and $\{\phi_k^B\}$ orthonormal. Indeed, this is an immediate consequence of the isomorphism between $\mathcal{H}_A \otimes \mathcal{H}_B$ and $\mathcal{L}(\mathcal{H}_A, \mathcal{H}_B)$, the set of linear operators from $\mathcal{H}_A$ to $\mathcal{H}_B$, and the singular value decomposition. It then follows that both ρ_A and ρ_B have the same non-zero eigenvalues μ_k^2, and hence $S(\rho_A) = S(\rho_B)$.

This motivates the process known as purification. Given a density matrix $\rho = \sum_k \lambda_k |\phi_k\rangle\langle\phi_k|$, one can define the bipartite state

$$|\psi\rangle = \sum_k \sqrt{\lambda_k}\, |\phi_k\rangle \otimes |\phi_k\rangle$$

whose reduced density matrix $\mathrm{Tr}_B\, |\psi\rangle\langle\psi|$ is ρ.

Therefore, every vector in an N-party quantum entropy cone Σ_N^Q can be obtained from the entropies of some reduced state of a $(N+1)$-party pure state $|\Psi\rangle$. In that case, we say that the entropy vector is realized by $|\Psi\rangle$.

In an abstract setting, we could define a cone $\widetilde{\Gamma}_{\mathcal{X}}^Q$ whose elements satisfy (+), (SSA) and the complementarity property $S(J) = S(J^c)$, and let Γ_N^Q be the cone of vectors which arise as subvectors of $\widetilde{\Gamma}_{N+1}^Q$. Although we will not need this level of abstraction, this correspondence is used in Section 5.

2.3 Group inequalities

Consider a (finite) group G and a family of subgroups $G_x < G$, $x \in \mathcal{X}$. Then, $H(J) = \log |G/G_J|$, with $G_J = \bigcap_{j\in J} G_j$ is a poly-matroid. In fact, Chan and Yeung [9] show that it is entropic because it can be realised by the random variables $X_j = gG_j \in G/G_j$ for a uniformly distributed $g \in G$. The fact that for two subgroups $G_1, G_2 < G$, the mappings

$$G/(G_1 \cap G_2) \longrightarrow G/G_1 \times G/G_2 \qquad \text{and} \qquad g(G_1 \cap G_2) \longmapsto (gG_1, gG_2),$$

are one-to-one [42], guarantees that indeed $H(X_J) = H(J)$.

Thus, the inequalities satisfied by poly-matroids, and more specifically entropic poly-matroids give rise to relations between the cardinalities of subgroups and their intersections in a generic group. Conversely, Chan and Yeung [9] have shown that every such relation for groups, is valid for all entropic poly-matroids. This result motivates the search for a similar, combinatorial or group theoretical, characterization of the von Neumann entropic poly-quantoids, and our interest in stabiliser states originally grew out of it.

However, it must be noted that if some subgroups of G are not general, but, e..g, normal subgroups as in Theorem 6 below, then the Chan-Yeung correspondence breaks down. In this case further inequalities hold for the group poly-matroid that are not satisfied by entropic poly-matroids.

3 Linear rank inequalities

3.1 The Ingleton inequality

The classic *Ingleton inequality*, when stated in information theoretical terms, and as manifestly balanced, reads

$$\mathrm{Ing}(AB:CD) \equiv I(A:B|C) + I(A:B|D) + I(C:D) - I(A:B) \geq 0, \tag{ING}$$

where A, B, C and D are elements (more generally pairwise disjoint subsets) of $\mathcal{X}$. It was discovered by Ingleton [22] as a constraint on linearly represented matroids.

Although this inequality does not hold universally, it is of considerable importance, and continues to be studied [32, 37, 41, 36], particularly when reformulated as an inequality for subgroup ranks. In Theorem 11 we show that (ING) always holds for a special class of states. Before doing that, we give some basic properties first. Observe that (ING) is symmetric with respect to the interchanges $A \leftrightarrow B$ and $C \leftrightarrow D$, so that it suffices to consider special properties only for A and D.

Because it is not always easy to see if a 4-party state ρ_{ABCD} is the reduction of a pure stabiliser state, it is worth listing some easily checked conditions under which (ING) holds.

▶ **Proposition 1.** *The Ingleton inequality* (ING) *holds if any one of the following conditions holds.*

(a) $\rho_{ABCD} = |\psi_{ABCD}\rangle\langle\psi_{ABCD}|$ *is any pure 4-party state.*

(b) $\rho_{ABCD} = \rho_{ABC} \otimes \rho_D$ *or* $\rho_A \otimes \rho_{BCD}$

(c) *The two-party component of the entropy vector for* (ρ_{ABCD}) *is symmetric under a partial exchange between* (A, B) *and* (C, D), i.e. *under any* one *(but not two) of the exchanges* $A \leftrightarrow C$, $B \leftrightarrow D$, $A \leftrightarrow D$ *or* $B \leftrightarrow C$.

Proof. To prove (a) it suffices to observe that

$$\begin{aligned} \mathrm{Ing}(AB:CD) &= I(A:B|C) + S(AD) + S(BD) - S(D) - S(ABD) \\ &\quad + S(C) + S(D) - S(CD) - S(A) - S(B) + S(AB) \\ &= I(A:B|C) + S(AD) + S(AC) - S(A) - S(ACD) \\ &= I(A:B|C) + I(C:D|A) \geq 0. \end{aligned}$$

To prove (b) observe that when $\rho_{ABCD} = \rho_{ABC} \otimes \rho_D$ then $I(A:B|D) = I(A:B)$ and $I(C:D) = 0$ so that (ING) follows immediately from (SSA). For $\rho_{ABCD} = \rho_A \otimes \rho_{BCD}$ the first, second and last terms in (ING) are zero so that it becomes simply $I(C:D) \geq 0$.

For (c) we observe that (ING) is equivalent to

$$I(B:C|A) + I(A:D|B) + R \geq 0 \quad \text{with} \quad R = S(BC) + S(AD) - S(CD) - S(AB). \tag{5}$$

The exchange $A \leftrightarrow C$ takes R to $-R$. Thus, if ρ_{ABCD} is symmetric under this exchange, then $R = 0$ and (ING) holds. The sufficiency of the other exchanges can be shown similarly. ◀

If (ING) holds, then all of the Matúš inequalities (3) also hold, since they add only conditional mutual informations $I(X:Y|Z) \geq 0$ to it. However, it is well-known that entropies do not universally obey the Ingleton inequality. A simple, well-known counterexample is given by independent and uniform binary variables C and D, and $A = C \vee D$, $B = C \wedge D$. Then the first three terms in (ING) vanish, so that $\mathrm{Ing}(AB:CD) = -I(A:B) < 0$.

To obtain a quantum state which violates Ingleton, let $|\psi\rangle = \frac{1}{\sqrt{2}}\big(|0000\rangle + |1111\rangle\big)$ and

$$\rho_{ABCD} = \tfrac{1}{2}|\psi\rangle\langle\psi| + \tfrac{1}{4}|1010\rangle\langle 1010| + \tfrac{1}{4}|1001\rangle\langle 1001|. \tag{6}$$

All the reduced states ρ_{ABC}, ρ_{BD}, etc. are separable and identical to those of the state

$$\rho_{ABCD} = \tfrac{1}{4}|0000\rangle\langle 0000| + \tfrac{1}{4}|1111\rangle\langle 1111| + \tfrac{1}{4}|1010\rangle\langle 1010| + \tfrac{1}{4}|1001\rangle\langle 1001|.$$

corresponding to the classical example above. Therefore (6) violates the Ingleton inequality, but still satisfies all of the Matúš inequalities. Note that the state $|\psi\rangle$ is maximally entangled wrt the splitting A and BCD. Additional quantum states with the same entropy vectors as classical states which violate Ingleton [32, 33] can be similarly constructed. However, we do not seem to know "genuinely quantum" counterexamples to the Ingleton inequality.

▶ **Question 2.** *Do there exist quantum states which violate Ingleton and are neither separable nor have the same entropy vectors as some classical state?*

3.2 Families of inequalities

When the subsystem C or D is trivial, the Ingleton inequality reduces to the 3-party SSA inequality, $I(A:B|C) \geq 0$; when subsystem A or B is trivial, it reduces to the 2-party subadditivity inequality $I(C:D) \geq 0$. This suggests that the Ingleton inequality is a member of a more general family of N-party inequalities. In 2011, Kinser [23] found the first such family, which can be written (for $N \geq 4$) as

$$K[N] = I(1:N|3) + H(1N) - H(12) - H(3N) + H(23) + \sum_{k=4}^{N} I(2:k-1|k) \geq 0. \qquad (7)$$

This is equivalent to the Ingleton inequality when $N = 4$.

▶ Remark. As in the proof of Proposition 1(c), it can be shown that Kinser's inequalities hold if ρ is symmetric with respect to the interchange $1 \leftrightarrow 3$ or $2 \leftrightarrow N$. They also hold if $\rho_{1,2,\ldots N} = \rho_2 \otimes \rho_{1,3,\ldots N}$ One can ask if part (a) of Theorem 1 can be extended to the new inequalities, *i.e.*, do they hold for hold for N-party pure quantum states?

3.3 Inequalities from common information

Soon after Kinser's work, another group [14] found new families of linear rank inequalities for poly-matroids for $N > 4$ that are independent of both Ingleton's inequality and Kinser's family. In the 5-party case, they found a set of 24 inequalities which generate all linear rank inequalities for poly-matroids. Moreover, they give an algorithm which allows one to generate many more families of linear rank inequalities based on the notion of common information, considered much earlier in [1, 2, 16] and used below. However, it was shown in [8] that there are N-party linear rank inequalities that cannot be obtained from the process described in [14].

▶ **Definition 3.** In a poly-matroid H on $\mathcal{X}$, two subsets A and B are said to *have a common information*, if there exists an extension of H to a poly-matroid on the larger set $\mathcal{X} \dot\cup \{\zeta\}$, such that $H(\{\zeta\} \cup A) - H(A) =: H(\zeta|A) = 0$, $H(\{\zeta\} \cup B) - H(B) =: H(\zeta|B) = 0$ and $H(\zeta) = I(A:B)$.

Here we used $H(Z|A) = H(AZ) - H(A)$ to denote the conditional entropy. For completeness we include a proof (courtesy of a Banff talk by Dougherty) of the following result, as well as a proof of Lemma 5 below, which appear in [14].

▶ **Proposition 4.** *Let h be a poly-matroid on $\mathcal{X}$, and $A, B, C, D \subset \mathcal{X}$ such that A and B have a common information. Then the Ingleton inequality (ING) holds for A, B, C and D.*

Proof. Let ζ be a common information of A and B. Then, using $H(F|A) \geq H(F|AC)$ in Lemma 5 below, and letting $F = \zeta$, gives

$$I(A:B|C) + H(\zeta|A) \geq I(\zeta:B|C).$$

Using this a total of six times, we obtain

$$\begin{aligned} I(A:B|C) + I(A:B|D) + I(C:D) + 2H(\zeta|A) + 2H(\zeta|B) \\ &\geq I(A:\zeta|C) + I(A:\zeta|D) + I(C:D) + 2H(\zeta|A) \\ &\geq I(\zeta:\zeta|C) + I(\zeta:\zeta|D) + I(C:D) \\ &= H(\zeta|C) + H(\zeta|D) + I(C:D) \geq H(\zeta|C) + I(\zeta:D) \geq I(\zeta:\zeta) = H(\zeta). \end{aligned}$$

Inserting the conditions for ζ being a common information, completes the proof. ◀

▶ **Lemma 5.** *In a poly-matroid H on a set $\mathcal{X}$ with subsets $A, B, C, F \subset \mathcal{X}$.*

$$I(A:B|C) + H(F|AC) \geq I(F:B|C) \tag{8}$$

Proof. By a direct application of the poly-matroid axioms:

$$\begin{aligned} I(A:B|C) + H(F|AC) - I(F:B|C) &= H(B|FC) - H(B|AC) + H(F|AC) \\ &= H(BCF) + H(ACF) - H(CF) - H(ABC) \qquad (9) \\ &\geq H(BCF) + H(ACF) - H(CF) - H(ABCF) \\ &= I(A:B|CF) \geq 0, \qquad (10) \end{aligned}$$

where we used only algebraic identities, SSA and monotonicity. ◀

In a linearly represented poly-matroid, (ING) is universally true: There, $H(J) = \dim V_J$, with $V_J = \sum_{j\in J} V_j$ for a family of linear subspaces $V_j \subset V$ of a vector space. The common information of any $A, B \subset \mathcal{X}$ is constructed by defining $V_\zeta = V_A \cap V_B$.

▶ **Theorem 6.** *Any linear rank inequality for a poly-matroid obtained using common information and the poly-matroid inequalities, also holds for a group poly-matroid when its defining subgroups are normal.*

Proof. It suffices to show that when $G_A, G_B \lhd G$ are normal subgroups for $A, B \subset \mathcal{X}$, then A and B have a common information given by $G_\zeta = G_A G_B \lhd G$ (the latter from the normality of G_A and G_B). The first two conditions for a common information are clearly satisfied, as $G_A, G_B \subset G_A G_B$, and the third follows from the well-known natural isomorphisms [42]

$$G/(G_A G_B) \approx (G/G_A)\Big/\big((G_A G_B)/G_A\big) \quad \text{and} \quad (G_A G_B)/G_A \approx G_B/(G_A \cap G_B),$$

which imply

$$\begin{aligned} H(\zeta) &= \log|G/(G_A G_B)| = \log|G/G_A| - \log|(G_A G_B)/G_A| \\ &= \log|G/G_A| + \log|G/G_B| - \log|G/(G_A \cap G_B)| = I(A:B). \end{aligned}$$ ◀

4 Entropies of stabiliser states

4.1 Stabiliser groups and stabiliser states

Motivated by the stabiliser states encountered in the extremal rays of Σ_2, Σ_3 and Σ_4, we now focus on (pure) stabiliser states, *i.e.* 1-dimensional quantum codes. Stabilizer codes have emerged in successively more general forms. We use the formulation described by Klappenecker and Rötteler [24, 25] (following Knill [26]) which relies on the notion of *abstract error group*: This is a finite subgroup $W < \mathcal{U}(\mathcal{H})$ of the unitary group of a (finite dimensional) Hilbert space $\mathcal{H}$, which satisfies the following axioms:

1. The center C of W consists only of multiples of the identity matrix ("scalars"): $C \subset \mathbf{C}\mathbb{1}$.
2. $\widehat{W} \equiv W/C$ is an abelian group of order $|\mathcal{H}|^2$, called the *abelian part* of W.
3. For all $g \in W \setminus C$, $\operatorname{Tr} g = 0$.

Note that conditions 1 and 2 imply that W is non-abelian; whereas condition 2 says that the non-commutativity is played out only on the level of complex phases: for $g, h \in W$,

$$gh = \omega(g,h)hg, \text{ with } \omega(g,h) \in \mathbf{C}.$$

Finally, condition 3 means that $g, h \in W$ in different cosets modulo C are orthogonal with respect to the Frobenius (or Hilbert-Schmidt) inner product: $\operatorname{Tr} g^\dagger h = 0$. It is known that $\widehat{W}$ is a direct product of an abelian group T with itself, such that $|T| = |\mathcal{H}|$.

▶ **Example 7** (Discrete Weyl-Heisenberg group)**.** Let $\mathcal{H}$ be a d-dimensional Hilbert space, with a computational orthonormal basis $\{|j\rangle\}_{j=0}^{d}$. Define discrete Weyl operators

$$X|j\rangle = |j+1\rangle \mod d, \qquad Z|j\rangle = e^{j\frac{2\pi i}{d}}|j\rangle.$$

They are clearly both of order d, and congruent via the discrete Fourier transform. The fundamental commutation relation, $XZ = e^{2\pi i/d}ZX$ ensures that the group W generated by X and Z is finite, and indeed an abstract error group with center $C = \left\{e^{j\frac{2\pi i}{d}} : j = 0, \ldots, d-1\right\}$.

Note that the tensor product of abstract error groups is again an abstract error group. Now, assume that each party $x \in \mathcal{X}$ of the composite quantum system can be associated with an abstract error group $W_x < \mathcal{U}(\mathcal{H}_x)$ of unitaries with center C_x, which satisfy $W_x \supset C_x \subset \mathbf{C}\mathbb{1}$, such that $\widehat{W_x} = W_x/C_x$ is abelian and has cardinality $d_x^2 \equiv |\mathcal{H}_x|^2$. Let $W \equiv \bigotimes_{x\in\mathcal{X}} W_x$ be the tensor product abstract error group, acting on $\mathcal{H} = \bigotimes_{x\in\mathcal{X}} \mathcal{H}_x$. For any subgroup $\Gamma < W$, we let $\widehat{\Gamma} = (C\Gamma)/C \simeq \Gamma/(\Gamma \cap C)$ denote the quotient of Γ by the center of W.

Stabiliser codes [17, 5] are subspaces of $\mathcal{H}$ which are simultaneous eigenspaces of abelian subgroups of W.

Consider a maximal abelian subgroup $G < W$, which contains the center $C = \bigotimes_{x\in\mathcal{X}} C_x < \mathbf{C}\mathbb{1}$ of W, so that $\widehat{G} = G/C$ has cardinality $\sqrt{|\widehat{W}|} = |\mathcal{H}| = \prod_{j=1}^{N} |\mathcal{H}_j|$. Since G is abelian it has a common eigenbasis, each state of which is called a stabiliser state $|\psi\rangle$.

More generally, let $G < W$ be any abelian subgroup of an abstract error group $W < \mathcal{U}(\mathcal{H})$. Since all $g \in G$ commute, they are jointly diagonalisable: let P be one of the maximal joint eigenspace projections. Then for $g \in G$, $gP = \chi(g)P$, for a complex number $\chi(g)$. Thus $\chi : G \longrightarrow \mathbf{C}$ is necessarily the character of a 1-dimensional group representation, which gives rise to the following expression for P:

$$P = \frac{1}{|G|} \sum_{g\in G} \overline{\chi(g)}\, g. \tag{11}$$

If $\chi(g_0) = 1$ and $g = c\, g_0$ is in the coset $g_0\, C$, then $c = \chi(g)$ and $\overline{\chi(g)}\, g = g_0$. Thus, $G_0 = \{g \in G : \chi(g) = 1\}$ is a subgroup of G isomorphic to $\widehat{G} = G/C$ and (11) can be rewritten as

$$P = \frac{1}{|G_0|} \sum_{g\in G_0} g. \tag{12}$$

Since $g^{-1} = g^\dagger$ this sum is self-adjoint, and

$$P^2 = \frac{1}{|G_0|^2} \sum_{g,h\in G_0} gh = \frac{1}{|G_0|} \sum_{g\in G_0} g = P,$$

so that (12) is indeed a projection.

Note: The above reasoning is true because we assumed that $\chi(g)$ records the eigenvalues of g on the eigenspace with projector P; as such, it has the property $\chi(t\mathbb{1}) = t$ for $t \in \mathbf{C}$. For a general character χ, however, only $G_0 < \chi^{-1}(1)$ holds.

Because of the importance of the case of rank one projections, we summarize the results above in the case of maximal abelian subgroups.

▶ **Theorem 8.** *Let G be a maximal abelian subgroup of an abstract error group W with center C. Any simultaneous eigenstate of G can be associated with a subgroup $G_0 \simeq G/C$ for which* $|\psi\rangle\langle\psi| = \frac{1}{|G_0|} \sum_{g\in G_0} g.$ ◀

▶ Remark. The use of the trivial representation is not essential in the expression above. It was used only to define G_0. Once this has been done, one can use the (1-dim) irreducible representations of G_0 to describe the orthonormal basis of stabiliser states associated with G. Let $\chi_k(g)$ denote the $d = |G_0|$ irreducible representations of G_0 and define

$$|\psi_k\rangle\langle\psi_k| = \tfrac{1}{|G_0|} \sum_{g\in G_0} \chi_k(g) g. \tag{13}$$

Then the orthogonality property of group characters implies that $\operatorname{Tr} |\psi_j\rangle\langle\psi_j||\psi_k\rangle\langle\psi_k| = |\langle\psi_j|\psi_k\rangle|^2 = \delta_{jk}$.

4.2 Entropies of stabiliser states

The next result seems to have been obtained independently by several groups [20, 10, 11].

▶ **Proposition 9.** *For a pure stabiliser state $\rho = |\psi\rangle\langle\psi|$ with associated error group $G < W$, and any $J \subset \mathcal{X}$, the entropy*

$$S(J) = S(\rho_J) = \log \frac{d_J}{|\widehat{G_J}|}. \tag{14}$$

Here, $d_J = \prod_{x\in J} d_x$ and

$$G_J \equiv \{\otimes_{x\in\mathcal{X}}\, g_x \in G : \forall x \notin J \;\; g_x = \mathbb{1}\} \subset G,$$

and $\widehat{G_J} = G_J/C_J$ is the quotient of G_J with respect to the center $C_J = G_J \cap C$.

Proof. It is enough to consider a bipartite system with local error groups W_A and W_B, by considering party A all systems in J, and B all systems in $\mathcal{X} \setminus J$. Then,

$$|\psi\rangle\langle\psi| = \frac{1}{|\widehat{G}|} \sum_{(g_A,g_B)\in\widehat{G}} g_A \otimes g_B.$$

Since $\operatorname{Tr} g_B = 0$ unless $g_B = \mathbb{1}$ and $|\widehat{G}| = d_A d_B$, this implies

$$\begin{aligned}
\rho_A = \operatorname{Tr}_B |\psi\rangle\langle\psi| &= \frac{1}{|\widehat{G}|} \sum_{(g_A,g_B)\in\widehat{G}} (\operatorname{Tr} g_B)\, g_A \\
&= \frac{1}{|\widehat{G}|} \sum_{g_A\in\widehat{G_A}} |\mathcal{H}_B|\, g_A \\
&= \frac{1}{|\mathcal{H}_A|} \sum_{g_A\in\widehat{G_A}} g_A = \frac{|\widehat{G_A}|}{d_A}\left(\frac{1}{|\widehat{G_A}|} \sum_{g_A\in\widehat{G_A}} g_A\right).
\end{aligned}$$

Since, $\operatorname{Tr}\rho_A = 1$, the last line implies that ρ_A is proportional to a projector of rank $\frac{d_A}{|\widehat{G_A}|}$. Thus, its entropy is simply $S(\rho_A) = \log \frac{d_A}{|\widehat{G_A}|}$. ◀

The following corollary is the key to our main result, Theorem 11.

▶ **Corollary 10.** *For a pure stabiliser state as in Proposition 9, the entropy of the reduced state ρ_J satisfies*

$$S(J) = S(\rho_J) = \log \frac{|\widehat{G}|}{|\widehat{G_{J^c}}|} - \log d_J = \log \frac{|\widehat{G}|}{|\widehat{G_J}|} - \log d_{J^c}. \tag{15}$$

Proof. As in Proposition 9, it suffices to consider the bipartite case. Since $|\psi\rangle\langle\psi|$ is pure,

$$S(\rho_A) = S(\rho_B) = \log \frac{d_B}{|\widehat{G_B}|} = \log \frac{d_A d_B}{|\widehat{G_B}|} - \log d_A.$$

Since $d_A d_B = |\widehat{G}|$ this gives the desired result. ◀

▶ **Theorem 11.** *Any pure stabiliser state $\rho = |\psi\rangle\langle\psi|$ on an 5-party system gives rise to 4-party reduced states whose entropies satisfy the Ingleton inequality.*

Proof. By Corollary 10, we have

$$S(J) = \log \frac{|\widehat{G}|}{|\widehat{G_{J^c}}|} - \sum_{x\in J} \log d_x \,. \tag{16}$$

The first term $H(J) = \log \frac{|\widehat{G}|}{|\widehat{G_{J^c}}|}$ is a Shannon entropy of the type used in [9]. To be precise, observe that $\widehat{G_{J^c}} = \bigcap_{x\in J} \widehat{G_{\mathcal{X}\setminus x}}$. Moreover, since $\widehat{G}$ and its subgroups $\widehat{G_{J^c}}$ are abelian, this implies that the entropy vector for each of the 4-party reduced states satisfies the Ingleton inequality. (This was observed in [9] and also follows from Theorem 6.)

To complete the argument, it suffices to observe that the Ingleton inequality is balanced, so that the Ingleton expression is identically zero for the sum-type "rank function" from the second term in (16), *i.e.* $h_0(J) \equiv \sum_{x\in J} \log d_x$ defines a poly-matroid satisfying (ING) with equality. ◀

Any linear combination of mutual informations and conditional mutual informations is a balanced expression (and vice versa, any balanced expression can be written as such a linear combination). Kinser's family of inequalities is balanced, which can be seen by inspection of (7). It also holds by construction for the inequalities obtained from [14, Thms. 3 and 4] and, more generally, any inequality obtained using a "common information" as in [14]. Therefore, we can conclude using the same argument as above that

▶ **Theorem 12.** *Any pure stabiliser state on an $(N+1)$-party system generate an N-party entropy vector which satisfies the Kinser [23] family* (7) *of inequalities, and more generally those of Dougherty et al. [14].*

A consequence of Theorem 11 is that the Matúš family of inequalities holds for stabiliser states; however, rays generated by the stabiliser state entropy vectors do not span the entropy cone $\overline{\Sigma}_4^Q$. In fact, from the proof of Theorem 11, we see that *every* balanced inequality that holds for the Shannon entropy, holds automatically for stabilizer quantum entropies.[1] Note also that apart from (MO), all other necessary entropy inequalities for the Shannon entropy are balanced [6].

5 The 4-party quantum entropy cone

By direct calculation using symbolic software, we can compute the extreme rays of 4-party poly-quantoids plus Ingleton inequalities. The results are given (up to permutation) as rays 0 to 6 in Table 1 below, as elements of the 5-party cone $\widetilde{\Gamma}_{4+1}^Q$ (on subsets of $\{a, b, c, d, e\}$) of vectors which satisfy (+) (SSA) and the complementarity property $S(J) = S(J^c)$ as described at the end of Section 2.2.

[1] We are grateful to D. Gross and M. Walter, whose paper [18] made us aware of this observation.

Table 1 Extreme rays of the 4-party quantum Ingleton cone.

subset \ ray	1	2	3	4	5	6	0
a	1	1	1	1	2	1	1
b	1	1	1	1	1	1	1
c	0	1	1	1	1	2	1
d	0	1	1	1	1	2	2
e ($\widehat{=}$ abcd)	0	0	0	1	1	2	2
ab	0	1	2	1	3	2	2
ac	1	1	2	1	3	3	2
ad	1	1	2	1	3	3	2
ae ($\widehat{=}$ bcd)	1	1	1	1	3	3	2
bc	1	1	2	1	2	3	2
bd	1	1	2	1	2	3	2
be ($\widehat{=}$ acd)	1	1	1	1	2	3	2
cd	0	1	2	1	2	2	2
ce ($\widehat{=}$ abd)	0	1	1	1	2	2	2
de ($\widehat{=}$ abc)	0	1	1	1	2	2	2

The following stabiliser states found by Ibinson [21] (some of which were known earlier) realise entropy vectors on the rays 1 through 6 shown in Table 1.

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}\big(|00\rangle + |11\rangle\big)_{ab}|000\rangle_{cde}, \tag{R1}$$

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}\big(|0000\rangle + |1111\rangle\big)_{abcd}|0\rangle_e, \tag{R2}$$

$$|\psi_3\rangle = \frac{1}{3}\sum_{i,j=0,1,2} |i\rangle_a|j\rangle_b|i\oplus j\rangle_c|i\oplus 2j\rangle_d|0\rangle_e, \tag{R3}$$

$$|\psi_4\rangle = \frac{1}{\sqrt{2}}\big(|00000\rangle + |11111\rangle\big)_{abcde}, \tag{R4}$$

$$|\psi_5\rangle = \frac{1}{\sqrt{2}}\big(|0\rangle_{a'}|0_L\rangle_{a''bcde} + |1\rangle_{a'}|1_L\rangle_{a''bcde}\big), \tag{R5}$$

$$|\psi_6\rangle = \frac{1}{\sqrt{27}}\sum_{i,j,k=0,1,2} |i\rangle_a|j\rangle_b|i\oplus j\rangle_{c'}|k\rangle_{c''}|i\oplus j\rangle_{d'}|k\rangle_{d''}|i\oplus j\rangle_{e'}|k\rangle_{e''}, \tag{R6}$$

where in eq. (R5), $|0_L\rangle$ and $|1_L\rangle$ are the logical 0 and 1 on the famous 5-qubit code [27, 3]. These are also extremal rays of the quantum entropy cone Σ_4^Q. In addition, ray 0 in Table 1 is realised by the (stabiliser!) state

$$|\psi_0\rangle = \frac{1}{2}\sum_{i,j=0,1} |i\rangle_A|j\rangle_B|i\oplus j\rangle_C|ij\rangle_D|ij\rangle_E. \tag{R0}$$

on $1+1+1+2+2$ qubits.

Let us call an N-party poly-quantoid *stabiliser-represented*, if it is in the closure of the cone generated by the entropy vectors of $(N+1)$-party stabiliser states in the sense used above. Then the above reasoning proves the following analogue of a theorem by Hammer, Romashchenko, Shen and Vereshchagin [19]:

▶ **Theorem 13.** *A 4-party poly-quantoid is stabiliser-represented if and only if it satisfies the Ingleton inequality (and all its permutations).* ◀

It seems reasonable to conjecture that the closure of the cone generated by the entropy vectors of stabiliser states is identical to that obtained when inequalities obtained from common information as in [14] are added to the classical ones. However, it is not even clear if stabiliser states satisfy the additional linear rank inequalities shown to exist in [8].

6 Conclusion

The difficult question of whether or not the quantum entropy satisfies inequalities beyond positivity and SSA remains open for four or more parties.

Do quantum states which do not satisfy Ingleton always lie within the classical part of the quantum entropy cone? We know that the quantum entropy cone $\overline{\Sigma}_{\mathcal{X}}^{Q}$ is strictly larger than the classical one $\overline{\Sigma}_{\mathcal{X}}^{C}$. Recall that $\Lambda_4^{C,Q}$ denotes the polyhedral cones formed from the classical inequalities (in each case) and the Ingleton inequality. We want to know whether or not $\overline{\Sigma}_4^Q \setminus \Lambda_4^Q$ is strictly larger than $\overline{\Sigma}_4^C \setminus \Lambda_4^C$, *i.e.*, are there quantum states whose entropy vectors do not satisfy the Ingleton inequality and are not equal to any vector in the closure of the classical entropy cone, $\overline{\Sigma}_4^C$? If the answer is negative, then 4-party quantum entropy vectors must also satisfy the new non-Shannon inequalities.

It seems that a better understanding of quantum states which do not satisfy (ING) may be the key to determining whether or not quantum states satisfy the classical non-Shannon inequalities.

This question extends naturally to the 5-party case, in which all linear rank inequalities are known from [14]. However, for more parties, one can ask the same question for both the cones associated with inequalities obtained using one common information as in [14], and for the cones obtained using all linear rank inequalities. Although we know from [7, 8] that additional inequalities are required, we do not even have explicit examples to consider.

Related work. After completion of the present research, we became aware of independent work by Gross and Walter [18], who use discrete phase space methods for stabilizer states to show that the entropies of stabilizer states satisfy all balanced classical entropy inequalities. Indeed, this can also be seen from our formula for the reduced state entropies in Corollary 10.

Acknowledgements. Portions of this work were done when FM, MBR and AW were participating in workshops at the IMS, National University of Singapore, and at the Center for Sciences "Pedro Pascual" in Benasque, Spain. This collaboration was a direct consequence of their participation in a workshop on matroids held at the Banff International Research Station in 2009.

The work of FM was partially supported by Grant Agency of the Czech Republic under Grant 13-20012S. The work of MBR was partially supported by NSF grant CCF-1018401. NL and AW acknowledge financial support by the European Commission (STREP "QCS" and Integrated Project "QESSENCE"), AW furthermore support by the ERC (Advanced Grant "IRQUAT"), a Royal Society Wolfson Merit Award and a Philip Leverhulme Prize. The Centre for Quantum Technologies is funded by the Singapore Ministry of Education and the National Research Foundation as part of the Research Centres of Excellence programme.

References

1 R. Ahlswede, P. Gács, "Spreading of sets in product spaces and hypercontraction of the Markov operator" *Ann. Prob.* **4**:925-939 (1976).

2 R. Ahlswede, J. Körner, "On Common Information and Related Characteristics of Correlated Information Sources" manuscript (1975); published in *General Theory of Information Transfer and Combinatorics*, LNCS Vol. 4123, Springer Verlag, 2006, pp. 664-677.

3 C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, W. K. Wootters, "Mixed-state entanglement and quantum error correction" *Phys. Rev. A* **54**(5):3824-3851 (1996).

4 J. Cadney, N. Linden, A. Winter, "Infinitely many constrained inequalities for the von Neumann entropy" *IEEE Trans. Inf. Theory* **58**, 3657 (2012). `arXiv[quant-ph]:1107.0624`.

5 A. R. Calderbank, E. M. Rains, P. W. Shor, N. J. A. Sloane, "Quantum Error Correction Via Codes Over GF(4)" *IEEE Trans. Inf. Theory* **44**(4):1369-1387 (1998).

6 T. H. Chan, "Balanced Information Inequalities" *IEEE Trans. Inf. Theory* **49**(12):3261-3267 (2003).

7 T. Chan, A. Grant, D. Kern, "Existence of new inequalities for representable poly-matroids" *Proc. ISIT 2010*, pp. 1364-1368 (2010). `arXiv[quant-ph]:0907.5030`.

8 T. Chan, A. Grant, D. Pflüger, "Truncation technique for characterizing linear poly-matroids" *IEEE Trans. Inf. Theory* **57**:6364-6378 (2011).

9 T. H. Chan, R. W. Yeung, "On a Relation Between Information Inequalities and Group Theory" *IEEE Trans. Inf. Theory* **48**(7):1992-1995 (2002).

10 D. Fattal, T. S. Cubitt, Y. Yamamoto, S. Bravyi, I. L. Chuang, "Entanglement in the stabiliser formalism" `arXiv:quant-ph/0406168` (2004).

11 M. Van den Nest, J. Dehaene, B. De Moor, "Local invariants of stabiliser codes" *Phys. Rev A* **70**:032323 (2004). `arXiv:quant-ph/0404106`.

12 R. Dougherty, C. Freiling, K. Zeger, "Six New Non-Shannon Information Inequalities" Proc. ISIT 2006, pp. 233-236 (2006).

13 R. Dougherty, C. Freiling, K. Zeger, "Networks, Matroids, and Non-Shannon Information Inequalities" *IEEE Trans. Inf. Theory* **53**(6):1949-1969 (2007).

14 R. Dougherty, C. Freiling, K. Zeger, "Linear rank inequalities on five or more variables" `arXiv[cs.IT]:0910.0284` (2009).

15 R. Dougherty, C. Freiling, K. Zeger, "Non-Shannon Information Inequalities in Four Random Variables" `arXiv[cs.IT]:1104.3602` (2011).

16 P. Gács, J. Körner, "Common information is far less than mutual information" *Problems of Contr. and Inf. Theory* **2**:149-162 (1973).

17 D. Gottesman, *Stabilizer Codes and Quantum Error Correction*, PhD thesis, Caltech, 1997.

18 D. Gross, M. Walter, "Stabilizer information inequalities from phase space distributions" `arXiv[quant-ph]:1302.6990` (2012).

19 D. Hammer, A. Romashchenko, A. Shen, N. Vereshchagin, "Inequalities for Shannon Entropy and Kolmogorov Complexity" *J. Comp. Syst. Sciences* **60**(2):442-464 (2000).

20 M. Hein, J. Eisert, H.J. Briegel "Multi-party entanglement in graph states" *Phys. Rev. A* **69**:062311 (2004). `arXiv:quant-ph/0307130`.

21 B. Ibinson, *Quantum Information and Entropy*, PhD thesis, University of Bristol, 2006 (unpublished). Available at URL `http://www.maths.bris.ac.uk/~csajw/BenIbinson.PhD-thesis-final.pdf`.

22 A. W. Ingleton, "Representation of matroids" in: *Combinatorial Mathematics and its Applications*, ed. D. J. A. Welsh, pp. 149-167 (Academic Press, 1971).

23 R. Kinser, "New Inequalities for subspace arrangements" *J. Comb Theory A* **118**:152-161 (2011).

24 A. Klappenecker, M. Rötteler, "Beyond Stabilizer Codes I: Nice Error Bases" *IEEE Trans. Inf. Theory* **48**(8):2392-2395 (2002).

25 A. Klappenecker, M. Rötteler, "Beyond Stabilizer Codes II: Clifford Codes" *IEEE Trans. Inf. Theory* **48**(8):2396-2399 (2002).

26 E. Knill, "Group Representations, Error Bases and Quantum Codes" LANL report LAUR-96-2807; `arXiv:quant-ph/9608049` (1996).

27 R. Laflamme, C. Miquel, J. P. Paz, W. H. Zurek, "Perfect Quantum Error Correcting Code" *Phys. Rev. Lett.* **77**(1):198-201 (1996).

28 S-Y.R. Li, R. Yeung, N. Cai, "Linear network coding" *IEEE Trans. Inf. Theory* **49**:371-381 (2003).

29 E. H. Lieb, M. B. Ruskai, "Proof of the strong subadditivity of quantum-mechanical entropy" *J. Math. Phys.* **14**(12):1938-1941 (1973).

30 E. H. Lieb, "Some Convexity and Subadditivity Properties of Entropy" Bull. Amer. Math. Soc. **81**(1):1-13 (1975).

31 N. Linden, A. Winter, "A New Inequality for the von Neumann Entropy" *Comm. Math. Phys.* **259**:129-138 (2005).

32 F. Matúš, M. Studený, "Conditional independences among four random variables I." *Comb. Prob. Comp.* **4**, 269-278 (1995).

33 F. Matúš, "Conditional independences among four random variables II." *Comb. Prob. Comp.* **4**:407-417 (1995); III **8**:269–276 (1999).

34 F. Matúš, "Infinitely Many Information Inequalities" Proc. ISIT 2007, pp. 41-44 (2007).

35 F. Matúš, "Two constructions on limits of entropy functions" *IEEE Trans. Inf. Theory* **53**:320-330 (2007).

36 F. Matúš, "Polymatroids and polyquantoids" Proc. WUPES 2012 (eds. J. Vejnarová and T. Kroupa), Mariánské Lázně, Prague, Czech Republic, pp. 126-136 (2012).

37 W. Mao, M. Thill, B. Hassibi, "On the Ingleton-Violating Finite Groups and Group Network Codes" `arXiv[cs.IT]:1202.5599` (2012).

38 J. G. Oxley, *Matroid Theory*, Oxford University Press, Oxford, 2006.

39 N. Pippenger, "The inequalities of quantum information theory" *IEEE Trans. Inf. Theory* **49**(4):773-789 (2003).

40 C. E. Shannon, "A mathematical theory of communication" *Bell System Technical Journal* **27**:379-423 & 623-656 (1948).

41 R. Stancu, F. Oggier "Finite Nilpotent and metacyclic groups never violate the Ingleton inequality" 2012 International Symposium on Network Coding (NetCod), pp. 25-30.

42 M. Suzuki, *Group Theory I*, Springer Verlag, Berlin New York, 1982.

43 R. W. Yeung, "A Framework for Linear Information Inequalities" *IEEE Trans. Inf. Theory* **43**(6):1924-1934 (1997).

44 Z. Zhang, R. W. Yeung, "A Non-Shannon-Type Conditional Inequality of Information Quantities" *IEEE Trans. Inf. Theory* **43**(6):1982-1986 (1997).

45 Z. Zhang, R. W. Yeung, "On Characterization of Entropy Function via Information Inequalities" *IEEE Trans. Inf. Theory* **44**(4):1440-1452 (1998).

Kitaev's $\mathbb{Z}_d$-Codes Threshold Estimates

Guillaume Duclos-Cianci and David Poulin

Département de Physique, Université de Sherbrooke, Sherbrooke, Québec, J1K 2R1, Canada
Guillaume.Duclos-Cianci@USherbrooke.ca
David.Poulin@USherbrooke.ca

Abstract

We study the quantum error correction threshold of Kitaev's toric code over the group $\mathbb{Z}_d$ subject to a generalized bit-flip noise. This problem requires novel decoding techniques, and for this purpose we generalize the renormalization group method we previously introduced in [5, 6] for $\mathbb{Z}_2$ topological codes.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases Quantum error-correction threshold, Topological stabilizer codes, Qudit stabilizer codes

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.285

1 Introduction

Kitaev's topological code (KTC) [11] on qubits is the archetypical topological code and has been extensively studied. As explained in Kitaev's original paper [11], this construction applies to any group. Much less is known about these generalizations, and in this paper we investigate the quantum error correction (QEC) thresholds of the KTCs built with the groups $\mathbb{Z}_d$, where $d \geq 2$. We label these as $\mathbb{Z}_d$-KTC, so the original code on qubits corresponds to $\mathbb{Z}_2$-KTC.

As explained in [4], $\mathbb{Z}_2$-KTC can be decoded by a binary perfect matching algorithm [7], since every particle is its own anti-particle in this model. Because this is not the case for $d > 2$, other techniques are required and for this purpose we generalize the renormalization group (RG) soft decoder that we introduced in [5, 6]. Our numerical simulations show that the threshold increases monotonically with d and appears to follow the general trend of the qudit hashing bound.

This paper is organized as follows. First, we introduce a generalized Pauli group (see [12, 9] for more details), stabilizer codes, and $\mathbb{Z}_d$-Kitaev's toric code. Next, we briefly review the decoding problem of these systems and show how the RG decoder applies in this case. Finally, we present the numerical results and close with a discussion.

2 $\mathbb{Z}_d$ generalization of Kitaev's toric code

In this section, we review the definition of $\mathbb{Z}_d$-KTC and show that many features of KTC on qubits extend to them. Since we will be working with qudits, we introduce a generalized Pauli group. The Hilbert space of a qudit, $\mathcal{H}_d$, is spanned by the states $\{|0\rangle, |1\rangle, \ldots, |d-1\rangle\}$. We define the operators X and Z such that

$$X|g\rangle = |g \oplus 1\rangle, \qquad Z|g\rangle = \omega^g |g\rangle, \tag{1}$$

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 285–293

Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

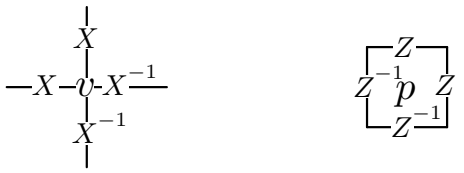


Figure 1 $\mathbb{Z}_d$-KTC stabilizer generators. To each vertex v, we associate an operator A_v(left) and to each plaquette p, we associate an operator B_p (right).

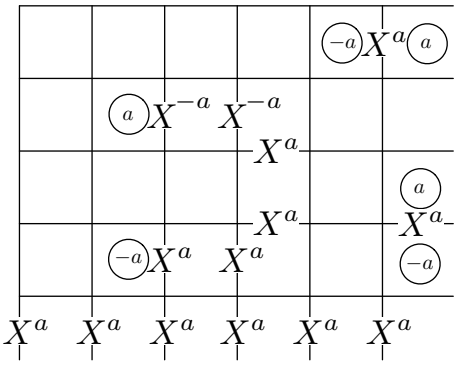


Figure 2 Plaquette defects created by the application of some power of X. The values a $(-a)$ in the plaquettes are such that the eigenvalue of the corresponding B_p is ω^a (ω^{-a}). By choosing appropriately the powers of X, we can build string operators with defects only on their endpoints. Non-trivial cocyles of X^a correspond to $\overline{X}^a$ logical operators.

where $0 \leq g < d$, "$\oplus$" denotes addition modulo d, and $\omega = \mathrm{e}^{i2\pi/d}$. The generalized Pauli group is generated by X, Z, and a phase, i.e., $\mathcal{P}_d = \langle \omega, X, Z \rangle$ if d is odd and $\mathcal{P}_d = \langle \omega^{1/2}, X, Z \rangle$ if d is even (XZ has order $2d$ in this case). From the definitions of Eq. (1), we deduce the following properties

$$\begin{aligned} X^a|g\rangle &= |g \oplus a\rangle, & ZX|g\rangle &= \omega XZ|g\rangle, \\ Z^a|g\rangle &= \omega^{ag}|g\rangle, & Z^aX^b|g\rangle &= \omega^{ab}X^bZ^a|g\rangle. \end{aligned} \tag{2}$$

Lastly, we define the n-qudit Pauli group $\mathcal{P}_d^n \equiv \mathcal{P}_d^{\otimes n}$ as the n-fold tensor product of $\mathcal{P}_d$.

The stabilizer group $\mathcal{S}$ is an ablian subgroup of $\mathcal{P}_d^n$. The code is defined as the simultaneous $+1$ eigenspace of all stabilizers. Note that even though the generalized Pauli operators are unitary, they are not hermitian in general so do not correspond to physical observables. However, the operator $\frac{1}{2}(s + s^\dagger)$ is hermitian and can be measured. Since s has eigenvalues ω^a, $\frac{1}{2}(s + s^\dagger)$ has eigenvalues $\frac{1}{2}(\omega^a + \omega^{-a}) = \cos(2\pi a/d)$ which are in one-to-one correspondence with the eigenvalues of s.

With these definitions in place, we present a generalization of KTC on qudits, which we call $\mathbb{Z}_d$-KTC, using Kitaev's original construction [11] on the cyclic groups $\mathbb{Z}_d$ with $d \geq 2$. The system is a square lattice of linear size L with periodic boundary conditions. Each edge is occupied by a qudit, so there are in total $n = 2L^2$ qudits. We define vertex operators A_v and plaquette operators B_p as shown in Fig. 1. There is one such operator for each vertex and each plaquette. We verify that they commute using the last line of Eq. (2). These operators generate the stabilizer group $\mathcal{S} = \langle A_v, B_p \rangle$ and the code is spanned by the simultaneous $+1$ eigenstates of the stabilizer generators.

Figure 2 illustrates how applying some power of X on a codestate creates defects on the lattice. Indeed, X^a applied on some qudit does not commute with the two plaquette

operators involving that qudit. The eigenvalues of the plaquettes to the north or east of the error will change from 1 to ω^a, and those of the plaquettes to the south or west will change from 1 to ω^{-a}. One can show that the defects thus created are topological charges; we associate the charge a to a plaquette defect corresponding to an eigenvalue ω^a of that plaquette. With this choice of labeling, the charge group restricted to plaquettes is $\mathbb{Z}_d$ with addition.

From these simple facts, it follows that string operators can be built with defects attached only to their endpoints (these strings actually live on the dual lattice, just like in KTC). This requires a careful choice of the powers of X on the qudits along the string such that the total charge in each plaquette is 0 except on its endpoints. For instance, one can adopt the convention that power a is used when heading north or east, and $-a$ when heading south or west. Moreover, we can verify that non-trivial cocycles (loops on the dual lattice, see Fig. 2) of any power of X obeying this convention commute with the stabilizer. These operators are not in the stabilizer as all the vertex generators of Fig. 1 are trivial cocyles. It follows that such operators, e.g. the one found at the bottom of Fig. 2, are logical operators (for any value of a).

A similar analysis holds for defects created by powers of Z operators. In this case, the defects live on vertices and string operators, on the direct lattice. Also, non-trivial cycles of any power of Z are logical operators. From the form of the logical operators, we directly deduce that there are two qudits encoded in the code space. Again, this is analogous to the case of KTC.

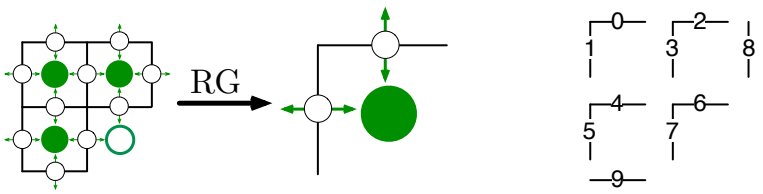
3 $\mathbb{Z}_d$-KTC decoding

We are now interested in the problem of error correcting $\mathbb{Z}_d$-KTCs for $d > 2$. In our study, we consider a simple noise model that generalizes the independent symmetric bit-flip channel to qudits[1]: with probability $1 - p_{\text{phys}}$, the qudit remains unaffected and with probability p_{phys}, we apply at random (uniformely distributed) one of $X, X^2, \ldots, X^{d-1}$. Suppose an error $E \in \mathcal{P}_d^n$ occurs on a code state. It creates defects on the lattice and by measuring the eigenvalues of every $\frac{1}{2}(A_v + A_v^\dagger)$ and $\frac{1}{2}(B_p + B_p^\dagger)$ we can learn the position and charge of each defect. The role of the decoder is to bring the system back in the code space by applying a correcting Pauli operator, $C \in \mathcal{P}_d^n$. However, care must be taken in choosing an appropriate correcting operation. Indeed, if the operator CE resulting from the combination of the error and the recovery is an element of $\mathcal{S}$, the state is unaffected. However, if CE is a non-trivial logical operator, then the system is returned to the code space but potentially in a different code state, so the information is corrupted.

Any operator $E \in \mathcal{P}_d^n$ creating the measured configuration of defects is a potential error. However, we classify these operators by their logical effect on the code space: two operators E_1, E_2 with the same configuration of defects are equivalent iff $E_2^\dagger E_1$ has a trivial effect on the code, i.e. $E_1 \sim E_2$ iff $E_2^\dagger E_1 \in \mathcal{S}$. Note that since E_1 and E_2 lead to the same defect configuration, $E_2^\dagger E_1$ creates no defect, or equivalently, E_1 creates some defects that $E_2^\dagger$ annihilates.

Given a measured defect configuration, the decoder seeks for the best correction among the set of all errors which would lead to this defect configuration. One strategy would be to identify the error from this set that has the largest probability $\mathcal{P}(E)$, where the probability

[1] This noise model can also be seen as emerging from a qudit depolarization channel that maps $\rho \to (1-q)\rho + q\frac{I}{d}$ when X and Z errors are treated independently, and $p_{\text{phys}} = q(1 - d^{-1})$.



(a) (b)

Figure 3 (a) The lattice is cut into unit cells containing ten qudits (edges). The renormalization process takes the defect configuration and the noise model on a unit cell as inputs and outputs a two-qudit distribution (white disks) which corresponds to a probability on the charge flow through the corresponding boundaries. Green disks represent plaquette operators. The plaquette corresponding to the green circle is replaced by the product of all four plaquettes of the unit cell, such that its eigenvalue gives the total charge of the cell. This value is only going to be used in the next round of RG (larger green disk). (b) Labeling convention for qudits in Eq. (3).

of an error is specified by the physical noise model, in our case the symmetric bit-flip channel. This turns out not to be optimal however, because some errors have equivalent effects on all code states. Thus, the decoder should instead seek for the most likely equivalence class of errors. The probability of an equivalence class of errors is obtained by summing over the probability of each error within a class. Given these probabilities, the optimal correction consists in applying the adjoint of any representative of the class with maximal probability.

4 RG decoder generalization to $\mathbb{Z}_d$-KTC

Unfortunately, the above procedure cannot be realized efficiently in general since the number of errors in each equivalence class scales exponentially with the system size. In [5, 6], we introduced a renormalization group soft decoder (RG decoder) that efficiently approximates the exact calculation (see [3] for a related scheme). The general idea is to cut the lattice into small unit cells (e.g. 2×2 sub-lattices) and to "distill" from each cell an effective two-qubit noise model, c.f. Fig. 3(a). This is realized by keeping track of the flow of charges through the cell and summing over the microscopic details leading to this flow. This has the effect of shrinking the lattice linear size by a constant factor (k for cells of size $k \times k$). Recursing on this process, one can shrink the lattice to a constant, manageable, size where the exact decoding can be performed. With appropriate simple modifications, this method can be used for charges over $\mathbb{Z}_d$.

There are two technical difficulties in realizing the above heuristic description, which are both caused by charge conservation. First, because the unit cells share boundaries, the flow of charge through one boundary of a cell should be equal and opposite to the flow of charge of the corresponding boundary of the neighbouring cell. Thus, the variable corresponding to charge flows in each cell are highly constrained. This problem is easily circumvented by keeping only track of the flow of charge through the northern and the western boundary of each cell, i.e. by eliminating this redundancy.

Second, the sum of the charge flow through the boundaries of a cell must be equal to its total charge, revealed by the syndrome measurement. This once again sets a hard constraint between the variables corresponding to the charge flows, which would in principle require a probability distribution that correlates all the variables of the system. This cannot be realized efficiently, so we must resort to some approximation. As a first approximation, we choose to ignore the cross-cell correlations, and keep only marginal probabilities on the flows

associated to a given cell (we keep a probability distribution that involves the northern and western boundary only). To diminish the effect of these correlations we are neglecting, we let the charge inside a unit cell fluctuate. For each unit cell, we measure all but one of the plaquettes it encloses. This remaining plaquette thus determines the total charge of the unit cell, and indeed we can substitute the corresponding stabilizer generator by a plaquette enclosing the entire unit cell (obtained by multiplying all the plaquette operators contained in the unit cell). This new stabilizer generator represents a renormalized charge.

This procedure is illustrated on Fig. 3(a) where green disks represent plaquettes that are measured and the green circle represents the plaquette that is left fluctuating. This green circle is replaced by the larger, renormalized green disk (on the right) that is used in the next RG step. The white disks on this figure each represent a probability distribution on charge flow, or equivalently a two-qudit probability distribution. Thus, after one round of RG, we are left with a smaller lattice and both renormalized charges and renormalized noise models.

Equation (3) lists a set of generators for all X operators living on a unit cell (see Fig. 3(b) for labelling). This basis will be used to decompose any X-type error contained on the unit cell. These operators are defined in accordance to the renormalization process itself as we now explain. The T_i operators are used to build a representative error with the appropriate defect configuration. Indeed, only the T_i operators of Eq. (3) do not commute with all three plaquette operators in the unit cell (green disks of Fig. 3(a)). Label the defect configuration on a unit cell as $\vec{a} = (a_0, a_1, a_2)$, where a_0 is the charge of the north-west plaquette, a_1 is the charge of the north-east one, and a_2 is the charge of the south-west one. Then, the Pauli operator $t(\vec{a}) = T_0^{a_0} T_1^{a_1} T_2^{a_2}$ creates the defect configuration $\vec{a}$. Moreover, given a defect configuration $\vec{a}$, every potential error has to contain this product in its decomposition on basis Eq. (3) since only the T_i operators do not commute with plaquettes. The L_i operators characterize the flow of charge through the northen and western boundaries, so the two-qudit ouput distribution of a RG round is precisely the probability distribution over these two operators. The S_i operators are stabilizer operators (or parts of stabilizer generators supported on the unit cell). They only deform strings without changing their defect configuration or their associated charge flow. Lastly, the E_i operators correspond to charge flowing through the southern and eastern boundaries into the plaquette operator that is left out. Thus, they are responsible for the charge fluctuation inside the unit cell and they are summed over.

$$\begin{aligned} S_0 &= X_0 X_2^{-1} X_3^{-1} & T_0 &= X_4 X_7^{-1} \\ S_1 &= X_1 X_4^{-1} X_5^{-1} & T_1 &= X_6 \\ S_2 &= X_3 X_4 X_6^{-1} X_7^{-1} & T_2 &= X_7^{-1} \\ E_0 &= X_6 X_8 & L_0 &= X_2 X_6 \\ E_1 &= X_7^{-1} X_9^{-1} & L_1 &= X_5 X_7 \end{aligned} \tag{3}$$

With these definitions, we can formally describe a RG round that starts with a defect configuration $\vec{a}$, and computes the marginal probability of each $l \in \langle L_0, L_1 \rangle$ conditioned on the measured defect configuration,

$$\mathcal{P}(l) = \sum_{e \in \langle E_0, E_1 \rangle} \sum_{s \in \langle S_0, S_1, S_2 \rangle} \mathcal{P}(tles), \tag{4}$$

where $t = T^{a_0}T^{a_1}T^{a_2}$ is given by the defect configuration and $\mathcal{P}(tles)$ is the probability assigned to the error $E = tles$ by the noise model. The complexity of decoding a unit cell is given by the number of operators that are considered in Eq. (4): $|\langle L_0, L_1\rangle| \cdot |\langle E_0, E_1\rangle| \cdot |\langle S_0, S_1, S_2\rangle|$. Since all L_i, E_i and S_i have order d, the complexity is the constant d^7. For different unit cell sizes, the complexity is still a power of d, but with a different exponent which depends on the number of qudits in the cell and the number of measured stabilizer generators. Moreover, the number of unit cells to decode in a given round of RG is given by $(L/k)^2$ where k and L are the linear sizes of the unit cell and the global lattice, respectively. Thus, the complexity of a step of RG goes as $d^c(L/k)^2$ for some constants c and k that depend on the choice of unit cell. Of course, the RG calculations on different cells can be executed in parallel.

The procedure we have described above to evade the correlations caused by local charge conservation is only a heuristic, and can be improved using belief propagation (BP). Roughly, the role of BP is to ensure consistency between the marginal probability of qubits located at the boundary of two or more unit cells, e.g. qudits 0, 1, 8 and 9 (see Fig. 3(b) for labeling). First, given a defect configuration inside a unit cell, one can compute the marginal error probability $\mathcal{P}_q(tles|_q)$ for each qudit q, obtained by taking a marginal of $\mathcal{P}(tles)$. These are called messages and denoted $m_q^{\text{out}}(p)$, where q labels a qudit and p is a one-qudit Pauli operator. These outgoing messages are then exchanged between neighbouring cells, and become incoming messages, e.g. a cell c sends to its northern neighbour c' the message m_0^{out} that becomes m_9^{in} in c', and receives from c' the message m_9^{out} that becomes m_0^{in} in c. Subsequent rounds of messages can be calculated using the received messages, following the prescription

$$m_q^{\text{out}}(p) \leftarrow \sum_{l,s,e} \delta(tles|_q, p) \frac{\mathcal{P}(tles)}{\mathcal{P}_q(tles|_q)} \prod_{q' \neq q} m_{q'}^{\text{in}}(tles|_{q'}), \tag{5}$$

Here, $q, q' \in \{0, 1, 8, 9\}$, $tles|_q$ is the restriction to qudit q of the Pauli operator $tles$ and $\mathcal{P}_q$ is the marginal on qudit q of the noise model as above. BP can be iterated a few times (e.g. three rounds) before executing a RG step. This has the effect of replacing Eq. (4) by

$$\mathcal{P}(l) = \sum_{e \in \langle E_0, E_1\rangle} \sum_{s \in \langle S_0 S_1 S_2\rangle} \mathcal{P}(tles) \prod_q m_q^{\text{in}}(tles|_q). \tag{6}$$

5 Numerical results

In this section, we present our numerical estimates of the thresholds of $\mathbb{Z}_d$-KTCs for $2 \leq d \leq 6$ subject to the generalized bit-flip noise model introduced in the previous section. The threshold is defined as the value of the physical noise rate p_{phys} below which the decoding error probability p_{dec} can be made arbitrarily small by increasing the lattice size L.

The simulations were performed as follows. For various values of d, L and p_{phys}, specifiying a $\mathbb{Z}_d$-KTC of linear size L subject to a noise of parameter p_{phys}, we performed a Monte Carlo simulation to estimate the decoding error probability p_{dec}. We used sample sizes of the order of 10^4. For a fixed value of d, we plotted estimates of p_{dec} vs p_{phys} for different values of L. We then used the fitting model $p_{\text{dec}} = (p_{\text{phys}} - p_{\text{th}})L^{1/\nu}$ (see [4, 10] for more details) to estimate the value of the threshold. As an example, we plotted the results and the fits for $\mathbb{Z}_3$-KTC on Fig. 4.

Repeating this for $3 \leq d \leq 6$ (2 was studied in [5, 6]), Fig. 5 shows p_{th} as a function of d. Heuristically, we did expect that the value of p_{th} increases with d. Indeed, if we imagine

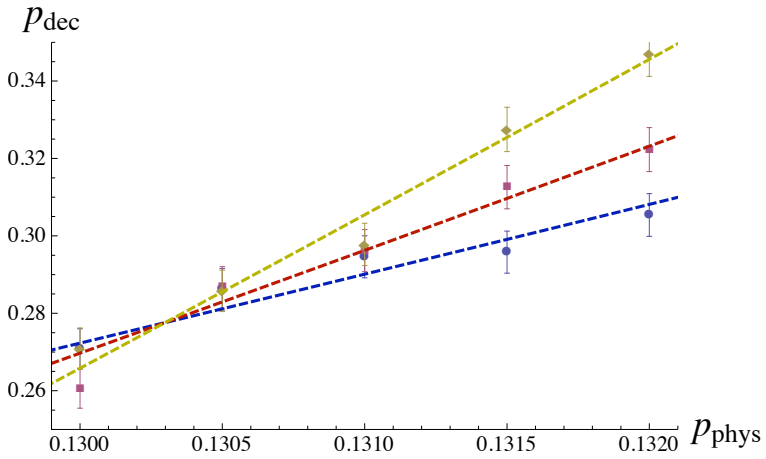


Figure 4 Threshold estimation for $\mathbb{Z}_3$-KTC. The x-axis represents physical error rate and the y-axis, decoding error rate. The blue dots, red squares and yellow diamonds correspond to $L = 32$, $L = 64$ and $L = 128$ respectively. The fitting curve used is $p_{dec} = (p_{phys} - p_{th})L^{1/\nu}$. In this case, we find $p_{th} = 0.13(0)$.

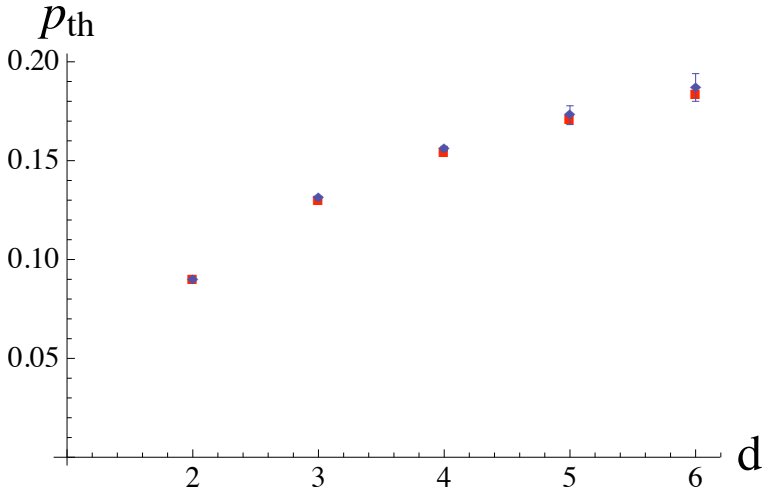


Figure 5 The blue diamonds are the values extracted by fitting the threshold values for $2 \leq n \leq 6$ (see Fig. 4 for example). The red squares are obtained via the generalized hashing bound (see text) rescaled by a common factor $\alpha = p_{th}(2)/C_2 \approx 0.81$. The error bars are (pessimistically) obtained e.g. by replacing each line in Fig. 4 by a stripe of width equal to the statistical error bars, and determining the values of p_{phys} above and below the crossing point where the strips cease to overlap. We do not report the fitting parameter ν because they are too sensitive to statistical fluctuations and therefore unreliable in our study.

simulating a qudit using $\log_2 d$ qubits, a fixed noise rate for increasing values of d translates into a decreased noise rate per qubit. Moreover, it was reported in [1] that the performance of BP for $\mathbb{Z}_d$-KTC, which is very poor in the qubit case, is greatly increased as d grows.

It is intringuing to note that for $\mathbb{Z}_2$-KTC subject to bit-flip or depolarizing noise, p_{th} is numerically very close to the hashing bound [4, 10, 2]. The hashing bound, obtained by a simple packing argument [8], states that for non-degenerate CSS codes,

$$0 \leq 1 - 2H_2(p), \tag{7}$$

where H_2 is the binary entropy: $H_2(p) = (1-p)\log_2(1-p) + p\log_2 p$. From Eq. (7), one can calculate the saturating point $C_2 \approx 0.110$ which is indeed quite close to the optimal threshold of the $\mathbb{Z}_2$-KTC subject to independent bit-flip and phase-flip errors, $p_{th}(2) \approx 0.109(4)$ [4, 10]. This near coincidence is intriguing given that topological codes are highly degenerate, so there is no reason they should obey the hashing bound. Of course, the decoder we are using here is sub-optimal, so the threshold we find $p_{th}(2) \approx 0.89(6)$ is a smaller fraction $\alpha = p_{\text{th}}(2)/C_2 \approx 0.81(4)$ of the hashing bound.

For qudits, the hashing bound is

$$0 \leq 1 - 2H_d(p) \quad \text{with} \quad H_d(p) = (1-p)\log(1-p) + p\log\frac{p}{d-1}. \tag{8}$$

In this case, we find $C_3 \approx 0.159, C_4 \approx 0.189$ and so on. Figure 5 shows the threshold $p_{\text{th}}(d)$ obtained with the RG decoder as well as a rescaled hashing bound αC_d where α is determined by the $\mathbb{Z}_2$ fit. The agreement is both unexplained and surprisingly good. Note also that even though our decoder is sub-optimal, $p_{th}(d+1) > C_d$ for all d we have studied, which strongly support the claim that the threshold increases with d.

6 Conclusion

In this paper, we presented a generalization of the renormalization group decoder of [5, 6] to Kitaev topological codes built with the groups $\mathbb{Z}_d$. Our numerical results show that the threshold value increases as a function of the local dimension d. Moreover, its behaviour is in very good agreement with a scaling predicted by the hashing bound. This trend could be confirmed by more accurate numerical estimates using a mapping to a statistical mechanics model, which does not require solving the decoding problem [4, 2]. A theoretical understanding of this behavior is also desirable. Lastly, estimating the threshold in the presence of measurement error and detailed syndrome measurement circuits on qudits remains an interesting open question.

Acknowledgements. We would like to thank Jonas Anderson for useful discussions regarding the generalized hashing bound. We also thank Simon Burton, Courtney Brell and Stephen Bartlett for enlightening discussions of Kitaev's construction [11]. Computational resources were provided by Calcul Québec and Compute Canada. This work was partially funded by NSERC and by Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior National Business Center contract D11PC20167. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright annotation thereon. Disclaimer: The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of IARPA, DoI/NBC, or the U.S. Government.

References

1 I. Andriyanova, D. Maurice, and J.-P. Tillich. New constructions of CSS codes obtained by moving to higher alphabets. *ArXiv e-prints*, 2012.

2 H. Bombin, R. S. Andrist, M. Ohzeki, H. G. Katzgraber, and M. A. Martin-Delgado. Strong Resilience of Topological Codes to Depolarization. *Physical Review X*, 2(2):021004, 2012.

3 S. Bravyi and J. Haah. Analytic and numerical demonstration of quantum self-correction in the 3D Cubic Code. *ArXiv e-prints*, 2011.

4 E. Dennis, A. Kitaev, A. Landahl, and J. Preskill. Topological quantum memory. *Journal of Mathematical Physics*, 43:4452, 2002.

5 G. Duclos-Cianci and D. Poulin. Fast decoders for topological quantum codes. *Physical Review Letters*, 104:050504, 2009.

6 G. Duclos-Cianci and D. Poulin. A renormalization group decoding algorithm for topological quantum codes. *Information Theory Workshop (ITW)*, page 1, 2010.

7 J. Edmonds. Paths, trees and flowers. *Canadian Journal of Mathematics*, 17:449–467, 1965.

8 A. Ekert and C. Macchiavello. Error Correction in Quantum Communication. *eprint arXiv:quant-ph/9602022*, 1996.

9 D. Gottesman. Fault-Tolerant Quantum Computation with Higher-Dimensional Systems. *Chaos, Solitons, and Fractals*, 10:1749–1758, 1999.

10 J.W. Harrington. Qecc : symplectic lattice and toric codes. *PhD Thesis - California Insitute of Technology*, 2004.

11 A. Kitaev. Fault-tolerant quantum computation by anyons. *ANNALS PHYS.*, 303:2, 2003.

12 E. Knill. Non-binary Unitary Error Bases and Quantum Codes. *eprint arXiv:quant-ph/9608048*, 1996.

Optimal Quantum Circuits for Nearest-Neighbor Architectures

David J. Rosenbaum

University of Washington
Department of Computer Science & Engineering
djr@cs.washington.edu

Abstract

We show that the depth of quantum circuits in the realistic architecture where a classical controller determines which local interactions to apply on the kD grid $\mathbb{Z}^k$ where $k \geq 2$ is the same (up to a constant factor) as in the standard model where arbitrary interactions are allowed. This allows minimum-depth circuits (up to a constant factor) for the nearest-neighbor architecture to be obtained from minimum-depth circuits in the standard abstract model. Our work therefore justifies the standard assumption that interactions can be performed between arbitrary pairs of qubits. In particular, our results imply that Shor's algorithm, controlled operations and fanouts can be implemented in constant depth, polynomial size and polynomial width in this architecture.

We also present optimal non-adaptive quantum circuits for controlled operations and fanouts on a kD grid. These circuits have depth $\Theta(\sqrt[k]{n})$, size $\Theta(n)$ and width $\Theta(n)$. Our lower bound also applies to a more general class of operations.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases 2D, Nearest Neighbor, Quantum Architecture, Quantum Complexity, Quantum Computation

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.294

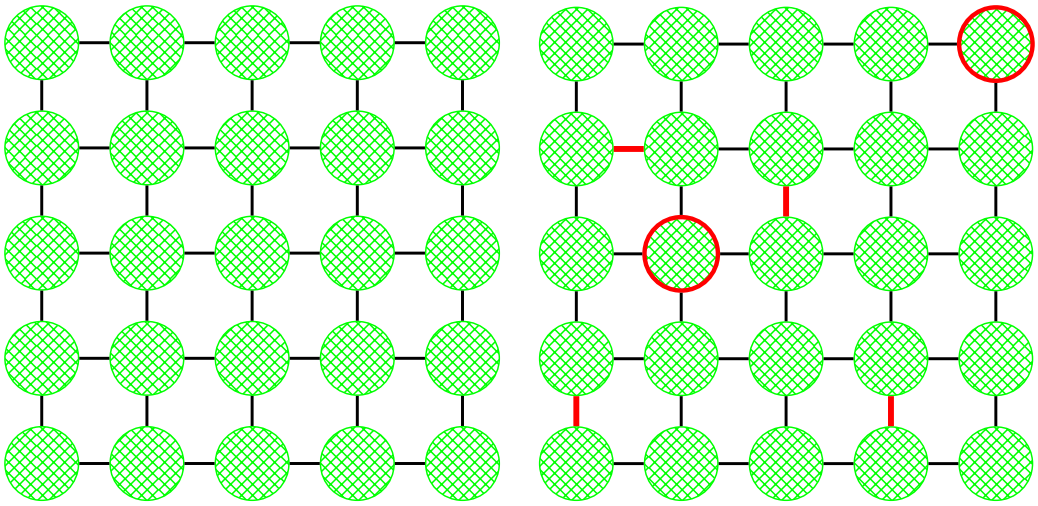
1 Introduction

Quantum algorithms are typically formulated at an abstract level and allow arbitrary one- and two-qubit interactions. However, in physical implementations of quantum computers, typically only local interactions between neighboring qubits are possible. This motivates the kD *nearest-neighbor two-qubit concurrent* (kD NTC) architecture [19] (cf. [5]) in which the qubits are arranged on the kD grid $\mathbb{Z}^k$; this is shown in Figure 1a for the case where $k = 2$. Operations may involve one or two qubits with the restriction that two-qubit operations may only be performed along an edge in the grid. Multiple operations may be performed concurrently as long as they are on disjoint sets of qubits; an example is shown in Figure 1b.

The idea of using a classical controller to determine which operations to apply at each step is implicit in the pre- and post-processing stages of Shor's algorithm [16] and is often assumed for fault-tolerant quantum computation. Since the classical controller can take intermediate measurement outcomes into account, this model includes the class of adaptive quantum circuits as a special case. It is potentially even more powerful since the classical controller can perform randomized polynomial-time computations to determine which operations to apply as well as perform pre- and post-processing. Since quantum operations are far more expensive than classical operations, we are primarily concerned with the depth of the quantum circuit and do not count the operations performed by the classical controller as long as they take polynomial time.

8th Conference on Theory of Quantum Computation, Communication and Cryptography.
Editors: Simone Severini and Fernando Brandao; pp. 294–307


Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany



(a) Interactions in the 2D NTC architecture: the grid lines indicate the two-qubit interactions which can be performed

(b) An example of concurrent interactions in the 2D NTC architecture: the components connected by the thick red edges indicate concurrent interactions and the thick red circles indicate single-qubit interactions

Figure 1 The 2D NTC architecture.

In this work, we study both the *classical-controller* kD *NTC* (kD CCNTC) architecture — a classical controller model where interactions are restricted to a kD grid — as well as the *non-adaptive* kD *NTC*[1] (NANTC) architecture where no classical controller is used and the operations applied cannot depend on intermediate measurement outcomes. The CCNTC model ignores the cost of offline computations performed by the classical controller and assumes that there are no classical locality restrictions. This is realistic since the clock rate for a classical computer is much faster than for a quantum computer. Because quantum computers are already forced to be parallel devices in order to perform operations fault tolerantly [1], the total runtime of a quantum circuit is proportional to the depth of the corresponding quantum circuit. The restriction that interactions are between neighbors on a kD grid comes from the underlying physical device: in most technologies, only qubits that are spatially close can interact.

We first show how to simulate the standard *classical controller abstract concurrent* (CCAC) architecture in kD CCNTC with constant factor overhead in the depth. We accomplish this using a 2D CCNTC teleportation scheme that allows arbitrary interactions on disjoint sets of qubits to be performed in constant depth.

▶ **Theorem 1.1.** *Suppose that C is a CCAC quantum circuit with depth d, size s and width n. Then C can be simulated in $O(d)$ depth, $O(sn)$ size and n^2 width in* 2D *CCNTC.*

This result justifies the standard assumption that non-local interactions can be performed efficiently. Simulating each of the d timesteps from the CCAC circuit in 2D CCNTC requires an $O(n)$ time classical computation; this can be reduced to $O(\log n)$ time if the classical controller is a parallel device or if it includes a simple classical circuit. Since the clock

[1] The original NTC architecture described by Van Meter and Itoh [19] is in fact NANTC; however, we prefer NANTC to avoid confusion with CCNTC where a classical controller is used.

speeds of classical devices are currently much faster than those of quantum devices, this overhead is not likely to be significant.

▶ **Corollary 1.2.** *Let $\mathcal{E}$ be a quantum operation on n qubits. Let d_1 and d_2 be the minimum depths*[2] *required to implement $\mathcal{E}$ with error at most ϵ using* $\mathsf{poly}(n)$ *size and* $\mathsf{poly}(n)$ *width in the CCAC and k*D *CCNTC models respectively where $k \geq 2$. Then $d_1 = \Theta(d_2)$.*

It is possible to implement Shor's algorithm [16] in constant depth in CCAC [3] which implies that it can also be implemented in constant depth in 2D CCNTC.

▶ **Corollary 1.3.** *Shor's algorithm can be implemented in constant depth, polynomial size and polynomial width in* 2D *CCNTC.*

Since controlled-U operations and fanouts can also be performed in constant depth and polynomial width in CCAC [8, 3, 17], we also have the following corollary.

▶ **Corollary 1.4.** *Controlled-U operations with n controls and fanouts with n targets can be implemented in constant depth,* $\mathsf{poly}(n)$ *size and* $\mathsf{poly}(n)$ *width in* 2D *CCNTC.*

Our main technical result allows any subset of qubits to be reordered in constant depth. Theorem 1.1 follows from this as a corollary.

▶ **Theorem 1.5.** *Suppose we have an $n \times n$ grid where all qubits except those in the first column are in the state $|0\rangle$. Let $T \subseteq \{0, \ldots, n-1\}$ and let $\pi : T \to \{0, \ldots, n-1\}$ be an injection such that for all $j \in T$ with $\pi(j) = 0$, $\{k \in T^c \mid k < j\} = \emptyset$. Set $m = |\{j \in T \mid \pi(j) \neq 0\}|$. Then we can move each qubit at $(0, j)$ to $(\pi(j), 0)$ for all $j \in T$ in $O(1)$ depth, $O(mn)$ size and $(m+1)n \leq n^2$ width in* 2D *CCNTC.*

Upper bounds for the depth of quantum circuits when converting between various architectures with no classical controller were previously studied by Cheung, Maslov and Severini [4]. Their results imply that CCAC can be simulated in kD CCNTC with $O(\sqrt[k]{n})$ factor depth overhead, $O(n)$ size overhead and no width overhead. In contrast to our results, their techniques are based on applying swap gates to move the interacting qubits next to each other and do not perform any measurements.

Implementations of Shor's algorithm in kD CCNTC with various super-constant depths were previously known for $k = 1$ and $k = 2$. Fowler, Devitt and Hollenberg [7] showed a 1D CCNTC circuit for Shor's algorithm which requires $O(n^3)$ depth, $O(n^4)$ size and $O(n)$ width where n is the number of bits in the integer which is being factored. Maslov [10] showed that any stabilizer circuit can be implemented in linear depth in 1D CCNTC from which the result of Fowler, Devitt and Hollenberg [7] can be recovered. Kutin [9] gave a more efficient 1D CCNTC circuit which uses $O(n^2)$ depth, $O(n^3)$ size and $O(n)$ width. For 2D CCNTC, Pham and Svore [12] showed an implementation of Shor's algorithm in polylogarithmic depth, polynomial size and polynomial width.

It was also previously known that controlled-U operations and fanouts can be implemented in constant depth, polynomial size and polynomial width in CCAC. This line of work was started by Moore [11] who showed that parity and fanout are equivalent and posed the question of whether fanout has constant-depth circuits. Høyer and Špalek [8] proved that if fanout has constant-depth circuits then controlled-U operations can also be implemented

[2] Here, we assume that there is a minimum depth required to implement $\mathcal{E}$ in CCAC when the size and width are $\mathsf{poly}(n)$.

in constant depth with inverse polynomial error. Browne, Kashefi and Predrix [3] showed that one-way quantum computation is equivalent to unitary quantum circuits with fanout. A consequence of this is that constant depth adaptive circuits for fanout can be used to implement controlled-U operations in constant depth in CCAC. Takahashi and Tani [17] reduced the size of this circuit by a polynomial and made it exact.

In many technologies, measurements are much more costly than unitary operations. For this reason, we also consider the non-adaptive kD NANTC model. Here, there is no classical controller and the operations applied depend only on the size of the input and not on intermediate measurement outcomes. Our result in this model is a characterization of the complexity of controlled-U operations and fanouts.

▶ **Theorem 1.6.** *The depth required for controlled-U operations with n controls and fanouts with n targets in kD NANTC is $\Theta(\sqrt[k]{n})$. Moreover, this depth can be achieved with size $\Theta(n)$ and width $\Theta(n)$.*

If the clock speeds of the quantum computer and its classical controller are comparable, then operations implemented using Theorem 1.6 are significantly faster than those implemented using Corollary 1.4. For this reason, Theorem 1.6 may become a better option as quantum computing technology matures.

The layout of our paper is as follows. In Section 2, we discuss definitions used in the rest of the paper and define the models of computation precisely. In Section 3, we review quantum teleportation and describe teleportation chains. In Section 4, we describe our 2D teleportation scheme and show that it allows arbitrary interactions to be implemented in constant depth in 2D CCNTC. In Section 5, we show an algorithm that implements controlled-U operations and fanouts for kD NANTC in depth $O(\sqrt[k]{n})$. In Section 6, we describe how our techniques can be applied to obtain kD NANTC quantum circuits for fanout with depth $O(\sqrt[k]{n})$. In Section 7, we prove a matching lower bound for a class of operations that includes controlled-U operations and fanouts.

2 Definitions

The one- and two-qubit operations that can be performed by the hardware are called the *basic operations*. We assume that the basic operations are a *universal gate set* so that any one- or two-qubit unitary can be constructed from the basic operations. We also assume that the basic operations include measurement in the computational basis.

It is useful to distinguish between physical and logical timesteps. During each *physical timestep*, we can perform any set of disjoint basic operations. During a *logical timestep*, we allow any set of disjoint t-qubit operations to be performed. In this work, we take $t = O(k)$ and assume k is constant.

▶ **Definition 2.1** (NANTC). *In the kD NANTC model, computation is performed by applying a sequence of sets of basic operations $S_1, \ldots, S_d$ to the kD grid of qubits. We require that the operations in the set S_i are disjoint and are either single-qubit operations or two-qubit operations between neighbors in the kD grid. The sequence of sets of operations must be randomized polynomial-time computable from the size n of the input.*

In the models where a classical controller is present, the classical controller is invoked after each physical timestep to determine which operations to apply at the next step.

▶ **Definition 2.2** (CCAC). *Let M be a randomized polynomial-time machine that takes the input x and the measurement outcomes from the first i physical timesteps and outputs a*

set $M_1, \ldots, M_\ell$ of disjoint basic operations to be applied to the qubits at the $i+1^{\text{th}}$ physical timestep. If no more physical timesteps are to be performed, then M outputs the special symbol $\boxdot$. Computation in the CCAC model is performed at physical timestep i by using M to compute the set of operations to apply and then applying them to the qubits.

The CCNTC model is similar except that it also requires that two-qubit operations are only performed between neighbors on the kD grid.

▶ **Definition 2.3** (CCNTC). *Let M be a randomized polynomial-time machine that takes the input x and the measurement outcomes from the first i physical timesteps and outputs a set $M_1, \ldots, M_\ell$ of disjoint basic operations to be applied to the kD grid of qubits at the $i+1^{\text{th}}$ physical timestep. We require that each M_i is either a single-qubit operation or a two-qubit operation between neighbors in the kD grid. If no more physical timesteps are to be performed, then M outputs the special symbol $\boxdot$. Computation in the CCNTC model is performed at physical timestep i by using M to compute the set of operations to apply and then applying them to the kD grid of qubits.*

In this paper, the machine M from Definitions 2.2 and 2.3 will be deterministic except for the pre- and post-processing stages of Shor's algorithm.

For NANTC, a *quantum circuit* is the sequence of basic operations $M_1, \ldots, M_\ell$ be applied to the kD grid of qubits. For the CCAC and CCNTC models, a *quantum circuit* is described by the machine M from Definitions 2.2 and 2.3. We now define three standard measures of cost in these models.

▶ **Definition 2.4.** *The depth of a quantum circuit is*

(a) *d for NANTC where $S_1, \ldots, S_d$ is the sequence of operations from Definition 2.1 for an input of size n*

(b) *$\max_{x\in\{0,1\}^n} \max_r d_{x,r}$ for CCAC and CCNTC where $d_{x,r}$ is the number of physical timesteps it takes for the machine M from Definitions 2.2 and 2.3 to output $\boxdot$ when the input is x and the random seed is r. The first max is taken is over all possible inputs x of length n and the second is over all possible random seeds r.*

We note that the depth only changes by a constant factor if we use logical timesteps instead of physical timesteps in the above definition. This is due to our assumption that any operation performed in a logical timestep acts on at most $O(k) = O(1)$ qubits.

▶ **Definition 2.5.** *The size of a quantum circuit is*

(a) *$\sum_i |S_i|$ for NANTC where $S_1, \ldots, S_d$ is the sequence of operations from Definition 2.1 for an input of size n*

(b) *$\max_{x\in\{0,1\}^n} \max_r s_{x,r}$ for CCAC and CCNTC where $S_{x,r}$ is the total number of operations applied when the input is x and the random seed is r. The first max is taken over all possible inputs x of length n and the second is over all possible random seeds r.*

In the next definition, we assume that the qubits are indexed by $\mathbb{N}$ for CCAC.

▶ **Definition 2.6.** *The width of a quantum circuit is*

(a) *the total number of qubits acted on by operations in the sets S_i for NANTC where $S_1, \ldots, S_d$ is the sequence of operations from Definition 2.1 for an input of size n*

(b) *$\max_{x\in\{0,1\}^n} |A_x|$ for CCAC where A_x is the smallest subset of $\mathbb{N}$ such that every qubit acted on is contained in A_x for input x and all random seeds r*

(c) *$\max_{x\in\{0,1\}^n} |A_x|$ for CCNTC where A_x is the smallest hypercube in $\mathbb{Z}^k$ such that every qubit acted on is contained in A_x for input x and all random seeds r*

Typically, the depth is the most important metric to optimize since it is proportional to the amount of time required to execute the quantum operations. The width is also important since the number of qubits is currently quite limited but the size is largely irrelevant. Moreover, if parallelism is properly exploited then we expect the size to be roughly the depth times the width.

3 Quantum teleportation

In this section we review quantum teleportation [2]. As we shall see, teleportation is a useful primitive that allows non-local interactions to be performed in a constant-depth circuit in kD CCNTC. Let us denote the states of the Bell basis by $|\Phi_0\rangle = \frac{|00\rangle+|11\rangle}{\sqrt{2}}$, $|\Phi_1\rangle = \frac{|01\rangle+|10\rangle}{\sqrt{2}}$, $|\Phi_2\rangle = \frac{|01\rangle-|10\rangle}{\sqrt{2}}$ and $|\Phi_3\rangle = \frac{|00\rangle-|11\rangle}{\sqrt{2}}$. Up to global phase, these can be written as $|\Phi_\ell\rangle^{AB} = \sigma_\ell^B |\Phi_0\rangle^{AB}$. Recall that in the quantum teleportation setting, Alice has a state $|\psi\rangle^S = \alpha|0\rangle^S + \beta|1\rangle^S$ that she wishes to send to Bob. The two parties are not allowed to send quantum states to each other but each have one qubit of a Bell state $\sigma_\ell^B |\Phi_0\rangle$ and can communicate classically.

To perform quantum teleportation, Alice performs a Bell measurement on the SA registers. If the measurement outcome is $|\Phi_k\rangle$, then a simple calculation shows that the resulting state is $|\Phi_k\rangle^{SA} \otimes \sigma_\ell\sigma_k |\psi\rangle^B$. Alice then sends the classical measurement outcome k to Bob; by applying the appropriate Pauli operation to his register B, Bob causes to overall state to become $|\Phi_k\rangle^{SA} \otimes |\psi\rangle^B$. Observe that Alice's state $|\psi\rangle$ has been recovered in Bob's register.

Let us now consider how quantum teleportation chains can be used in the 1D CCNTC model to perform non-local operations in constant depth. Suppose that we have a qubit in the state $|\psi\rangle^S$ along with m Bell states $|\Phi_{\ell_j}\rangle^{A_jB_j}$. These are arranged on a line so that the overall state is $|\psi\rangle^S \bigotimes_{j=1}^m |\Phi_{\ell_j}\rangle^{A_jB_j}$. Our goal is to move qubit S to B_m. One way to do this is to first teleport S to B_1 by performing a Bell measurement on SA_1. We then store the measurement outcome k_1 but do not apply the correcting Pauli operation; at this point, the state of B_1 is $\sigma_{\ell_1}\sigma_{k_1}|\psi\rangle$. Continuing this process, we obtain the state $\bigotimes_{j=1}^m |\Phi_{k_j}\rangle \prod_{j=m}^1 (\sigma_{\ell_j}\sigma_{k_j}) |\psi\rangle^{B_m}$. Since $\prod_{j=m}^1 (\sigma_{\ell_j}\sigma_{k_j})$ is just a Pauli operation, we obtain the state $\bigotimes_{j=1}^m |\Phi_{k_j}\rangle |\psi\rangle^{B_m}$ in a single quantum operation. The crucial point here is that all of the Bell measurements are performed on disjoint pairs of qubits so they can all be done in parallel as in one-way quantum computation [14, 13] and [18]. Thus, we can perform a non-local interaction of arbitrary distance in constant depth. It is important to note that this is not possible without a classical controller since otherwise there is no way to compute the correcting Pauli operation.

4 Depth complexity in kD CCNTC

In this section, we show that an arbitrary set of CCAC interactions corresponding to basic operations can be performed in constant depth in 2D CCNTC. We assume that there are n qubits on which the interactions are to be performed and store these in the first column of a 2D $n \times n$ CCNTC grid. Since we must handle interactions between qubits that are not neighbors, we may as well assume that the original n qubits are stored in the first column. The remaining columns are used as ancillas to implement teleportation chains. We teleport each of the n qubits horizontally to the right so that interacting pairs are in adjacent columns. Since these teleportations are on disjoint sets of qubits, they can be performed in parallel as in [14, 13, 18]. A second set of vertical teleportation chains is then used to move all the qubits down to the first row. At this point, the interacting qubits are neighbors so

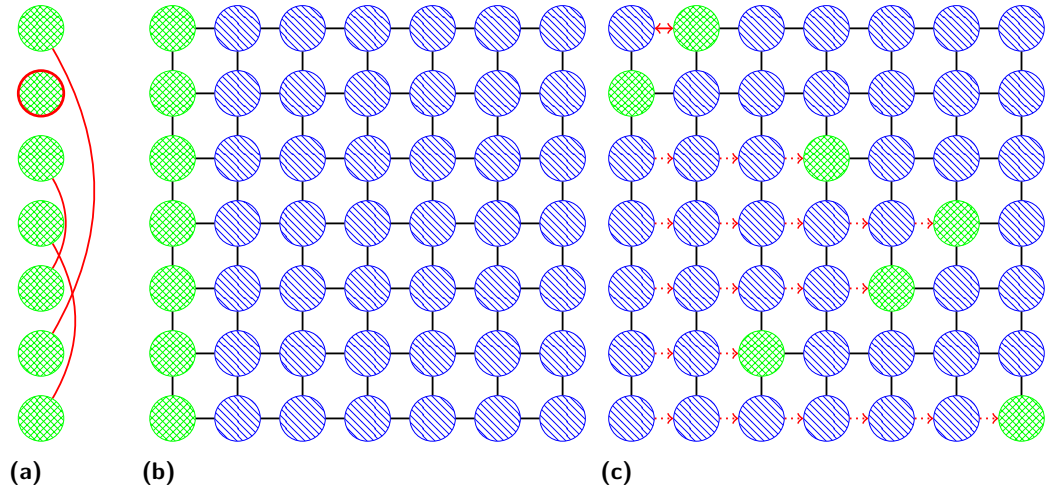


Figure 2 Performing an arbitrary set of interactions in 2D CCNTC. The qubits crosshatched green are the data qubits and the qubits shaded with diagonal downward blue lines are ancilla qubits.

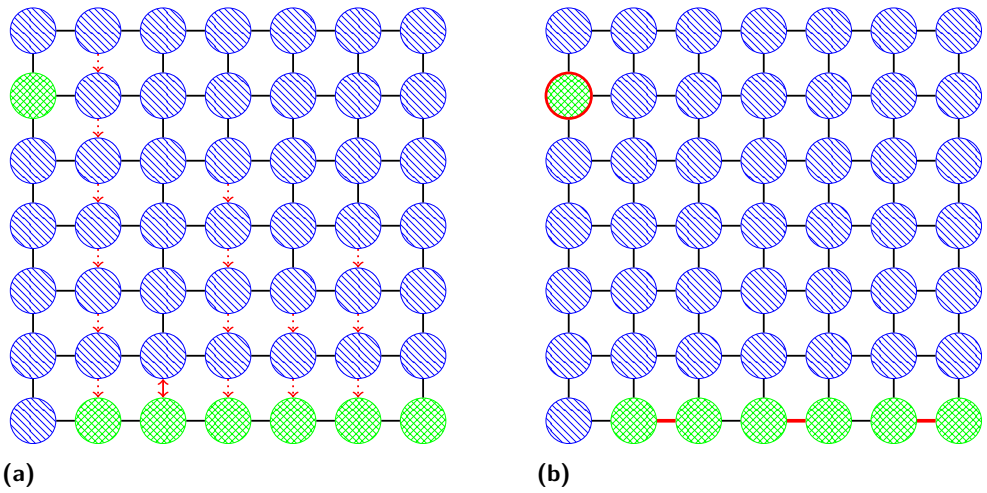


Figure 3 Performing an arbitrary set of interactions in 2D CCNTC.

the interactions may be implemented directly. We then perform the reverse teleportations to move the qubits back to their original positions.

4.1 An example of arbitrary interactions in 2D CCNTC

We show an example in Figure 2. The desired interactions are shown in Figure 2a. The layout of the data qubits in the 2D grid is shown in Figure 2b; the ancilla qubits are used to implement the teleportation chains and are initially set to $|0\rangle$. We start by horizontally teleporting the qubits that interact to adjacent columns in Figure 2c where the teleportation chains are denoted by the dotted red arrows. The red double arrow indicates a swap operation; this is just a less expensive way of achieving the same result when the qubits are neighbors. The next step is to vertically teleport the data qubits down to the first row as shown in Figure 3a. Finally, all interacting qubits are now neighbors so we perform the

desired interactions in Figure 3b. The final reverse teleportations are not shown but can be obtained by reversing the arrows in Figures 2c and 3a.

4.2 An algorithm for performing arbitrary interactions in 2D CCNTC

It is easy to generalize the approach of Figure 2 to show that the qubits in the first column can be reordered arbitrarily in constant depth. The pseudocode is the obvious generalization of Figure 2 (see the full version of our paper [15]).

▶ **Theorem 1.5.** *Suppose we have an $n \times n$ grid where all qubits except those in the first column are in the state $|0\rangle$. Let $T \subseteq \{0, \ldots, n-1\}$ and let $\pi : T \to \{0, \ldots, n-1\}$ be an injection such that for all $j \in T$ with $\pi(j) = 0$, $\{k \in T^c \mid k < j\} = \emptyset$. Set $m = |\{j \in T \mid \pi(j) \neq 0\}|$. Then we can move each qubit at $(0, j)$ to $(\pi(j), 0)$ for all $j \in T$ in $O(1)$ depth, $O(mn)$ size and $(m+1)n \leq n^2$ width in* 2D *CCNTC.*

We note that the teleportation chains in our scheme require an $O(n)$ time classical computation to determine the correcting Pauli matrix (see Section 3). Since this computation simply involves multiplying $O(n)$ Pauli matrices, it can be done more efficiently in $O(\log n)$ time by arranging the multiplications in a binary tree. The $O(\log n)$ runtime requires either that the classical controller is a parallel device or that it includes a special classical circuit for computing the correcting Pauli operation. Since classical operations are much faster than quantum operations on current devices, this overhead is unlikely to be a problem.

From this, it follows that any set of one- and two-qubit operations on disjoint sets of qubits can be performed in constant depth in 2D. This implies that any CCAC circuit can be simulated with constant factor depth overhead in 2D CCNTC.

▶ **Theorem 1.1.** *Suppose that C is a CCAC quantum circuit with depth d, size s and width n. Then C can be simulated in $O(d)$ depth, $O(sn)$ size and n^2 width in* 2D *CCNTC.*

The rest of our results for kD CCNTC follow from Theorem 1.1. Let $\mathcal{D}_n$ denote the set of all $n \times n$ density matrices. A general quantum operation is represented as a completely positive trace preserving (CPTP) map $\mathcal{E} : \mathcal{D}_n \to \mathcal{D}_n$. Obviously, any circuit in the 2D CCNTC model can also be applied when arbitrary interactions are allowed. The following corollary is immediate.

▶ **Corollary 1.2** (continuing from p. 296)**.** *Let $\mathcal{E} : \mathcal{D}_n \to \mathcal{D}_n$ be a CPTP map and let $\epsilon \geq 0$. Let d_1 and d_2 be the minimum depths required to implement $\mathcal{E}$ with error at most ϵ in the CCAC and kD CCNTC models respectively where $k \geq 2$. Then $d_1 = \Theta(d_2)$.*

It is known that Shor's algorithm can be implemented in constant depth, polynomial size and polynomial width in CCAC [3] from which we obtain another corollary.

▶ **Corollary 1.3.** *Shor's algorithm can be implemented in constant depth, polynomial size and polynomial width in* 2D *CCNTC.*

Because controlled-U operations and fanouts with unbounded numbers of control qubits or targets can be performed in constant depth, polynomial size and polynomial width in CCAC [8, 3, 17], we have the following result.

▶ **Corollary 1.4.** *Controlled-U operations with n controls and fanouts with n targets can be implemented in constant depth,* $\mathsf{poly}(n)$ *size and* $\mathsf{poly}(n)$ *width in* 2D *CCNTC.*

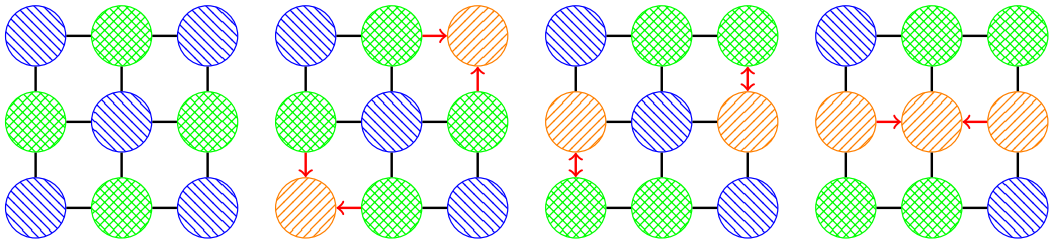


Figure 4 A controlled operation on a 3×3 grid. The qubits crosshatched green are the data qubits, the qubits shaded with diagonal upward orange lines are ancilla qubits which store intermediate data and the qubits shaded with diagonal downward blue lines are ancilla qubits which are currently unused.

5 Controlled operations in kD NANTC

In this section, we show how to control a single-qubit U operation by n controls using $O(\sqrt[k]{n})$ operations in kD NANTC. We start with an $m \times m$ grid; for reasons that will become clear later, we require that m is odd. The control qubits are placed such that they are not at adjacent grid points; the central 3×3 square has no controls except when $m = 3$. This is illustrated in Figures 4a and 5a for the cases where $m = 3$ and $m = 5$. Let c be the center of the grid which corresponds to the target qubit. The circuit works by considering each square ring in the grid with center c (i.e., a set of points in the grid that all have the same distance to the center under the ℓ_∞ norm). We start with the outermost such ring and propagate its control values into the next ring. At each such step, some of the control values are combined so that all the values can fit into the smaller ring. This continues until we reach a 3×3 ring at which point we apply a special sequence of operations to finish applying the controlled operation to the central qubit. Each stage can be implemented in constant depth so the overall depth is $O(\sqrt{n})$.

5.1 The base case: the 3×3 grid

We now describe how this circuit works in greater detail. First, consider the case where $m = 3$. The grid starts as shown in Figure 4a; note that we do not force the central 3×3 square to be devoid of controls in this case since this is the entire grid. All ancilla qubits start in the state $|0\rangle$. We start by setting the lower left and upper right corner ancilla qubits to the ANDs of their neighboring controls as shown in Figure 4b. Both of these operations are disjoint, so this can be done in one logical timestep. The next step is to swap these two corner qubits with the vertical middle qubits so they can interact with the central target qubit; this is done in Figure 4c. Finally, we apply a U operation to the target qubit and control by the two middle qubits in Figure 4d.

At this point, the target qubit has the desired value; however, there are two other ancilla qubits in Figure 4d that must have their values uncomputed. This is done by applying the operations of Figures 4b–c in reverse order.

5.2 An example of the general case: the 5×5 grid

We now consider an example of the general case where $m = 5$ as shown in Figure 5a. The first step is to propagate the values of the outer ring inwards; since the inner ring is 3×3,

there are no controls in the inner ring so this can be done as shown in Figure 5b. We then rotate the inner ring as in Figure 5c. At this point, the remaining operations to perform are the same as in the 3×3 case and are shown in Figures 5d–f. At this point the target qubit has the desired value so we uncompute the intermediate ancillas by applying the operations of Figures 5b–e in reverse order.

The same idea applies to an $m \times m$ grid except that when the inner rings have controls (i.e. for $m \geq 7$), the controls from the outer ring must be combined with those in the inner ring at the same time they are propagated inwards. See the full version of our paper [15] for examples of the 7×7 and 9×9 cases.

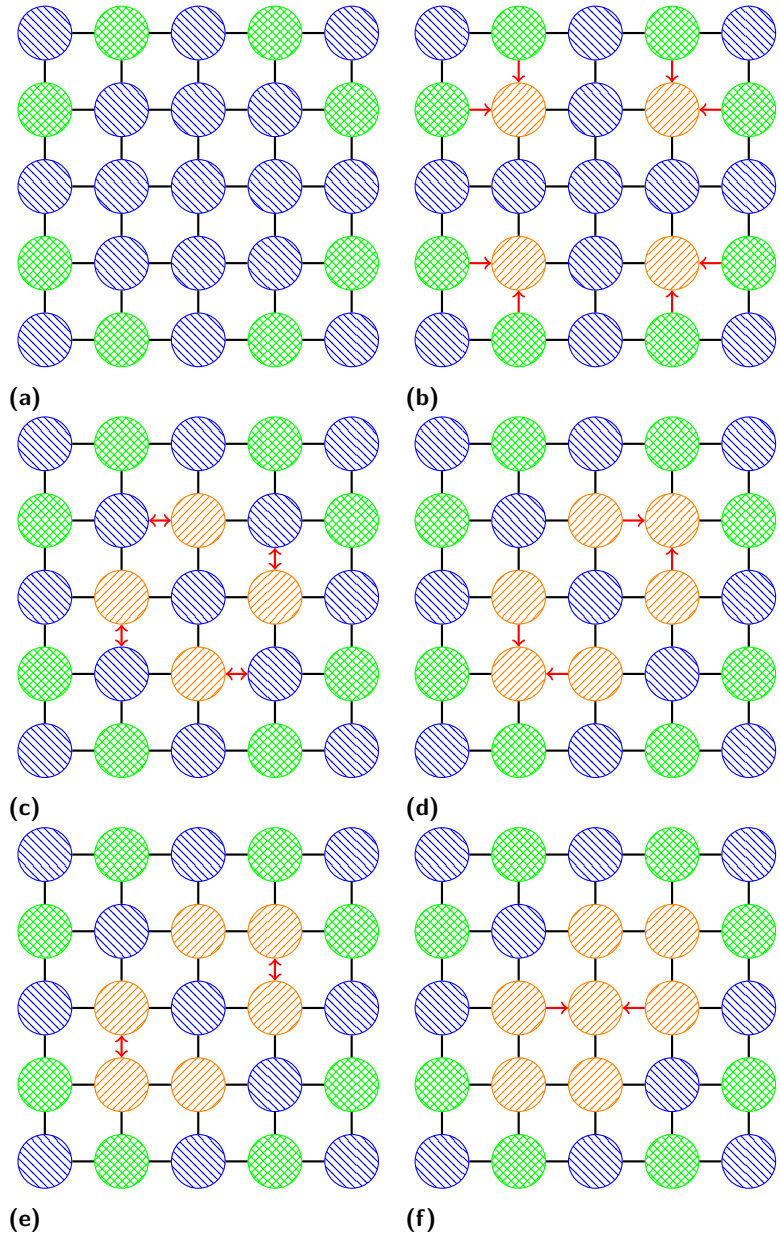


Figure 5 A controlled operation on a 5×5 grid. See Figure 4 for the meaning of the colors and shading used.

5.3 An algorithm for controlled-U operations in $O(\sqrt{n})$ depth in 2D NANTC

We now present the algorithm used in Figures 4 – 5 for the general $m \times m$ grid. Consider an odd $m > 3$. We denote the coordinates of the qubits on this grid by (x, y) where $0 \leq x, y < m$. Let G be the set $\{0, \ldots, m-1\}^2$ of all points on the grid and let $c = ((m-1)/2, (m-1)/2)$ be the central point. As discussed previously, the geometry induced by the ℓ_∞ norm is useful for reasoning about this grid. From now on, all distances in this subsection are understood to be with respect to the ℓ_∞ norm.

We will say that the k^{th} *ring* is the set of points that have distance $(m-1)/2 - k$ to c so the zeroth ring is outermost; we denote by $R_k = (r_0^k, \ldots, r_{\ell_k}^k)$ the points of the k^{th} ring where r_0^k is the bottom left corner and the rest of the points are in clockwise order.

The ring R_k contains $4\left(\frac{m-1}{2} - k\right)$ controls so the entire grid has $n = 4\sum_{3<m-2k\leq m}\left(\frac{m-1}{2} - k\right) = (1/2)(m^2 - 9/2)$ controls for $m > 3$. In the case where $m = 3$, there are 4 controls. Thus, it is indeed the case that the depth is $O(\sqrt{n})$.

Writing out the explicit pseudocode is straightforward (see the full version of our paper [15]).

From this, we obtain the following theorem.

▶ **Theorem 5.1.** *Controlled-U operations with n controls have depth $O(\sqrt{n})$, size $O(n)$ and width $O(n)$ in* 2D *NANTC.*

5.4 Generalization to kD NANTC

In this section, we discuss how the circuit can be generalized to k dimensions. The algorithm works in the same way except the ring R_k is replaced by the grid points on the surface of the hypercube formed by the points at ℓ_∞ distance $(m-1)/2 - k$ from the center c of the grid. We proceed as before and propagate the controls on R_k into R_{k+1} until we obtain a grid of width 3. Since the number of controls on a kD grid of length m is $O(m^k)$, we obtain a circuit of depth $O(\sqrt[k]{n})$ for implementing a controlled-U operation with n controls. The constant depends on k, but we assumed that k is constant in Section 2. From this, we obtain the following result.

▶ **Theorem 5.2.** *Controlled-U operations with n controls have depth $O(\sqrt[k]{n})$, size $O(n)$ and width $O(n)$ in k*D *NANTC.*

6 Fanout operations

In this section, we describe quantum circuits for fanout. In this case, we have a single control qubit and our goal is to XOR it into each of the target qubits. The construction of fanout circuits is adapted from that of controlled operations; the circuits are the same except that the qubit that was the target becomes the control qubit and qubits that were the controls become the targets. Let n be the number of targets. In the case of the circuit of Section 5, we simply apply all operations in reverse order and replace each Toffoli gate $y \leftarrow y \oplus x_1 \wedge \ldots \wedge x_n$ with a fanout operation $x_j \leftarrow x_j \oplus y$ for all $1 \leq j \leq n$. This yields a kD NANTC fanout circuit of depth $O(\sqrt[k]{n})$. We have shown the following.

▶ **Theorem 6.1.** *fanouts to n targets have depth $O(\sqrt[k]{n})$, size $O(n)$ and width $O(n)$ in k*D *NANTC.*

7 Optimality

In this section, we prove that the depth, size and width of the circuits of Theorem 5.1 (and its kD generalization) are optimal for NANTC. A similar lower bound for addition is discussed in [6]. These lower bounds hold regardless of where the controls and target qubits are located on the kD grid. They also hold for a more general class of operations that contains the controlled-U operations and fanouts.

Since each qubit is acted on by a constant number of operations in Theorem 5.1, the size of the circuit is $O(n)$. This is clearly optimal since any circuit that implements a controlled operation must act on each of the controls.

▶ **Theorem 7.1.** *Any NANTC quantum circuit that implements a non-trivial controlled-U operation with n controls has size $\Omega(n)$.*

The *trace norm* of a density matrix ρ (denoted $\|\rho\|_{\mathrm{tr}}$) is equal to $(1/2)\operatorname{tr}|\rho|$ (the $(1/2)$ factor ensures that $\|\rho-\sigma\|_1$ is the probability of distinguishing ρ and σ with the best possible measurement). Consider a general quantum operation $\mathcal{E}:\mathcal{D}_n\to\mathcal{D}_n$ represented as a CPTP map. We will use an operator version of the trace norm defined by $\|\mathcal{E}\|_{\mathrm{tr}}=\sup_{\rho\in\mathcal{D}}\|\mathcal{E}(\rho)\|_1$; if $\mathcal{E}_1$ and $\mathcal{E}_2$ are two CPTP maps then $\|\mathcal{E}_1-\mathcal{E}_2\|_{\mathrm{tr}}$ is the probability of distinguishing between them on the worst possible input. Thus, it is a measure of how much these operations differ. We will also make use of the partial trace. If x is a qubit, then we will denote the partial trace over all qubits except x by $\mathrm{tr}_{\neg x}=\mathrm{tr}_{\mathbb{Z}^k\setminus\{x\}}$.

Controlled-U operations are special case of a more general class of operations.

▶ **Definition 7.2.** *Let $\mathcal{E}:\mathcal{D}_n\to\mathcal{D}_n$ be a CPTP map. We say that $\mathcal{E}$ is ϵ-input sensitive if there exists a qubit y such that for $\Omega(n)$ qubits x, there exists a CPTP map $\mathcal{F}:\mathcal{D}_n\to\mathcal{D}_n$ acting only on x such that $\|\mathrm{tr}_{\neg y}(\mathcal{E}\mathcal{F}-\mathcal{E})\|_{\mathrm{tr}}\geq\epsilon$.*

Intuitively, an ϵ-input sensitive operation is a generalization of a Toffoli gate where modifying some input qubit x yields a different value on the output with probability ϵ. Similarly, we can define ϵ-output sensitive operations which are generalizations of fanout.

▶ **Definition 7.3.** *Let $\mathcal{E}:\mathcal{D}_n\to\mathcal{D}_n$ be a CPTP map. We say that $\mathcal{E}$ is ϵ-output sensitive if there exists a qubit x such that for $\Omega(n)$ qubits y, there exists a CPTP map $\mathcal{F}:\mathcal{D}_n\to\mathcal{D}_n$ acting only on x such that $\|\mathrm{tr}_{\neg y}(\mathcal{E}\mathcal{F}-\mathcal{E})\|_{\mathrm{tr}}\geq\epsilon$.*

We say that $\mathcal{E}$ is *ϵ-sensitive* if it is ϵ-input or ϵ-output sensitive. A family $\{\mathcal{E}:\mathcal{D}_n\to\mathcal{D}_n\}$ of CPTP maps is ϵ-sensitive if every $\mathcal{E}_n$ is ϵ-sensitive. Our lower bounds will apply to all families of ϵ-sensitive operations. All proofs will be for the case of ϵ-input sensitive operations but the argument of ϵ-output sensitive operations is all but identical.

▶ **Theorem 7.4.** *Let $\{\mathcal{E}_n:\mathcal{D}_n\to\mathcal{D}_n\}$ be a family of ϵ-sensitive operations. Then any family of kD NANTC circuits $\{C_n\}$ such that $\|\mathcal{E}_n-C_n\|_{\mathrm{tr}}<\epsilon/2$ for all n has size $\Omega(n)$.*

Proof. Suppose that C_n has size $o(n)$. Assume $\mathcal{E}_n$ is ϵ-input sensitive and choose a qubit y as in definition Definition 7.2 (the case where it is ϵ-output sensitive is very similar). There are $\Omega(n)$ qubits x such that there exists a CPTP map $\mathcal{F}:\mathcal{D}_n\to\mathcal{D}_n$ acting only on x such that $\|\mathrm{tr}_{\neg y}(\mathcal{E}_n\mathcal{F}-\mathcal{E}_n)\|_{\mathrm{tr}}\geq\epsilon$. For large n, there is such an x which is not acted on by C_n. Then $\mathrm{tr}_{\neg y}C_n\mathcal{F}=\mathrm{tr}_{\neg y}C_n$.Now $\|\mathrm{tr}_{\neg y}(C_n-\mathcal{E}_n)\|_{\mathrm{tr}}\geq\big|\|\mathrm{tr}_{\neg y}(C_n\mathcal{F}-\mathcal{E}_n\mathcal{F})\|_{\mathrm{tr}}-\|\mathrm{tr}_{\neg y}(\mathcal{E}_n\mathcal{F}-\mathcal{E}_n)\|_{\mathrm{tr}}\big|$ $>\epsilon/2$ which is a contradiction. ◀

We call a controlled-U operation *non-trivial* if $U\neq I$. It is easy to prove the following.

▶ **Lemma 7.5.** *Non-trivial controlled-U operations and fanouts are* 1*-sensitive.*

We now obtain a corollary of Theorem 7.4 of which Theorem 7.1 is a special case.

▶ **Corollary 7.6.** *Let* $\{\mathcal{E}_n : \mathcal{D}_n \to \mathcal{D}_n\}$ *denote a family of controlled-U operations or fanouts. Any family of* kD *NANTC circuits* $\{C_n\}$ *such that* $\|C_n - \mathcal{E}_n\|_{\text{tr}} < 1/2$ *has size* $\Omega(n)$*.*

This shows that the circuits of Theorem 5.1 (and its kD generalization) have optimal size. Next, we will show that ϵ-sensitive kD NTC circuits have depth $\Omega(\sqrt[k]{n})$. For this we require the following easy lemma.

▶ **Lemma 7.7.** *For any subset* $S \subseteq \mathbb{Z}^k$ *and any* $x \in \mathbb{Z}^k$*, there exists a subset* $T \subseteq S$ *of size* $\Omega(|S|)$ *such that for all* $y \in T$*,* $\|x - y\|_1 = \Omega(\sqrt[k]{|S|})$*.*

We are now ready to prove our depth lower bound.

▶ **Theorem 7.8.** *Let* $\{\mathcal{E}_n : \mathcal{D}_n \to \mathcal{D}_n\}$ *be a family of* ϵ*-sensitive operations. Then any family of* kD *NANTC circuits* $\{C_n\}$ *such that* $\|\mathcal{E}_n - C_n\|_{\text{tr}} < \epsilon/2$ *for all* n *has depth* $\Omega(\sqrt[k]{n})$*.*

Proof. Suppose $\{C_n\}$ has depth $t = o(\sqrt[k]{n})$. Assume that $\mathcal{E}_n$ is ϵ-input sensitive (the case where it is ϵ-output sensitive is very similar) and choose a qubit y as in Definition 7.2. There is a set S of $\Omega(n)$ qubits such that for each $x \in S$, there exists a CPTP map $\mathcal{F} : \mathcal{D}_n \to \mathcal{D}_n$ acting only on x with $\|\text{tr}_{\neg y}(\mathcal{E}_n\mathcal{F} - \mathcal{E}_n)\|_{\text{tr}} \geq \epsilon$. Let $c > 0$ be the hidden constant in the expression $\Omega(\sqrt[k]{|S|})$ from Lemma 7.7. For sufficiently large n, the depth of C_n is strictly less than $c\sqrt[k]{n}$. Let G_i be the set of disjoint one- and two-qubit operations that are performed at timestep $1 \leq i \leq t$ in C_n. For an operation $M \in G_i$, let us say that M is *active* if

(a) M acts non-trivially on y or

(b) there is an operation $M' \in G_j$ with $i < j \leq t$ such that M' is active and M and M' act non-trivially on a common qubit

Let us say that a qubit x *influences* y if there exists an active operation $M \in G_i$ that acts non-trivially on x. Suppose x influences y after t timesteps. Because all operations act on pairs of adjacent qubits, the ℓ_1 distance between x and y is at most t. By Lemma 7.7, there exists a subset T of S of size $\Omega(n)$ such that $\|x - y\|_1 \geq c\sqrt[k]{n}$ for all $x \in T$. Because $t < c\sqrt[k]{n}$, x does not influence y for $x \in T$. Let us fix some $x \in T$. Choosing a $\mathcal{F}$ acting only on x as in Definition 7.2, we have $\|\text{tr}_{\neg y}(C_n - \mathcal{E}_n)\|_{\text{tr}} \geq \left|\|\text{tr}_{\neg y}(C_n\mathcal{F} - \mathcal{E}_n\mathcal{F})\|_{\text{tr}} - \|\text{tr}_{\neg y}(\mathcal{E}_n\mathcal{F} - \mathcal{E}_n)\|_{\text{tr}}\right| > \epsilon/2$ which is a contradiction. ◀

By Lemma 7.5, we obtain the following corollary.

▶ **Corollary 7.9.** *Let* $\{\mathcal{E}_n : \mathcal{D}_n \to \mathcal{D}_n\}$ *denote a family of controlled-U operations or fanouts. Any family of* kD *NANTC circuits* $\{C_n\}$ *such that* $\|C_n - \mathcal{E}_n\|_{\text{tr}} < 1/2$ *has depth* $\Omega(\sqrt[k]{n})$*.*

From Theorems 5.2 and 6.1 and Corollaries 7.6 and 7.9, we conclude that the circuits of Theorem 5.1 and its kD generalization are optimal in their depth, size and width.

▶ **Theorem 1.6.** *The depth required for controlled-U operations with* n *controls and fanouts with* n *targets in* kD *NANTC is* $\Theta(\sqrt[k]{n})$*. Moreover, this depth can be achieved with size* $\Theta(n)$ *and width* $\Theta(n)$*.*

Acknowledgments. I thank Paul Beame and Aram Harrow for useful discussions and feedback and the anonymous reviewers for helpful comments. Aram Harrow suggested the use of teleportation chains as a primitive. Paul Pham suggested applying the technique of Theorem 5.1 to fanouts. I was funded by the DoD AFOSR through an NDSEG fellowship. Partial support was provided by IARPA under the ORAQL project.

References

1. D. Aharonov, M. Ben-Or, R. Impagliazzo, and N Nisan. Limitations of Noisy Reversible Computation. *arXiv e-prints*, 1996.
2. Charles H. Bennett, Gilles Brassard, Claude Crépeau, Richard Jozsa, Asher Peres, and William K. Wootters. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Physical Review Letters*, 70, 1993.
3. Dan E. Browne, Elham Kashefi, and Simon Perdrix. Computational depth complexity of measurement-based quantum computation. In *In Proceedings of the Fifth Conference on the Theory of Quantum Computation, Communication and Cryptography*, 2010.
4. Donny Cheung, Dmitri Maslov, and Simone Severini. Translation techniques between quantum circuit architectures. In *Workshop on Quantum Information Processing*, 2007.
5. B.-S. Choi and R. Van Meter. An $\Theta(\sqrt{n})$-depth quantum adder on a 2D NTC quantum computer architecture. *arXiv:1008.5093*, 2010.
6. Byung-Soo Choi and Rodney Van Meter. On the effect of quantum interaction distance on quantum addition circuits. *ACM Journal on Emerging Technologies in Computing Systems*, 7:11:1–11:17, 2011.
7. A. G. Fowler, S. J. Devitt, and L. C. L. Hollenberg. Implementation of Shor's Algorithm on a Linear Nearest Neighbour Qubit Array. *arXiv e-prints*, 2004.
8. Peter Høyer and Robert Špalek. Quantum fan-out is powerful. *Theory of Computing*, 1:81–103, 2005.
9. S. A. Kutin. Shor's algorithm on a nearest-neighbor machine. *arXiv e-prints*, 2006.
10. D. Maslov. Linear depth stabilizer and quantum Fourier transformation circuits with no auxiliary qubits in finite-neighbor quantum architectures. *arXiv e-prints*, 2007.
11. C. Moore. Quantum Circuits: Fanout, Parity, and Counting. *arXiv e-prints*, 1999.
12. P. Pham and K. M. Svore. A 2D Nearest-Neighbor Quantum Architecture for Factoring. *arXiv e-prints*, 2012.
13. R. Raussendorf, D. E. Browne, and H. J. Briegel. The one-way quantum computer–a non-network model of quantum computation. *arXiv e-prints*, 2002.
14. Robert Raussendorf and Hans J Briegel. A one-way quantum computer. *Physical Review Letters*, 86:5188–5191, 2001.
15. D. Rosenbaum. Optimal Quantum Circuits for Nearest-Neighbor Architectures. *arXiv:1205:0036*, 2012.
16. Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *Annual Symposium on Foundations of Computer Science*, 1994.
17. Y. Takahashi and S. Tani. Constant-Depth Exact Quantum Circuits for the OR and Threshold Functions. *arXiv e-prints*, 2011.
18. B. M. Terhal and D. P. DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *arXiv e-prints*, 2002.
19. Rodney Van Meter and Kohei M. Itoh. Fast quantum modular exponentiation. *Phys. Rev. A*, 71:052320, 2005.

Access Structure in Graphs in High Dimension and Application to Secret Sharing

Anne Marin[1], Damian Markham[2], and Simon Perdrix[3]

1 **LTCI, INFRES, Telecom ParisTech, France**
anne.marin@telecom-paristech.fr
2 **CNRS / LTCI, INFRES, Telecom ParisTech, France**
damian.markham@telecom-paristech.fr
3 **CNRS / LIG, Grenoble University, France**
Simon.Perdrix@imag.fr

Abstract

We give graphical characterisation of the access structure to both classical and quantum information encoded onto a multigraph defined for prime dimension q, as well as explicit decoding operations for quantum secret sharing based on graph state protocols. We give a lower bound on k for the existence of a $((k,n))_q$ scheme and prove, using probabilistic methods, that there exists α such that a random multigraph has an accessing parameter $k \leq \alpha n$ with high probability.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Quantum Secret Sharing, Graph State, Multigraph, Access Structure

Digital Object Identifier 10.4230/LIPIcs.TQC.2013.308

1 Introduction

In this work we consider encoding, and accessing, both quantum and classical information onto graph states of qudits - multipartite entangled states which are one to one corresponding to multigraphs (which we will consider as simple graphs with multiple edges). We are particularly interested in using these states for secret sharing.

Secret sharing is an important cryptographic primitive, which was first put forward classically in [33], and then extended to the quantum realm in [19, 9]. The aim of the protocol is for a dealer to distribute a secret (quantum or classical) to a set of players, in such a way that only authorized sets of players can access the secret, and unauthorized sets of players cannot (there may be sets of players which are neither authorized nor unauthorized). The sets of authorized and unauthorized players is called the access structure. Any secret sharing scheme of n players can be loosely paramaterised by two numbers, k and k', such that any subset of k players is an authorized set, whereas any subset of k' players or less is unauthorized. We call such paramaterised schemes (k, k', n) ramp schemes. In the case when $k' = k - 1$, we say it is a threshold scheme, and simplify the notation to (k, n).

In this work we consider two classes of quantum schemes, one class using quantum channels to distribute classical secrets, denoted CQ schemes [19], and the other sharing quantum secrets [9, 19], denoted QQ schemes. The notation CQ and QQ used here follows the work [30, 26, 28], where both classes were phrased in the same language of graph states (first for qubits [30] then qudits [26, 28]). The equivalence of both schemes was shown in [28]. Using the graph state formalism can be useful both practically - since graph states are amongst the most well developed multipartite entangled states experimentally - and theoretically, since graph states are rich in their uses in quantum information, and allow for

graphical characterization of information flow, and access of information. The connection between error correction and secret sharing was understood early on [9], and implies that for general access structures it is necessary to use high dimensional states to encode the secret [30, 28]. In [24] an entirely graphical description of the access structure was given for the graph state protocols on qubits. This has led to many applications, for instance in proving lower and upper bounds on what k and k' are possible in ramp schemes. We are naturally interested in doing the same for higher dimensional versions.

The first result of this paper is to extend to higher dimension the characterisation of the access structure in a graph, previously done in [24] for 2-dimensional system. By gathering the graphical conditions and previous results, we show that the accessibility problem to quantum information can be reduced to study the classical information's one in both a set of player and its complementary (which was proved in [24, 21] for 2 dimensional system). Finally we use this result for the decoding phase of both QQ and CQ protocols, as we know [28] that a CQ authorised is a QQ authorised set and vice versa. In the last part, we study the existence, as a function of k, of a $((k,n))_q$ scheme (this will be defined explicitly later, but can be understood as the underlying graph encoding which gives rise to $(k, n-k, n)$ QQ secret sharing schemes). We derive a lower bound over k, that is, there exists α such that every $(k, n-k, n)$ QQ secret sharing must satisfies $k > \alpha n$, and we use probabilistic method to find $c < 1$ such that a $((cn, n))_q$ scheme exists with high probability.

2 Background

2.1 Qudit graph states, $\mathbb{F}_q^*$-graphs, and multigraphs

The *qudit graph state* formalism [32, 14, 27, 1] consists of representing a quantum state using a weighted undirected graph where every vertex represents a q-dimensional quantum system and every edge, which has assigned an element from the finite field $\mathbb{F}_q$, represents intuitively the entanglement between the elementary systems (a formal definition is given in Definition 1). Such graphs, labeled with elements of a finite field $\mathbb{F}_q$, are known as $\mathbb{F}_q^*$-*graphs* [23] and can be interpreted as edge-colored graphs. In this paper, we consider q prime, and choose to interpret $\mathbb{F}_q^*$-graphs as multigraphs i.e., graphs with possibly parallel edges between pairs of vertices. Albeit equivalent to the other interpretation of $\mathbb{F}_q^*$-graphs, we believe that the multigraph interpretation is relevant in the context of qudit graph states for secret sharing protocols, in particular for the graphical characterisation of authorised and unauthorised sets of players (see Lemmas 5 and 7).

▶ **Definition 1** (q-multigraphs)**.** Given a prime number q, a q-multigraph G is a pair (V, Γ) where V is the finite set of vertices and $\Gamma : V \times V \to \mathbb{F}_q$ is the adjacency matrix of G: for any $u, v \in V$, $\Gamma(u, v)$ is the multiplicity of the edge (u, v) in G.

The term multigraph is used for q-multigraph when q is clear from the context or irrelevant. In this paper, we consider undirected simple multigraphs $G = (V, \Gamma)$ i.e., for any vertices $u, v \in V$, $\Gamma(u, v) = \Gamma(v, u)$ and $\Gamma(u, u) = 0$. For our characterizations of encoding and accessing later on, it will be useful to introduce further concepts. We will see several examples of them along the way, but for now we state definitions. Given a set V of vertices, a vector $D : V \to \mathbb{F}^q$ represents a multiset of vertices of V: for every $v \in V$, $D(v) \in \mathbb{F}_q$ is the multiplicity of v in D. $\sup(D) = \{v \in V \mid D(v) \neq 0 \bmod q\}$ is the support of D. For any multigraph $G = (V, \Gamma)$ and any multiset of vertices $D : V \to \mathbb{F}_q$, the matrix product $\Gamma.D$ is the multiset of neighbours of D: for any $v \in V$, v is a neighbour of D with multiplicity $(\Gamma.D)(v) = \sum_{u \in V} \Gamma(u, v).D(u) \bmod q$. In particular, for any vertex u, $\Gamma.\{u\}$ is the multiset

of neighbours of u. We call $G[D] = (V', \Gamma')$ the sub-multigraph of $G = (V, \Gamma)$ induced by the multiset $D : V \to \mathbb{F}_q$, where $V' = V \cap \sup(D)$ and $\Gamma' : V' \times V' \to \mathbb{F}_q = (u, v) \mapsto D(u).\Gamma(u, v).D(v) \bmod q$. Notice that the multiplicity of an edge in $G[D]$ is the multiplicity of this edge in the original graph G times the multiplicity in D of the two vertices connected by this edge. For any $A, B \subseteq V$, $\Gamma[A, B]$ denotes the submatrix of Γ whose columns correspond to the vertices in A and rows to the vertices in B. $\Gamma[A, B]$ represents the edges which have one end in A and the other one in B.

▶ **Definition 2** (Qudit Graph State)**.** Given a q-multigraph $G = (V, \Gamma)$ with $V = \{v_1, \ldots, v_n\}$, let $|G\rangle \in \mathbb{C}^{q^n}$ be its associated qudit graph state defined as

$$|G\rangle = \frac{1}{\sqrt{q^n}} \sum_{x=(x_1,\cdots,x_n)\in\mathbb{F}_q^n} \omega^{|G[x]|}|x\rangle$$

where ω is the q^{th} root of unity and $|G[x]|$ is the number of edges of the sub-multigraph $G[x] = (V_x, \Gamma_x)$ induced by x, where $V_x = \{v_i \in V, x_i \neq 0\}$ and $\Gamma_x : V_x \times V_x \to \mathbb{F}_q = (v_i, v_j) \mapsto x_i x_j \Gamma(v_i, v_j)$.

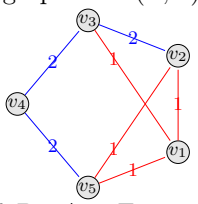
Qudit graph states satisfy the following fundamental fixpoint property. Given a q-multigraph $G = (V, \Gamma)$, $|G\rangle$ is the unique quantum state (up to a global phase) such that, for any $u \in V$,

$$X_u Z_{\Gamma.\{u\}}|G\rangle = |G\rangle \tag{1}$$

where $\Gamma.\{u\}$ is the multiset of neighbours of u, $X = |b\rangle \mapsto |b + 1 \bmod q\rangle$, $Z = |b\rangle \mapsto \omega^b|b\rangle$ are Pauli operators, and for any mulitset $D : V \to \mathbb{F}_q$, $Z_D := \bigotimes_{v\in V} Z_v^{D(v)}$.

▶ **Example 3.** We define the 3-multigraph $G = (V, \Gamma)$ by $V = \{v_1, v_2, v_3, v_4, v_5\}$,

$$\Gamma = \begin{bmatrix} 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 2 & 0 & 1 \\ 1 & 2 & 0 & 2 & 0 \\ 0 & 0 & 2 & 0 & 2 \\ 1 & 1 & 0 & 2 & 0 \end{bmatrix}$$



Let $A = \{v_1, v_2\}$ be a subset of V, and $D : A \to \mathbb{F}_3$ a multiset such that $D(v_1) = 2$, $D(v_2) = 1$. That is $D = \{v_1, v_1, v_2\}$. Then, with previous definitions, the graph induced by D is $G[D] =$

The multiset of neighbours of A is $\{v_1, v_2, v_5, v_5\}$. The multiset of neighbours of D is $\{v_1, v_2, v_2, v_3\}$.

2.2 Local complementation and cut rank

The *local complementation* [5] is a graph transformation which is incredibly useful for the study of graph states [35]. Indeed, if two graphs G and G' are locally equivalent (i.e. one can transform G into G' by means of a series of local complementations), they represent the same entanglement (i.e. there exists a local unitary transformation U such that $|G'\rangle = U|G\rangle$) [35]. Local complementation is extended to multigraphs as follows [23]: Given a q-multigraph $G = (V, \Gamma)$, $u \in V$ and $\lambda \in \mathbb{F}_q$, the *λ-local complementation at u of G* is the q-multigraph $G \star^\lambda u = (V, \Gamma')$ such that $\forall v, w \in V$, $v \neq w$, $\Gamma'(v, w) = \Gamma(v, w) + \lambda.\Gamma(v, u).\Gamma(u, w) \bmod q$. Keet et al. [26] have proved that for any q-multigraph $G = (V, \Gamma)$, any $u \in V$ and any $\lambda \in \mathbb{F}_q$, there exists a local unitary transformation U such that $|G \star^\lambda u\rangle = U|G\rangle$.

The *cut rank* [31] is a set function which associates with every set B of vertices the rank of the matrix describing the edges of the cut $(B, V\backslash B)$: Given a multigraph $G = (V, \Gamma)$, let $\Gamma[B] := \Gamma[B, V\backslash B]$ be the cut matrix of the cut $(B, V\backslash B)$, moreover for any $A, B \subseteq V$, let $\mathrm{rk}_G(A, B) := \mathrm{rank}(\Gamma[A, B])$ and $\mathrm{cutrk}_G(B) := \mathrm{rk}_G(B, V\backslash B)$ be the cut rank of B. Notice that $\mathrm{rk}_G(A, B) = \mathrm{rk}_G(B, A)$ and $\mathrm{cutrk}_G(B) = \mathrm{cutrk}_G(V\backslash B)$.

We point out in this paper that the cut rank, which is known to be invariant by local complementation [23], is a key parameter of q-multigraphs for the study of secret sharing protocols with qudit graph states. Indeed, Theorem 9 states that the capability of a set of players to reconstruct a quantum secret is characterised by the discrete derivative of the cut rank function. Notice that the cut-rank of a bipartition is nothing but the Schmidt measure of entanglement of this bipartition in the corresponding graph state. This is shown for the qubit case in [17], and easily extends to the qudit case. As a consequence, Theorem 9 characterises the accessibility of a set of players as the derivative of the Schmidt measure of entanglement.

2.3 Description of the encoding:

We now introduce the encoding of classical and quantum information onto graph states (CQ and QQ respectively), which will be the starting point for the secret sharing protocols defined in section 4. For ease of notation we present the CQ encoding as deterministic, and in one basis. When used in the full CQ protocol this is randomised by measurement and choice of basis (described fully in section 4). The ability of players to access encoded information (both classical and quantum) is fully described in graph theoretical language in section 3.

CQ encoding:

Given a multigraph $G = (V, \Gamma)$ of order n and a distinguished non isolated vertex $d \in V$, the corresponding CQ encoding of a classical secret $s \in \mathbb{F}_q$ among $n-1$ players consists of the dealer preparing the state

$$|s_L\rangle := Z^s_{\Gamma.\{d\}}|G\backslash d\rangle$$

and sending one qudit to each player, where $G\backslash d = (V\backslash\{d\}, \Gamma[V\backslash\{d\}, V\backslash\{d\}])$ is the multigraph obtained by removing the vertex d and all its incident edges.

In the CQ protocol (described in section 4) the secret s is randomised by measurement on the dealer's vertex d of the full graph state $|G\rangle$, and further, the encoding is randomised by choice of measurement basis - the dealer chooses at random $t \in T$, $T \subseteq \mathbb{F}_q$ and $|T| \geq 2$, and measures his qudit in the associated complementary basis X^tZ. Measuring in this t basis will correspond exactly to using the above CQ encoding of the same secret value s onto the complementary multigraph $G\star^t d$.

QQ encoding:

Given a multigraph $G = (V, \Gamma)$ of order n and a distinguished non isolated vertex $d \in V$, the corresponding QQ encoding on a qudit graph state for sharing an arbitrary quantum secret $|\phi\rangle = \sum_{j=0}^{q-1} s_j|j\rangle \in \mathbb{C}^q$ among $n-1$ players consists, for the dealer, in preparing the state

$$|\phi_L\rangle = \sum_{j=0}^{q-1} s_j Z^j_{\Gamma.\{d\}}|G\backslash d\rangle = \sum_{j=0}^{q-1} s_j|j_L\rangle$$

and in sending one qudit of $|\phi_L\rangle$ to each player.

Notice that the preparation consists in applying the map $|j\rangle \mapsto Z^j_{\Gamma.\{d\}}|G\backslash d\rangle$ which is an isometry as long as d is not an isolated vertex in G. We describe encoding procedures in appendix A.

The accessing structure of the protocols (i.e. the description of the sets of players which can recover the secret, as well as those which have no information about the secret) is given in the next section which provides a graphical characterisation of the accessing structure for the secret sharing protocols using these encodings. Moreover, the operations the authorised sets of players have to perform to reconstruct the secret are also described in the next section.

3 Access Structure in a Graph in Higher Dimension:

3.1 Classical Information

In this section, we show, when the secret is classical, that the protocol is perfect (i.e. every set of players is either able to recover the secret or has no information about the secret), and that the accessing structure is graphically characterised by a simple rank-based function:

▶ **Theorem 4.** *Given a q-multigraph $G = (V, \Gamma)$ and a distinguished vertex $d \in V$, a set $B \subseteq V\backslash\{d\}$ of players can recover a classical secret for the corresponding CQ encoding if and only if $\pi_G(B, d) = 1$, where*

$$\pi_G(B, d) := \mathrm{cutrk}_G(B) - \mathrm{cutrk}_{G\backslash d}(B)$$

A graphical interpretation of Theorem 4 is that a set B is accessible if and only if the presence of the 'dealer vertex' d increases the rank of the cut between B and the rest of the vertices.

The rest of the section is dedicated to the proof of Theorem 4.

First, we prove that a set B of players can recover a classical secret if, roughly speaking, there exists a multiset D of them which is not 'seen' from outside except by the 'dealer':

▶ **Lemma 5.** *Given a q-multigraph $G = (V, \Gamma)$ and $d \in V$, a set $B \subseteq V\backslash\{d\}$ of players can recover a classical secret for the corresponding CQ encoding if there exists a multiset $D : B \to \mathbb{F}_q$ such that* $\mathrm{sup}(\Gamma[B, V\backslash B].D) = \{d\}$ *i.e.,*

- *the number of neighbours of d in D is not congruent to* $0 \bmod q$*;*
- *$\forall u \in V\backslash(B \cup \{d\})$, the number of neighbours of u in D is congruent to* $0 \bmod q$*.*

Proof. Given $B \subseteq V$ and $D : B \to \mathbb{F}_q$ such that $\mathrm{sup}(\Gamma[B, V\backslash B].D) = \{d\}$. W.l.o.g. we assume the multiplicity of d in $\Gamma.D$ is 1 (otherwise we consider the multiset $D' = u \mapsto (\Gamma.D)(d)^{-1}.D(u)$ instead of D). The players in B can recover the secret by measuring an appropriate product of stabilizers. Indeed, there exists $r \in \mathbb{F}_q$ such that $\prod_{u\in B}(X_u Z_{\Gamma.\{u\}})^{D(u)} = \omega^r X_D Z_{\Gamma.D} = Z_d \omega^r X_D Z_{\Gamma[V,V\backslash\{d\}].D}$. As $\prod_{u\in B}(X_u Z_{\Gamma.\{u\}})^{D(u)}|G\rangle = |G\rangle$, we deduce that $\omega^r X_D Z_{\Gamma[V,V\backslash\{d\}].D}|G\backslash d\rangle = |G\backslash d\rangle$. If the classical secret is $s \in \mathbb{F}_q$, $\omega^r X_D Z_{\Gamma[V,V\backslash\{d\}].D} Z^s_{\Gamma.\{d\}}|G\backslash d\rangle = \omega^{r-s} Z^s_{\Gamma.\{d\}} X_D Z_{\Gamma[V,V\backslash\{d\}].D}|G\backslash d\rangle = \omega^{-s} Z^s_{\Gamma.\{d\}}|G\backslash d\rangle$. So if the players in B measure according to $\omega^r X_D Z_{\Gamma[V,V\backslash\{d\}].D}$, they get the outcome $-s \bmod q$, so they recover the classical secret s. ◀

Lemma 5 provides a sufficient condition for a set of players to be able to reconstruct a classical secret. Notice that this reconstruction is nothing but a Pauli measurement, so it can be done by means of local Pauli measurements and classical communications.

▶ **Corollary 6.** *Given a q-multigraph $G = (V, \Gamma)$, $d \in V$, and $B \subseteq V\backslash\{d\}$, if $\pi_G(B, d) = 1$ then B can reconstruct a classical secret for the corresponding CQ encoding.*

Proof. Let $F = V\backslash(B \cup \{d\})$. According to lemma 5, B can recover a classical secret if there exists $D : B \to \mathbb{F}_q$ such that $\mathrm{sup}(\Gamma[B, V\backslash B].D) = \{d\}$. W.l.o.g. we can assume

that the multiplicity of d in $\Gamma[B, V\backslash B].D$ is one. So B can recover a classical secret if the system $\left(\frac{\Gamma[B,\{d\}]}{\Gamma[B,F]}\right).x = \left(\frac{1}{0}\right)$ has a non zero solution, which is equivalent to $\text{rank}\left(\frac{\Gamma[B,\{d\}]}{\Gamma[B,F]}\right) = \text{rank}\left(\frac{\Gamma[B,\{d\}] \mid 1}{\Gamma[B,F] \mid 0}\right)$. Using the last column of the right-side matrix to cancel terms of the row $\Gamma[B,\{d\}]$, we are finally reduced to $\text{rank}\left(\frac{\Gamma[B,\{d\}]}{\Gamma[B,F]}\right) = 1 + \text{rank}(\Gamma[B,F])$ i.e., $\text{cutrk}_G(B) - \text{rk}_G(B,F) = 1 = \pi_G(B,d)$. ◀

In the following, a sufficient condition for a set of players to have no information about the secret is introduced: roughly speaking, a multiset of players D which includes the dealer d, can 'hide' the secret to every player who is connected to D with a number of edges congruent to 0 modulo q:

▶ **Lemma 7.** *Given a q-multigraph $G = (V,\Gamma)$ and $d \in V$, a set $B \subseteq V\backslash\{d\}$ has no information about a classical secret for the corresponding CQ encoding if there exists $D : V\backslash B \to \mathbb{F}_q$, such that $D(d) \neq 0 \bmod q$ and $\Gamma[V\backslash B, B].D = 0$ i.e.,*

- *the multiplicity of d in D is not congruent to $0 \bmod q$;*
- *$\forall u \in B$, the number of neighbours of u in D is congruent to $0 \bmod q$.*

Proof. W.l.o.g. we assume $D(d) = 1 \bmod q$. Notice that $R|G\backslash d\rangle\langle G\backslash d|R^\dagger = |G\backslash d\rangle\langle G\backslash d|$ with $R = \prod_{u\in V\backslash(B\cup\{d\})} \left(X_u Z_{\Gamma[V\backslash\{d\},V\backslash\{d\}].\{u\}}\right)^{D(u)}$. Moreover $R.Z_{\Gamma.\{u\}}$ is only acting on $V\backslash(B\cup\{d\})$, so the reduced density matrix for the players in B is

$$\begin{aligned} & Tr_{V\backslash(B\cup\{d\})}(Z^s_{\Gamma.\{d\}}|G\backslash d\rangle\langle G\backslash d|{Z^\dagger}^s_{\Gamma.\{d\}}) \\ = \; & Tr_{V\backslash(B\cup\{d\})}((Z_{\Gamma.\{d\}}R)^s|G\backslash d\rangle\langle G\backslash d|(Z_{\Gamma.\{d\}}R)^{\dagger s}) \\ = \; & Tr_{V\backslash(B\cup\{d\})}(|G\backslash d\rangle\langle G\backslash d|) \end{aligned}$$

which does not depend on the secret, so the players in B have no information about the secret. ◀

▶ **Corollary 8.** *Given a q-multigraph $G = (V,\Gamma)$, $d \in V$, and $B \subseteq V\backslash\{d\}$, if $\pi_G(B,d) = 0$ then B has no information about the classical secret for the corresponding CQ encoding.*

Proof. Let $F = V\backslash(B\cup\{d\})$. According to lemma 7, B has no information about classical secret if there exists $D : V\backslash B \to \mathbb{F}_q$ such that $D(d) = 1 \bmod q$ and $\Gamma[V\backslash B, B].D = 0$, so if $\Gamma[F,B].C = -\Gamma[V,B]\{d\}$, where $C : F \to \mathbb{F}_q = u \mapsto D(u)$ is the restriction of D to F. As a consequence, B has no information about classical secret if the system $\Gamma[F,B].x = -\Gamma[V,B]\{d\}$ has a non zero solution, which is equivalent to find a non zero solution to the system $\Gamma[F,B].x = \Gamma[V,B]\{d\}$, so if $\text{rank}(\Gamma[F,B]) = \text{rank}(\Gamma[V\backslash B,B])$ i.e., $\pi_G(B,d) = 0$. ◀

Proof of Theorem 4. The proof of Theorem 4 follows from Corollaries 6 and 8 and the fact that for every B, $0 \leq \pi_G(B,d) \leq 1$. It proves that the encoding is perfect i.e., every set of players is either able to reconstruct the secret (when $\pi_G(B,d) = 1$) or has no information about the secret (when $\pi_G(B,d) = 0$). ◀

3.2 Quantum Information

In the following we prove that the accessibility of a set a players is characterised by the derivative of the cut-rank function with respect to the dealer.

▶ **Theorem 9.** *Given a q-multigraph G with a distinguished dealer $d \in V(G)$, a set $B \subseteq V(G)\backslash\{d\}$ of players can recover a quantum secret in the corresponding QQ encoding iff*

$$\partial_d \mathrm{cutrk}_G(B) = -1$$

where $\partial_d \mathrm{cutrk}_G(B) = \mathrm{cutrk}_G(B \cup \{d\}) - \mathrm{cutrk}_G(B)$ is the discrete derivative of cutrk_G in B with respect to d.

Proof. It is known that B can access a quantum secret in G iff B can access a classical secret in two mutual unbiased bases, say in G and $G *^1 d$ [28]. Moreover B can access a classical secret in G iff $\pi_G(B, d) = 1$, where $\pi_G(B, d) = \mathrm{cutrk}_G(B) - \mathrm{rk}_G(B, V\backslash(B \cup \{d\}))$.
($\Rightarrow$) If B can access a quantum secret, B can access a classical secret and $V\backslash(B \cup \{d\})$ has no information about a quantum secret [9], which implies that $V\backslash(B \cup \{d\})$ cannot access a classical secret. Thus $\pi_G(B, d) = 1$ and $\pi_G(V\backslash(B \cup \{d\}), \{d\}) = 0$. As a consequence $\pi_G(B, d) - \pi_G(V\backslash(B \cup \{d\}), \{d\}) = 1$, so $1 = \mathrm{cutrk}(B) - \mathrm{rk}_G(B, V\backslash(B \cup \{d\})) - \mathrm{cutrk}(V\backslash(B \cup \{d\})) + \mathrm{rk}_G(V\backslash(B \cup \{d\}), B) = \mathrm{cutrk}(B) - \mathrm{cutrk}(V\backslash(B \cup \{d\})) = \mathrm{cutrk}(B) - \mathrm{cutrk}(B \cup \{d\})$.
($\Leftarrow$) If $\mathrm{cutrk}_G(B) = \mathrm{cutrk}_G(B \cup \{d\}) + 1$, then $\pi_G(B, \{d\}) = 1$, so B can access a classical secret in G. Moreover, since the cut rank is invariant by local complementation [23], $\mathrm{cutrk}_{G\star^1 d}(B) = \mathrm{cutrk}_{G\star^1 d}(B \cup \{d\}) + 1$, so B can also access a classical secret in $G\star^1 d$. ◀

Notice that for any set B of players, $\partial_d \mathrm{cutrk}_G(B) \in \{-1, 0, 1\}$: if $\partial_d \mathrm{cutrk}_G(B) = -1$, B can recover the quantum secret; if $\partial_d \mathrm{cutrk}_G(B) = 1$ they have no information since $V\backslash(B \cup \{d\})$ can recover the quantum secret; and if $\partial_d \mathrm{cutrk}_G(B) = 0$ they have some partial information about the secret.

Since the cut rank function is submodular [31], its derivative is monotonic (decreasing): if $B \subseteq B'$, $\partial_d \mathrm{cutrk}_G(B) \geq \partial_d \mathrm{cutrk}(B')$. Indeed, if B can recover the secret, any superset B' of B can recover it too; and if B' has no information about the secret, any subset B of B' has no information too.

4 Application to CQ and QQ protocols

We now see how the encoding of section 2.3, and the results on access structures in section 3 can be used to provide secret sharing protocols. Following the prescription of [28] (based on [30, 26], see also [29]) we will now introduce two protocols, one for sharing classical secrets over a quantum channel (CQ) and one for sharing a quantum secret (QQ), both based on a graph state associated with a multigraph. Both protocols can be understood as using the graph state as a channel between the dealer (associated with vertex d) and the players (the remaining vertices). In the CQ case this channel is used to perform an Ekert-like key distribution protocol between the dealer and authorised players, so that when completed the dealer and authorised players will share a random 'dit' string which is unknown to anybody else. In the QQ case the channel is used to teleport the secret to the players such that only authorised sets of players can access the information (the QQ encoding in section 2.3 can be understood as this teleportation, see Appendix A). More details on the protocols and their relation to each other as well as error correction can be found in [28].

4.1 Detailed protocols

Before we write the full protocols out, we first review some background on the graph state formalism, which will be the key in seeing how the stabilisers can be used to specify how authorised sets can access the information, given the satisfaction of the conditions outlined in the previous section.

Given a multigraph $G = (V, \Gamma)$, we begin with an illustrative expansion of the graph state $|G\rangle_V$ according to the d, $V\backslash\{d\}$ partition.

$$|G\rangle = \frac{1}{\sqrt{q^n}} \sum_{x\in\mathbb{F}_q^n} \omega^{|G[x]|}|x\rangle_V = \frac{1}{\sqrt{q}} \sum_s |s\rangle_d Z^s_{\Gamma.\{d\}}|G\backslash d\rangle_{V\backslash\{d\}}$$
$$= \frac{1}{\sqrt{q}} \sum_s |s\rangle_d |s_L\rangle_{V\backslash\{d\}}$$
$$= \frac{1}{\sqrt{q}} \sum_s |s(t)\rangle_d |s_L(t)\rangle_{V\backslash\{d\}},$$

for any $t \in \mathbb{F}_q$, where the second line follows from definitions in section 2.3, corresponding to the CQ encoding achieved by the dealer measuring in the Z basis. The third line corresponds to when the dealer measures in bases X^tZ (explained in more detail later), where they are defined as $|s(0)\rangle = |s\rangle$, and $|s(t)\rangle = \frac{1}{\sqrt{q}} \sum_{j=0}^{q-1} \omega^{\frac{j(j-t)}{2t} - st^{-1}j}|j\rangle$ for $t = 1...q-1$, so that $X^tZ|s(t)\rangle = \omega^s|s(t)\rangle$, and further $|s(0)_L\rangle = |s_L\rangle = Z^s_{\Gamma.\{d\}}|G\backslash d\rangle$ and $|s_L(t)\rangle := \frac{1}{\sqrt{q}} \sum_{k=0}^{q-1} \omega^{\frac{-k(k-t)}{2t} + st^{-1}k}|k_L\rangle$ for $t = 1...q-1$. The state $|s(t)_L\rangle_{V\backslash\{d\}}$ is equivalent to the CQ encoding of i on graph $G *^t d$ [26].

We now look at how the conditions for access arrived at in section 3 can be used, along with the stabiliser (or "fixed point") condition (1), to eventually see how authorised sets can access the information in the CQ and QQ protocols. We start with the QQ case, which is enough to imply the CQ case (see [28]). Suppose a set of players $B \subset V\backslash\{d\}$ has access to quantum information in a graph $G = (V, \Gamma)$. We proved with Theorem 9 that B can access QQ encoded quantum information in G if and only if B can access the CQ encoded classical information in G and $V\backslash(B\cup\{d\})$ cannot. Thus, by rewriting lemma 5 and 7 applied to B and $V\backslash(B\cup\{d\})$, we have: there exists $D : B \to \mathbb{F}_q$ and $C : B\cup\{d\} \to \mathbb{F}_q$ such that $C(d) = 1$ and $\begin{cases} \sup(\Gamma[B, V\backslash B].D) = \{d\} & \text{(A)} \\ \Gamma[B\cup\{d\}, V\backslash(B\cup\{d\})].C = 0 & \text{(B)} \end{cases}$

Now, call $K_i = X_i Z_{\Gamma.\{i\}}$ and $k_i = X_i Z_{\Gamma[V\backslash\{d\}, V\backslash\{d\}]\{i\}}$ (these are the fixpoint operators, or stabilisers for graphs G and $G\backslash d$ respectively according to (1)).

First we have $K_C = K_d \prod_{i\in B} K_i^{C(i)} = X_d Z_d^{\beta.} Z_{\Gamma.\{d\}} \prod_{i\in B} k_i^{C(i)}$ with $\beta = \Gamma.C(d)$. Then $Z_{\Gamma.\{d\}} \prod_{i\in B} k_i^{C(i)} = \omega^\lambda \prod_{i\in B} X_i^{C(i)} Z_i^{\Gamma.C(i)}$, with $\lambda = \sum_{i,j\in B\cup\{d\}, j<i} \Gamma(j,i)C(j)C(i)$.

Next K_D satisfies $K_D = \prod_{i\in B} K_i^{D(i)} = Z_d^\alpha \prod_{i\in B} k_i^{D(i)}$, with $\alpha = \Gamma.D(d) \neq 0$ since (A), and $\prod_{i\in B} k_i^{D(i)} = \omega^{\lambda'} \prod_{i\in B} X_i^{D(i)} Z_i^{\Gamma.D(i)}$, $\lambda' = \sum_{i,j\in B, j<i} \Gamma(j,i)D(j)D(i)$.

Later we will suppose $\alpha = 1$ (change D to $\alpha^{-1}.D$ if necessary).

Hence $K_C^t {K_D}^{1-t\beta}|G\rangle = \omega^{\frac{t(t-1)}{2}\beta} X_d^t Z_d.[Z_{\Gamma.\{d\}} \prod_{i\in B} k_i^{C(i)}]^t [\prod_{i\in B} k_i^{D(i)}]^{1-t\beta}|G\rangle = |G\rangle$ which is a stabiliser / fixpoint equation involving operators only in d and B which will be used to inform which measurements should be made to recover the secret in the CQ case, and how to find the QQ decoding operation. We can rewrite this as follows

$[Z_{\Gamma.\{d\}} \prod_{i\in B} k_i^{C(i)}]^t [\prod_{i\in B} k_i^{D(i)}]^{1-t\beta} = \omega^c \prod_{i\in B} X_i^{x_i(t)} Z_i^{z_i(t)}$ with

$$x_i(t) = tC(i) + (1-t\beta)D(i) \tag{2}$$
$$z_i(t) = t\Gamma.C(i) + (1-t\beta)\Gamma.D(i), \tag{3}$$
$$c = t^2\lambda' + (1-t\beta)^2\lambda + t(1-t\beta) \sum_{i,j\in B} \Gamma(i,j)C(i)D(j) \tag{4}$$

and we further define

$$f_t(r) := -r - c - \frac{t(t-1)}{2}\beta. \tag{5}$$

We can then see that given the state $|G\rangle_V$, if the dealer measures X^tZ, getting result $\omega^{s(t)}$

and each player i in B measures its qudit in the $X^{x_i(t)}Z^{z_i(t)}$ bases, denoting their results $m_i(t)$, if we define $m(t) = f_t^{-1}(\sum_i m_i(t))$, then the fixpoint stabiliser conditions imply $m(t) = s(t)$. This will be the basis of the CQ accessing strategy.

For the QQ accessing, we define operators U_B and V_B only acting on B such that $U_B := \prod_{i\in B} k_i^{-D(i)}$, which satisfies $U_B|s_L\rangle = \omega^s|s_L\rangle$ and $V_B := Z_{\Gamma.\{d\}}\prod_{i\in B} k_i^{C(i)-\beta D(i)}$, which satisfies $V_B|s_L\rangle = |(s+1)_L\rangle$.
We also define the extended Bell basis as the following bipartite states over a system $\{a,b\}$: $\forall k,l \in F_q$, $|\beta_{k,l}\rangle_{ab} = Z_a^k X_b^l \sum_{i\in\mathbb{F}_q} \frac{|ii\rangle_{ab}}{\sqrt{q}}$. The result (k,l) of a measurement over $\{a,b\}$ in the Bell basis yield the state as $|\beta_{k,l}\rangle_{ab}$.

CQ Protocol: Let T be a subset of $\{0,\cdots,q-1\}, |T| \geq 2$

1. The dealer prepares the graph state $|G\rangle = \frac{1}{\sqrt{q}}\sum_{i=0}^{q-1} |i(t)\rangle_d |i'_L(t)\rangle_{V\setminus\{d\}}$ and sends one qudit of the state to each player.
2. The dealer randomly measures its qudit among the bases: $\{X^tZ\}_{t\in T}$ and denotes the result $\omega^{s(t)}$. That leaves the state over the players on $|i(t)_L\rangle_{V\setminus\{d\}}$.
3. A player $b \in B$ randomly chooses $t' \in T$ and send t' to the other players in B using their private channel.
4. Each player i in B measures its qudit in the $X^{x_i(t')}Z^{z_i(t')}$ bases (see (2),(3)) and sends the result $\omega^{m_i(t')} \in \{1,\omega,..,\omega^{q-1}\}$ to b.
5. b computes $m(t') = f_{t'}^{-1}(\sum_i m_i(t'))$ (see (5)).
6. Repeat step 1. 2. 3. $p \to \infty$ times. The list of measurement results $s(t)$ and $m(t')$ are the raw keys of the dealer and players B respectively.
7. SECURITY TEST: Follow standard QKD security steps. Through public discussion between d and B first sift the key followed by standard error correction and privacy amplification to generate a secure key (see [28] and [34]).

Correctness : After the QKD security steps the dealer and the authorised set B will be able to share a perfectly secure random key. Furthermore, QQ unauthorised sets for the same graph will not be able to establish such a key (see [28] for proofs).

QQ Protocol: Let $|\zeta\rangle_S = \sum_{i=0}^{q-1} s_i|i\rangle_S \in \mathbb{C}^q$ be a quantum secret.

1. A dealer prepares the state $\frac{1}{\sqrt{q}}\sum_{i=0}^{q-1} s_i Z^i_{\Gamma.\{d\}}|G\backslash d\rangle_{V\setminus\{d\}}$
2. The dealer sends one qudit of the resultant state to each player.
3. (measurement) The authorized set B uses two ancillas qudits $\{a_1,a_2\}$ prepared in the Bell pair state $|\beta_{00}\rangle_{a_1a_2}$, and performs the following two commuting projective measurement on $\{B\,a_1\}$, $V_B^{-1}X_{a_1}^{-1}$ and $U_BZ_{a_1}^{-1}$ on , with result denoted k and l respectively.
4. (correction) B applies Z^kX^{-l} over the second ancilla $\{a_2\}$.

Correctness: U_B and V_B satisfy $U_B|i_L\rangle_{V\setminus\{d\}} = \omega^i|i_L\rangle_{V\setminus\{d\}}$, and $V_B|i_L\rangle_{V\setminus\{d\}} = |(i+1)_L\rangle_{V\setminus\{d\}}$ $\forall i \in \mathbb{F}_q$. We can rewrite the state over $V\setminus\{d\}\cup\{a_1,a_2\}$ as:

$$\begin{aligned}
& \textstyle\sum_{i\in\mathbb{F}_q} s_i|i_L\rangle_{V\setminus\{d\}} \sum_{j\in\mathbb{F}_q} \frac{|jj\rangle_{a_1a_2}}{\sqrt{q}}. \\
= \; & \textstyle\frac{1}{\sqrt{q}}\sum_{l\in\mathbb{F}_q} I_{V\setminus\{d\}}X_{a_1}^lX_{a_2}^l \sum_{i\in\mathbb{F}_q} |i_Li\rangle_{V\setminus\{d\}a_1}s_i|i\rangle_{a_2} \\
= \; & \textstyle\frac{1}{\sqrt{q}}\sum_{l\in\mathbb{F}_q} I_{V\setminus\{d\}}X_{a_1}^lX_{a_2}^l \sum_{k\in\mathbb{F}_q}\sum_{i\in\mathbb{F}_q} \omega^{k.i}\frac{|i_Li\rangle_{V\setminus\{d\}a_1}}{q}\sum_j \omega^{-k.j}s_j|j\rangle_{a_2} \\
= \; & \textstyle\frac{1}{\sqrt{q}}\sum_{l\in\mathbb{F}_q} I_{V\setminus\{d\}}X_{a_1}^lX_{a_2}^l \sum_{k\in\mathbb{F}_q} U_B^kI_{a_1}Z_{a_2}^{-k}\sum_{i\in\mathbb{F}_q}\frac{|i_Li\rangle_{V\setminus\{d\}a_1}}{q}\sum_j s_j|j\rangle_{a_2} \\
= \; & \textstyle\frac{1}{q}\sum_{l,k\in\mathbb{F}_q} U_B^kX_{a_1}^l\sum_{i\in\mathbb{F}_q}\frac{|i_Li\rangle_{V\setminus\{d\}a_1}}{\sqrt{q}}X_{a_2}^lZ_{a_2}^{-k}\sum_{j\in\mathbb{F}_q}s_j|j\rangle_{a_2})
\end{aligned}$$

Table 1 List of typical subsets B of 4 players in the Reed Solomon Graph State described in Fig 1a. For each B, $B \cup \{u \in V\backslash B \mid \sum_{v\in B} D(v).\Gamma(u,v) \neq 0 \bmod q\} = B \cup \{d\} = B \cup \{d\} \cup \{u \in V\backslash(B\cup\{d\}) \mid \sum_{v\in B\cup\{d\}} C(v).\Gamma(v,u) \neq 0 \bmod q\}$, meaning that B can access quantum information, whereas $V\backslash(B\cup\{d\})$, that is all subset of 3 players, cannot. (The remaining subsets are covered by symmetry.)

B	$(D(b))_{b\in Bs}$	$(C(b))_{b\in d\cup B}$	B	$(D(b))_{b\in Bs}$	$(C(b))_{b\in d\cup B}$
$\{v_7,v_1,v_2,v_3\}$	$(1,0,0,0)$	$(1,0,4,3,6)$	$\{v_6,v_1,v_2,v_3\}$	$(1,0,0,0)$	$(1,0,2,2,1)$
$\{v_5,v_1,v_2,v_3\}$	$(1,0,0,0)$	$(1,0,3,4,1)$	$\{v_4,v_1,v_2,v_3\}$	$(1,0,0,0)$	$(1,0,4,6,2)$
$\{v_6,v_7,v_2,v_3\}$	$(3,1,0,0)$	$(1,0,0,1,3)$	$\{v_6,v_7,v_1,v_2\}$	$(1,4,0,0)$	$(1,0,0,3,6)$
$\{v_6,v_7,v_1,v_3\}$	$(4,1,0,0)$	$(1,0,0,5,5)$	$\{v_5,v_7,v_2,v_3\}$	$(3,4,0,0)$	$(1,0,0,6,1)$
$\{v_5,v_7,v_1,v_2\}$	$(1,1,0,0)$	$(1,0,0,2,1)$	$\{v_5,v_7,v_1,v_3\}$	$(4,3,0,0)$	$(1,0,0,1,4)$
$\{v_4,v_7,v_2,v_3\}$	$(4,1,0,0)$	$(1,0,0,2,2)$	$\{v_4,v_7,v_1,v_3\}$	$(3,4,0,0)$	$(1,0,0,6,1)$
$\{v_5,v_6,v_1,v_2\}$	$(3,1,0,0)$	$(1,0,0,1,3)$	$\{v_5,v_6,v_1,v_3\}$	$(1,6,0,0)$	$(1,0,0,4,3)$
$\{v_5,v_6,v_7,v_3\}$	$(2,2,1,0)$	$(1,0,0,0,2)$	$\{v_5,v_6,v_7,v_2\}$	$(4,1,1,0)$	$(1,0,0,0,5)$
$\{v_5,v_6,v_7,v_1\}$	$(5,6,1,0)$	$(1,0,0,0,6)$	$\{v_4,v_5,v_7,v_3\}$	$(1,1,1,0)$	$(1,0,0,0,6)$
$\{v_4,v_5,v_7,v_1\}$	$(4,6,1,0)$	$(1,0,0,0,3)$	$\{v_4,v_5,v_7,v_2\}$	$(6,1,4,0)$	$(1,0,0,0,3)$
$\{v_4,v_5,v_6,v_7\}$	$(5,6,1,2)$	$(1,0,0,0,0)$			

As $V_B^{-1}X_{a_1}^{-1}(U_B^k X_{a_1}^l \sum_{i\in\mathbb{F}_q} \frac{|i_L i\rangle_{V\backslash\{d\}a_1}}{\sqrt{q}}) = \omega^k U_B^k X_{a_1}^l \sum_{i\in\mathbb{F}_q} \frac{|i_L i\rangle_{V\backslash\{d\}a_1}}{\sqrt{q}}$ and $U_B Z_{a_1}^{-1}(U_B^k X_{a_1}^l \sum_{i\in\mathbb{F}_q} \frac{|i_L i\rangle_{V\backslash\{d\}a_1}}{\sqrt{q}}) = \omega^l U_B^k X_{a_1}^l \sum_{i\in\mathbb{F}_q} \frac{|i_L i\rangle_{V\backslash\{d\}a_1}}{\sqrt{q}}$, the projective measurement according to $V_B^{-1}X_{a_1}^{-1}$ and $U_B Z_{a_1}^{-1}$ reveals the syndrome (k,l), such that the correction $Z^k X^{-l}$ over the ancilla $\{a_2\}$ leaves the state as $\sum_i s_i|i\rangle_{a_2}$.

4.2 Example

We illustrate the use of characterisation of the access structure in a multigraph with a Reed Solomon Graph State that allows a quantum secret (or equivalently a random key of *dits*) to be shared between a dealer and all subset of at least $\frac{n+1}{2}$ players among a set of n players over a field of q elements, with $q \geq n$. We refer to [29], [9] for more details about Reed Solomon Graph for secret sharing.

We saw $B \subset V\backslash\{d\}$ can access quantum information with respect to d in G iff there exist $D : B \to \mathbb{F}_q$ and $C : B\cup\{d\} \to \mathbb{F}_q$ such that $C(d) = 1$ and

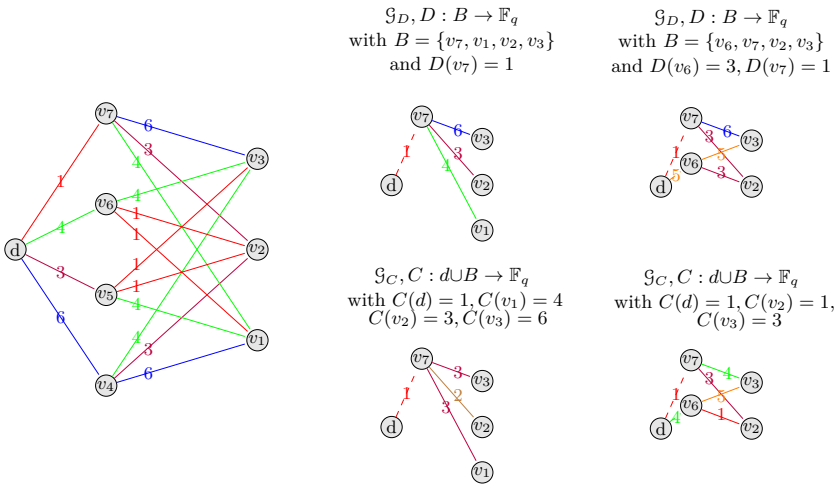
$$\sup(\Gamma[B, V\backslash B].D) = \{d\}$$
$$\Gamma[B\cup\{d\}, V\backslash(B\cup\{d\})].C = 0$$

We rewritte these conditions in the following way: $B \subset V\backslash\{d\}$ can access quantum information in G iff there exist $D : B \to \mathbb{F}_q$ and $C : B \cup \{d\} \to \mathbb{F}_q$ such that $C(d) = 1$ and

$$\begin{cases} B \cup \{u \in V\backslash B \mid \sum_{v\in B} D(v).\Gamma(u,v) \neq 0 \bmod q\} = B \cup \{d\}. & (5) \\ B \cup \{d\} \cup \{u \in V\backslash(B\cup\{d\}) \mid \sum_{v\in B\cup\{d\}} C(v).\Gamma(v,u) \neq 0 \bmod q\} = B \cup \{d\} & (6) \end{cases}$$

For $A : V \to \mathbb{F}_q$, we call $\mathcal{G}_A = (V_A, \Gamma_A)$ the subgraph induced by A and its neighbours such that: $V_A = \sup(A) \cup \{v \in V\backslash\sup(A) \mid \Gamma[\sup(A), V\backslash\sup(A)].A(v) \neq 0 \bmod q\}$ and $\forall v_i \in \sup(A)$, $\begin{cases} \Gamma_A(v_i,v_j) = A(v_i)A(v_j).\Gamma(v_i,v_j) \text{ if } v_j \in \sup(A) \\ \Gamma_A(v_i,v_j) = A(v_i)\Gamma(v_i,v_j) \text{ if } v_j \in V_A\backslash\sup(A) \end{cases}$

For example, let $G = (V,\Gamma)$, $d \in V$, $|V| = 8$, be the $(4,3,7)_7$ Reed Solomon Graph State given in Fig 1a. Such a graph can be used by dealer d to encode any quantum secret $|\xi\rangle \in \mathbb{C}^7$ and share it between 7 players such that all subset of at least 4 players can recover the secret, whereas any subset of 3 players or less cannot have any information about it. We can reprove this result using the previous graph characterisation, that is by checking if conditions (5) (6) are satisfied in a basis G for all subset $B \subset V\backslash\{d\}$ of 4 players. In fig 1b, we give the relevant induced graphs for two different subsets B. And in table 1 we give a list of relevant multiset $D : B \to \mathbb{F}_q$ and $C : B\cup\{d\} \to \mathbb{F}_q$ for typical subsets B of four players.



(a) $(4, 3, 7)_7$ Reed Solomon QQ Graph scheme.

(b) Examples of subgraphs $\mathcal{G}_D$ associated with subset B showing that B can access classical information in G and of subgraphs $\mathcal{G}_C$ associated with $d \cup B$ showing that $V \backslash B$ cannot have classical information in G.

Figure 1 Checking quantum accessibility in a $(4, 3, 7)_7$ Reed Solomon Graph.

5 Existence of $((k, n))_q$ scheme

In this section, we focus on the properties of the secret sharing scheme realised by a given $\mathbb{F}_q$-graph, as well as the existence of $\mathbb{F}_q$-graphs realising a given secret sharing protocol. A $\mathbb{F}_q$-graph G of order n with a particular dealer d is said to realise a $((k, n))_q$ scheme if $k - 1 = \max_{B \subseteq V \backslash \{d\}}(\partial_d \text{cutrk}_G(B) \geq 0)$. In other words, G realises a $((k, n))_q$ scheme if all sets of at least k players can recover a quantum secret and there exists a set of $k - 1$ players which cannot. A $\mathbb{F}_q$-graph which realises an $((k, n))_q$ scheme can be used as an $(k, k' \geq n - k, n)_q$ CQ protocol or $(k, n - k, n)_q$ QQ protocol as described in section 4 (note that they can also be used for $(k, k' \geq n - k, n)_q$ schemes to share a quantum secret using hybrid protocols (e.g. [4, 21, 11, 12])).

5.1 Finding new schemes

Theorem 9 offers a combinatorial characterisation of quantum accessibility, and raises as a consequence several questions about the complexity of deciding: (i) whether a given set of players can access a quantum secret in a given q-multigraph? (ii) whether a given q-multigraph realises a $((k, n))$ protocol? (iii) whether, given q, n and k, there exists an $\mathbb{F}_q$-graphs realising a $((k, n))$ protocol?

Problem i Given an $\mathbb{F}_q$-graph G of order n with a particular dealer d and a set B of k players, deciding whether B can access a quantum secret consists of deciding whether $\partial_d \text{cutrk}_G(B) = -1$. This can be decided efficiently since $\partial_d \text{cutrk}_G(B)$ is computed in $O(nk^{1.38})$ operations using the Gaussian elimination for computing the rank [6, 20].

Problem ii Given a $\mathbb{F}_q$-graph G of order n and $\alpha \in [0, 1]$, deciding whether G is a $((\alpha n, n))$ scheme can be done by enumerating all the $\binom{n}{\alpha n}$ sets of players of size αn and for each of them deciding whether they can access a quantum secret. It leads to $O(n^{2.38} 2^{nH_2(\alpha)})$

operations. This problem is NP-complete, as it has been shown to be NP complete when $q = 2$ [7], and also hard in terms of parameterised complexity as it is hard for $W[1]$ [7].

Problem *iii* Given n, α, and q, deciding whether there exists a $((\alpha n, n))$ $\mathbb{F}_q$-graph? A brute-force approach consists in enumerating all the $q^{\frac{n(n-1)}{2}}$ $\mathbb{F}_q$-graphs of order n and then decide whether they realise a $((\alpha n, n))$ protocol. It leads to $O(q^{\frac{n(n-1)}{2}} n^{2.38} 2^{nH_2(\alpha)})$ operations. This can be implemented for small values of n only and permits to prove that there is no $(4, 3, 7)_3$ QQ secret sharing with qutrit graph state.

Solving problem i can be done with the similar algortihm C of [13]. Note that for one thing, the later is more general and can be applied to input states (that is quantum secrets) and to multigraphs of arbitrary dimension (not necessarily prime number). For another thing, it concerns rather the access to partial information. Also it is not optimised for problem i of our particular interest.

In the following sections, we develop a different approach for deciding the existence of $((\alpha n, n))$ $\mathbb{F}_q$-graphs realising. We show an upper and a lower bound on the minimal α such that there exists an $\mathbb{F}_q$-graph realising a $((\alpha n, n))$ protocol. The upper bound (Theorem 11) is based on non constructive probabilistic methods, whereas the lower bound (Theorem 14) is based on a counting argument.

5.2 Existence of q-multigraphs realising $((\alpha n, n))_q$ schemes

In this section, we prove a Gilbert-Varshamov-like result: for any α such that $H_{q^2}(1-\alpha) < \frac{1}{2}$ there exists a q-multigraph realising a $((\alpha n, n))_q$ scheme. The proof is using probabilistic methods and is, as a consequence, non constructive. However, we prove that a random q-multigraph satisfies such $((\alpha n, n))_q$ scheme with high probability as long as $H_{q^2}(1-\alpha) < \frac{1}{2}$.

▶ **Lemma 10.** *For any q-multigraph $G = (V, \Gamma)$ of order n, and any $\alpha \in [0.5, 1]$, if for any multiset $C : V \to \mathbb{F}_q$, $|\sup(C) \cup \sup(\Gamma.C)| > (1-\alpha)n$ then for any $d \in V$ and any $B \subseteq V \backslash \{d\}$ such that $|B| \geq \alpha n$, $\partial_d \mathrm{cutrk}_G(B) = -1$.*

Proof. For any $B \subseteq V$ such that $|B| \geq \alpha n$, $\ker(\Gamma[V \backslash B]) = \{0\}$, otherwise there would be a multiset C such that $\sup(C) \subseteq V \backslash B$ and $|\sup(C) \cup \sup(\Gamma.C)| \leq (1-\alpha)n$. So for any $B \subseteq V$ such that $|B| \geq \alpha n$, $\mathrm{cutrk}_G(B) = n - |B|$. As a consequence, for any $d \in V$ and any $B \subseteq V \backslash \{d\}$ such that $|B| \geq \alpha n$, $\partial_d \mathrm{cutrk}_G(B) = n - |B \cup \{d\}| - (n - |B|) = -1$. Thus $\partial_d \mathrm{cutrk}_G(B) = -1$ ◀

A random $\mathbb{F}_q$-graph $G(n, 1/q)$ is a $\mathbb{F}_q$-graph of order n such that, for every pair of vertices u and v, the number of edges between u and v is chosen uniformly at random in $\mathbb{F}_q$.

▶ **Theorem 11.** *Given $q \geq 2$, and $\alpha \in [0.5, 1]$ such that $H_{q^2}(1-\alpha) < \frac{1}{2}$, for any $n \in \mathbb{N}$, a random q-multigraph $G(n, 1/q)$ realises a $((\alpha n, n))_q$ scheme with probability $1 - 2^{\Omega(n)}$, where d is any vertex of $G(n, 1/q)$.*

Proof. Let $\mathcal{C}_\alpha = \{C : V \to \mathbb{F}_q, |\sup(C)| \leq (1-\alpha)n\}$. For any $C \in \mathcal{C}_\alpha$, let A_C be the (bad) event $|\sup(C) \cup \sup(\Gamma.C)| \leq (1-\alpha)n$.
For any $C \in \mathcal{C}_\alpha$, $Pr(A_C) = \frac{1}{q^{(1-c)n}} \sum_{k=0}^{(1-\alpha-c)n} \binom{(1-c)n}{k} (q-1)^k$ where $c = |\sup(C)|/n$, and $\sum_{C \in \mathcal{C}_\alpha} Pr(A_C) = \sum_{j=0}^{(1-\alpha)n} f(j)$ with $f(j) = \sum_{C s.t. |\sup(C)|=j} Pr(A_C)$.
In the following, we show an upperbound on $f(k)$. For any $c \in [0, 0.5]$, $f(cn) = \binom{n}{cn}(q-1)^{cn} \frac{1}{q^{(1-c)n}} \sum_{k=0}^{(1-\alpha-c)n} \binom{(1-c)n}{k} (q-1)^k \leq \frac{(q-1)^{cn}}{q^{(1-c)n}} 2^{nH_2(c) + (1-c)nH_2(\frac{1-\alpha-c}{1-c})} (q-1)^{(1-\alpha-c)n} = 2^{ng(c)}$ where $g(c) = H_2(c) + (1-c)H_2(\frac{\alpha}{1-c}) + (1-\alpha)\log_2(q-1) - (1-c)\log_2(q)$. $g'(c) =$

$-\log_2(c)+\log_2(1-\alpha-c)+\log_2(q)$, so $g'(c)=0 \iff c=\frac{q}{q+1}(1-\alpha)$. As a consequence, $g(c) \leq g(\frac{q}{q+1}(1-\alpha)) = -\alpha\log_2(\alpha)-(1-\alpha)\log_2(\alpha)+(1-\alpha)\log_2(q^2-1)-\log_2(q) = \log_2(q)(2H_{q^2}(1-\alpha)-1)$. Thus, $\sum_{C\in\mathcal{C}_\alpha} Pr(A_C) \leq (1-\alpha)nq^{n[2H_{q^2}(1-\alpha)-1]}$, so, thanks to the union bound, $Pr(\bigcap_{C\in\mathcal{C}_\alpha}\overline{A_C}) \geq 1-(1-\alpha)nq^{n[2H_{q^2}(1-\alpha)-1]} = 1-2^{\Omega(n)}$ when $2H_{q^2}(1-\alpha)-1<0$. So according to lemma 10, $\kappa_Q(G,d)\leq \alpha n$ for any vertex d when $H_{q^2}(1-\alpha)<\frac{1}{2}$. ◀

Theorem 11 extends the upper bound of the binary case ($q=2$) [21]. Notice that even if a random $\mathbb{F}_q$-graph realises a $((\alpha n,n))_q$ scheme with probability almost 1, double checking whether a (randomly chosen) $\mathbb{F}_q$-graph actually realises a $((\alpha n,n))_q$ scheme is a hard task (see Problem (*ii*) in section 5.1).

5.3 Lower bound on quantum accessibility

The no cloning theorem implies that for any $((\alpha n,n))$ secret sharing protocol, $\alpha>0.5$. In the following we improve this lower bound for secret sharing schemes based on qudit graph states. The lower bound on α depends on the dimension q (see Theorem 14), the value of the lower bound is plotted for small values of q in figure 2.

The lower bound is based on the properties of the *kernel with respect to the dealer* defined as follows:

▶ **Definition 12.** Given a q-multigraph G, for any $d\in V(G)$ and any $B\subseteq V(G)\backslash\{d\}$, let $\mathcal{S}_d(B)=\ker(\Gamma_G[B\cup\{d\}])\backslash\ker(\Gamma_G[B])$ be the kernel of B with respect to d.

▶ **Lemma 13.** *Given a q-multigraph G, for any $d\in V(G)$ and any $B\subseteq V(G)\backslash\{d\}$, if $\partial_d \mathrm{cutrk}_G(B)=-1$, there exists $C\in\mathcal{S}_d(B)$ such that*

$$|\sup(C)| < \frac{q}{q+1}\mathrm{cutrk}_G(B)$$

Proof. Since $\mathrm{cutrk}_G(B\cup\{d\})-\mathrm{cutrk}_G(B)=-1$, $\dim(\ker(\Gamma_G[B\cup\{d\}]))-\dim(\ker(\Gamma_G[B]))=2$. Moreover, $\ker(\Gamma_G[B])\subseteq\ker(\Gamma_G[B\cup\{d\}])$, so $|\mathcal{S}_d(B)|=(q^2-1).q^t$ where $t=\dim(\ker(\Gamma_G[B]))$. Let $M=\binom{I}{M'}$ a matrix in standard form (or reduced column echelon form) generating $\Gamma_G[B\cup\{d\}]$. Since $|\mathcal{S}_d(B)|=(q^2-1).q^t$ and $|\ker(\Gamma_G[B\cup\{d\}])|=q^{t+2}$, there exist two columns C_1 and C_2 of M such that $\forall(x,y)\in[0,q-1]^2\backslash\{(0,0)\}$, $x.C_1+y.C_2\in\mathcal{S}_d(B)$. Notice that since M is in standard form, $|\sup(C_1)\cup\sup(C_2)|\leq|B|+1-t$. Moreover for any $v\in\sup(C_1)\cup\sup(C_2)$, v has a zero multiplicity in $q-1$ vectors of the q^2-1 linear combinations $x.C_1+y.C_2$ for $x,y\in[0,q-1]\backslash\{(0,0)\}$, so $\sum_{(x,y)\in[0,q-1]^2\backslash\{(0,0)\}}|\sup(x.C_1+y.C_2)| = (q^2-1-(q-1)).|\sup(C_1)\cup\sup(C_2)|$, so there exists $C\in\mathcal{S}_d(B)$ such that $|\sup(C)|\leq \frac{q^2-q}{q^2-1}(|B|+1-t)=\frac{q}{q+1}(\mathrm{cutrk}_G(B)+1)<\frac{q}{q+1}\mathrm{cutrk}_G(B)$. ◀

▶ **Theorem 14.** *If a q-multigraph G of order n realises a $((\alpha n,n))_q$ scheme, then*

$$\binom{n}{\frac{(1-\alpha)qn}{q+1}}\binom{\alpha n}{(2\alpha-1)n} \geq \frac{(2\alpha-1)(1-\alpha)}{2}\binom{n}{\alpha n}$$

Asymptotically, as n tends to infinity, α satisfies:

$$H_2(\frac{\alpha q+1}{q+1})+\alpha H_2(\frac{1-\alpha}{\alpha}) \geq H_2(\alpha)$$

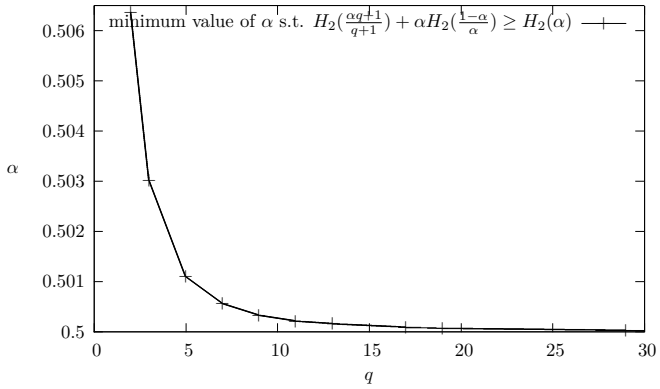


Figure 2 Lower bound on the accessibility to quantum information in a $((k,n))_q$ scheme. There is no $((k,n))_q$ scheme with $k \leq \alpha n$

Proof. Given B_0 of size αn, according to lemma 13 there exists $C_0 \in \mathcal{S}_d(B_0)$ such that $|\sup(C_0)| < \frac{q}{q+1}(1-\alpha)n$. Notice that the set $\sup(C_0) \cup \sup(\Gamma_G.C_0)$ has some partial information about the secret so $|\sup(C_0) \cup \sup(\Gamma_G.C_0)| \geq (1-\alpha)n$. Moreover for any B of size αn, if $C_0 \in \mathcal{S}_q(B)$ then $\sup(C) \cup \sup(\Gamma_G.C) \subseteq B$. So there are at most $\binom{n-1-(1-\alpha)n}{\alpha n-(1-\alpha)n} = \binom{\alpha n-1}{(2\alpha-1)n}$ sets $B \subseteq V \backslash \{d\}$ of size αn such that $C_0 \in \mathcal{S}_d(B)$. For any B of size αn there is a C which support is of size at most $\frac{q}{q+1}(1-\alpha)n - 1$, any every such C is associated with at most $\binom{\alpha n-1}{(2\alpha-1)n}$ such Bs, so a counting argument implies $\binom{n-1}{\alpha n} \leq \binom{\alpha n-1}{(2\alpha-1)n} \sum_{i=1}^{\frac{q}{q+1}(1-\alpha)n-1} \binom{n-1}{i}$. Moreover, $\sum_{i=1}^{\frac{q}{q+1}(1-\alpha)n-1} \binom{n-1}{i} \leq \frac{1+\alpha q}{q(2\alpha-1)} \binom{n-1}{\frac{(1-\alpha)qn}{q+1}-1} = \frac{(1-\alpha)(1+\alpha q)}{(2\alpha-1)(q+1)} \binom{n}{\frac{(1-\alpha)qn}{q+1}}$. So, $\frac{\binom{n}{\alpha n}}{\binom{\alpha n-1}{(2\alpha-1)n}} = \frac{\alpha}{(1-\alpha)^2} \frac{\binom{n-1}{\alpha n}}{\binom{\alpha n-1}{(2\alpha-1)n}} \leq \frac{\alpha(1+\alpha q)}{(2\alpha-1)(1-\alpha)(q+1)} \binom{n}{\frac{(1-\alpha)qn}{q+1}} \leq \frac{2}{(2\alpha-1)(1-\alpha)} \binom{n}{\frac{(1-\alpha)qn}{q+1}}$.

Since $2^{n(H_2(p)+o(1))} \leq \binom{n}{pn} \leq 2^{nH_2(p)}$, asymptotically, as n tends to infinity, α satisfies the equation $H_2(\frac{\alpha q+1}{q+1}) + \alpha H_2(\frac{1-\alpha}{\alpha}) \geq H_2(\alpha)$. ◀

6 Discussion

In this work we have studied the encoding of classical and quantum information onto graph states of qudits, and its application for secret sharing schemes. We have given complete graphical characterization of which sets of vertices (players) can access the information, and shown how this can be done both for classical and quantum information. Using this characterization we have given bounds on which protocols are possible and how difficult the access structure is to calculate given a graph.

Whilst we have focused on the application of our results for secret sharing, there may be applications to other quantum information protocols. Indeed, the QQ encoding defined in section 2.3 is exactly the same encoding procedure used in measurement based quantum computing and error correction, so we can expect that these results have implications in both these domains. Furthermore, quantum secret sharing is intimately linked to error correction [28, 9]. All secret sharing schemes are error correcting schemes, and the QQ protocols presented here are equivalent to all possible stabilizer codes [28]. Thus, the existence of $((\alpha n, n))$ protocols is an existence statement about error correcting protocols too, and the no goes on secret sharing imply no-goes for all stabilizer codes - so that there are no stabilizer codes with parameters violating our lower bounds.

Acknowledgements. The authors want to thank Mehdi Mhalla and David Cattanéo for fruitful discussions. This work has been funded by the ANR-10-JCJC-0208 CausaQ grant, the FREQUENCY (ANR-09-BLAN-0410), HIPERCOM (2011-CHRI-006) projects, and by the Ville de Paris Emergences program, project CiQWii.

References

1 M. Bahramgiri,S. Beigi, *Graph states under the action of local Clifford group in non-binary case* arXiv:quant-ph/0610267 (2006).
2 S. Beigi, I. Chuang, M. Grassl, P. Shor, B. and Zeng, *Graph concatenation for quantum codes.* J. Math. Phys. 52, 022201 (2011).
3 M. Ben-Or, C. Crépeau, D. Gottesman, A. Hassidim, A. Smith, *Secure Multiparty Quantum Computation with (Only) a Strict Honest.* Proc. 47th Annual IEEE Symposium on the Foundations of Computer Science (FOCS '06), pp. 249-260 (2006).
4 A. Broadbent, P. Chouha, A. Tapp, *The GHZ state in secret sharing and entanglement simulation.* Third International Conference on Quantum, Nano and Micro Technologies, ICQNM'09, 59-62 (2009).
5 A. Bouchet *Circle Graph Obstructions* Journal of Combinatorial Theory, Series B, Vol 60, 1,pp 107-144 (1994).
6 J.R. Bunch, J.E. Hopcroft. *Triangular Factorization and Inversion by Fast Matrix Multiplication.* Mathematics of Computation, 28(125):231236, (1974).
7 David Cattanéo, Simon Perdrix. *Parametrized Complexity of Weak Odd Domination Problems.* arXiv:1206.4081 (2012).
8 M. Christandl, A. Winter, *Uncertainty, Monogamy and Locking of Quantum Correlations.* IEEE Trans Inf Theory, vol 51, no 9, pp 3159-3165 (2005).
9 R. Cleve, D. Gottesman, H.K. Lo, *How to share a quantum secret.* Phys. Rev. Lett. **83**, 648 (1999).
10 A.K. Ekert, *Quantum cryptography based on Bell's theorem.* Phys. Rev. Lett., **67**, 6, pp. 661–663 (1991).
11 B. Fortescue, G. Gour, *Reducing the quantum communication cost of quantum secret sharing.* IEEE Trans. Inf. Th. 58(10), pp. 6659 - 6666 (2012)
12 V. Gheorghiu, *Generalized Semi-Quantum Secret Sharing Schemes.* Phys. Rev. A 85, 052309 (2012)
13 V. Gheorghiu, S.Y. Looi, R.B. Griffiths, *Location of quantum information in additive graph codes* Phys. Rev. A, **81**, 3, pp. 032326, (2010).
14 M. Grassl, A. Klappenecker, M. Rötteler, *Graphs, Quadratic forms and Quantum Codes*, IEEE International Symposium on Information Theory (ISIT 2002), p.45 (2002).
15 S. Gravier, J. Javelle, M. Mhalla, S. Perdrix, *On Weak Odd Domination and Graph-based Quantum Secret Sharing.* arXiv:1112.2495 (2011).
16 S. Gravier, J. Javelle, M. Mhalla, S. Perdrix, *Optimal accessing and non-accessing structures for graph protocols.* arXiv:1109.6181 (2011).
17 M. Hein, J. Eisert, H. J. Briegel. *Multiparty entanglement in graph states.* Phys. Rev. A 69, 062311 (2004).
18 M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. Van den Nest, H. J. Briegel, *Entanglement in graph states and its applications* in Quantum Computers, Algorithms and Chaos, Proceedings of the International School of Physics Enrico Fermi, Vol. 162 (2006).
19 M. Hillery, V. Bužek, A. Berthiaume, *quantum secret sharing.* Phys. Rev. A **59**, 1829 (1999).
20 O.H. Ibarra, S. Moran, R. Hui. *A Generalization of the Fast LUP Matrix Decomposition Algorithm and Applications.* Journal of Algorithms, 3(1):4532656, (1982).
21 J. Javelle, M. Mhalla, S. Perdrix, *New Protocols and Lower Bound for Quantum Secret Sharing with Graph States.* TQC'12. LNCS Vol 7582, pp 1-12 (2013).
22 J. Javelle, M. Mhalla, S. Perdrix, *On the Minimum Degree up to Local Complementation: Bounds and Complexity.* WG'12. LNCS Vol 7551, pp 138-147 (2012).

23 M.M. Kanté, M. Rao. *The Rank-Width of Edge-Coloured Graphs.* Theory of Computing Systems, 1-46, (2012).

24 E. Kashefi, D. Markham, M. Mhalla, and S. Perdrix. *Information flow in Secret Sharing Protocols. Electronic Proceedings in Theoretical Computer Science*, 9:87–97 (2009).

25 A. Karlsson, M. Koashi, N. Imoto, *Quantum entanglement for secret sharing and secret splitting. Phys. Rev. A* **59**, 162–168, (1999).

26 A. Keet, B. Fortescue, D. Markham and B. C. Sanders, *Quantum secret sharing with qudit graph states.* Phys. Rev. A **82**, 062315 (2010).

27 A. Ketkar, A. Klappenecker, S. Kumar, P. K. Sarvepalli, *Nonbinary Stabilizer Codes Over Finite Fields.* IEEE Trans. Inf. Th. 52, 4892 (2005).

28 A. Marin, D. Markham, *On the equivalence between sharing quantum and classical secrets, and error correction.* arxiv:1205.4182 (2012)

29 A. Marin, D. Markham, *High dimensional CSS code and application to secret sharing.* in preparation (2013).

30 D. Markham, B. C. Sanders, *Graph State for Quantum Secret Sharing.* Phys. Rev. A, **78**, (2008).

31 S. Oum, P. Seymour. *Approximating rank-width and clique-width quickly.* Journal ACM Transactions on Algorithms (TALG), vol 5, 1-20 (2008).

32 D. Schlingemann, R. F. Werner, *Quantum error-correcting codes associated with graphs.* Phys. Rev. A, vol. 65, p. 012308 (2001).

33 A. Shamir, *How to share a secret.* Communications of the ACM, **22**, 612–613 (1979).

34 L. Sheridan, V. Scarani, *Security proof for quantum key distribution using qudit systems.* Phys. Rev. A **82**, 030301(R) (2010).

35 M. Van den Nest, J. Dehaene, B. De Moor *Graphical description of the action of local Clifford transformations on graph states.* Physical Review A (69) 022316 (2004).

A Appendix-QQ Encoding-Decoding Operations

The QQ encoding-decoding can basically be done by three typical ways. The first method is based on projective Bell measurements (possibly extended to a $|B|+1$ length state) and the two last one are accessible by local measurements and/or series of two qudit control operations, which should finally result in a similar experimental complexity. We briefly describe the three encoding methods $E1, E2, E3$ and decoding $D2, D3$. ($D1$ has been done in section 4.1). For a graph $G=(V,\Gamma)$, with $d,u \in V$ such that $\Gamma(d,u)\neq 0$, $W:=V\backslash\{d\}$, a quantum secret $|\xi\rangle_S := \sum_{i=0}^{q-1} s_i|i\rangle_S$, we write $\bar{X} := Z_{\Gamma.\{d\}}$ and $\bar{Z} := (X_u Z_{\Gamma.\{u\}})^{-\Gamma(u,d)^{-1}}$, as they act like logical operators over the bases states over W, that is $\bar{Z}|i_L\rangle = \omega^i|i_L\rangle$, $\bar{X}|i_L\rangle = |(i+1)_L\rangle$ with notation of 4.1.

E1 $|\xi\rangle|G\rangle = \sum_{i\in\mathbb{F}_q} s_i|i\rangle_S \sum_{j\in\mathbb{F}_q} \frac{|j\rangle_D|j_L\rangle_W}{\sqrt{q}}$
$= \frac{1}{\sqrt{q}}\sum_{i,j\in\mathbb{F}_q}|i\rangle_S|j\rangle_D s_i|j_L\rangle_W = \frac{1}{\sqrt{q}}\sum_{l\in\mathbb{F}_q} I_S X_D^l \bar{X}_W^l(\sum_{i\in\mathbb{F}_q}|i\rangle_S|i\rangle_D s_i|i_L\rangle_W)$
$= \frac{1}{\sqrt{q}}\sum_{l\in\mathbb{F}_q} I_S X_D^l \bar{X}_W^l(\sum_{k\in\mathbb{F}_q}\sum_{i\in\mathbb{F}_q}\omega^{k.i}\frac{|i\rangle_I|i\rangle_D}{q}\sum_{j\in\mathbb{F}_q}\omega^{-k.j}s_j|j_L\rangle_W)$
$= \frac{1}{\sqrt{q}}\sum_{l\in\mathbb{F}_q} I_S X_D^l \bar{X}_W^l(\sum_{k\in\mathbb{F}_q} Z_I^k I_D \bar{Z}_W^{-k}\sum_{i\in\mathbb{F}_q}\frac{|i\rangle_I|i\rangle_D}{q}\sum_{j\in\mathbb{F}_q}s_j|j_L\rangle_W)$
$= \frac{1}{q}\sum_{l,k\in\mathbb{F}_q} Z_S^k X_D^l(\sum_{i\in\mathbb{F}_q}\frac{|i\rangle_S|i\rangle_D}{\sqrt{q}})\bar{X}_W^l\bar{Z}_W^{-k}\sum_{j\in\mathbb{F}_q}s_j|j_L\rangle_W)$

so that applying the correction: $\bar{Z}^k\bar{X}^{-l}$ over $V\backslash\{d\}$, according to the syndrom (l,k) of a Bell measurement over $\{S,D\}$, leaves the state over W as $\sum_{i\in\mathbb{F}_q} s_i|i_L\rangle$.

E2 $C\bar{X}_{dW}|\xi\rangle|0_L\rangle = \sum_{i\in\mathbb{F}_q} s_i|i\rangle\bar{X}^i|0_L\rangle = \sum_{i\in\mathbb{F}_q} s_i|ii_L\rangle = \sum_{i\in\mathbb{F}_q} s_i X^i|0\rangle|i_L\rangle$
$= \sum_{i\in\mathbb{F}_q} s_i X^i \sum_{j\in\mathbb{F}_q}\frac{|b_j\rangle}{\sqrt{q}}|i_L\rangle = \sum_{i\in\mathbb{F}_q} s_i X^i\sum_j\frac{Z^{-j}|b_0\rangle}{\sqrt{q}}|i_L\rangle$
$= \frac{1}{\sqrt{q}}\sum_{i,j} s_i\omega^{i.j}Z^{-j}|b_0\rangle|i_L\rangle = \sum_{j\in\mathbb{F}_q}\frac{|b_j\rangle}{\sqrt{q}}(\bar{Z}^j\sum_{i\in\mathbb{F}_q}s_i|i_L\rangle)$

where $|b_j\rangle = Z^{-j}|+\rangle$ constitutes the X basis, so that applying the correction $\bar{Z}^{-j}$ over W,

according to the result j of a X_d measurement, leaves the state to distribute as $\sum_i s_i|i_L\rangle$ (see also [2]).

E3 $C\bar{Z}_{dW}H_dC\bar{X}_{dW}|\xi\rangle|0_L\rangle = C\bar{Z}_{dW}H_d(\sum_i s_i|i\rangle|i_L\rangle) = C\bar{Z}_{dW}(\sum_i s_i Z_d^{-i}|+\rangle|i_L\rangle)$
$= \frac{1}{\sqrt{q}}C\bar{Z}_{dW}(\sum_{i,k} s_i|k\rangle\bar{Z}^{-k}|i_L\rangle) = |+\rangle\sum_i s_i|i_L\rangle)$.

The same process can be done for the decoding by an authorised set B, where the operators U_B and V_B defined in 4.1 will act as $\bar{Z}$ and $\bar{X}$ operators respectively. An ancilla qudit $\{a\}$ is prepared in the state $|+\rangle_a$ by B.

D2 ${CV^{-1}}_{aB}|+\rangle_a(\sum_{j\in\mathbb{F}_q} s_j|j_L\rangle_W) = \frac{1}{\sqrt{q}}\sum_{k\in\mathbb{F}_q} X_a^{-k}(\sum_{i\in\mathbb{F}_q} s_i|i\rangle_a)|k_L\rangle_W$.

D3 $CU_{aB}.H_a.\frac{1}{\sqrt{q}}\sum_{k\in\mathbb{F}_q} X^{-k}(\sum_{i\in\mathbb{F}_q} s_i|i\rangle_a)|k_L\rangle_W = \sum_{i\in\mathbb{F}_q} s_i|b_i\rangle_a\sum_{k\in\mathbb{F}_q^m}\frac{|k_L\rangle_W}{\sqrt{q}}$.