

15th Conference on the Theory of Quantum Computation, Communication and Cryptography

TQC 2020, June 9–12, 2020, Riga, Latvia

Edited by

Steven T. Flammia



LIPICs



Editors

Steven T. Flammia 

University of Sydney, Australia
steven.flammia@sydney.edu.au

ACM Classification 2012

Theory of computation → Quantum computation theory; Theory of computation → Quantum complexity theory; Theory of computation → Quantum information theory; Theory of computation → Quantum communication complexity; Hardware → Quantum communication and cryptography; Hardware → Quantum error correction and fault tolerance

ISBN 978-3-95977-146-7

Published online and open access by

Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at <https://www.dagstuhl.de/dagpub/978-3-95977-146-7>.

Publication date

June, 2020

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 <https://portal.dnb.de>.

License

This work is licensed under a Creative Commons Attribution 3.0 Unported license (CC-BY 3.0):
<https://creativecommons.org/licenses/by/3.0/legalcode>.



In brief, this license authorizes each and everybody to share (to copy, distribute and transmit) the work under the following conditions, without impairing or restricting the authors' moral rights:

- Attribution: The work must be attributed to its authors.

The copyright is retained by the corresponding authors.

Digital Object Identifier: 10.4230/LIPIcs.TQC.2020.0

ISBN 978-3-95977-146-7

ISSN 1868-8969

<https://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.

Editorial Board

- Luca Aceto (*Chair*, Gran Sasso Science Institute and Reykjavik University)
- Christel Baier (TU Dresden)
- Mikolaj Bojanczyk (University of Warsaw)
- Roberto Di Cosmo (INRIA and University Paris Diderot)
- Javier Esparza (TU München)
- Meena Mahajan (Institute of Mathematical Sciences)
- Dieter van Melkebeek (University of Wisconsin-Madison)
- Anca Muscholl (University Bordeaux)
- Luke Ong (University of Oxford)
- Catuscia Palamidessi (INRIA)
- Thomas Schwentick (TU Dortmund)
- Raimund Seidel (Saarland University and Schloss Dagstuhl – Leibniz-Zentrum für Informatik)

ISSN 1868-8969

<https://www.dagstuhl.de/lipics>

Contents

Preface	
<i>Steven T. Flammia</i>	0:vii
Conference Organization	
.....	0:ix
Exponential Quantum Communication Reductions from Generalizations of the Boolean Hidden Matching Problem	
<i>João F. Doriguello and Ashley Montanaro</i>	1:1–1:16
Improved Approximate Degree Bounds for k -Distinctness	
<i>Nikhil S. Mande, Justin Thaler, and Shuchen Zhu</i>	2:1–2:22
Building Trust for Continuous Variable Quantum States	
<i>Ulysse Chabaud, Tom Douce, Frédéric Grosshans, Elham Kashefi, and Damian Markham</i>	3:1–3:15
Unclonable Quantum Encryption via Oracles	
<i>Anne Broadbent and Sébastien Lord</i>	4:1–4:22
Quasirandom Quantum Channels	
<i>Tom Bannink, Jop Briët, Farrokh Labib, and Hans Maassen</i>	5:1–5:20
Towards Quantum One-Time Memories from Stateless Hardware	
<i>Anne Broadbent, Sevag Gharibian, and Hong-Sheng Zhou</i>	6:1–6:25
Beyond Product State Approximations for a Quantum Analogue of Max Cut	
<i>Anurag Anshu, David Gosset, and Karen Morenz</i>	7:1–7:15
Simpler Proofs of Quantumness	
<i>Zvika Brakerski, Venkata Koppula, Umesh Vazirani, and Thomas Vidick</i>	8:1–8:14
Quantum Algorithms for Computational Geometry Problems	
<i>Andris Ambainis and Nikita Larka</i>	9:1–9:10
Quantum Coupon Collector	
<i>Srinivasan Arunachalam, Aleksandrs Belovs, Andrew M. Childs, Robin Kothari, Ansis Rosmanis, and Ronald de Wolf</i>	10:1–10:17
Fast and Effective Techniques for T-Count Reduction via Spider Nest Identities	
<i>Niel de Beaudrap, Xiaoning Bian, and Quanlong Wang</i>	11:1–11:23
A Device-Independent Protocol for XOR Oblivious Transfer ¹	
<i>Srijita Kundu, Jamie Sikora, and Ernest Y.-Z. Tan</i>	12:1–12:15

¹ Note of the publisher: Unfortunately, this article was accidentally skipped in the first version of the conference proceedings published on June 8, 2020 and was subsequently published on August 19, 2020.



■ Preface

The 15th Conference on the Theory of Quantum Computation, Communication and Cryptography was hosted by the University of Latvia, and held online from June 9–12, 2020.

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 2019, University of Maryland, USA
- TQC 2018, University of Technology Sydney, Australia
- TQC 2017, Université Pierre et Marie Curie, France
- TQC 2016, Freie Universität Berlin, Germany
- TQC 2015, Université libre de Bruxelles, Brussels, Belgium
- TQC 2014, National University of Singapore, Singapore
- TQC 2013, University of Guelph, Canada
- TQC 2012, 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 Elena Kirshanova (Immanuel Kant Baltic Federal University), Thomas Monz (University of Innsbruck), Xin Wang (Baidu Research), and Henry Yuen (University of Toronto).

The conference was possible thanks to financial support from the European Regional Development Fund (project 1.1.1.5/18/I/016), Baidu, and the University of Latvia.

We wish to thank the members of the Program Committee and all subreviewers for their precious help. Our warm thanks also go to the members of the Local Organizing Committee, for their considerable efforts in organizing the conference. We would like to thank Michael Wagner (Dagstuhl Publishing) for his 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!

April 2020
Steven T. Flammia

■ Conference Organization

Local Organizing Committee

- Andris Ambainis (chair)
Latvia
- Kaspars Čikste
Latvia
- Jelena Polakova
Latvia
- Juris Smotrovs
Latvia
- Aleksandrs Rivoss
Latvia
- Dace Sostoka
Latvia

Program Committee

- Victor Albert
Caltech
- Itai Arad
Technion
- Rotem Arnon-Friedman
Berkeley
- Salman Beigi
IPM, Tehran
- Chris Chubb
Sherbrooke
- Richard Cleve
Waterloo
- Elizabeth Crosson
New Mexico
- Gemma De las Cuevas
Innsbruck
- Lidia del Rio
ETH Zurich
- Steven Flammia (chair)
University of Sydney
- Keisuke Fujii
Osaka
- David Gosset (co-chair)
Waterloo
- Markus Grassl
Gdańsk
- Min-Hsiu Hsieh
University of Technology, Sydney
- Shelby Kimmel
Middlebury
- Martin Kliesch
Düsseldorf
- Cécilia Lancien
Toulouse
- Angelo Lucia
Caltech
- Prabha Mandayam
IIT Madras
- Ashley Montanaro
Bristol
- Hui Khoo Ng
NUS
- Mark Wilde
Louisiana State
- Xiaodi Wu
Maryland
- Sisi Zhou
Chicago

Steering Committee

- Gorjan Alagic
Maryland
- Andris Ambainis
Latvia
- Anne Broadbent (chair)
Ottawa
- Aram Harrow
MIT
- Stacey Jeffery
QuSoft, CWI
- Laura Mančinská
Copenhagen
- Marco Tomamichel
UTS

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).
Editor: Steven T. Flammia



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



Exponential Quantum Communication Reductions from Generalizations of the Boolean Hidden Matching Problem

João F. Doriguello¹ 

School of Mathematics, University of Bristol, United Kingdom

Quantum Engineering Centre for Doctoral Training, University of Bristol, United Kingdom

<http://www.joadoriguello.com>

joao.doriguellodiniz@bristol.ac.uk

Ashley Montanaro

School of Mathematics, University of Bristol, United Kingdom

ashley.montanaro@bristol.ac.uk

Abstract

In this work we revisit the Boolean Hidden Matching communication problem, which was the first communication problem in the one-way model to demonstrate an exponential classical-quantum communication separation. In this problem, Alice's bits are matched into pairs according to a partition that Bob holds. These pairs are compressed using a Parity function and it is promised that the final bit-string is equal either to another bit-string Bob holds, or its complement. The problem is to decide which case is the correct one. Here we generalize the Boolean Hidden Matching problem by replacing the parity function with an arbitrary function f . Efficient communication protocols are presented depending on the sign-degree of f . If its sign-degree is less than or equal to 1, we show an efficient classical protocol. If its sign-degree is less than or equal to 2, we show an efficient quantum protocol. We then completely characterize the classical hardness of all symmetric functions f of sign-degree greater than or equal to 2, except for one family of specific cases. We also prove, via Fourier analysis, a classical lower bound for any function f whose pure high degree is greater than or equal to 2. Similarly, we prove, also via Fourier analysis, a quantum lower bound for any function f whose pure high degree is greater than or equal to 3. These results give a large family of new exponential classical-quantum communication separations.

2012 ACM Subject Classification Theory of computation \rightarrow Communication complexity; Theory of computation \rightarrow Quantum complexity theory

Keywords and phrases Communication Complexity, Quantum Communication Complexity, Boolean Hidden Matching Problem

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.1

Related Version A full version of the paper is available at <https://arxiv.org/abs/2001.05553>.

Funding João F. Doriguello: EPSRC grant EP/L015730/1.

Ashley Montanaro: QuantERA ERA-NET Cofund in Quantum Technologies implemented within the European Union's Horizon 2020 Programme (QuantAlgo project); the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 817581); and EPSRC grants EP/L021005/1, and EP/R043957/1.

Acknowledgements We would like to thank Ronald de Wolf for pointing out Ref. [24], and Makrand Sinha for useful discussions about the hypercontractive inequality.

¹ Corresponding author



© João F. Doriguello and Ashley Montanaro;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 1; pp. 1:1–1:16



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

One of the main aims of the field of quantum information and quantum computation is to establish the superiority of quantum computers and quantum resources over their classical counterparts. While in some areas this superiority is based on a belief in the impossibility of classical computers or classical resources solving particular tasks, e.g. the efficiency of Shor's algorithm [25] coming from the belief that there is no efficient classical factoring algorithm, in other areas like communication complexity one can establish unconditional exponential separations between classical and quantum performances.

Communication complexity is a model of computation first introduced by Yao [28]. In this model, two parties (normally called Alice and Bob) hold each a piece of data and want to solve some computational task that jointly depends on their data. More specifically, if Alice holds some information x and Bob holds some information y , they want to solve some function $f(x, y)$ or relational problem with several valid outputs for each x and y . In order to do so, they will need to communicate between themselves, and their goal is to solve the problem with minimal communication. The protocol that Alice and Bob employ could be *two-way*, where they take turns sending messages to each other; *one-way*, where Alice sends a single message to Bob who then outputs the answer; or *simultaneous*, where Alice and Bob each pass one message to a third party (the referee) who outputs the answer. Apart from these different types of communication settings, one is also interested in the error of a protocol when solving a communication problem: the zero-error *communication complexity* is the worst-case communication of the best protocol that gives a correct output with probability 1 for every input (x, y) ; the bounded-error *communication complexity* is the worst-case communication cost of the best protocol that gives a correct output with probability $1 - \epsilon$ for every input (x, y) , with $\epsilon \in [0, 1/2]$.

An interesting extension of the original communication model is the model of *quantum communication complexity* [8], also introduced by Yao [29]. In this model, Alice and Bob each has a quantum computer and they exchange qubits instead of bits and/or make use of shared entanglement. The use of quantum resources can drastically reduce the amount of communication in solving some problems in comparison to the classical communication model.

Exponential quantum-classical separations are known in the two-way (e.g. [22]), one-way (e.g. [4, 15]) and simultaneous (e.g. [9, 12]) models. Indeed, it is even known that one-way quantum communication can be exponentially more efficient than two-way classical communication [14, 23]. However, surprisingly few examples of such exponential separations are known, compared (for example) with the model of query complexity in which Shor's algorithm operates.

The Hidden Matching problem [4] was the first problem to exhibit an exponential separation between the bounded-error classical communication complexity and the bounded-error quantum communication complexity in the one-way model. The problem can be efficiently solved by one quantum message of $\log n$ qubits, while any classical one-way protocol needs to send $O(\sqrt{n})$ bits to solve it. The hardness of the problem is essentially one-way: it could be efficiently solved by having Bob send a classical message of $\log n$ bits to Alice. The Hidden Matching problem is a relational problem. In the same paper [4] the authors proposed a Boolean version of the problem, the Boolean Hidden Matching problem (which is a partial Boolean function), and conjectured that the same quantum-classical gap holds for it as well, which was later proven to be true by Gavinsky *et al.* [15]. Generalizing this separation is the focus of this work.

1.1 Hidden matching problems

Throughout the paper, $[n] = \{1, 2, \dots, n\}$ and given $x, y \in \{-1, 1\}^n$, we denote by $x \circ y$ the Hadamard (elementwise) product of x and y , and by \bar{x} the complement of x , such that $x \circ \bar{x} = 1^n$.

The Hidden Matching (HM_n^α) and Boolean Hidden Matching (BHM_n^α) problems are defined with respect to some $\alpha \in (0, 1]$. Alice is given a string $x \in \{-1, 1\}^n$ and Bob is given a sequence $M \in \mathcal{M}_{\alpha n/2}$ of $\alpha n/2$ disjoint pairs $(i_1, j_1), (i_2, j_2), \dots, (i_{\alpha n/2}, j_{\alpha n/2}) \in [n]^2$. Such a sequence is called an α -matching, and $\mathcal{M}_{\alpha n/2}$ denotes the family of all α -matchings – i.e. partial matchings of a fixed size in the complete graph on n vertices. Together x and M induce a string $z \in \{-1, 1\}^{\alpha n/2}$ defined by the parities of the $\alpha n/2$ edges, i.e., $z_\ell = x_{i_\ell} x_{j_\ell}$ for $\ell = 1, \dots, \alpha n/2$. Then the HM_n^α and BHM_n^α problems are defined as follows.

► **Definition 1 (The Hidden Matching problem (HM_n^α)).** Let $n \in \mathbb{N}$ be even and $\alpha \in (0, 1]$. Alice receives $x \in \{-1, 1\}^n$ and Bob receives $M \in \mathcal{M}_{\alpha n/2}$. Their goal is to output a tuple $\langle i, j, b \rangle$ such that $(i, j) \in M$ and $b = x_i x_j$.

► **Definition 2 (The Boolean Hidden Matching problem (BHM_n^α)).** Let $n \in \mathbb{N}$ be even and $\alpha \in (0, 1]$. Alice receives $x \in \{-1, 1\}^n$ and Bob receives $M \in \mathcal{M}_{\alpha n/2}$ and $w \in \{-1, 1\}^{\alpha n/2}$. It is promised that $z \circ w = b^{\alpha n/2}$ for some $b \in \{-1, 1\}$. Their goal is to output b .

Given inputs x and M , it is clear that there are many possible correct outputs for the HM_n^α problem ($\alpha n/2$ correct outputs, actually), making it a relational problem. On the other hand, the BHM_n^α is a partial Boolean function due to the promise statement.

Bar-Yossef *et al.* [4] gave a simple quantum protocol to solve the HM_n^1 problem with just $O(\log n)$ qubits of communication⁵, while proving that any classical protocol needs to communicate at least $\Omega(\sqrt{n})$ bits in order to solve it. Similarly with the BHM_n^α problem, Gavinsky *et al.* [15] demonstrated the same exponential classical-quantum communication gap for any $\alpha \leq 1/2$ (note that the definition of α they use differs from ours by a factor of 2). As HM_n^α is at least as difficult as BHM_n^α , their result implies the same lower bound for HM_n^α . The approach taken by Gavinsky *et al.* in proving the classical lower bound is particularly interesting in that it uses the Fourier coefficients inequality of Kahn, Kalai, and Linial [17], which is proven via the Bonami-Beckner inequality [7, 5]. We also mention that Fourier analysis had been previously used in communication complexity by Raz [21] and Klauck [18].

A slightly weaker separation ($O(\log n)$ vs. $\Omega(n^{7/16})$) for a closely related problem was shown in [19] using similar techniques. The BHM_n^α problem was generalized by Verbin and Yu [26] to a problem that they named Boolean Hidden Hypermatching (BHH_n^t). In this problem, instead of having the bits from Alice matched in pairs, they are now matched in tuples of t elements. In other words, a bit from the final string z is obtained by XORing t bits from Alice's string. More precisely, Alice is given a string $x \in \{-1, 1\}^n$ and Bob is given a sequence $M \in \mathcal{M}_{n/t}$ of n/t disjoint tuples $(M_{1,1}, \dots, M_{1,t}), \dots, (M_{n/t,1}, \dots, M_{n/t,t}) \in [n]^t$ called a hypermatching, where $\mathcal{M}_{n/t}$ denotes the family of all hypermatchings. Both x and M induce a string $z \in \{-1, 1\}^{n/t}$ defined by the parities of the n/t edges, i.e., $z_\ell = \prod_{j=1}^t x_{M_{\ell,j}}$ for $\ell = 1, \dots, n/t$. The BHH_n^t problem is defined as follows.

► **Definition 3 (The Boolean Hidden Hypermatching problem (BHH_n^t)).** Let $n, t \in \mathbb{N}$ be such that $2t \mid n$. Alice receives $x \in \{-1, 1\}^n$ and Bob receives $M \in \mathcal{M}_{n/t}$ and $w \in \{-1, 1\}^{n/t}$. It is promised that $z \circ w = b^{n/t}$ for some $b \in \{-1, 1\}$. Their goal is to output b .

⁴ Throughout this paper we shall use $\{-1, 1\}$ instead of $\{0, 1\}$ for convenience.

⁵ Their protocol extends easily to the more general HM_n^α problem.

Verbin and Yu proved a classical lower bound of $\Omega(n^{1-1/t})$ communication for every bounded-error one-way protocol, showing the increasing hardness of the problem with t , as one should expect since the BHH_n^t problem can be reduced from the BHM_n problem (we will show how this is done in detail later). The authors subsequently used this problem to prove various streaming lower bounds, i.e., lower bounds on the space required of streaming algorithms (algorithms that read the input from left to right, use a small amount of space, and approximate some function of the input). However, no efficient quantum protocol was proposed for solving the BHH_n^t problem for $t > 2$. It was only later that Shi, Wu and Yu [24] showed that such efficient quantum protocols do not exist. More specifically, they proved a quantum lower bound of $\Omega(n^{1-2/t})$ communication for every bounded-error one-way protocol for the BHH_n^t problem. Their proof is similar to the ones used in the classical lower bound, the difference lying in the use of Fourier analysis of *matrix-valued* functions and the matrix-valued Hypercontractive Inequality of Ben-Aroya, Regev, and de Wolf [6].

Note that the lower bound of Verbin and Yu does not use an α parameter, unlike the lower bound of [15]. However, their lower bound requires n/t to be even, otherwise Alice can just send the parity of her bit-string. (The result of [15] can be extended to hold for any $\alpha < 1$ fairly straightforwardly, but achieving a strong lower bound for $\alpha = 1$ requires some more work.)

1.2 Our Results

This paper focuses on the study of a broad generalization of the BHH_n^t problem. In the (Boolean) Hidden Matching and Boolean Hidden Hypermatching problems, the task Alice and Bob want to solve can be viewed as rearranging Alice's data according to some permutation that Bob holds, and "compressing" the data to a final bit-string by applying some Boolean function to the bits. Then Alice and Bob's goal is to determine some information about this final bit-string. The way this compression was originally done was via the Parity function, but, apart from the obvious reason that Parity gives the desired classical-quantum communication gap and, less obviously, leads to a clear proof, there is no particular need to restrict to this function in order to arrive at the final bit-string. This observation leads to a generalization of the Boolean Hidden Hypermatching problem, which we named the f -Boolean Hidden Partition (f -BHP $_n^{\alpha,t}$) problem, where $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ is the Boolean function used to compress Alice's bits.

Given $y \in \{-1, 1\}^n$, we define by $y^{(j;t)} = (y_{(j-1)t+1}, y_{(j-1)t+2}, \dots, y_{jt}) \in \{-1, 1\}^t$ the j -th block of size t from y , with $t|n$ and $j = 1, \dots, n/t$. When the size of the block is clear from the context, we shall simply write $y^{(j)}$.

The f -Boolean Hidden Partition problem is defined as follows. Alice is given a bit-string $x \in \{-1, 1\}^n$, and Bob is given a permutation $\sigma \in S_n$ and a bit-string $w \in \{-1, 1\}^{\alpha n/t}$, where $\alpha \in (0, 1]$ is fixed. Given a Boolean function $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$, we can define the map $B_f : \{-1, 1\}^n \rightarrow \{-1, 1\}^{\alpha n/t}$ by $B_f(x) = (f(\sigma(x)^{(1)}), \dots, f(\sigma(x)^{(\alpha n/t)}))$, where $\sigma(x)_i = x_{\sigma^{-1}(i)}$. Hence x and σ induce a bit-string given by $B_f(x)$, each of whose bits is obtained by applying f to a block of the permuted bit-string $\sigma(x)$. The f -BHP $_n^{\alpha,t}$ problem can be defined as follows.

► **Definition 4 (The f -Boolean Hidden Partition problem (f -HMP $_n^{\alpha,t}$)).** Let $n, t \in \mathbb{N}$ be such that $t|n$ and $\alpha \in (0, 1]$. Alice receives $x \in \{-1, 1\}^n$ and Bob receives $\sigma \in S_n$ and $w \in \{-1, 1\}^{\alpha n/t}$. It is promised that there exists $b \in \{-1, 1\}$ such that $B_f(x) \circ w = b^{\alpha n/t}$. The problem is to output b .

The adoption of the word “Partition” instead of “(Hyper)Matching” from previous works comes from our decision to view the problem in terms of a hidden partition that Bob holds, instead of an α -(Hyper)Matching. Bob shuffles Alice’s data according to some permutation, and then just partitions the resulting data in adjacent blocks of size t and uses f to get the final bit-string. Obviously both views are equivalent, but we think that the permutation approach eases the analysis of the problem.

Our aim is to study the f -Boolean Hidden Partition problem in terms of the function f . It should be clear that for some functions the problem is hard to solve classically, e.g. when f is the Parity function and we recover the usual Boolean Hidden Hypermatching problem. On the other hand, for some functions it becomes easily solvable, e.g. when f is the AND function (Alice needs only to send the position of any 0 in her string). We would like to characterize for which functions the problem can be efficiently solved classically, i.e., with $O(\log n)$ bits of communication, and for which functions it is hard to solve classically, i.e., requires $\Omega(n^a)$ bits of communication for some $a \in (0, 1]$. And the same question applies to quantum communication complexity: we would like to determine for which functions the problem admits or not an efficient quantum communication protocol. Given this characterization, we can check for which functions there is an exponential classical-quantum communication gap.

We partially prove that the whole f -BHP $_{n,t}^{\alpha}$ problem can be fully characterized by just one quantity: the *sign-degree* of the function f . A polynomial $p : \{-1, 1\}^t \rightarrow \mathbb{R}$ is said to *sign-represent* f if $f(x) = \text{sgn}(p(x))$. If $|p(x)| \leq 1$ for all x , we say that p is *normalized*. The *bias* of a normalized polynomial p is defined as $\beta = \min_x |p(x)|$. The sign-degree ($sdeg(f)$) of f is the minimum degree of polynomials that sign-represent it. In Appendix A we prove the following upper bounds on the classical and quantum communication complexity of the f -Boolean Hidden Partition problem based on the sign-degree:

► **Theorem 5.** *Let $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be a Boolean function. If $sdeg(f) \leq 1$, then there exists a bounded-error classical protocol that solves the f -BHP $_{n,t}^{\alpha}$ problem with error probability ϵ and $O\left(\left(\frac{t}{\alpha\beta}\right)^2 \log \frac{1}{\epsilon} \log n\right)$ bits of communication, where β is the maximal bias of a polynomial of degree $sdeg(f)$ that sign-represents f .*

► **Theorem 6.** *Let $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be a Boolean function. If $sdeg(f) \leq 2$, then there exists a bounded-error quantum protocol that solves the f -BHP $_{n,t}^{\alpha}$ problem with error probability ϵ and $O\left(\left(\frac{t}{\alpha\beta}\right)^2 \log \frac{1}{\epsilon} \log n\right)$ qubits of communication, where β is the maximal bias of a polynomial of degree $sdeg(f)$ that sign-represents f .*

Note that the bias β can be very small, but can also be lower-bounded in terms only of t : indeed, it is shown in [10] that β is lower-bounded by $t^{-O(t^{sdeg(f)})}$. In this work we will usually assume that $t = O(1)$, so $\beta = \Omega(1)$. We assume throughout that Alice and Bob do not have access to shared randomness or entanglement. The classical complexity in the above theorem can actually be improved to an additive dependence on $\log n$ via applying Newman’s Theorem [20] to a protocol with shared randomness, but at the expense of making the protocol less intuitive.

The classical upper bound stated above comes from the observation that, if f has a sign-representing polynomial p of degree 1, it is possible to determine whether $f(z) = 1$ with probability $> 1/2$ by only evaluating f on one uniformly random bit of z , by writing down a probabilistic procedure whose expectation on z mimics $p(z)$. So Alice sends a few uniformly random bits to Bob, who matches them to blocks in his partition, and evaluates f on the corresponding blocks with success probability $> 1/2$ for each block. Only a few repetitions are required to determine whether $f(x) = w$ or $f(x) = \bar{w}$ with high probability.

On the other hand, to obtain the quantum upper bound we use the idea of *block-multilinear* polynomials from [1, 2], and some auxiliary results also from [2]. The idea is that Alice sends a superposition of her bits, and Bob, after collapsing the state onto one of the blocks from his partition (say block j), applies a controlled unitary operator that describes a block-multilinear polynomial \tilde{p} of degree 2, which is produced from a sign-representing polynomial p for f of degree 2. A Hadamard test is used to return an output with probability depending (roughly speaking) on $\tilde{p}(\sigma(x)^{(j)}, \sigma(x)^{(j)})$, which in turn is equal to $p(\sigma(x)^{(j)})$ according to a theorem from [2]. The Hadamard test then outputs 1 with probability greater than $1/2$ if $f(x^{(j)}) = 1$ and 0 with probability greater than $1/2$ if $f(x^{(j)}) = -1$.

We remark that both of these protocols actually solve a natural generalization of the Hidden Matching problem [4] (i.e. they output the result of evaluating $f(x^{(j)})$ for Bob's block j , where j is arbitrary), which is at least as hard as the f -Boolean Hidden Partition problem. However, unlike the Hidden Matching problem, the output is not correct with certainty, but only with probability strictly greater than $1/2$.

In Section 2 we reduce the f -Boolean Hidden Partition problem from the Boolean Hidden Matching problem and prove that for almost all symmetric Boolean function f with $\text{sdeg}(f) \geq 2$ the f -BHP $_{n,t}^{\alpha,t}$ problem require at least $\Omega(\sqrt{n})$ bits of communication. The only functions for which the reduction does not work are the Not All Equal functions on an odd number of bits, i.e., $\text{NAE} : \{-1, 1\}^t \rightarrow \{-1, 1\}$, defined by $\text{NAE}(x) = -1$ if $|x| \in \{0, t\}$ and $\text{NAE}(x) = 1$ otherwise, with t odd.

► **Theorem 7.** *Let $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be a symmetric Boolean function with $\text{sdeg}(f) \geq 2$. If f is not the NAE function on an odd number of bits, then any bounded-error classical communication protocol for solving the f -BHP $_{n,t}^{\alpha,t}$ problem needs to communicate at least $\Omega(\sqrt{n}/(\alpha t))$ bits.*

Finally, we generalize the Fourier analysis methods from [15, 26, 24] to prove a partial result on the hardness of the f -BHP $_{n,t}^{\alpha,t}$ problem, both classically and quantumly. Ideally we would like to prove that any bounded-error classical and quantum protocols would need to communicate $\Omega(n^{1-1/d})$ bits and $\Omega(n^{1-2/d})$ qubits, respectively, where $\text{sdeg}(f) = d$. What we obtained is this result but with d being the *pure high degree* of f . A Boolean function f is said to have pure high degree ($\text{phdeg}(f)$) d if $\hat{f}(S) = 0$ for all $|S| = 0, 1, \dots, d-1$, where $\hat{f}(S) = \langle f, \chi_S \rangle = \frac{1}{2^n} \sum_{x \in \{-1, 1\}^n} f(x) \chi_S(x)$ is the Fourier transform of f and $\chi_S(x) = \prod_{i \in S} x_i$, with $S \subseteq [n]$, is a character function. It is possible to prove that $\text{phdeg}(f) \leq \text{sdeg}(f)$, so our result is a step towards proving a lower bound for all functions with sign degree ≥ 2 .

► **Theorem 8.** *Let $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be a Boolean function. If $\text{phdeg}(f) = d \geq 2$, then, for sufficiently small $\alpha > 0$ that does not depend on n , any bounded-error classical communication protocol for solving the f -BHP $_{n,t}^{\alpha,t}$ problem needs to communicate at least $\Omega(n^{1-1/d})$ bits.*

► **Theorem 9.** *Let $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be a Boolean function. If $\text{phdeg}(f) = d \geq 3$, then, for sufficiently small $\alpha > 0$ that does not depend on n , any bounded-error quantum communication protocol for solving the f -BHP $_{n,t}^{\alpha,t}$ problem needs to communicate at least $\Omega(n^{1-2/d})$ qubits.*

The above lower bounds are proved in [11]. The classical proof follows the general idea from [15, 26], but the technical execution was substantially changed by borrowing ideas from [24]. First, we apply Yao's minimax principle [27], which says that it suffices to prove a lower bound for a *deterministic* protocol under a hard probability distribution on Alice and

Bob's inputs. We choose Alice's input x and Bob's input σ independently and uniformly over $\{-1, 1\}^n$ and \mathbb{S}_n (the set of all permutations on $[n]$), respectively. The input distribution is completed by choosing $w = B_f(x)$ with probability $1/2$ and $w = \overline{B_f(x)}$ with probability $1/2$.

Alice sends a message to Bob. If the length of the message sent is c , then the inputs for which Alice could have sent that specific message define a set A of about 2^{n-c} x 's. From Bob's perspective, he knows that the random variable X corresponding to Alice's bit-string is uniformly distributed in a set A and he knows his permutation σ , hence his knowledge of the random variable $B_f(X)$ is described by the distributions

$$p_\sigma(z) = \frac{|\{x \in A | B_f(x) = z\}|}{|A|} \text{ and } q_\sigma(z) = \frac{|\{x \in A | B_f(x) = \bar{z}\}|}{|A|}.$$

It is well known that the best success probability for distinguishing two distributions q_1 and q_2 with one sample is $1/2 + \|q_1 - q_2\|_{\text{td}}/4$. Therefore the bias of the protocol, i.e., the protocol's successful probability minus a half, is equal to the total variation distance between p_σ and q_σ . Differently from the approach of [15, 26], and following [24], we directly upper bound the expectation of the bias over Bob's permutation. By demanding a small distributional error, we arrive at the desired communication lower bound. Upper bounding the bias is done via Fourier analysis, using the inequality of Kahn, Kalai, and Linial [17].

The quantum proof follows the same idea from [24]. Yao's minimax principle is still applied and the "hard" input distribution is still uniform on Alice's input $x \in \{-1, 1\}^n$, Bob's input $\sigma \in \mathbb{S}_n$ and the function value $b \in \{-1, 1\}$, which fixes Bob's second input $w = B_f(x) \circ b^{\alpha n/t}$. The best strategy for Bob in determining b conditioned on his input (σ, w) is no more than the chance to distinguish between two subsets of Alice's messages, where a message corresponds to a quantum state ρ_x , selected according to b . In other words, no more than the chance to distinguish between the following $\rho_0^{\sigma, w}$ and $\rho_1^{\sigma, w}$, each appearing with probability $\Pr[b = 0 | \sigma, w]$ and $\Pr[b = 1 | \sigma, w]$, respectively,

$$\rho_0^{\sigma, w} = \frac{\sum_{x \in \{-1, 1\}^n} \Pr[x, 0, \sigma, w] \rho_x}{\Pr[x, 0, \sigma, w]} \text{ and } \rho_1^{\sigma, w} = \frac{\sum_{x \in \{-1, 1\}^n} \Pr[x, 1, \sigma, w] \rho_x}{\Pr[x, 1, \sigma, w]}.$$

It is known that any protocol that tries to distinguish two quantum states ρ_0 and ρ_1 appearing with probability p and $1 - p$, respectively, by a POVM has bias at most $\|p\rho_0 - (1 - p)\rho_1\|_{\text{tr}}/2$ [16]. The bias is then upper bounded by using Fourier analysis of matrix-valued functions, in particular by the matrix-valued hypercontractive inequality of Ben-Aroya, Regev, and de Wolf [6].

The difference between the classical and quantum lower bound proofs was considerably reduced in our paper, e.g., the classical proof now relies less on the use of the Parseval's identity. Still some differences persist. Apart from the obvious generalization of Fourier analysis to matrix-valued functions, the Fourier analysis in the quantum lower bound proof is performed directly on the encoding messages and not on the pre-images of a fixed encoding message, since there is no clear quantum analogue of conditioning on a message. The main technical difficulty we faced compared to [15, 26] is that the Fourier coefficients of Bob's distributions $p_\sigma(z)$ and $q_\sigma(z)$ are not nicely related to just one Fourier coefficient of the characteristic function of A any more, but instead to a more complicated sum of many coefficients. This requires us to carefully bound various combinatorial terms occurring in the proof and to use our freedom to choose α fairly small.

In Section 3 we analyse the limitations of our techniques and show that under the uniform distribution, which was used as the "hard" distribution during the proof of Theorem 8, we cannot obtain a lower bound depending on the sign degree instead of the pure high degree.

We finally remark that the one-way communication complexity separations we found can easily be used to obtain corresponding separations in the streaming model, similarly to [15, 26].

2 Reductions from the Boolean Hidden Matching problem

As mentioned before, in [15] it was proved that the Boolean Hidden Partition problem using PARITY on 2 bits (aka the BHM problem) is hard to solve, i.e., $R^1(\text{BHM}) = \Omega(\sqrt{n/\alpha})$. With this result alone it is possible to prove that the f -Boolean Hidden Partition problem for almost any symmetric Boolean function with $\text{sdeg}(f) \geq 2$ is at least as hard to solve. This can be achieved via a simple reduction from the BHM problem to the f -BHP $_{n,t}^{\alpha,t}$ problem with symmetric functions, which we shall show in this section.

For this section, in a slight abuse of notation we define $|x| = |\{i : x_i = -1\}|$ to be the “Hamming weight” of x . Let $s, t \in \mathbb{N}$, with $s \leq t$. Consider a symmetric Boolean function $f_s : \{-1, 1\}^t \rightarrow \{-1, 1\}$ such that (without loss of generality) $f_s(1^n) = 1$ and

$$f_s(x) = \begin{cases} +1 & \text{if } 0 \leq |x| \leq \theta_1 \text{ or } \theta_{2i} < |x| \leq \theta_{2i+1}, i = 1, 2, \dots, \lfloor s/2 \rfloor, \\ -1 & \text{if } \theta_{2j-1} < |x| \leq \theta_{2j}, j = 1, 2, \dots, \lfloor (s+1)/2 \rfloor, \end{cases} \quad (1)$$

where $\theta_k \in \mathbb{N}$ for $k = 1, \dots, s+1$ and $0 \leq \theta_1 < \dots < \theta_s < \theta_{s+1} = t$ and $\theta_{k+1} - \theta_k \geq 1$ for all $k = 1, \dots, s$. The following result from [3] tells us that $\text{sdeg}(f_s) = s$.

► **Lemma 10.** (Lemma 2.6 from [3]) *If f is a symmetric function, then $\text{sdeg}(f)$ is equal to the number of times f changes sign when expressed as a univariate function in $\sum_i x_i$.*

In order to reduce f_s -BHP $_{n,t}^{\alpha,t}$ from BHM we first need to reduce the function f_s from PARITY, i.e., we want that $\forall x' \in \{-1, 1\}^2, \exists x \in \{-1, 1\}^t$ such that $f_s(x) = \text{PARITY}(x')$. The key combinatorial step to achieve this is shown in the next Lemma.

► **Lemma 11.** *Let $f_s : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be the symmetric Boolean function from Eq. 1 with $s \geq 2$ such that either $2|t$ or $\theta_2 - \theta_1 < t - 1$. Then there exists $a, b \in \mathbb{N}$ such that $\forall x' \in \{-1, 1\}^2, \exists x \in \{-1, 1\}^t$ such that $f_s(x) = \text{PARITY}(x')$ and $|x| = a|x'| + b$.*

Proof. The condition that $\forall x' \in \{-1, 1\}^2, \exists x \in \{-1, 1\}^t$ such that $f_s(x) = \text{PARITY}(x')$ and $|x| = a|x'| + b$ is equivalent to

$$\begin{cases} |x'| = 0 \implies f_s(b) = 1, \\ |x'| = 1 \implies f_s(a+b) = -1, \\ |x'| = 2 \implies f_s(2a+b) = 1. \end{cases} \quad (2)$$

We divide the proof into two cases: either there exists $k^* \in \{1, \dots, s-1\}$ such that $\theta_{k^*+1} - \theta_{k^*}$ is odd or there does not exist such a k^* . Suppose first that such k^* exists. Without loss of generality we can assume that $f_s(x) = -1$ for $\theta_{k^*} < |x| \leq \theta_{k^*+1}$, otherwise we just flip the values of f_s . Then we set

$$\begin{cases} a = (\theta_{k^*+1} - \theta_{k^*} + 1)/2, \\ b = \theta_{k^*}. \end{cases}$$

First, $a, b \in \mathbb{N}$. Second, $a + b = (\theta_{k^*+1} + \theta_{k^*} + 1)/2$, hence $\theta_{k^*} < a + b \leq \theta_{k^*+1}$, since $\theta_{k^*+1} - \theta_{k^*} \geq 1$. And third, $2a + b = \theta_{k^*+1} + 1 \leq \theta_{k^*+2}$. Therefore all conditions from Eqs. 2 are satisfied.

Now suppose that for all $k = 1, \dots, s-1$ we have $2|(\theta_{k+1} - \theta_k)$. Define the bit $\delta = [\theta_1 \neq 0]$ and set

$$\begin{cases} a = (\theta_2 - \theta_1 + 2)/2, \\ b = \theta_1 - \delta. \end{cases}$$

First, $a, b \in \mathbb{N}$ (note that $\delta = 1 \implies \theta_1 > 0$). Second, $a + b = (\theta_2 + \theta_1 + 2 - 2\delta)/2$, hence $\theta_1 < a + b \leq \theta_2$, since $\theta_2 - \theta_1 \geq 2$ by hypothesis. And third, $2a + b = \theta_2 + 2 - \delta \leq t$ since $\theta_2 - \theta_1 < t - 1$ and $\theta_2 < t$ (so that $\theta_2 = t - 1 \implies \delta = 1$). Therefore all conditions from Eqs. 2 are satisfied. \blacktriangleleft

If $2 \nmid t$ and $\theta_2 - \theta_1 = t - 1$, then our conditions give us

$$\begin{cases} b = 0, \\ 0 < a < t, \\ 2a = t, \end{cases}$$

and we see that the condition $2a = t$ cannot be fulfilled by $a \in \mathbb{N}$. This case corresponds to the symmetric Boolean function Not All Equal (NAE), defined by $\text{NAE}(x) = 1$ if $|x| \in \{0, t\}$ and $\text{NAE}(x) = -1$ otherwise, with t odd.

Given the reduction above from PARITY to f_s , we can construct our reduction from the BHM problem to the f_s -BHP $_{n,t}^{\alpha,t}$ problem.

► Theorem 7. *Let $f_s : \{-1, 1\}^t \rightarrow \{-1, 1\}$ be the symmetric Boolean function from Eq. 1 with $s \geq 2$ such that either $2|t$ or $\theta_2 - \theta_1 < t - 1$. Then $R^1(f_s\text{-BHP}_{n,t}^{\alpha,t}) = \Omega(\sqrt{n/(\alpha t)})$.*

Proof. Suppose by contradiction that $R^1(f_s\text{-BHP}_{n,t}^{\alpha,t}) = o(\sqrt{n/(\alpha t)})$, i.e., there exists a protocol Π that solves $f_s\text{-BHP}_{n,t}^{\alpha,t}$ with $o(\sqrt{n/(\alpha t)})$ bits of communication. We are going to show that such protocol would allow Alice and Bob to solve the BHM problem with $o(\sqrt{n/\alpha})$ bits of communication, which leads to a contradiction.

Let $a, b \in \mathbb{N}$ be the numbers used in reducing f_s from PARITY in Lemma 11. Alice increases her bit string $x \in \{-1, 1\}^n$ as follows: she makes a copies of x , obtaining $x^a \in \{-1, 1\}^{an}$, where $x^a = xx \cdots x$ represents x repeated a times. She then adds $bn/2$ times the bit 1, obtaining $x^a 1^{bn/2}$. Finally, she adds $(t - 2a - b)n/2$ times the bit -1 , to finally obtain $x_f = x^a 1^{bn/2} (-1)^{(t-2a-b)n/2}$. Note that $x_f \in \{-1, 1\}^{nt/2}$.

Bob, on the other hand, increases his permutation $\sigma \in S_n$ to a new permutation $\sigma_f \in S_{nt/2}$. In order to describe how he does this, we ease the notation by referring to the j -th block $(\pi^{-1}((j-1)t+1), \dots, \pi^{-1}(jt))$ of a given permutation π as $(B_{j,1}, \dots, B_{j,t})$. With this notation, the j -th block $(B_{j,1}, B_{j,2})$ of the permutation σ is mapped to the j -th block

$$\left(B_{j,1}, B_{j,2}, n + B_{j,1}, n + B_{j,2}, \dots, (a-1)n + B_{j,1}, (a-1)n + B_{j,2}, \right. \\ \left. an + j, an + j + \frac{n}{2}, \dots, an + j + (t-2a-1)\frac{n}{2} \right)$$

of the new permutation σ_f . Note that the new block has t elements, as expected.

Consider the block strings $\sigma_f(x_f)^{(j;t)} \in \{-1, 1\}^t$ and $\sigma(x)^{(j;2)} \in \{-1, 1\}^2$, with $j = 1, \dots, n/2$. By construction we have that $|\sigma_f(x_f)^{(j;t)}| = a|\sigma(x)^{(j;2)}| + b$ and, according to Lemma 11, we get $f_s(\sigma_f(x_f)^{(j;t)}) = \text{PARITY}(\sigma(x)^{(j;2)})$ for all $j = 1, \dots, n/2$. Hence we see that every instance of the problem BHM : $\{-1, 1\}^n \rightarrow \{-1, 1\}$ is mapped to an instance of the problem $f_s\text{-BHP}_{n,t}^{\alpha,t} : \{-1, 1\}^{nt/2} \rightarrow \{-1, 1\}$. Therefore we could map the BHM problem into the $f_s\text{-BHP}_{n,t}^{\alpha,t}$ problem and use the protocol Π in order to solve it with $o(\sqrt{n/(\alpha t)})$ bits of communication, which is impossible. Thus $R^1(f_s\text{-BHP}_{n,t}^{\alpha,t}) = \Omega(\sqrt{n/(\alpha t)})$. \blacktriangleleft

3 Limitations of proof technique

Theorem 8 guarantees the classical hardness of the f -BHP $_{n,t}^{\alpha,t}$ problem if f has pure high degree ≥ 2 , and not sign degree ≥ 2 , which would be a stronger result. To arrive at this result, we used the uniform distribution as a “hard” distribution for Yao’s principle. In this section we shall prove that under the uniform distribution we cannot obtain a better result. More specifically, we shall prove that under the uniform distribution there is an efficient bounded-error classical protocol for solving the f -BHP $_{n,t}^{\alpha,t}$ problem if $phdeg(f) \leq 1$.

► **Theorem 12.** *Under the uniform distribution for Alice and Bob’s inputs, if $phdeg(f) \leq 1$ then $R^1(f\text{-BHP}_{n,t}^{\alpha,t}) = O\left(\frac{t^2}{\alpha} \log n\right)$.*

Proof. Let $F = \{i \in [t] \mid \widehat{f}(\{i\}) \neq 0\}$. Given that $phdeg(f) \leq 1$, this set is non-empty. Consider the following protocol: Alice picks a subset $I \subseteq [n]$ of indices uniformly at random using shared randomness, where $|I|$ will be determined later, and sends the indices and corresponding bitvalues to Bob. Let $\{x_i\}_{i \in I}$ be the bitvalues sent, and let $j(i) = \lceil \sigma(i)/t \rceil$ and $k(i) \equiv \sigma(i) \pmod t$ for all $i \in I$, where $\sigma \in S_n$ is Bob’s permutation. The probability that none of the indices sent by Alice are matched to a non-zero Fourier coefficient according to Bob’s permutation, within one of the $\alpha n/t$ blocks he has, is

$$\Pr_{\sigma}[k(i) \notin F, \forall i \in I] \leq \left(1 - \alpha \frac{|F|}{t}\right)^{|I|} \leq e^{-\alpha |I| |F|/t}$$

which we can make almost arbitrarily small by choosing $|I|$ to be sufficiently large. (Note that the first inequality above would be an equality if we chose the elements of I with replacement, and choosing them without replacement cannot make $\Pr[k(i) \notin F, \forall i \in I]$ higher.) Hence with high probability $I \cap F \cap [\alpha n/t] \neq \emptyset$. Choose some $\ell \in I \cap F \cap [\alpha n/t]$. Bob computes $\text{sgn}[\widehat{f}(\{k(\ell)\})] \cdot \sigma(x)_{k(\ell)}^{(j(\ell))} \cdot w_{j(\ell)}$: if it is $+1$, then he outputs that $B_f(x) = w$, and if it is -1 , then he outputs that $B_f(x) = \bar{w}$.

To see why the protocol works, we calculate the probability that $\text{sgn}[\widehat{f}(\{k(\ell)\})] \cdot \sigma(x)_{k(\ell)}^{(j(\ell))}$ is equal to $f(\sigma(x)^{(j(\ell))})$.

$$\begin{aligned} \Pr_x \left[\text{sgn}[\widehat{f}(\{k(\ell)\})] \sigma(x)_{k(\ell)}^{(j(\ell))} = f(\sigma(x)^{(j(\ell))}) \right] &= \\ &= \frac{1}{2} + \frac{1}{2^{t+1}} \sum_{x \in \{-1,1\}^t} \text{sgn}[\widehat{f}(\{k(\ell)\})] \sigma(x)_{k(\ell)}^{(j(\ell))} f(\sigma(x)^{(j(\ell))}) \\ &= \frac{1}{2} + \frac{1}{2} \text{sgn}[\widehat{f}(\{k(\ell)\})] \cdot \widehat{f}(\{k(\ell)\}) \\ &= \frac{1}{2} + \frac{1}{2} |\widehat{f}(\{k(\ell)\})|, \end{aligned}$$

which is greater than $1/2$ and where we used in the first line that the distribution on Alice’s inputs is uniform. Therefore, by a union bound, for sufficiently large $|I| = O\left(\frac{t}{\alpha} \log \frac{1}{|\widehat{f}(\{k(\ell)\})|}\right)$, the overall success probability of the protocol (i.e. $I \cap F \cap [\alpha n/t] \neq \emptyset$ and Bob’s output equals f) is strictly greater than $1/2$. Since $|\widehat{f}(\{k(\ell)\})| \geq 2^{1-t}$ (as it is nonzero and is an average of $2^t \pm 1$ ’s), this gives us the final overhead of $O(t^2/\alpha)$. ◀

4 Conclusions

We proposed a very broad generalization of the famous Boolean Hidden (Hyper)Matching problem, which we called the f -Boolean Hidden Partition (f -BHP $^{\alpha,t}_n$) problem. Instead of using the Parity function to arrive at the final bit-string that Alice and Bob wish to explore, we use a generic Boolean function f . We partially characterize the communication complexity of the whole problem in terms of one property of f : its sign degree. We proved that if $sdeg(f) \leq 1$, then there exists an efficient bounded-error classical protocol that solves the f -BHP $^{\alpha,t}_n$ with $O(\log n)$ bits. Similarly to the classical case, we proved that if $sdeg(f) \leq 2$, then there exists an efficient bounded-error quantum protocol that solves the f -BHP $^{\alpha,t}_n$ with $O(\log n)$ qubits. We then pursued a classical-quantum communication gap by proving classical and quantum lower bounds for cases of the problem where $sdeg(f) \geq 2$. First we noted that the f -BHP $^{\alpha,t}_n$ problem is hard for almost all symmetric functions with $sdeg(f) \geq 2$ via a simple reduction from the Boolean Hidden Matching problem. And second we generalized previous communication complexity lower bounds based on Fourier analysis to prove that functions with $phdeg(f) = d \geq 2$ lead to a classical $\Omega(n^{1-1/d})$ communication cost and functions with $phdeg(f) = d \geq 3$ lead to a quantum $\Omega(n^{1-2/d})$ communication cost for the f -BHP $^{\alpha,t}_n$ problem.

It is known that $phdeg(f) \leq sdeg(f)$, but our lower bounds are probably not tight for *all* functions with sign degree ≥ 2 . We proved that this is an inherent limitation of the chosen distribution for Alice and Bob's inputs during the proof, since under the uniform distribution it is possible to solve the problem with $O(\log n)$ bits of communication if $phdeg(f) \leq 1$. We then make the following conjectures.

► **Conjecture 13.** $R_\epsilon^1(f\text{-BHP}_n^{\alpha,t}) = \Omega(n^{1-1/d})$ if $sdeg(f) = d \geq 2$.

► **Conjecture 14.** $Q_\epsilon^1(f\text{-BHP}_n^{\alpha,t}) = \Omega(n^{1-2/d})$ if $sdeg(f) = d \geq 3$.

A proof of these results would require a non-uniform distribution on Alice and Bob's inputs.

We hope that these conjectures help motivate the development of necessary quantum lower bound techniques.

References

- 1 Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. *SIAM Journal on Computing*, 47(3):982–1038, 2018. [arXiv:1411.5729](#).
- 2 Scott Aaronson, Andris Ambainis, Janis Iraids, Martins Kokainis, and Juris Smotrovs. Polynomials, quantum query complexity, and Grothendieck's inequality. In *31st Conference on Computational Complexity*, 2016. [arXiv:1511.08682](#).
- 3 James Aspnes, Richard Beigel, Merrick Furst, and Steven Rudich. The expressive power of voting polynomials. *Combinatorica*, 14(2):135–148, 1994.
- 4 Ziv Bar-Yossef, Thathachar S. Jayram, and Iordanis Kerenidis. Exponential separation of quantum and classical one-way communication complexity. In *Proceedings of the thirty-sixth annual ACM symposium on Theory of computing*, pages 128–137. ACM, 2004.
- 5 William Beckner. Inequalities in Fourier analysis. *Ann. of Math.*, 102:159–182, 1975.
- 6 Avraham Ben-Aroya, Oded Regev, and Ronald de Wolf. A hypercontractive inequality for matrix-valued functions with applications to quantum computing and LDCs. In *2008 49th Annual IEEE Symposium on Foundations of Computer Science*, pages 477–486. IEEE, 2008. [arXiv:0705.3806](#).
- 7 Aline Bonami. Étude des coefficients Fourier des fonctions de $L^p(G)$. In *Annales de l'institut Fourier*, volume 20(2), pages 335–402, 1970.

- 8 Harry Buhrman, Richard Cleve, Serge Massar, and Ronald de Wolf. Nonlocality and communication complexity. *Reviews of modern physics*, 82(1):665, 2010. [arXiv:0907.3584](#).
- 9 Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, 2001. [quant-ph/0102001](#).
- 10 Harry Buhrman, Nikolay Vereshchagin, and Ronald de Wolf. On computation and communication with small bias. In *Proc. 22nd Annual IEEE Conf. Computational Complexity*, pages 24–32, 2007.
- 11 João F. Doriguello and Ashley Montanaro. Exponential quantum communication reductions from generalizations of the boolean hidden matching problem. *arXiv preprint arXiv:2001.05553*, 2020.
- 12 João Fernando Doriguello and Ashley Montanaro. Quantum sketching protocols for Hamming distance and beyond. *Phys. Rev. A*, 99:062331, 2019. [arXiv:1810.12808](#).
- 13 Devdatt P. Dubhashi and Alessandro Panconesi. *Concentration of measure for the analysis of randomized algorithms*. Cambridge University Press, 2009.
- 14 Dmitry Gavinsky. Classical interaction cannot replace a quantum message. In *Proc. 40th Annual ACM Symp. Theory of Computing*, pages 95–102, 2008. [quant-ph/0703215](#).
- 15 Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. *SIAM Journal on Computing*, 38(5):1695–1708, 2008. [quant-ph/0611209](#).
- 16 Carl Wilhelm Helstrom. *Quantum detection and estimation theory*. Academic press, 1976.
- 17 Jeff Kahn, Gil Kalai, and Nathan Linial. The influence of variables on boolean functions. In *Proc. 29th Annual Symp. Foundations of Computer Science*, pages 68–80. IEEE, 1988.
- 18 Hartmut Klauck. Lower bounds for quantum communication complexity. In *Proceedings 42nd IEEE Symposium on Foundations of Computer Science*, pages 288–297. IEEE, 2001. [quant-ph/0106160](#).
- 19 Ashley Montanaro. A new exponential separation between quantum and classical one-way communication complexity. *Quantum Inf. Comput.*, 11(7&8):574–591, 2011. [arXiv:1007.3587](#).
- 20 Ilan Newman. Private vs. common random bits in communication complexity. *Information Processing Letters*, 39(2):67–71, 1991.
- 21 Ran Raz. Fourier analysis for probabilistic communication complexity. *Computational Complexity*, 5(3-4):205–221, 1995.
- 22 Ran Raz. Exponential separation of quantum and classical communication complexity. In *STOC*, volume 99, pages 358–367. Citeseer, 1999.
- 23 Oded Regev and Bo’az Klartag. Quantum one-way communication can be exponentially stronger than classical communication. In *Proc. 43rd Annual ACM Symp. Theory of Computing*, pages 31–40, 2011. [arXiv:1009.3640](#).
- 24 Yaoyun Shi, Xiaodi Wu, and Wei Yu. Limits of quantum one-way communication by matrix hypercontractive inequality, 2012.
- 25 Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Comput.*, 26:1484–1509, 1997. [quant-ph/9508027](#).
- 26 Elad Verbin and Wei Yu. The streaming complexity of cycle counting, sorting by reversals, and other problems. In *Proceedings of the twenty-second annual ACM-SIAM symposium on Discrete Algorithms*, pages 11–25. Society for Industrial and Applied Mathematics, 2011.
- 27 Andrew Chi-Chih Yao. Probabilistic computations: Toward a unified measure of complexity. In *18th Annual Symposium on Foundations of Computer Science (sfcs 1977)*, pages 222–227. IEEE, 1977.
- 28 Andrew Chi-Chih Yao. Some complexity questions related to distributive computing (preliminary report). In *Proceedings of the eleventh annual ACM symposium on Theory of computing*, pages 209–213. ACM, 1979.
- 29 Andrew Chi-Chih Yao. Quantum circuit complexity. In *Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science*, pages 352–361. IEEE, 1993.

A Proof of Upper Bounds

In this and the following appendices, denote by $R_\epsilon^1(\mathcal{P})$ and $Q_\epsilon^1(\mathcal{P})$ the classical and quantum communication cost of the protocol \mathcal{P} in bits and qubits, respectively, and denote by $R_\epsilon^1(f) = \min_{\mathcal{P}} R_\epsilon^1(\mathcal{P})$ and $Q_\epsilon^1(f) = \min_{\mathcal{P}} Q_\epsilon^1(\mathcal{P})$ the minimum classical and quantum communication cost, respectively, over all one-way protocols \mathcal{P} without shared randomness that solve a communication problem f with failure probability $0 < \epsilon < 1/2$.

A.1 Classical Upper Bound

Consider the f -BHP $_n^{\alpha,t}$ problem for $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ with $\text{sdeg}(f) \leq 1$. Now let $p : \{-1, 1\}^t \rightarrow [-1, 1]$ be a normalized sign-representing polynomial for f . Hence we can write

$$p(x) = \alpha_0 + \sum_{i=1}^t \alpha_i x_i$$

with $(\alpha_i)_{i=0}^t \in \mathbb{R}$. Let $\beta = \min_x |p(x)|$ be the bias of p .

► **Theorem 5.** $R_\epsilon^1(f\text{-BHP}_n^{\alpha,t}) = O\left(\left(\frac{t}{\alpha\beta}\right)^2 \log \frac{1}{\epsilon} \log n\right)$ if $\text{sdeg}(f) \leq 1$.

Proof. Consider the following protocol: Alice picks $m = O\left(\left(\frac{t}{\alpha\beta}\right)^2 \log \frac{1}{\epsilon}\right)$ bits from x uniformly at random (with replacement) and sends them to Bob, together with their indices. Let I and $\{x_i\}_{i \in I}$ be the indices and bitvalues sent, respectively. Let $j(i) = \lceil \sigma(i)/t \rceil$ and $k(i) \equiv \sigma(i) \bmod t$ for all $i \in I$, where $\sigma \in S_n$ is Bob's permutation. Define the random variable $X(i) = (\alpha_{k(i)}x_i + \alpha_0/t)w_{j(i)}$ if $\sigma(i) \in [\alpha n/t]$ and $X(i) = 0$ if $\sigma(i) \notin [\alpha n/t]$, where α_0 and α_k are the zeroth order and x_k 's coefficients, respectively, from the sign-representing polynomial p , and define $X = \sum_{i \in I} X(i)$. Bob then computes $\text{sgn}(X)$. If the sign is 1, then he outputs $B_f(x) = w$, and if the sign is -1 , then he outputs $B_f(x) = \bar{w}$.

To see why the protocol works, we calculate the expectation value of random variable X .

$$\begin{aligned} \mathbb{E}[X] &= m \cdot \mathbb{E}_i[X(i)] \\ &= \alpha m \cdot \mathbb{E}_i[(\alpha_{k(i)}x_i + \alpha_0/t)w_{j(i)}] \\ &= \alpha m \cdot \mathbb{E}_j[\mathbb{E}_k[\alpha_k \sigma(x)_k^{(j)} + \alpha_0/t]w_j] \\ &= \alpha m \cdot \mathbb{E}_j\left[\frac{p(\sigma(x)^{(j)})}{t}w_j\right] \\ &= \alpha m \frac{t}{n} \sum_{j=1}^{n/t} \frac{p(\sigma(x)^{(j)})}{t}w_j \\ &= \frac{\alpha m}{n} \left[\sum_{j:w_j=1} p(\sigma(x)^{(j)}) - \sum_{j:w_j=-1} p(\sigma(x)^{(j)}) \right]. \end{aligned}$$

If $f(\sigma(x)^{(j)}) = w_j$, then $w_j = 1 \implies p(\sigma(x)^{(j)}) \geq \beta > 0$ and $w_j = -1 \implies p(\sigma(x)^{(j)}) \leq -\beta < 0$. Therefore

$$\mathbb{E}[X] \geq \frac{\alpha m}{n} \left[\sum_{j:w_j=0} \beta - \sum_{j:w_j=1} -\beta \right] = \alpha m \frac{\beta}{t}.$$

1:14 Generalized Boolean Hidden Matching Problem

If, on the other hand, $f(\sigma(x)^{(j)}) = -w_j$, then $w_j = 1 \implies p(\sigma(x)^{(j)}) \leq -\beta < 0$ and $w_j = -1 \implies p(\sigma(x)^{(j)}) \geq \beta > 0$. Therefore

$$\mathbb{E}[X] \leq \frac{\alpha m}{n} \left[\sum_{j:w_j=0} -\beta - \sum_{j:w_j=1} \beta \right] = -\alpha m \frac{\beta}{t}.$$

By using a Chernoff bound [13] of the type $\Pr[X > \mathbb{E}[X] + u], \Pr[X < \mathbb{E}[X] - u] \leq e^{-2u^2/m}$ with $u > 0$ and setting $u = \pm \mathbb{E}[X] > 0$, we can make

$$\Pr[X > 0 \mid B_f(x) = \bar{w}], \Pr[X < 0 \mid B_f(x) = w] \leq \epsilon$$

by taking $m = O((\frac{t}{\alpha\beta})^2 \log \frac{1}{\epsilon})$. Therefore Alice and Bob can decide if $B_f(x) = w$ or $B_f(x) = \bar{w}$ with error probability ϵ and $O((\frac{t}{\alpha\beta})^2 \log \frac{1}{\epsilon} \log n)$ bits of communication. ◀

A.2 Quantum Upper Bound

Consider the f -BHP $_n^{\alpha,t}$ problem for $f : \{-1, 1\}^t \rightarrow \{-1, 1\}$ with $\text{sdeg}(f) = 2$. Let $p : \{-1, 1\}^t \rightarrow [-1, 1]$ be a normalized sign-representing polynomial for f . Let $\beta = \min_x |p(x)|$ be the bias of p .

We say that a polynomial q of degree k is block-multilinear if its variables x_1, \dots, x_N can be partitioned into k blocks R_1, \dots, R_k , such that every monomial of q contains exactly one variable from each block. As a special case, a block-multilinear polynomial q of degree 2 can be written as

$$q(x_1, \dots, x_n, y_1, \dots, y_m) = \sum_{\substack{i \in [n] \\ j \in [m]}} a_{ij} x_i y_j$$

with variables in the first block labeled as x_1, \dots, x_n and the variables in the second block labeled as y_1, \dots, y_m . Defining the matrix $A = (a_{ij})_{i \in [n], j \in [m]}$, then

$$q(x, y) = x^T A y$$

for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. We say that q is *bounded* if $|q(x, y)| \leq 1$ for all $x \in \{-1, 1\}^n, y \in \{-1, 1\}^m$. This translates to

$$\max_{\substack{x \in \{-1, 1\}^n \\ y \in \{-1, 1\}^m}} \left| \sum_{\substack{i \in [n] \\ j \in [m]}} a_{ij} x_i y_j \right| \leq 1,$$

i.e., $\|A\|_{\infty \rightarrow 1} \leq 1$.

In order to prove the quantum upper bound, we will need the following results. In what comes, define $\tilde{x} = (1, x_1, \dots, x_t)$.

► **Lemma 15** ([2]). *Given a $m \times m$ complex matrix M , there exists a unitary U (on a possibly larger space with basis $|1\rangle, \dots, |k\rangle$ for some $k \geq m$) such that, for any unit vector $|y\rangle = \sum_{i=1}^m \alpha_i |i\rangle$, $U|y\rangle = \frac{M|y\rangle}{\|M|y\rangle\|} + |\phi\rangle$, where $|\phi\rangle$ consists of basis states $|i\rangle$, $i > m$ only.*

► **Theorem 16** ([2]). *Let $p : \{-1, 1\}^t \rightarrow [-1, 1]$ be a sign-representing polynomial for f with $\text{sdeg}(f) = 2$. Then there is a block-multilinear polynomial $\tilde{p} : \mathbb{R}^{2(t+1)} \rightarrow \mathbb{R}$ such that $\tilde{p}(\tilde{x}, \tilde{x}) = p(x)$ for any $x \in \{-1, 1\}^t$, and $|\tilde{p}(y)| \leq 3$ for any $y \in \{-1, 1\}^{2(t+1)}$.*

Let $\tilde{p} : \mathbb{R}^{2(t+1)} \rightarrow \mathbb{R}$ be the block-multilinear polynomial of degree 2 obtained from the sign-representing polynomial p of f according to Theorem 16. It can be written as

$$\tilde{p}(x, y) = \sum_{i, j \in [t+1]} a_{ij} x_i y_j = x^T A y, \quad (3)$$

where $A = (a_{ij})_{i, j \in [t+1]}$.

With these in hands, we present our upper bound.

► **Theorem 6.** $Q_\epsilon^1(f\text{-BHP}_n^{\alpha, t}) = O\left(\left(\frac{t}{\alpha\beta}\right)^2 \log \frac{1}{\epsilon} \log n\right)$ if $sdeg(f) \leq 2$.

Proof. Consider the following protocol: Alice sends to Bob $m = O\left(\left(\frac{t}{\alpha\beta}\right)^2 \log \frac{1}{\epsilon}\right)$ copies of the quantum state of $O(\log n)$ qubits

$$|\psi_A\rangle = \frac{1}{\sqrt{n + n/t}} \left(\sum_{i=1}^n x_i |i\rangle + \sum_{i=1}^{n/t} |n + i\rangle \right).$$

Bob measures each of them by using the POVM

$$\left\{ |n + j\rangle\langle n + j| + \sum_{i=(j-1)t+1}^{jt} |\sigma^{-1}(i)\rangle\langle\sigma^{-1}(i)| \right\}_{j \in [n/t]},$$

where $\sigma \in S_n$ is his permutation, and attaches a qubit in the state $|+\rangle$ to each of the final states. Let $I \subseteq [n/t]$ be the sequence of indices from his measurements. Then his final state is

$$|\psi_B\rangle = \bigotimes_{j \in I} |+\rangle |\psi^{(j)}\rangle,$$

where

$$|\psi^{(j)}\rangle = \frac{1}{\sqrt{t+1}} \left(|n + j\rangle + \sum_{i=(j-1)t+1}^{jt} x_{\sigma^{-1}(i)} |\sigma^{-1}(i)\rangle \right).$$

Let A be the $(t+1) \times (t+1)$ matrix from the representation of \tilde{p} according to Eq. 3. Lemma 15 guarantees the existence of a unitary U_j such that $U_j |\psi^{(j)}\rangle = \frac{A|\psi^{(j)}\rangle}{\|A\|} + |\phi^{(j)}\rangle$, with $\langle \phi^{(j)} | \psi^{(j)} \rangle = 0$. Bob then applies a controlled U_j gate onto each $|+\rangle_j |\psi^{(j)}\rangle$ to obtain

$$\bigotimes_{j \in I} C U_j |\psi_B\rangle = \bigotimes_{j \in I} \left(\frac{1}{\sqrt{2}} |0\rangle |\psi^{(j)}\rangle + \frac{1}{\sqrt{2}} |1\rangle U_j |\psi^{(j)}\rangle \right)$$

and then performs a Hadamard gate on the first qubit of each of the subsystems I and measures them. Let $m_j \in \{0, 1\}$ be the result of the measurement for block $j \in I$. Define the random variable $X(j) = -(-1)^{m_j} w_j$ if $j \in [\alpha n/t]$ and $X(j) = 0$ if $j \notin [\alpha n/t]$, and define $X = \sum_{j \in I} X(j)$. Bob then computes $\text{sgn}(X)$: if $\text{sgn}(X) > 0$, he outputs that $B_f(x) = w$, and if $\text{sgn}(X) < 0$, he outputs that $B_f(x) = \bar{w}$.

To see why the protocol works, first note that the probability of measuring 1 is

$$\begin{aligned} \Pr[1] &= \frac{1}{2} \left(1 + \langle \psi^{(j)} | U | \psi^{(j)} \rangle \right) = \frac{1}{2} \left(1 + \frac{\langle \psi^{(j)} | A | \psi^{(j)} \rangle}{\|A\|} \right) \\ &= \frac{1}{2} \left(1 + \frac{\tilde{p}(\widetilde{\sigma(x)^{(j)}}, \widetilde{\sigma(x)^{(j)})}}{\|A\|(t+1)} \right) = \frac{1}{2} \left(1 + \frac{p(\sigma(x)^{(j)})}{\|A\|(t+1)} \right). \end{aligned}$$

1:16 Generalized Boolean Hidden Matching Problem

The remainder of the argument is similar to the classical upper bound proof. Recalling that $m = |I|$, the expectation value of X is

$$\begin{aligned}\mathbb{E}[X] &= m \cdot \mathbb{E}_j[X(j)] \\ &= \alpha m \cdot \mathbb{E}_j[-(-1)^{m_j} w_j] \\ &= \alpha m \frac{t}{n} \sum_{j=1}^{n/t} (\Pr[m_j = 1] - \Pr[m_j = 0]) w_j \\ &= \alpha m \frac{t}{n} \left[\sum_{j:w_j=1} \frac{p(\sigma(x)^{(j)})}{\|A\|(t+1)} - \sum_{j:w_j=-1} \frac{p(\sigma(x)^{(j)})}{\|A\|(t+1)} \right].\end{aligned}$$

If $f(\sigma(x)^{(j)}) = w_j$, then $w_j = 1 \implies p(\sigma(x)^{(j)}) \geq \beta > 0$ and $w_j = -1 \implies p(\sigma(x)^{(j)}) \leq -\beta < 0$. Therefore

$$\mathbb{E}[X] \geq \alpha m \frac{t}{n} \frac{1}{\|A\|(t+1)} \left[\sum_{j:w_j=1} \beta - \sum_{j:w_j=-1} -\beta \right] = \frac{\alpha m \beta}{\|A\|(t+1)}.$$

If, on the other hand, $f(\sigma(x)^{(j)}) = -w_j$, then $w_j = 1 \implies p(\sigma(x)^{(j)}) \leq -\beta < 0$ and $w_j = -1 \implies p(\sigma(x)^{(j)}) \geq \beta > 0$. Therefore

$$\mathbb{E}[X] \leq \alpha m \frac{t}{n} \frac{1}{\|A\|(t+1)} \left[\sum_{j:w_j=1} -\beta - \sum_{j:w_j=-1} \beta \right] = -\frac{\alpha m \beta}{\|A\|(t+1)}.$$

By using a Chernoff bound [13] of the type $\Pr[X > \mathbb{E}[X] + u], \Pr[X < \mathbb{E}[X] - u] \leq e^{-2u^2/m}$ with $u > 0$ and setting $u = \pm \mathbb{E}[X] > 0$, we can make

$$\Pr[X > 0 \mid B_f(x) = \bar{w}], \Pr[X < 0 \mid B_f(x) = w] \leq \epsilon$$

by taking $m = O((\frac{t}{\alpha\beta})^2 \log \frac{1}{\epsilon})$, where we use that $\|A\| \leq \|A\|_{\infty \rightarrow 1} \leq 3$ according to Theorem 16 (note that $\frac{\|Ax\|_2}{\|x\|_2} \leq \frac{\|Ax\|_1}{\|x\|_\infty}$, and taking maximums over all x on both sides gives $\|A\| \leq \|A\|_{\infty \rightarrow 1}$). Therefore Alice and Bob can decide if $B_f(x) = w$ or $B_f(x) = \bar{w}$ with error probability ϵ and $O((\frac{t}{\alpha\beta})^2 \log \frac{1}{\epsilon} \log n)$ qubits of communication. \blacktriangleleft

Improved Approximate Degree Bounds for k -Distinctness

Nikhil S. Mande

Georgetown University, Washington DC, USA
nikhil.mande@georgetown.edu

Justin Thaler

Georgetown University, Washington DC, USA
justin.thaler@georgetown.edu

Shuchen Zhu

Georgetown University, Washington DC, USA
shuchen.zhu@georgetown.edu

Abstract

An open problem that is widely regarded as one of the most important in quantum query complexity is to resolve the quantum query complexity of the k -distinctness function on inputs of size N . While the case of $k = 2$ (also called Element Distinctness) is well-understood, there is a polynomial gap between the known upper and lower bounds for all constants $k > 2$. Specifically, the best known upper bound is $O\left(N^{(3/4)-1/(2^{k+2}-4)}\right)$ (Belovs, FOCS 2012), while the best known lower bound for $k \geq 2$ is $\tilde{\Omega}\left(N^{2/3} + N^{(3/4)-1/(2k)}\right)$ (Aaronson and Shi, J. ACM 2004; Bun, Kothari, and Thaler, STOC 2018).

For any constant $k \geq 4$, we improve the lower bound to $\tilde{\Omega}\left(N^{(3/4)-1/(4k)}\right)$. This yields, for example, the first proof that 4-distinctness is strictly harder than Element Distinctness. Our lower bound applies more generally to approximate degree.

As a secondary result, we give a simple construction of an approximating polynomial of degree $\tilde{O}(N^{3/4})$ that applies whenever $k \leq \text{polylog}(N)$.

2012 ACM Subject Classification Theory of computation \rightarrow Quantum query complexity

Keywords and phrases Quantum Query Complexity, Approximate Degree, Dual Polynomials, k -distinctness

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.2

Related Version A full version of the paper is available at <https://arxiv.org/abs/2002.08389>

Funding *Justin Thaler*: Supported by the National Science Foundation CAREER award (grant CCF-1845125).

Shuchen Zhu: Supported by the National Science Foundation CAREER award (grant CCF-1845125).

Acknowledgements JT is grateful to Robin Kothari for extremely useful suggestions and discussions surrounding Theorem 2, and to Mark Bun for essential discussions regarding Theorem 19. SZ would like to thank Yao Ji for several helpful conversations.

1 Introduction

In quantum query complexity, a quantum algorithm is given query access to the bits of an unknown input x , and the goal is to compute some (known) function f of x while minimizing the number of bits of x that are queried. In contrast to classical query complexity, quantum query algorithms are allowed to make queries in superposition, and the algorithm is not charged for performing unitary operations that are independent of x . Quantum query



© Nikhil S. Mande, Justin Thaler, and Shuchen Zhu;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 2; pp. 2:1–2:22

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

complexity is a rich model that allows for the design of highly sophisticated algorithms and captures much of the power of quantum computing. Indeed, most quantum algorithms were discovered in or can easily be described in the query setting.

An open problem that is widely regarded as one of the most important in quantum query complexity [18] is to resolve the complexity of the k -distinctness function. For this function, the input x specifies a list of N numbers from a given range of size R ,¹ and the function evaluates to TRUE² if there is any range item that appears k or more times in the list. The case $k = 2$ corresponds to the complement of the widely-studied *Element Distinctness* function, whose complexity is known to be $\Theta(N^{2/3})$ [4, 1].

For general values of k , the best known upper bound on the quantum query complexity of k -distinctness is $O\left(N^{3/4-1/(2^{k+2}-4)}\right)$, due to a highly sophisticated algorithm of Belovs [8]. For a long time, the best known lower bound on the quantum query complexity of k -distinctness was $\Omega(N^{2/3})$ for any $k \geq 2$, due to Aaronson and Shi [1], with refinements given by Kutin [15] and Ambainis [2]. This lower bound is tight for $k = 2$ (matching Ambainis' upper bound [4]), but it is not known to be tight for any $k > 2$. Recently, Bun, Kothari, and Thaler [11] proved a lower bound of $\tilde{\Omega}(N^{3/4-1/(2k)})$ for constant k .³ This improved over the prior lower bound of $\Omega(N^{2/3})$ for any constant $k \geq 7$. Furthermore, combined with Belovs' upper bound, this established that for sufficiently large constants k , the exponent in the quantum query complexity of k -distinctness approaches $3/4$ from below. However, the precise rate at which the quantum query complexity approaches $N^{3/4}$ remains open: there is a polynomial gap between the upper and lower bounds for any constant k , and indeed there is a qualitative difference between the inverse-exponential dependence on k in the exponent of $N^{3/4-1/(2^{k+2}-4)}$ (the known upper bound), and the inverse-linear dependence in the known lower bound of $N^{3/4-1/(2k)}$.

Main Result

This paper improves the lower bound from $\tilde{\Omega}(N^{3/4-1/(2k)})$ to $\tilde{\Omega}(N^{3/4-1/(4k)})$. While this bound is qualitatively similar to the lower bound of [11], it offers a polynomial improvement for every constant $k \geq 4$. Perhaps more significantly, for $k \in \{4, 5, 6\}$, it is the first improvement over Aaronson and Shi's $\Omega(N^{2/3})$ lower bound that has stood for nearly 20 years.

Approximate Degree

The ϵ -error approximate degree of a Boolean function $f: \{-1, 1\}^n \rightarrow \{-1, 1\}$, denoted $\widetilde{\deg}_\epsilon(f)$, is the least degree of a real polynomial p such that $|p(x) - f(x)| \leq \epsilon$ for all $x \in \{-1, 1\}^n$. The standard setting of the error parameter is $\epsilon = 1/3$, and the $(1/3)$ -approximate degree of f is denoted $\widetilde{\deg}(f)$ for brevity. As famously observed by Beals et al. [6], the quantum query complexity of a function f is lower bounded by (one half times) the approximate degree of f . Hence, any lower bound on the approximate degree of f implies that (up to a factor of 2) the same lower bound holds for the quantum query complexity of f . As with prior lower bounds for k -distinctness [1, 15, 2, 11], our k -distinctness lower bound is in fact an approximate

¹ For purposes of this introduction, N and R are assumed to be of the same order of magnitude (up to a factor depending on k alone). For simplicity throughout this section, we state our bounds purely in terms of N , leaving unstated the assumption that R and N are of the same order of magnitude.

² Throughout this manuscript, we associate -1 with logical TRUE and $+1$ with logical FALSE.

³ Throughout this manuscript, \tilde{O} , $\tilde{\Omega}$ and $\tilde{\Theta}$ notations are used to hide factors that are polylogarithmic in N .

degree lower bound (on the natural Boolean function induced by k -distinctness on $N \lceil \log_2 R \rceil$ bits, where R denotes the size of the range). Our analysis is a substantial refinement of the lower bound analysis of Bun et al. [11].

► **Theorem 1** (Informal version of Theorem 17 and Corollary 18). *For any constant $k \geq 2$, the approximate degree and quantum query complexity of the k -distinctness function with domain size N and range size $R \geq N$ is $\tilde{\Omega}(N^{3/4-1/(4k)})$.*

A Secondary Result: The Approximate Degree for Super-Constant Values of k

Recall that for constant k , the best known approximate degree upper bound for k -distinctness is $O(N^{3/4-1/(2^{k+2}-4)})$ [8]. For non-constant values of k , the upper bound implied by Belovs' algorithm grows exponentially with k . That is, the Big-Oh notation in the upper bound hides a leading factor of at least 2^{ck} for some positive constant c .⁴ Consequently Belovs' bound is $N^{3/4+\Omega(1)}$ for any $k \geq \Omega(\log N)$. Furthermore, the bound becomes vacuous (i.e., linear in N) for $k \geq c \log N$ for a large enough constant $c > 0$.

Our secondary result improves this state of affairs by giving a $\tilde{O}(N^{3/4})$ approximate degree upper bound that holds for any value of k that grows at most polylogarithmically with N .

► **Theorem 2** (Informal). *For any $k \leq \text{polylog}(N)$, the approximate degree of k -distinctness is $\tilde{O}(N^{3/4})$.*

We mention that for any $k \geq 2$, the approximating polynomials for k -distinctness that follow from prior works [4, 8, 24] are quite complicated, and in our opinion there has not been a genuinely simple construction of any $O(N^{3/4})$ -degree approximating polynomials recorded in the literature, even for the case of $k = 2$ (i.e., Element Distinctness). Accordingly, we feel that Theorem 2 has didactic value even for constant values of k (though the $\tilde{O}(N^{3/4})$ approximate degree upper bound that it achieves is not tight for any constant $k \geq 2$).

To clarify, Theorem 2 does *not* yield a quantum query upper bound, only an approximate degree upper bound. It remains an interesting open question whether the quantum query complexity of k -distinctness is sublinear in N for all $k = \text{polylog}(N)$ (see Section 1.1 for further discussion).

Our proof of Theorem 2 is a simple extension of a result of Sherstov [24, Theorem 1.3] that yielded an $O(N^{3/4})$ approximate degree upper bound for a different function called Surjectivity.⁵ A formal statement and proof can be found in the full version of this paper.

1.1 Discussion and Open Problems

The most obvious and important open question is to finish resolving the approximate degree and quantum query complexity of k -distinctness for any $k > 2$. Currently, the upper and lower bounds qualitatively differ in their dependence on k , with the upper bound having an exponent of the form $3/4 - \exp(-O(k))$ and the lower bound having an exponent of the form $3/4 - \Omega(1/k)$. It seems very likely that major new techniques will be needed to

⁴ Belovs' approximate degree upper bound was recently reproved by Sherstov [24], who made the exponential dependence on k explicit (see, e.g., [24, Theorem 6.6]). To clarify, Belovs' result is in fact a quantum query upper bound, which in turn implies an approximate degree upper bound. Sherstov's proof avoids quantum algorithms, and hence does not yield a quantum query upper bound.

⁵ Surjectivity is the function that interprets its input as a list of N numbers from a given range of size R , and evaluates to TRUE if and only if every range element appears at least once in the list.

qualitatively change the form of *either* the upper or lower bound. In particular, on the lower bounds side, our analysis is based on a variant of a technique called *dual block composition* (see Section 1.2), and we suspect that we have reached the limit of what is provable for k -distinctness using this technique and its variants.

We remark here that Liu and Zhandry [18] recently showed that the quantum query complexity of a certain *search* version of k -distinctness (defined over randomly generated inputs) is $\Theta(n^{1/2-1/(2^k-1)})$. This inverse-exponential dependence on k is tantalizingly reminiscent of Belovs' upper bound for k -distinctness. This may be construed as mild evidence that $3/4 - \exp(-O(k))$ is the right qualitative bound for k -distinctness itself.

A very interesting intermediate goal is to establish any polynomial improvement over the long-standing $\Omega(n^{2/3})$ lower bound for 3-distinctness. This would finally establish that 3-distinctness is strictly harder than Element Distinctness (such a result is now known for all $k \geq 4$ due to Theorem 1).

It would also be interesting to resolve the quantum query complexity of k -distinctness for $k = \text{polylog}(N)$. Although this question may appear to be of specialized interest, we believe that resolving it could shed light on the relationship between approximate degree and quantum query complexity. Indeed, while any quantum algorithm for a function f can be turned into an approximating polynomial for f via the transformation of Beals et al. [6], no transformation in the reverse direction is possible in general [3]. This can be seen, for example, because the quantum query complexity of Surjectivity is known to be $\Omega(N)$ [7, 25], but its approximate degree is $O(N^{3/4})$ [24, 11]. Nonetheless, approximate degree and quantum query complexity turn out to coincide for most functions that arise naturally (Surjectivity remains the only function that exhibits a separation, without having been specifically constructed for that purpose). In our opinion, this phenomenon remains mysterious, and it would be interesting to demystify it. For example, could one identify special properties of approximating polynomials that would permit a reverse-Beals-et-al. transformation to turn that polynomial into a quantum query algorithm?⁶ Perhaps an $\tilde{O}(N^{3/4})$ upper bound for $(\text{polylog}(N))$ -distinctness could be derived in this manner. Such an upper bound (even for $(\log N)$ -distinctness) would yield improved quantum query upper bounds for min-entropy estimation [17]. On the other hand, due to our Theorem 2, any $N^{3/4+\Omega(1)}$ lower bound for $(\text{polylog}(N))$ -distinctness would require moving beyond the polynomial method.⁷

1.2 Overview of the Lower Bound

Throughout this subsection we assume that $k \geq 2$ is an arbitrary but fixed constant.

Let THR_N^k denote the function on N -bit inputs that evaluates to -1 on inputs of Hamming weight at least k , and evaluates to 1 otherwise. For $N \leq n$, let $(\{-1, 1\}^n)^{\leq N}$ denote the subset of $\{-1, 1\}^n$ consisting of all inputs of Hamming weight at most N . For any function $f_n: \{-1, 1\}^n \rightarrow \{-1, 1\}$,⁸ let $f_n^{\leq N}$ denote the partial function obtained by restricting the domain of f to $(\{-1, 1\}^n)^{\leq N}$, and let $\deg(f_n^{\leq N})$ denote the least degree of a real polynomial p such that $|p(x) - f_n(x)| \leq 1/3$ for all $x \in (\{-1, 1\}^n)^{\leq N}$.

⁶ There are works in this general direction, notably [5], which shows that a certain technical refinement of approximate degree, called approximation by completely bounded forms, characterizes quantum query complexity. But to our knowledge these works have not yielded any novel quantum query upper bounds for any specific function.

⁷ We remark that the positive-weights adversary method is also incapable of proving such a result due to the certificate complexity barrier.

⁸ Throughout, we use subscripts where appropriate to clarify the number of bits over which a function is defined.

Simplifying very slightly, prior work by Bun and Thaler [13] (building on an important lemma of Ambainis [2]) implied that for $k \geq 2$ the approximate degree of k -distinctness is equivalent to $\widetilde{\deg}(f_{RN}^{\leq N})$ for $f = \text{OR}_R \circ \text{THR}_N^k$. Here, $g_n \circ h_m$ denotes the function on $n \cdot m$ bits obtained by block-composing g and h , i.e., $g \circ h$ evaluates h on n disjoint inputs and feeds the outputs of all n copies of h into g .

Bun et al. [11] proved their $\tilde{\Omega}(N^{3/4-1/(2k)})$ lower bound for $\widetilde{\deg}(f_{RN}^{\leq N})$ via the *method of dual polynomials*. This is a technique for proving approximate degree lower bounds that works by constructing an explicit solution to a certain linear program capturing the approximate degree of any function. Specifically, a dual witness to the fact that $\widetilde{\deg}(f_{RN}^{\leq N}) \geq d$ is a function $\psi: \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ satisfying the following properties (this dual formulation is standard, and can be found, for example, in [21]).

First, ψ must be uncorrelated with all polynomials p of degree at most d , i.e., $\langle \psi, p \rangle = 0$ for all such polynomials p , where $\langle \psi, p \rangle = \sum_{x \in \{-1, 1\}^{RN}} \psi(x)p(x)$. Such a ψ is said to have *pure high degree* at least d . Second, ψ must be well-correlated with f , i.e., $\langle \psi, f \rangle \geq (1/3) \cdot \|\psi\|_1$, where $\|\psi\|_1 := \sum_{x \in \{-1, 1\}^{RN}} |\psi(x)|$. Finally, ψ must equal 0 on inputs in $\{-1, 1\}^{RN} \setminus \left(\{-1, 1\}^{RN}\right)^{\leq N}$.

To simplify greatly, Bun et al. [11] constructed their dual witness for $\left(\text{OR}_R \circ \text{THR}_N^k\right)^{\leq N}$ roughly as follows. They took a dual witness Ψ for the fact that $\widetilde{\deg}(\text{OR}_R) = \Omega(R^{1/2})$ [19, 28, 12] and a dual witness ϕ for the fact that THR_N^k also has large approximate degree, and they combined Ψ and ϕ in a certain manner (introduced in prior works [27, 23, 16]) to get a dual witness for the composed function $\left(\text{OR}_R \circ \text{THR}_N^k\right)^{\leq N}$. The technique used to combine Ψ and ϕ is often called *dual block composition*, and is denoted $\Psi \star \phi$.⁹ Dual block composition is defined as follows (below, each $x_i \in \{-1, 1\}^N$):

$$(\Psi \star \phi)(x_1, \dots, x_R) = 2^R \cdot \Psi(\text{sgn}(\phi(x_1)), \dots, \text{sgn}(\phi(x_R))) \cdot \prod_{i=1}^R |\phi(x_i)| / \|\phi\|_1.$$

Here, $\text{sgn}(r)$ equals -1 if $r < 0$ and equals $+1$ if $r > 0$.¹⁰ To show that $\Psi \star \phi$ is a dual witness for the fact that the approximate degree of $\left(\text{OR}_R \circ \text{THR}_N^k\right)^{\leq N}$ is at least d , it is necessary to show that $\Psi \star \phi$ has pure high degree at least d , and that $\Psi \star \phi$ is well-correlated with $\left(\text{OR}_R \circ \text{THR}_N^k\right)^{\leq N}$. It is known that pure high degree increases multiplicatively under the \star operation, and hence the pure high degree calculation for $\Psi \star \phi$ is straightforward. In contrast, the correlation calculation is the key technical challenge and bottleneck in the analysis of [11]. Our key improvement over their work is to modify the construction of the dual witness in a manner that allows for an improved correlation bound.

At a high level, what we do is replace the dual block composition $\Psi \star \phi$ from the construction of [11] with a *variant* of dual block composition introduced by Sherstov [22]. Sherstov introduced this variant to address the correlation issues that arise when attempting

⁹ To clarify, this entire outline is a major simplification of the actual dual witness construction in [11]. The details provided in the outline of this introduction are chosen to highlight the key technical issues that we must address in this work. Amongst other simplifications in this outline, the actual dual witness from [11] is not $\Psi \star \phi$, but rather a “post-processed” version of $\Psi \star \phi$, where the post-processing step is used to ensure that the dual witness evaluates to 0 on all inputs of Hamming weight more than N .

¹⁰ It is irrelevant how one defines $\text{sgn}(0)$ because if $\phi(x_i) = 0$ for any i , the product $\prod_{i=1}^R |\phi(x_i)| / \|\phi\|_1$ forces $\Psi \star \phi$ to 0. For this reason, the remainder of the discussion in this section implicitly assumes that $\phi(x_i) \neq 0$ for all $i \in \{1, \dots, R\}$.

to use dual block composition to prove approximate degree lower bounds for composed functions, and he used it to prove direct sum and direct product theorems for approximate degree.¹¹ However, we have to modify even Sherstov’s variant of dual block composition in significant ways to render it useful in our context. We now attempt to give an informal sense of our modification and why it is necessary.

For block-composed functions $g \circ h$, the rough idea of any proof attempting to show that $\langle \Psi \star \phi, g \circ h \rangle$ is large is to hope that the following approximate equality holds:

$$\langle \Psi \star \phi, g \circ h \rangle \approx \langle \Psi, g \rangle. \quad (1)$$

If Equation (1) holds even approximately, then the correlation analysis of $\Psi \star \phi$ is complete, since the assumption that Ψ is a dual witness for the high approximate degree of g implies that the right hand side is large.

Equation (1) in fact holds with *exact* equality if ϕ agrees in sign with h at all inputs, i.e., if $\langle \phi, h \rangle = \|\phi\|_1$ [23, 16]. Unfortunately, the fact that ϕ is a dual witness for the large approximate degree of h implies only a much weaker lower bound on $\langle \phi, h \rangle$, namely that

$$\langle \phi, h \rangle \geq (1/3) \cdot \|\phi\|_1. \quad (2)$$

In general, Equation (2) is not enough to ensure that Equation (1) holds even approximately.

A rough intuition for why Equation (1) may fail to hold is the following. The definition of $\Psi \star \phi$ feeds $(\text{sgn}(\phi(x_1)), \dots, \text{sgn}(\phi(x_R)))$ into Ψ . One can think of $\text{sgn}(\phi(x_i))$ as ϕ ’s “prediction” about $h(x_i)$, and the fact that $\langle \phi, h \rangle \geq (1/3) \cdot \|\phi\|_1$ means that for an x_i chosen at random from the probability distribution $|\phi|/\|\phi\|_1$, this prediction is correct with probability at least $2/3$. Unfortunately, there are values of x_i for which $\text{sgn}(\phi(x_i)) \neq h(x_i)$, meaning that ϕ ’s predictions can sometimes be wrong. In this case, when feeding $\text{sgn}(\phi(x_i))$ into Ψ , dual block composition is “feeding an error” into Ψ , and this can cause $\Psi \star \phi$ to “make more errors” (i.e., output a value on an input that disagrees in sign with $g \circ h$ on that same input) than Ψ itself.

That is, there are two reasons $\Psi \star \phi$ may make an error: either Ψ itself may make an error (let us call this Source 1 for errors), and/or one or more copies of ϕ may make an error (let us call this Source 2 for errors).¹² The first source of error is already fully accounted for in the right hand side of Equation (1). The second source of error is not, and this is the reason that Equation (1) may fail to hold even approximately.

Roughly speaking, while Equation (2) guarantees that $\text{sgn}(\phi(x_i))$ is not “an error” for each i with good probability (i.e., probability at least $2/3$), that still means that with very high probability, $\text{sgn}(\phi(x_i))$ will be in error (i.e., not equal to $h(x_i)$) for a *constant fraction* of blocks $i \in \{1, \dots, R\}$. Any one of these errors could be enough to cause a Source 2 error.

Fortunately for us, $g = \text{OR}_R$ has low (-1) -certificate complexity, meaning that on inputs x in $\text{OR}_R^{-1}(-1)$, to certify that indeed $x \in \text{OR}_R^{-1}(-1)$, it is sufficient to identify just one coordinate of x that equals -1 . This renders certain kinds of sign-errors made by ϕ benign. Specifically, letting $S = \{x: \phi(x) < 0\}$ and $E^- = S \cap f^{-1}(1)$ denote the false-negative errors made by ϕ , the low (-1) -certificate complexity of OR_R means that it is okay if “a constant

¹¹ Variants of dual block composition related to the one introduced in [22] have played important roles in other recent works on approximate degree lower bounds, e.g., [14, 26].

¹² There may be inputs $x = (x_1, \dots, x_n)$ to $\Psi \star \phi$ that could be classified as *both* Source 1 and Source 2 errors. For purposes of this high-level introduction, it is not important whether such inputs get classified as Source 1 or Source 2 errors for $\Psi \star \phi$.

fraction of the negative values output by ϕ are in error". That is, so long as

$$\left(\sum_{E^-} |\phi(x)| \right) / \left(\sum_{x \in S} |\phi(x)| \right) = 1 - \Omega(1), \quad (3)$$

the contribution of "false negative errors made by ϕ " to actual Source 2 errors made by $\Psi \star \phi$ is low.

However, the situation is starkly different for "false positive errors" made by ϕ ; while OR_R has certificates of size 1 for inputs in $\text{OR}_R^{-1}(-1)$, the certificate complexity of the (unique) input in $\text{OR}_R^{-1}(+1)$ is n . That is, letting $T = \{x: \phi(x) > 0\}$ and $E^+ = T \cap f^{-1}(-1)$, for Equation (1) to hold even approximately for $g = \text{OR}_R$, it is essential that

$$\left(\sum_{E^+} |\phi(x)| \right) / \left(\sum_{x \in T} |\phi(x)| \right) \ll 1/R. \quad (4)$$

Accordingly, Bun et al. [11] obtain their lower bound for k -distinctness by using a dual witness ϕ for $h = \text{THR}_N^k$ that satisfies Equation (4). Using a dual with such few false positive errors causes [11] to lose an additive $1/(2k)$ term in the exponent of N in their final degree bound, relative to what they would obtain if Equation (2) were sufficient to ensure that Equation (1) approximately held.

As previously mentioned, Sherstov [22] introduced a variant of dual block composition intended to handle Source 2 errors that might have otherwise rendered Equation (1) false. Specifically, Sherstov proposed multiplying $(\Psi \star \phi)(x)$ by a low-degree polynomial $p_\eta(x)$ intended to "kill" any inputs x that may contribute Source 2 errors (here, η is a parameter, and we will explain shortly how the value of η is ultimately chosen). Specifically, p_η "counts" the number of blocks x_i of x such that $\text{sgn}(\phi(x_i)) \neq h(x_i)$, and p_η is defined (through polynomial interpolation) to evaluate to 0 if this number is any integer between 1 and η . This has the effect of eliminating all Source 2 errors made by $\Psi \star \phi$ on inputs x for which at most η copies of ϕ make an error. That is, p_η kills all inputs x in the set $U_\eta := \{x = (x_1, \dots, x_R): \text{sgn}(\phi(x_i)) \neq h(x_i) \text{ for between 1 and } \eta \text{ values of } i\}$. Note that multiplying $\Psi \star \phi$ by p_η has the additional, unfortunate effect of distorting the values that $\Psi \star \phi$ takes on other inputs; bounding the effect of this distortion is one challenge that Sherstov's analysis (as well as our own analysis in this work) has to address.

The intuition is that, so long as most Source 2 errors made by $\Psi \star \phi$ are caused by inputs in the set U_η , then multiplying $\Psi \star \phi$ by p_η should eliminate the otherwise devastating effects of most Source 2 errors. So the remaining challenge is to choose a dual witness ϕ for h guaranteeing that indeed most Source 2 errors are caused by inputs in U_η . More precisely, ϕ must be chosen to ensure that, with respect to the product distribution $\prod_{i=1}^R |\phi(x_i)| / \|\phi\|_1$, it is very unlikely that more than η copies of ϕ make an error on their input x_i .

To this end, it is implicit in Sherstov's analysis that Equation (1) approximately holds with $(\Psi \star \phi) \cdot p_\eta$ in place of $\Psi \star \phi$ so long as

$$\left(\sum_{x \in E^- \cup E^+} |\phi(x)| \right) / \|\phi\|_1 \ll \eta/R. \quad (5)$$

Notice that this is exactly Equation (4), except that the right hand side has crucially increased by a factor of η (also, Equation (5) counts both false-positive and false-negative errors, as opposed to just false-positive errors, which is a key discrepancy that we address below). The bigger that η is set, the less stringent is the requirement of Equation (5). However, it turns

out that, in order to ensure that $(\Psi \star \phi) \cdot p_\eta$ has pure high degree close to that of $\Psi \star \phi$ itself, η must be set to a value that is noticeably smaller than the pure high degree of Ψ . Ultimately, to obtain the strongest possible results, η gets set to some constant $C < 1$ times the pure high degree of Ψ .

In order to bring Sherstov's ideas to bear on k -distinctness, we have to modify his construction as follows. The key issue (alluded to above) is that Sherstov's construction is not targeted at functions $g \circ h$ where g has low (-1) -certificate complexity, and it is essential that we exploit this low certificate complexity in the correlation analysis to improve on the k -distinctness lower bound from [11]. Essentially, we modify Sherstov's definition of p_η to "ignore" all false negative errors (which as explained above are benign in our setting because $g = \text{OR}_R$ has low (-1) -certificate complexity). Rather we have p_η only "count" the false positive errors and kill any inputs where this number is between 1 and η .

We are able to show that with this modification, it is sufficient to choose a dual witness ϕ for THR_N^k satisfying

$$\left(\sum_{E^+} |\phi(x)| \right) / \left(\sum_{x \in T} |\phi(x)| \right) \ll \eta/R. \quad (6)$$

We end up setting $\eta \approx O(\sqrt{R})$ for our lower bound, hence the denominator on the right hand side of this inequality represents a quadratic improvement compared to that on the right hand side of Equation (4). This improvement ultimately enables us to improve the lower bound from $\tilde{\Omega}(N^{3/4-1/(2k)})$ to $\tilde{\Omega}(N^{3/4-1/(4k)})$.

The actual calculations required to establish the sufficiency of Equation (6) are quite involved, and we provide a more detailed proof overview in Section 3 to help the reader make sense of them.

2 Preliminaries

Let N, n and m be positive integers, $N \leq n$. For $z \in \{-1, 1\}^n$, let $|z|$ represent the *Hamming weight* of z , i.e., the number of -1 's in z . Define $(\{-1, 1\}^n)^{\leq N} := \{x \in \{-1, 1\}^n : |x| \leq N\}$. For any function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, denote by $f^{\leq N}$ the partial function that is defined on $(\{-1, 1\}^n)^{\leq N}$ and agrees with f on all such inputs. Define $\text{sgn} : \mathbb{R} \rightarrow \{-1, 1\}$ by $\text{sgn}(x) = 1$ for all non-negative x , and -1 otherwise. For any function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, define $\|f\|_1 := \sum_{x \in \{-1, 1\}^n} |f(x)|$. All logarithms in this paper are base 2 unless otherwise specified. Let $\mathbf{1}^n$ (respectively, $-\mathbf{1}^n$) denote the n -bit string $(1, 1, \dots, 1)$ (respectively, $(-1, -1, \dots, -1)$). We use the notation $[n]$ to denote the set $\{1, 2, \dots, n\}$.

Define the function $\text{OR}_N : \{-1, 1\}^N \rightarrow \{-1, 1\}$ to equal 1 if $x = \mathbf{1}^N$, and -1 otherwise. Define the *Threshold* function $\text{THR}_N^k : \{-1, 1\}^N \rightarrow \{-1, 1\}$ to equal 1 for inputs of Hamming weight less than k , and -1 otherwise. Given any functions $f_n : \{-1, 1\}^n \rightarrow \{-1, 1\}$ and $g_m : \{-1, 1\}^m \rightarrow \{-1, 1\}$, we define the function $f_n \circ g_m : \{-1, 1\}^{nm} \rightarrow \{-1, 1\}$ as

$$f_n \circ g_m(x_{11}, \dots, x_{1m}, x_{21}, \dots, x_{2m}, \dots, x_{n1}, \dots, x_{nm}) = f_n(g_m(x_1), g_m(x_2), \dots, g_m(x_n)),$$

$x_i \in \{-1, 1\}^m$ for all $i \in [n]$. We drop subscripts when the arities of the constituent functions are clear.

For any function $\psi : \{-1, 1\}^m \rightarrow \mathbb{R}$ such that $\|\psi\|_1 = 1$, let μ_ψ be the distribution on $\{-1, 1\}^m$, defined by $\mu_\psi(x) = |\psi(x)|$. Any function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ has a unique multilinear representation $f = \sum_{S \subseteq [n]} \hat{f}(S) \chi_S$, where for any $S \subseteq [n]$, the function $\chi_S : \{-1, 1\}^n \rightarrow \{-1, 1\}$ is defined by $\chi_S(x) = \prod_{i \in S} x_i$. Hence, $\|\hat{f}\|_1 = \sum_{S \subseteq [n]} |\hat{f}(S)|$. It follows that for any function $\phi : \{-1, 1\}^n \rightarrow \mathbb{R}$, there exists a unique multilinear polynomial $\tilde{\phi} : \mathbb{R}^n \rightarrow \mathbb{R}$ such that $\tilde{\phi}(x) = \phi(x)$ for all $x \in \{-1, 1\}^n$.

► **Definition 3** (*k*-distinctness). For integers k, N, R with $k \leq N$, define the function $\text{DIST}_{N,R}^k : [R]^N \rightarrow \{-1, 1\}$ by $\text{DIST}_{N,R}^k(s_1, \dots, s_N) = -1$ iff there exists an $r \in [R]$ and distinct indices i_1, \dots, i_k such that $s_{i_1} = \dots = s_{i_k} = r$. When necessary, the domain of the function can be viewed as $\{-1, 1\}^{N \log R}$.

► **Definition 4** (Approximate degree). For any function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$, any integer $N \leq n$, and any $\epsilon \in [0, 1]$, define the ϵ -approximate degree of $f^{\leq N}$ to be

$$\widetilde{\deg}_\epsilon(f^{\leq N}) = \min_{\substack{p: |p(x) - f(x)| \leq \epsilon \\ \forall x \in \{-1, 1\}^n, |x| \leq N}} \deg(p).$$

When the subscript is dropped, ϵ is assumed to equal $1/3$. When the superscript is dropped in $f^{\leq N}$, then N is assumed to equal n .

Note that this definition places no constraints on an approximating polynomial on inputs outside the promise domain.

We require the following relation between approximate degree of k -distinctness and a related Boolean function; this relationship follows from [10, Proposition 21 and Corollary 26].

▷ **Claim 5** ([10]). Let $N, R \in \mathbb{N}$ and $2 \leq k \leq N$ be integer. Then for any $\epsilon > 0$,

$$\widetilde{\deg}_\epsilon(\text{DIST}_{N,R+N}^k) = \tilde{\Omega}(\widetilde{\deg}_\epsilon(\text{OR}_R \circ \text{THR}_N^k)^{\leq N}). \quad (7)$$

We also require the following error reduction theorem for approximate degree.

► **Lemma 6** ([9]). Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be any (possibly partial) Boolean function and let $0 < \epsilon < 1$. Then, $\widetilde{\deg}_\epsilon(f) = \deg(f) \cdot O(\log(1/\epsilon))$.¹³

► **Definition 7** (Correlation). Consider any function $f : \{-1, 1\}^n \rightarrow \mathbb{R}$ and $\psi : \{-1, 1\}^n \rightarrow \mathbb{R}$. Define the correlation between f and ψ to be $\langle f, \psi \rangle = \sum_{x \in \{-1, 1\}^n} f(x)\psi(x)$.

► **Definition 8** (Pure high degree). For $\phi : \{-1, 1\}^n \rightarrow \mathbb{R}$, we say that the pure high degree of ϕ , which we denote by $\text{phd}(\phi)$, is d if $d \geq 0$ is the largest integer for which $\langle \phi, p \rangle = 0$ for any polynomial $p : \{-1, 1\}^n \rightarrow \mathbb{R}$ of degree strictly less than d .

By linear programming duality, we have the following standard equivalence between lower bounds on approximate degree and existence of “dual polynomials”. See, for example, [10].

► **Lemma 9**. Let $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ be any function. For any integer $0 \leq j \leq n$, we have $\widetilde{\deg}_\epsilon(f^{\leq j}) \geq d$ if and only if there exists a “dual polynomial” $\phi : \{-1, 1\}^n \rightarrow \mathbb{R}$ satisfying the following properties: $\phi(x) = 0$ for all $|x| > j$, $\langle f, \phi \rangle > \epsilon$, $\sum_{x \in \{-1, 1\}^n} |\phi(x)| = 1$, and $\text{phd}(\phi) \geq d$. We say that ϕ is a dual polynomial witnessing the fact that $\widetilde{\deg}_\epsilon(f^{\leq j}) \geq d$. For brevity, when ϵ and d are clear from context, we say that ϕ is a dual polynomial for $f^{\leq j}$.

Špalek [28] exhibited an explicit dual witness for OR (the existence of a dual witness for OR was already implicit from the work of Nisan and Szegedy [19]).

▷ **Claim 10** (Implicit in [19]). There exists a constant $c \in (0, 1]$ such that for any integer $n \geq 0$, there exists a function $\theta : \{-1, 1\}^n \rightarrow \mathbb{R}$ satisfying $\|\theta\|_1 = 1$, $\text{phd}(\theta) \geq c\sqrt{n}$, and $\langle \theta, \text{OR}_n \rangle \geq 3/5$.

¹³The statement in [9] only deals with total functions. It can be seen that the proof works for partial functions too.

2:10 Improved Approximate Degree Bounds for k -Distinctness

Towards proving approximate degree lower bounds for composed functions, one might hope to combine dual polynomials of the constituent functions in some way to obtain a dual polynomial for the composed function. A series of works [27, 16, 23] introduced the notion of “dual block composition”, which is a powerful method of combining dual witnesses.

► **Definition 11** (Dual block composition). *Let $\theta : \{-1, 1\}^n \rightarrow \mathbb{R}$, $\phi : \{-1, 1\}^m \rightarrow \mathbb{R}$ be any functions satisfying $\|\theta\|_1 = \|\phi\|_1 = 1$ and $\text{phd}(\phi) \geq 1$. Let $x = (x_1, \dots, x_n)$ where each $x_i \in \{-1, 1\}^m$. Define the dual block composition of θ and ϕ , denoted $\theta \star \phi$, to be*

$$\theta \star \phi(x) = 2^n \theta(\text{sgn}(\phi(x_1)), \dots, \text{sgn}(\phi(x_n))) \prod_{i=1}^n |\phi(x_i)|.$$

We now define a simple but important function ϕ that we use in our construction of a dual witness for $\text{DIST}_{N,R}^k$. This function was first used in the context of dual block composition by Bun and Thaler [12].

▷ **Claim 12** ([12]). Define $\phi : \{-1, 1\}^n \rightarrow \mathbb{R}$ as $\phi(x) = -1/2$ if $x = -\mathbf{1}^n$, $\phi(x) = 1/2$ if $x = \mathbf{1}^n$ and $\phi(x) = 0$ otherwise. Then, $\text{phd}(\phi) = 1$.

Next we require a lemma, implicit in a result of Razborov and Sherstov [20] (also see [13] for a formulation similar to the one we require).

► **Lemma 13** (Implicit in [20]). *Let $N \geq R$ be positive integers, $\Delta \in \mathbb{R}^+$, and $\theta : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ be any polynomial such that*

$$\sum_{x \notin (\{-1, 1\}^{RN})^{\leq N}} |\theta(x)| \leq (2NR)^{-\Delta}. \quad (8)$$

For any positive integer $D < \Delta$, there exists a function $\nu : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ such that $\text{phd}(\nu) > D$, $\|\nu\|_1 \leq 1/10$, and $|x| > N \Rightarrow \nu(x) = \theta(x)$.

Lemma 13 helps us convert a dual polynomial θ with little mass on large Hamming weight inputs to a dual polynomial $(\theta - \nu)/\|\theta - \nu\|_1$ with no mass on large Hamming weight inputs without affecting the pure high degree by much.

► **Definition 14.** *For $\eta_i \in [0, 1]$, let $\Pi(\eta_1, \dots, \eta_n)$ be the product distribution on $\{-1, 1\}^n$ where the i th bit of the string equals -1 with probability η_i , and 1 with probability $1 - \eta_i$.*

For any Boolean function $f : \{-1, 1\}^m \rightarrow \{-1, 1\}$ and function $\psi : \{-1, 1\}^m \rightarrow \mathbb{R}$, $\|\psi\|_1 = 1$, let

$$\epsilon_{f,\psi}^+ := \Pr_{\mu_\psi}[f(x)\psi(x) < 0 | \psi(x) > 0], \quad \epsilon_{f,\psi}^- := \Pr_{\mu_\psi}[f(x)\psi(x) < 0 | \psi(x) < 0], \quad \epsilon_{f,\psi} = \epsilon_{f,\psi}^+ + \epsilon_{f,\psi}^-. \quad (9)$$

► **Definition 15.** *For any functions $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ and $\psi : \{-1, 1\}^n \rightarrow \mathbb{R}$, let*

$$E^+(f, \psi) := \{x \in \{-1, 1\}^n : f(x)\psi(x) < 0, \psi(x) > 0\},$$

$$E^-(f, \psi) := \{x \in \{-1, 1\}^n : f(x)\psi(x) < 0, \psi(x) < 0\}.$$

We define the false positive error between f and ψ to be $\delta_{f,\psi}^+ := \sum_{x \in E^+(f,\psi)} |\psi(x)|$ and false negative error to be $\delta_{f,\psi}^- := \sum_{x \in E^-(f,\psi)} |\psi(x)|$.

Given any function $f : \{-1, 1\}^m \rightarrow \{-1, 1\}$ and $\psi : \{-1, 1\}^m \rightarrow \mathbb{R}$, $\|\psi\|_1 = 1$, let $\epsilon^+ = \epsilon_{f,\psi}^+$ and $\epsilon^- = \epsilon_{f,\psi}^-$ as defined in Equation (9). Define the function $\alpha_{f,\psi} : \{-1, 1\}^m \rightarrow \mathbb{R}$ as

$$\alpha_{f,\psi}(x) := \begin{cases} 1 =: a^+ & \text{if } \psi(x)f(x) > 0, \psi(x) > 0 \\ \frac{1-2\epsilon^+ + \epsilon^-}{1-\epsilon^-} =: a^- & \text{if } \psi(x)f(x) > 0, \psi(x) < 0 \\ -1 & \text{if } \psi(x)f(x) < 0, \psi(x) > 0 \\ 1 & \text{if } \psi(x)f(x) < 0, \psi(x) < 0. \end{cases} \quad (10)$$

For the remaining sections, for $z_i \in \{-1, 1\}$, $a^{z_i} = a^+$ if $z_i = 1$, and $a^{z_i} = a^-$ if $z_i = -1$.

► **Lemma 16** ([22, Lemma 3.1]). *For any $\tau_1, \dots, \tau_n \in [0, 1]$, define $\nu = \Pi(\tau_1, \dots, \tau_n)$ and $\tau = \max\{\tau_1, \dots, \tau_n\}$. For any $\eta = 0, 1, \dots, n-1$, let $p_\eta : [-1, 1]^n \rightarrow \mathbb{R}$ be the unique degree- η multilinear polynomial that satisfies*

$$p_\eta(z) = (-1)^\eta \prod_{i=1}^{\eta} (|z| - i), \forall z \in \{-1, 1\}^n. \quad (11)$$

Then,

$$p_\eta(\mathbf{1}^n) = \eta!, \quad (12)$$

$$\|\hat{p}\|_1 \leq \eta! \binom{n+\eta}{\eta}, \quad (13)$$

$$\mathbb{E}_\nu[|p_\eta(z)|] \leq p_\eta(\mathbf{1}^n) \nu(\mathbf{1}^n) (1 + A), \quad \text{where } A := \binom{n}{\eta+1} \frac{\tau^{\eta+1}}{(1-\tau)^n}. \quad (14)$$

Furthermore, $p_\eta(z) \geq 0$ for all $z \in \{-1, 1\}^n$ provided that η is even.

3 Detailed Outline of Proof of Main Theorem

Our main theorem is as follows.

► **Theorem 17.** *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2} R)$,*

$$\widetilde{\deg}(\text{DIST}_{N,R+N}^k) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \quad (15)$$

Ambainis [2] showed that the approximate degree¹⁴ of functions that are symmetric (both with respect to range elements and with respect to domain elements) is the same for all range sizes greater than or equal to N . As a corollary, we obtain the following.

► **Corollary 18.** *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2} R)$,*

$$\widetilde{\deg}(\text{DIST}_{N,N}^k) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \quad (16)$$

¹⁴There are several different conventions used in the literature when defining the domain of functions such as k -distinctness. The convention used by Ambainis [2] considers the input to be specified by $N \cdot R$ variables $y_{1,1}, \dots, y_{N,R}$, where $y_{i,j} = -1$ if and only if the i th list item in the input equals range element j (i.e., it is promised that for each i , $y_{i,j} = -1$ for exactly one j). We use the convention that the input is specified by $N \lceil \log_2 R \rceil$ bits. It is well-known (and not hard to show) that conversion between the two conventions affects approximate degree by at most a factor of $\lceil \log_2 R \rceil$.

To prove Theorem 17, Claim 5 implies that it suffices to prove a lower bound on $\widetilde{\deg}(\text{OR}_R \circ \text{THR}_N^k)^{\leq N}$.

► **Theorem 19.** *For $R \in \mathbb{N}$ sufficiently large, $2 \leq k \leq \frac{\log R}{4}$, and some $N = \Theta(k^{k/2}R)$,*

$$\widetilde{\deg}((\text{OR}_R \circ \text{THR}_N^k)^{\leq N}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \quad (17)$$

Note that the theorems above continue to yield non-trivial lower bounds for some values of $k = \omega(1)$. However for ease of exposition, we assume throughout this section that $k \geq 2$ is an arbitrary but fixed constant.

Towards proving Theorem 19, we construct a dual witness Γ satisfying the following four conditions.

- **Normalization:** $\|\Gamma\|_1 = 1$,
- **Pure high degree:** There exists a $D = \tilde{\Omega}\left(R^{\frac{3}{4} - \frac{1}{4k}}\right)$ such that for every polynomial $p : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ of degree less than D , we have $\langle p, \Gamma \rangle = 0$,
- **Correlation:** $\langle \Gamma, (\text{OR}_R \circ \text{THR}_N^k) \rangle > 1/3$,
- **Exponentially little mass on inputs of large Hamming weight:**

$$\sum_{x \notin \{-1, 1\}^{RN} : |x| \leq N} |\Gamma(x)| \leq (2NR)^{-\tilde{\Omega}\left(R^{\frac{3}{4} - \frac{1}{4k}}\right)}.$$

Next, Lemma 13 implies existence of a function ν that equals Γ on $x \notin \{-1, 1\}^{RN} : |x| \leq N$, has pure high degree $\tilde{\Omega}\left(R^{\frac{3}{4} - \frac{1}{4k}}\right)$, and $\|\nu\|_1 \leq 1/10$. The function $\mathcal{W} : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ defined by $\mathcal{W}(x) := \frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1}$ places no mass on inputs of Hamming weight larger than N and satisfies $\|\mathcal{W}\|_1 = 1$, $\langle \mathcal{W}, (\text{OR}_R \circ \text{THR}_N^k) \rangle > 7/33$, and $\text{phd}(\mathcal{W}) = \tilde{\Omega}\left(R^{\frac{3}{4} - \frac{1}{4k}}\right)$. Theorem 19 then follows by Lemma 9 and Lemma 6.

In the next subsection we provide a sketch of how we construct such a dual witness Γ and where our approach differs from [11].

3.1 Our Construction of Γ

Our construction of Γ is based on three dual witnesses θ, ϕ and ψ . The function θ is constructed as in Claim 10 with $n = R/4^k$. The function ϕ is defined on 4^k inputs, and is defined as in Claim 12. Our ψ is a fairly straightforward modification of [10, Proposition 55], that has a larger pure high degree, at the cost of a worse false positive error. A little more formally, our functions θ, ϕ, ψ have ℓ_1 -norm equal to 1, and additionally ψ satisfies the properties described in the following claim, with $T = \sqrt{R}$.

► **Claim 20 (Modification of [10, Proposition 55]).** Let $k, T, N \in \mathbb{N}$ with $2 \leq k \leq T \leq N$, and let ω_T be as constructed in Claim 27, with constants c_1, c_2 . Define¹⁵ $\psi : \{-1, 1\}^N \rightarrow \mathbb{R}$ by $\psi(x) = \omega_T(|x|)/\binom{N}{|x|}$ for $x \in \{-1, 1\}^N : |x| \leq T$ and $\psi(x) = 0$ otherwise. Then

$$\delta_{\text{THR}_N^k, \psi}^+ \leq \frac{1}{48 \cdot 4^k \sqrt{N} \log N} \quad (18)$$

$$\delta_{\text{THR}_N^k, \psi}^- \leq \frac{1}{2} - \frac{2}{4^k} \quad (19)$$

¹⁵ Note that we suppress the dependence of ψ on T for convenience.

$$\|\psi\|_1 = 1 \quad (20)$$

For any polynomial $p: \{-1, 1\}^N \rightarrow \mathbb{R}$,

$$\deg(p) < c_1 \sqrt{4^{-k} k^{-1} T N^{-1/(2k)} \log^{-1} N} \implies \langle \psi, p \rangle = 0 \quad (21)$$

$$\text{For all } t \in [n], \quad \sum_{|x|=t} |\psi(x)| \leq \frac{(2k)^k \exp\left(-c_2 t / \sqrt{4^k k T N^{1/(2k)} \log N}\right)}{t^2}. \quad (22)$$

The false positive error between THR_N^k and ψ is $\tilde{O}(1/\sqrt{N})$ (as compared to $O(1/N)$ in [11]). The pure high degree of ψ is $\tilde{\Omega}(R^{1/4} N^{-1/(4k)})$ (as compared to $\tilde{\Omega}(R^{1/4} N^{-1/(2k)})$ in [11]). ψ satisfies a “weak decay condition”, viz. $\sum_{|x|=t} |\psi(x)| \leq \sigma \exp(-\beta t)/t^2$ for some constant σ (for general k , the value of σ only depends on k), and $\beta = \tilde{\Omega}(R^{1/4} N^{1/(4k)})$ (as compared to $\beta = \tilde{\Omega}(R^{1/4} N^{1/(2k)})$ in [11]).

If we were to define $\Gamma = \theta \star \phi \star \psi$, all the analyses from [11] would work, except for the correlation analysis, which fails. To fix this, our main technical contribution is to not use dual block composition, but rather a variant of it inspired by a result of Sherstov [22]. Our function Γ takes the form $\Gamma = \theta \bullet (\phi \star \psi)$, where \bullet denotes our variant of dual block composition. In a little more detail, $\Gamma(x_1, \dots, x_{R/4^k})$ equals $\theta \bullet (\phi \star \psi)(x)$, which equals

$$\frac{1}{p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+)} \cdot (\theta \star (\phi \star \psi))(x_1, \dots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \dots, \alpha(x_{R/4^k})),$$

for $\epsilon^+ = \epsilon_{\phi \star \psi, \text{OR}_{4^k} \circ \text{THR}_N^k}^+$, $\epsilon^- = \epsilon_{\phi \star \psi, \text{OR}_{4^k} \circ \text{THR}_N^k}^-$, η is a parameter that we set later, p_η is defined as in Lemma 16, and α in a function whose definition we elaborate on later in this section.

We first give a very high-level idea of how we prove the required properties of Γ , and then elaborate on the definitions of η, p_η and α .

- **Normalization:** Following along similar lines as [22, Claim 6.2], we prove that $\|\Gamma\|_1 = 1$ by modifying the proof that dual block composition preserves ℓ_1 -norm, crucially exploiting properties of p_η and α (see Claim 33).
- **Pure high degree:** Using our definition of p_η , and α , one can show (Claim 34) that the pure high degree of $\theta \bullet (\phi \star \psi)$ is at least $(\text{phd}(\theta) - \eta)\text{phd}(\phi \star \psi)$. The value of η is chosen to be $\text{phd}(\theta)/2$ so that this quantity is the same order of magnitude as $\text{phd}(\theta)\text{phd}(\phi \star \psi) = \text{phd}(\theta)\text{phd}(\psi)$, which is $\tilde{\Omega}(R^{3/4} N^{-1/(4k)})$.
- **Exponentially little mass on inputs of large Hamming weight:** Since ψ satisfies $\sum_{|x|=t} |\psi(x)| \leq \sigma \exp(-\beta t)/t^2$ for some constant σ and $\beta = \tilde{\Omega}(R^{1/4} N^{1/(4k)})$, Claim 29 implies that $\theta \star (\phi \star \psi) = (\theta \star \phi) \star \psi$ places exponentially small (in $R^{\frac{3}{4} - \frac{1}{4k}}$) mass on inputs in $\{-1, 1\}^{RN}$ of Hamming weight larger than N . By the definition of Γ , it suffices to show that the maximum absolute value of $\frac{p_\eta(\alpha(x_1), \dots, \alpha(x_{R/4^k}))}{p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+)}$ is at most exponentially large in $R^{\frac{3}{4} - \frac{1}{4k}}$, for which we require Claim 30.
- **Correlation:** Conceptually, the function $p_\eta: \{-1, 1\}^{R/4^k} \rightarrow \mathbb{R}$ can be viewed as one that “corrects” $\theta \star (\phi \star \psi)$: it “counts” the number of false positives fed to it by $\phi \star \psi$, and changes the output of $\theta \star (\phi \star \psi)$ to 0 on inputs where this number is any integer between 1 and η . The function $\alpha: \{-1, 1\}^N \rightarrow \mathbb{R}$ acts as the function that, in a sense, indicates whether or not $\phi \star \psi$ is making a *false positive* error.
- **Detecting errors:** The function α takes three possible output values: it outputs -1 for $x \in E^+(\text{OR}_{4^k} \circ \text{THR}_N^k, \phi \star \psi)$ and outputs either 1 or a value very close to 1 for $x \notin E^+(\text{OR}_{4^k} \circ \text{THR}_N^k, \phi \star \psi)$. This definition of α is our biggest departure from

Sherstov's construction in [22]; Sherstov defined α to output -1 for *both* false-positive and false-negative errors, whereas our α only outputs -1 for false-positive errors.

- **Zeroing out errors:** Define the function p_η to be (the unique multilinear extension of) the function that outputs 0 if its input has Hamming weight between 1 and η . Recall that our construction considers the dual witness

$$\frac{1}{p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+)} \cdot (\theta \star (\phi \star \psi))(x_1, \dots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \dots, \alpha(x_{R/4^k})),$$

and the purpose of multiplying $\theta \star (\phi \star \psi)$ by p_η is for p_η to zero out most inputs in which one or more false-positive errors are being fed by $\phi \star \psi$ into θ (see Definition 11). Unfortunately, p_η is nonzero on inputs of Hamming weight more than η . Hence, in terms of the correlation analysis, a key question that must be addressed is: what fraction of the ℓ_1 -mass of $\theta \star (\phi \star \psi)$ is placed on inputs where more than η copies of $\phi \star \psi$ make a false-positive error? We need this fraction to be very small, because multiplying by p_η fails to zero out such inputs.

Note that under the distribution defined by $|\phi \star \psi|$, the *expected* number of false positive errors fed into θ is $(R/4^k) \cdot \epsilon^+$. Since we have set $\eta = O(\sqrt{R/(4 \cdot 4^k)})$, it suffices to have $\epsilon^+ \ll 1/(c\eta)$ for some large enough constant c to conclude that with high probability (over the distribution $|\phi \star \psi|$), the number of false positive errors fed into θ is at most a small constant times η . It turns out that this value of ϵ^+ is indeed attained by $\phi \star \psi$, since the false positive error between THR_N^k and ψ was set to be $\tilde{O}(1/\sqrt{N}) = \tilde{O}(1/\sqrt{R})$ to begin with. Thus, with high probability, multiplying $\theta \star (\phi \star \psi)$ by p_η successfully zeros out all but an exponentially small fraction of the errors made by $\theta \star (\phi \star \psi)$ that can be attributed to false-positive errors made by $\phi \star \psi$. This intuitive proof outline is formalized in Claim 21, which in turn is a formalization of Equation (1) that holds with the setting of parameters mentioned above.

The key technical lemma that we use for the correlation analysis is the following, and a sketch of its proof is deferred to Appendix B.

▷ **Claim 21.** Let m, n be any positive integers, $\eta < n$ be any even positive integer, and $f : \{-1, 1\}^m \rightarrow \{-1, 1\}$ be any function. Let $\zeta : \{-1, 1\}^n \rightarrow \mathbb{R}$ be such that $\langle \zeta, \text{OR}_n \rangle > \delta$ and $\|\zeta\|_1 = 1$, and $\xi : \{-1, 1\}^m \rightarrow \mathbb{R}$ be any function such that $\|\xi\|_1 = 1$ and $\text{phd}(\xi) \geq 1$. Let $p_\eta : \{-1, 1\}^n \rightarrow \mathbb{R}$ be as defined in Lemma 16, let $\alpha = \alpha_{f, \xi} : \{-1, 1\}^m \rightarrow \mathbb{R}$ be as defined in Equation (10), and consider the distribution μ_ξ over $\{-1, 1\}^{nm}$. Let $\epsilon^+ = \epsilon_{f, \xi}^+$, $\epsilon^- = \epsilon_{f, \xi}^-$, $\epsilon = \epsilon^+ + \epsilon^-$, and $A = \binom{n}{\eta+1} \frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^n}$. If $A < 1$, then,

$$\langle \text{OR} \circ f, (\zeta \star \xi)(p_\eta \circ \alpha) \rangle \geq p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) \cdot \left(\delta - \left(2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} (1 - A) \right) \right). \quad (23)$$

4 Proof of Theorem 19

Due to space constraints, we omit some proofs henceforth. The reader is referred to the full version for complete proofs.

Towards proving Theorem 19, it suffices to exhibit a dual polynomial (see Lemma 9) that has ℓ_1 -norm 1, sufficiently large pure high degree, good correlation with $(\text{OR}_R \circ \text{THR}_N^k)^{\leq N}$, and places no mass outside $(\{-1, 1\}^{RN})^{\leq N}$. We first define a function Γ (Definition 23) that satisfies the first three properties above, and additionally satisfies a strong decay condition as we described in Section 3.1. In Section 4.1 we use Γ to construct a dual polynomial \mathcal{W} , via Lemma 13, satisfying all the requisite properties. We now set several key variables.

- Let R be sufficiently large and fix $k \leq (\log R)/4$. Set $T = \sqrt{R}$, $\eta = \left(\frac{c}{2}\sqrt{\frac{R}{4^k}}\right) - 1$ where $c \in (0, 1]$ is the constant from Claim 10 (assume without loss of generality that η is even), $\sigma = (2k)^k$, $c_1, c_2 \in (0, 1]$ are constants fixed in the next bullet point, $\beta = \frac{c_2}{\sqrt{4^k k T N^{1/(2k)} \log N}}$, $\Delta = \frac{\beta \sqrt{\sigma} R}{4 \ln^2 R} = \frac{c_2 R}{4 \ln^2 R} \sqrt{\frac{(2k)^k}{4^k k T N^{1/(2k)} \log N}}$, $N = \lceil 20\sqrt{\sigma} R \rceil$.
- Let $\omega_T : [T] \cup \{0\} \rightarrow \mathbb{R}$ be a function that satisfies the conditions in Claim 27 and let c_1, c_2 be the constants for which the claim holds. Let $\psi : \{-1, 1\}^N \rightarrow \mathbb{R}$ be defined by $\psi(x) = \omega_T(|x|)/\binom{N}{|x|}$ if $|x| \leq T$, and 0 otherwise so that ψ satisfies the conditions in Claim 20.
- Let $\theta : \{-1, 1\}^{R/4^k} \rightarrow \mathbb{R}$ be any function satisfying the conditions in Claim 10 for $n = R/4^k$ (note that $R/4^k > 0$ since $k < (\log R)/2$), and let $\phi : \{-1, 1\}^{4^k} \rightarrow \mathbb{R}$ be the function defined in Claim 12 with $n = 4^k$.
- Let $p_\eta : \{-1, 1\}^{R/4^k} \rightarrow \mathbb{R}$ be as defined in Lemma 16 and $\alpha := \alpha_{\phi \star \psi, \text{OR}_{4^k} \circ \text{THR}_N^k} : \{-1, 1\}^{4^k N} \rightarrow \mathbb{R}$ be as defined in Equation (10).
- Let $\epsilon^+ := \epsilon_{\text{OR}_{4^k} \circ \text{THR}_N^k, \phi \star \psi}^+$, $\epsilon^- := \epsilon_{\phi \star \psi}^-$, and $\epsilon := \epsilon^+ + \epsilon^-$.

We first show that the function $\phi \star \psi$ has large correlation with $\text{OR}_{4^k} \circ \text{THR}_N^k$, via an analysis that is essentially the same as in [10, Proposition 55].

▷ Claim 22.

$$\epsilon_{\text{OR}_{4^k} \circ \text{THR}_N^k, \phi \star \psi}^+ \leq \frac{1}{24\sqrt{R} \log R}, \quad \epsilon_{\text{OR}_{4^k} \circ \text{THR}_N^k, \phi \star \psi}^- \leq e^{-4}.$$

We next define the function Γ .

▶ **Definition 23.** Let $\Gamma : \{-1, 1\}^{NR} \rightarrow \mathbb{R}$ be defined by

$$\Gamma(x_1, \dots, x_{R/4^k}) := \frac{(\theta \star (\phi \star \psi))(x_1, \dots, x_{R/4^k}) \cdot p_\eta(\alpha(x_1), \dots, \alpha(x_{R/4^k}))}{p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+)}, \quad (24)$$

where each $x_i \in \{-1, 1\}^{4^k N}$.

▷ Claim 24.

$$\|\Gamma\|_1 = 1, \quad (25)$$

$$\text{phd}(\Gamma) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{3/4-1/(4k)}\right), \quad (26)$$

$$\langle \Gamma, (\text{OR}_R \circ \text{THR}_N^k) \rangle > 1/3 \quad (27)$$

$$\sum_{x \notin (\{-1, 1\}^{RN})^{\leq N}} |\Gamma(x)| \leq (2NR)^{-2(\Delta - \sqrt{R})}. \quad (28)$$

Sketch of Proof of Claim 24

We require certain properties of dual block composition, and of the functions p_η and α , which are listed in Appendix A and Appendix B, respectively.

- The fact that $\|\Gamma\|_1 = 1$ follows from the definition of Γ and Claim 33.
- By the definition of Γ , we have $\text{phd}(\Gamma) = \text{phd}((\theta \star (\phi \star \psi))(p_\eta \circ \alpha))$. By Claim 34, this is at least $(\text{phd}(\theta) - \eta) \cdot (\text{phd}(\phi \star \psi))$. Next, using the facts that $\text{phd}(\psi) = 1$ (Claim 12), multiplicativity of pure high degree under dual block composition (Equation (45)), and our choices of parameters, it can be shown that $\text{phd}(\Gamma) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\sqrt{\log R}} \cdot R^{3/4-1/(4k)}\right)$.

- Recall from our choice of parameters and Claim 22 that $\epsilon^+ \leq \frac{1}{24\sqrt{R} \log R}$ and $\epsilon^- \leq e^{-4}$. Define $A = \binom{R/4^k}{\eta+1} \frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^{R/4^k}}$. The above upper bounds on ϵ^+ and ϵ^- , and standard computations reveal that $A < 1/16$. Hence the conditions of Claim 21 are satisfied with the parameters fixed in the beginning of this section. Using Claim 21 with $\delta > 3/5$ and the above upper bounds on ϵ^+ and ϵ^- , we are able to show that $\langle \Gamma, (\text{OR}_R \circ \text{THR}_N^k) \rangle > 1/3$.
- We first show, using Lemma 16 and Lemma 26, that $p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) \geq (1 - \epsilon^+)^{R/4^k} \eta!$. Standard computations reveal that, for our choice of parameters, this quantity is at least 1. Hence, it suffices to show that $\sum_{x \notin \{-1, 1\}^{RN} \leq N} |(\theta \star (\phi \star \psi)) \cdot (p_\eta \circ \alpha)(x)| \leq (2NR)^{-2(\Delta - \sqrt{R})}$. Next we observe that, using Claim 29 with $\Phi = \theta \star \phi$ and associativity of dual block composition (Equation (46)), that $\sum_{x \notin \{-1, 1\}^{RN} \leq N} |((\theta \star \phi) \star \psi)(x)| \leq (2NR)^{-2\Delta}$. Since $\alpha(y) \in [-1, 1]$ for all $y \in [-1, 1]^{4^k N}$ (Equation (10)), it suffices to show a suitable bound on $\max_{y \in [-1, 1]^{R/4^k}} |p_\eta(y)|$, which we are able to do using Claim 30.

4.1 Final Dual Polynomial

We now prove Theorem 19.

Proof of Theorem 19. We exhibit a function $\mathcal{W} : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ satisfying

$$\mathcal{W}(x) = 0, \forall x \notin (\{-1, 1\}^{RN})^{\leq N}, \quad (29)$$

$$\|\mathcal{W}\|_1 = 1 \quad (30)$$

$$\langle \mathcal{W}, (\text{OR}_R \circ \text{THR}_N^k) \rangle > 7/33, \quad (31)$$

$$\text{phd}(\mathcal{W}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right). \quad (32)$$

The theorem then follows by Lemma 9 and Lemma 6. Towards the construction of such a \mathcal{W} , first note that by Equation (28) and Lemma 13 there exists a function $\nu : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ that satisfies the following properties.

$$|x| > N \Rightarrow \nu(x) = \Gamma(x), \quad (33)$$

$$\text{phd}(\nu) \geq 2(\Delta - \sqrt{R}) - 1, \quad (34)$$

$$\|\nu\|_1 \leq 1/10. \quad (35)$$

Define $\mathcal{W} : \{-1, 1\}^{RN} \rightarrow \mathbb{R}$ by

$$\mathcal{W}(x) := \frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1}. \quad (36)$$

Clearly Equation (29) and Equation (30) are satisfied. We show in Appendix C that the function \mathcal{W} also satisfies Equation (31), and Equation (32). ◀

References

- 1 Scott Aaronson and Yaoyun Shi. Quantum lower bounds for the collision and the element distinctness problems. *J. ACM*, 51(4):595–605, 2004. doi:10.1145/1008731.1008735.
- 2 Andris Ambainis. Polynomial degree and lower bounds in quantum complexity: Collision and element distinctness with small range. *Theory of Computing*, 1(1):37–46, 2005. doi:10.4086/toc.2005.v001a003.
- 3 Andris Ambainis. Polynomial degree vs. quantum query complexity. *Journal of Computer and System Sciences*, 72(2):220–238, 2006.

- 4 Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007. doi:10.1137/S0097539705447311.
- 5 Srinivasan Arunachalam, Jop Briët, and Carlos Palazuelos. Quantum query algorithms are completely bounded forms. *SIAM Journal on Computing*, 48(3):903–925, 2019.
- 6 Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald De Wolf. Quantum lower bounds by polynomials. *Journal of the ACM (JACM)*, 48(4):778–797, 2001.
- 7 Paul Beame and Widad Machmouchi. The quantum query complexity of AC^0 . *Quantum Information & Computation*, 12(7-8):670–676, 2012.
- 8 Aleksandrs Belovs. Learning-graph-based quantum algorithm for k-distinctness. In *53rd Annual IEEE Symposium on Foundations of Computer Science, FOCS 2012, New Brunswick, NJ, USA, October 20-23, 2012*, pages 207–216, 2012. doi:10.1109/FOCS.2012.18.
- 9 Harry Buhrman, Ilan Newman, Hein Röhrig, and Ronald de Wolf. Robust polynomials and quantum algorithms. *Theory Comput. Syst.*, 40(4):379–395, 2007.
- 10 Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *CoRR*, abs/1710.09079, version 3, 2017.
- 11 Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: tight quantum query bounds via dual polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2018, Los Angeles, CA, USA, June 25-29, 2018*, pages 297–310, 2018.
- 12 Mark Bun and Justin Thaler. Hardness amplification and the approximate degree of constant-depth circuits. In *Automata, Languages, and Programming - 42nd International Colloquium, ICALP 2015, Kyoto, Japan, July 6-10, 2015, Proceedings, Part I*, pages 268–280, 2015.
- 13 Mark Bun and Justin Thaler. A nearly optimal lower bound on the approximate degree of AC^0 . In *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2017, Berkeley, CA, USA, October 15-17, 2017*, pages 1–12, 2017.
- 14 Mark Bun and Justin Thaler. The large-error approximate degree of ac^0 . In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*. Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik, 2019.
- 15 Samuel Kutin. Quantum lower bound for the collision problem with small range. *Theory of Computing*, 1(1):29–36, 2005. doi:10.4086/toc.2005.v001a002.
- 16 Troy Lee. A note on the sign degree of formulas. *CoRR*, abs/0909.4607, 2009.
- 17 Tongyang Li and Xiaodi Wu. Quantum query complexity of entropy estimation. *IEEE Trans. Inf. Theory*, 65(5):2899–2921, 2019.
- 18 Qipeng Liu and Mark Zhandry. On finding quantum multi-collisions. In *Advances in Cryptology - EUROCRYPT 2019 - 38th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Darmstadt, Germany, May 19-23, 2019, Proceedings, Part III*, pages 189–218, 2019. doi:10.1007/978-3-030-17659-4_7.
- 19 Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational Complexity*, 4:301–313, 1994.
- 20 Alexander A. Razborov and Alexander A. Sherstov. The sign-rank of AC^0 . *SIAM J. Comput.*, 39(5):1833–1855, 2010. doi:10.1137/080744037.
- 21 Alexander A. Sherstov. The pattern matrix method. *SIAM J. Comput.*, 40(6):1969–2000, 2011.
- 22 Alexander A. Sherstov. Strong direct product theorems for quantum communication and query complexity. *SIAM J. Comput.*, 41(5):1122–1165, 2012.
- 23 Alexander A Sherstov. The intersection of two halfspaces has high threshold degree. *SIAM Journal on Computing*, 42(6):2329–2374, 2013.
- 24 Alexander A Sherstov. Algorithmic polynomials. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, pages 311–324, 2018.
- 25 Alexander A. Sherstov. The power of asymmetry in constant-depth circuits. *SIAM J. Comput.*, 47(6):2362–2434, 2018.

- 26 Alexander A Sherstov and Justin Thaler. Vanishing-error approximate degree and QMA complexity. *arXiv preprint arXiv:1909.07498*, 2019.
- 27 Yaoyun Shi and Yufan Zhu. Quantum communication complexity of block-composed functions. *Quantum Information & Computation*, 9(5):444–460, 2009.
- 28 Robert Špalek. A dual polynomial for OR. *CoRR*, abs/0803.4516, 2008.

A Preliminaries

► **Definition 25.** For any integer $n > 0$, any function $\psi : \{-1, 1\}^m \rightarrow \mathbb{R}$ such that $\|\psi\|_1 = 1$, and any $w \in \{-1, 1\}$, let μ_w be the probability distribution μ_ψ conditioned on the event that $\text{sgn}(\psi(x)) = w$. For any $z \in \{-1, 1\}^n$, let μ_z denote the probability distribution $(\mu_\psi)^{\otimes n}$ conditioned on the event that $\text{sgn}(\psi(x_i)) = z_i$ for all $i \in [n]$.

We omit the dependence of μ_z on ψ since ψ will typically be clear from context. Note that μ_z as defined above is a product distribution given by

$$\mu_z(x_1, \dots, x_n) = \prod_{i=1}^n \mu_{z_i}(x_i). \quad (37)$$

► **Lemma 26.** Let n be any positive integer, $p : \{-1, 1\}^n \rightarrow \mathbb{R}$ be a multilinear polynomial, and $\eta_1, \dots, \eta_n \in [0, 1]$. For $x = (x_1, \dots, x_n)$ drawn from the product distribution $\Pi(\eta_1, \dots, \eta_n)$ defined in Definition 14, we have

$$\mathbb{E}_{\Pi(\eta_1, \dots, \eta_n)}[p(x_1, \dots, x_n)] = p(1 - 2\eta_1, \dots, 1 - 2\eta_n). \quad (38)$$

A.1 Dual Polynomials and Dual Block Composition

Bun et al. [11] exhibited a dual witness for the approximate degree of the k -threshold function. Their dual witness additionally satisfies a decay condition, meaning that it places very little mass on inputs of large Hamming weight. The following claim is a mild modification of [10, Proposition 54].

► **Claim 27 (Modification of [10, Proposition 54]).** Let $k, T, N \in \mathbb{N}$ with $2 \leq k \leq T$. There exist constants $c_1, c_2 \in (0, 1]$ and a function $\omega_T : [T] \cup \{0\} \rightarrow \mathbb{R}$ such that all of the following hold.

$$\sum_{\omega_T(t) > 0, t \geq k} |\omega_T(t)| \leq \frac{1}{48 \cdot 4^k \sqrt{N} \log N}. \quad (39)$$

$$\sum_{\omega_T(t) < 0, t < k} |\omega_T(t)| \leq \left(\frac{1}{2} - \frac{2}{4^k} \right). \quad (40)$$

$$\|\omega_T\|_1 := \sum_{t=0}^T |\omega_T(t)| = 1. \quad (41)$$

For all polynomials $q : \mathbb{R} \rightarrow \mathbb{R}$,

$$\deg(q) < c_1 \sqrt{4^{-k} k^{-1} T N^{-1/(2k)} \log^{-1} N} \implies \sum_{t=0}^T \omega_T(t) q(t) = 0. \quad (42)$$

$$\text{For all } t \in [T], |\omega_T(t)| \leq \frac{\sigma \exp(-\beta t)}{t^2} \quad \text{for } \sigma = (2k)^k, \quad \beta = c_2 / \sqrt{4^k k T N^{1/(2k)} \log N}. \quad (43)$$

Sherstov [23] showed that dual block composition (see Definition 11) preserves ℓ_1 -norm and that pure high degree is multiplicative (also see [16]). Bun and Thaler [13] observed that dual block composition is associative.

► **Lemma 28.** *Let $\phi : \{-1, 1\}^{m_\phi} \rightarrow \mathbb{R}, \theta : \{-1, 1\}^{m_\theta} \rightarrow \mathbb{R}$ be any functions. Then, **Preservation of ℓ_1 -norm:** If $\|\theta\|_1 = 1, \|\phi\|_1 = 1$ and $\langle \phi, 1 \rangle = 0$, then*

$$\|\theta \star \phi\|_1 = 1. \quad (44)$$

Multiplicativity of pure high degree:

$$\text{phd}(\theta) > D, \text{phd}(\phi) > d \implies \text{phd}(\theta \star \phi) > Dd. \quad (45)$$

Associativity: For every $\psi : \{-1, 1\}^{m_\psi} \rightarrow \mathbb{R}$, we have

$$(\phi \star \theta) \star \psi = \phi \star (\theta \star \psi). \quad (46)$$

It was shown in [10] that for any dual polynomial Φ , and ψ as constructed in Claim 20, the dual block composed function $\Phi \star \psi$ satisfies a “strong dual decay” condition.¹⁶

► **Claim 29** ([10, Proposition 31]). Let R be sufficiently large and $k \leq T \leq R$ be any positive integer. Fix $\sigma = (2k)^k$ and let $N = \lceil 20\sqrt{\sigma}R \rceil$. Let $\Phi : \{-1, 1\}^R \rightarrow \mathbb{R}$ be any function with $\|\Phi\|_1 = 1$ and $\psi : \{-1, 1\}^N \rightarrow \mathbb{R}$ as defined in Claim 20. Then

$$\sum_{x \notin \{-1, 1\}^{RN} \leq N} |(\Phi \star \psi)(x)| \leq (2NR)^{-2\Delta} \quad (47)$$

for some $\Delta \geq \frac{\beta\sqrt{\sigma}R}{4\ln^2 R}$ for $\beta = c_2/\sqrt{4^k k T N^{1/(2k)} \log N}$.

B Properties of Auxiliary Functions

It is easy to show that any multilinear polynomial $p : \mathbb{R}^n \rightarrow \mathbb{R}$ satisfies $\max_{y \in [-1, 1]^n} |p(y)| \leq \|\hat{p}\|_1$. When applied to the function in Lemma 16, we obtain

► **Claim 30.** For p_η defined as in Lemma 16, $\max_{y \in [-1, 1]^n} |p_\eta(y)| \leq \eta! \binom{n+\eta}{\eta}$.

We now state the setting for our next few claims.

Assumptions for Claim 31, Claim 32, Claim 33: Let m, n be any positive integers, $\eta < n$ be any even positive integer, and $f : \{-1, 1\}^m \rightarrow \{-1, 1\}$ be any function. Let $\zeta : \{-1, 1\}^n \rightarrow \mathbb{R}$ be such that $\langle \zeta, \text{OR}_n \rangle > \delta$ and $\|\zeta\|_1 = 1$, and $\xi : \{-1, 1\}^m \rightarrow \mathbb{R}$ be any function such that $\|\xi\|_1 = 1$ and $\text{phd}(\xi) \geq 1$. Let $p_\eta : \{-1, 1\}^n \rightarrow \mathbb{R}$ be as defined in Lemma 16, let $\alpha = \alpha_{f, \xi} : \{-1, 1\}^m \rightarrow \mathbb{R}$ be as defined in Equation (10), and consider the distribution μ_ξ over $\{-1, 1\}^{nm}$. Let $\epsilon^+ = \epsilon_{f, \xi}^+, \epsilon^- = \epsilon_{f, \xi}^-, \epsilon = \epsilon^+ + \epsilon^-$, and $A = \binom{n}{\eta+1} \frac{(\epsilon^+)^{\eta+1}}{(1-\epsilon^+)^n}$.

► **Claim 31.**

$$\begin{aligned} & \zeta(\mathbf{1}^n) \mathbb{E}_{x \sim \mu_{\mathbf{1}^n}} [p_\eta(\alpha(x_1), \dots, \alpha(x_n)) \text{OR}(f(x_1), \dots, f(x_n))] \\ & \geq p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) (\zeta(\mathbf{1}^n) - |\zeta(\mathbf{1}^n)| 2A). \end{aligned} \quad (48)$$

¹⁶They in fact showed that $\Psi \star \psi$ satisfies this strong decay condition for *any* ψ satisfying a corresponding “weak decay” condition. However for this paper, we only require this statement for ψ as constructed in Claim 20.

▷ Claim 32.

$$\begin{aligned} & \sum_{z \neq \mathbf{1}^n} \zeta(z) \mathbb{E}_{\mu_z} [p_\eta(\alpha(x_1), \dots, \alpha(x_n)) \text{OR}(f(x_1), \dots, f(x_n))] \\ & \geq p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) \left(\sum_{z \neq \mathbf{1}^n} \zeta(z) \text{OR}(z) - \left(2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} (1 - A) \right) \sum_{z \neq \mathbf{1}^n} |\zeta(z)| \right). \end{aligned} \quad (49)$$

Due to space constraints we do not prove Claim 31 and Claim 32 here, and refer the reader to the full version for these proofs. We now prove Claim 21 using Claim 31 and Claim 32.

Proof of Claim 21.

$$\begin{aligned} \langle \text{OR} \circ f, (\zeta \star \xi)(p_\eta \circ \alpha) \rangle &= \sum_{x \in \{-1, 1\}^{mn}} (\text{OR} \circ f)(x) (\zeta \star \xi)(p_\eta \circ \alpha)(x) \\ &= \sum_{x \in \{-1, 1\}^{mn}} \text{OR}(f(x_1), \dots, f(x_n)) \\ & \quad \cdot 2^n \zeta(\text{sgn}(\xi(x_1)), \dots, \text{sgn}(\xi(x_n))) p_\eta(\alpha(x_1), \dots, \alpha(x_n)) \prod_{i=1}^n |\xi(x_i)| \quad \text{by Definition 11} \\ &= \sum_{z \in \{-1, 1\}^n} \zeta(z) \left(\sum_{x: \text{sgn}(\xi(x_i)) = z_i \forall i \in [n]} 2^n p_\eta(\alpha(x_1), \dots, \alpha(x_n)) \right. \\ & \quad \left. \text{OR}(f(x_1), \dots, f(x_n)) \prod_{i=1}^n |\xi(x_i)| \right) \\ &= \sum_{z \in \{-1, 1\}^n} \zeta(z) \mathbb{E}_{\mu_z} [p_\eta(\alpha(x_1), \dots, \alpha(x_n)) \text{OR}(f(x_1), \dots, f(x_n))] \\ & \text{by Definition 25 and } \Pr_{x_i \sim \mu_\xi} [\text{sgn}(x_i) = 1] = \Pr_{x_i \sim \mu_\xi} [\text{sgn}(x_i) = -1] = 1/2 \text{ since } \text{phd}(\xi) \geq 1 \\ & \geq p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) (\zeta(\mathbf{1}^n) \text{OR}(\mathbf{1}^n) - 2|\zeta(\mathbf{1}^n)|A \\ & \quad + \sum_{z \neq \mathbf{1}^n} \zeta(z) \text{OR}(z) - \left(2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} (1 - A) \right) \sum_{z \neq \mathbf{1}^n} |\zeta(z)|) \\ & \quad \text{by Claim 31, 32 and } \text{OR}(\mathbf{1}^n) = 1 \\ & \geq p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) \left(\delta - \max \left\{ 2A, 2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} (1 - A) \right\} \right) \\ & \quad \text{since } \|\zeta\|_1 = 1 \text{ and } \langle \zeta, \text{OR} \rangle > \delta \\ & \geq p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+) \left(\delta - \left(2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} (1 - A) \right) \right), \\ & \text{where the last inequality holds as } \left(2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} (1 - A) \right) - 2A = (1 - A) \left(2 - 2 \frac{1 - \epsilon}{1 - \epsilon^+} \right) > 0, \\ & \text{since } \frac{1 - \epsilon}{1 - \epsilon^+} < 1, \text{ and } A < 1. \quad \triangleleft \end{aligned}$$

Finally, we require a closed form expression for $\|(\zeta \star \xi)(p_\eta \circ \alpha)\|_1$.

▷ Claim 33.

$$\|(\zeta \star \xi)(p_\eta \circ \alpha)\|_1 = p_\eta(1 - 2\epsilon^+, \dots, 1 - 2\epsilon^+). \quad (50)$$

The proof of the claim follows along the lines as that of [22, Claim 6.2].

▷ **Claim 34.** Let $\Psi : \{-1, 1\}^n \rightarrow \mathbb{R}$, $\Lambda : \{-1, 1\}^m \rightarrow \mathbb{R}$, and $f : \{-1, 1\}^m \rightarrow \mathbb{R}$ be any functions. For any positive integer η , let $\alpha = \alpha_{f, \Lambda} : \{-1, 1\}^m \rightarrow \mathbb{R}$ be as defined in Equation (10), and $p_\eta : \{-1, 1\}^n \rightarrow \mathbb{R}$ defined in Lemma 16. Then

$$\text{phd}((\Psi \star \Lambda) \cdot (p_\eta \circ \alpha)) > (\text{phd}(\Psi) - \eta) \cdot \text{phd}(\Lambda). \quad (51)$$

The proof follows along the same lines as that of [22, Equation (6.7)] and we omit it.

C Main Theorem

Recall from the proof of Theorem 19 in Section 4.1 that it remains to show $\langle \mathcal{W}, (\text{OR}_R \circ \text{THR}_N^k) \rangle > 7/33$ and $\text{phd}(\mathcal{W}) = \Omega\left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4} - \frac{1}{4k}}\right)$.

Remaining proof of Theorem 19. To justify Equation (31), we have

$$\begin{aligned} \langle \mathcal{W}, \text{OR}_R \circ \text{THR}_N^k \rangle &= \frac{1}{\|\Gamma - \nu\|_1} \left(\langle \Gamma, \text{OR}_R \circ \text{THR}_N^k \rangle - \langle \nu, \text{OR}_R \circ \text{THR}_N^k \rangle \right) && \text{by Equation (36)} \\ &\geq \frac{1}{\|\Gamma - \nu\|_1} \left(1/3 - \langle \nu, \text{OR}_R \circ \text{THR}_N^k \rangle \right) && \text{by Claim 24} \\ &\geq \frac{1}{\|\Gamma - \nu\|_1} \{1/3 - \|\nu\|_1\} \\ &\geq \frac{1}{\|\Gamma - \nu\|_1} \frac{7}{30} && \text{by Equation (35)} \\ &\geq \frac{7}{33}. && \text{since } \|\Gamma - \nu\|_1 \leq \frac{11}{10} \text{ by triangle inequality} \end{aligned}$$

We have from Equation (36) that

$$\text{phd}(\mathcal{W}) = \text{phd}\left(\frac{\Gamma(x) - \nu(x)}{\|\Gamma - \nu\|_1}\right) \quad (52)$$

$$= \text{phd}(\Gamma(x) - \nu(x)) \quad (53)$$

$$\geq \min\{\text{phd}(\Gamma), \text{phd}(\nu)\}. \quad (54)$$

From Equation (34) we have

$$\text{phd}(\nu) \geq 2(\Delta - \sqrt{R}) - 1 \quad (55)$$


$$\begin{aligned} &= 2 \left(\frac{c_2 R}{4 \ln^2 R} \sqrt{\frac{(2k)^k}{4^k k T N^{1/(2k)} \log N}} - \sqrt{R} \right) - 1 && \text{substituting the value of } \Delta \\ &\geq 2 \left(\frac{c_2}{4} \cdot \frac{1}{\log^2 R \sqrt{\log N}} \cdot \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{1/2}} \cdot \frac{R^{3/4}}{N^{1/(4k)}} - \sqrt{R} \right) - 1 \\ &\quad \text{taking } T = \sqrt{R} \text{ and } \ln R < \log R \\ &= 2 \left(\frac{c_2}{4} \cdot \frac{1}{\log^2 R \sqrt{k \log R}} \cdot \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{1/2}} \cdot \frac{R^{3/4}}{20^{1/(4k)} 2^{1/8} k^{1/8} R^{1/(4k)}} - \sqrt{R} \right) - 1 \\ &\quad \text{substituting the value of } N \text{ and using } k \log R > \log N \text{ for sufficiently large } R \\ &\geq 2 \left(\frac{c_2}{2^{25/24}} \cdot \frac{1}{\log^2 R \cdot \sqrt{\log R}} \cdot \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{9/8} \cdot 20^{1/(4k)}} \cdot R^{3/4 - 1/(4k)} - \sqrt{R} \right) - 1 \end{aligned} \quad (56)$$

2:22 Improved Approximate Degree Bounds for k -Distinctness

$$\begin{aligned}
&\geq 2 \left(\frac{c_2}{3} \cdot \frac{1}{\log^{5/2} R} \cdot \frac{1}{2^{9/8} \cdot 20^{1/(4k)}} \cdot R^{3/4-1/(4k)} - \sqrt{R} \right) - 1 \\
&\hspace{15em} \text{since } \left(\frac{k}{2}\right)^{k/2} \frac{1}{k^{9/8}} \geq \frac{1}{2^{9/8}} \text{ for all } k \geq 2 \\
&\geq \frac{c_2}{180} \cdot \frac{1}{\log^{5/2} R} \cdot R^{3/4-1/(4k)} - 1 \\
&\hspace{10em} \text{since } \frac{c_2}{3} \cdot \frac{1}{\log^{5/2} R} \cdot R^{3/4-1/(4k)} > 2\sqrt{R} \text{ for } k \geq 2, \text{ for sufficiently large } R \\
&= \Omega \left(\frac{1}{\log^{5/2} R} \cdot R^{3/4-1/(4k)} \right). \tag{57}
\end{aligned}$$

Therefore by Claim 24 and Equation (54), we have $\text{phd}(\mathcal{W}) = \Omega \left(\frac{1}{4^k k^2} \cdot \frac{1}{\log^{5/2} R} \cdot R^{\frac{3}{4}-\frac{1}{4k}} \right)$, justifying Equation (32) and finishing the proof. \blacktriangleleft

Building Trust for Continuous Variable Quantum States

Ulysse Chabaud¹ 

Laboratoire d'Informatique de Paris 6, CNRS, Sorbonne Université,
4 place Jussieu, 75005 Paris, France
ulysse.chabaud@gmail.com

Tom Douce

School of Informatics, University of Edinburgh,
10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom

Frédéric Grosshans 

Laboratoire d'Informatique de Paris 6, CNRS, Sorbonne Université,
4 place Jussieu, 75005 Paris, France
Laboratoire Aimé Cotton, CNRS, Université Paris-Sud, ENS Cachan, Université Paris-Saclay, 91405
Orsay Cedex, France

Elham Kashefi

Laboratoire d'Informatique de Paris 6, CNRS, Sorbonne Université,
4 place Jussieu, 75005 Paris, France
School of Informatics, University of Edinburgh,
10 Crichton Street, Edinburgh, EH8 9AB, United Kingdom

Damian Markham

Laboratoire d'Informatique de Paris 6, CNRS, Sorbonne Université,
4 place Jussieu, 75005 Paris, France

Abstract

In this work we develop new methods for the characterisation of continuous variable quantum states using heterodyne measurement in both the trusted and untrusted settings. First, building on quantum state tomography with heterodyne detection, we introduce a reliable method for continuous variable quantum state certification, which directly yields the elements of the density matrix of the state considered with analytical confidence intervals. This method neither needs mathematical reconstruction of the data nor discrete binning of the sample space and uses a single Gaussian measurement setting. Second, beyond quantum state tomography and without its identical copies assumption, we promote our reliable tomography method to a general efficient protocol for verifying continuous variable pure quantum states with Gaussian measurements against fully malicious adversaries, i.e., making no assumptions whatsoever on the state generated by the adversary. These results are obtained using a new analytical estimator for the expected value of any operator acting on a continuous variable quantum state with bounded support over the Fock basis, computed with samples from heterodyne detection of the state.

2012 ACM Subject Classification Theory of computation → Quantum information theory

Keywords and phrases Continuous variable quantum information, reliable state tomography, certification, verification

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.3

Related Version A full version of the paper is available at <https://arxiv.org/abs/1905.12700>

Acknowledgements We thank N. Treps, V. Parigi, and especially M. Walschaers for stimulating discussions. We also thank A. Leverrier for interesting discussion on de Finetti reductions, and useful comments on previous versions of this work. This work was supported by the ANR project ANR-13-BS04-0014 COMB.

¹ Corresponding author



1 Introduction

Out of the many properties featured by quantum physics, the impossibility to perfectly determine an unknown state [8] is specially interesting. This property is at the heart of quantum cryptography protocols such as quantum key distribution [3]. On the other hand, it makes certification of the correct functioning of quantum devices a challenge, since the output of such devices can only be determined approximately, through repeated measurements over numerous copies of the output states. With rapidly developing quantum technologies for communication, simulation, computation and sensing, the ability to assess the correct functioning of quantum devices is of major importance, for near-term systems, the so-called Noisy Intermediate-Scale Quantum (NISQ) devices [23], and for the more sophisticated devices.

Depending on the desired level of trust, various methods are available for certifying the output of quantum devices. In the following, the task of checking the output state of a quantum device is denoted *tomography* for state independent methods, when i.i.d. behaviour is assumed, *certification* for a given a target state, when i.i.d. behaviour is assumed, and *verification* for a given target state, with no assumption whatsoever, and in particular without the i.i.d. assumption.

Quantum state tomography [9] is an important technique which aims at reconstructing a good approximation of the output state of a quantum device by performing multiple rounds of measurements on several copies of said output states. Given an ensemble of identically prepared systems, with measurement outcomes from the same observable, one can build up a histogram, from which a probability density can be estimated. According to Born's rule, this probability density is the square modulus of the state coefficients, taken in the basis corresponding to the measurement. However, a single measurement setting cannot yield the full state information since the phase of its coefficients are then lost. Many sets of measurements on many subensembles must be performed and combined to reconstruct the density matrix of the state. The data do not yield the state directly, but rather indirectly through data analysis. Quantum state tomography assumes an *independent and identically distributed* (i.i.d.) behaviour for the device, i.e., that the density matrix of the output state considered is the same at each round of measurement. This assumption may be relaxed with a tradeoff in the efficiency of the protocol [7].

A certification task corresponds to a setting where one wants to benchmark an industrial quantum device, or check the output of a physical experiment. On the other hand, a verification task corresponds to a cryptographic scenario, where the device to be tested is untrusted, or the quantum data is given by a potentially malicious party, for example in the context of delegated quantum computing. In the latter case, the task of quantum verification is to ensure that either the device behaved properly, or the computation aborts with high probability. While delegated computing is a natural platform for the emerging NISQ devices, one can provide a physical interpretation to this adversarial setting by emphasising that we aim for deriving verification schemes that make no assumptions whatsoever about the noise model of the underlying systems. Various methods for verification of quantum devices have been investigated, in particular for discrete variable quantum information [14], and they provide different efficiencies and security parameters depending on the computational power of the verifier. The common feature for all of these approaches is to utilise some basic obfuscation scheme that allows to reduce the problem of dealing with a fully general noise model, or a fully general adversarial deviation of the device, to a simple error detection scheme [27].

In this work, we consider the setting of quantum information with continuous variables [18], in which quantum states live in an infinite-dimensional Hilbert space. Using continuous variable systems for quantum computation and more general quantum information processing is a powerful alternative to the discrete variable case. *Firstly*, it is compatible with standard network optics technology, where more efficient measurements are available. *Secondly*, it allows for unprecedented scaling in entanglement, with entangled states of up to tens of thousands of subsystems reported [30] generated deterministically.

A continuous variable quantum process or state can be described by a quasi-probability distribution in phase space, often the Wigner function [28], but also the Husimi Q function or the Glauber–Sudarshan P function [5]. This allows for a simple and experimentally relevant classification of quantum states: those with a Gaussian quasiprobability distributions are called Gaussian states, and the others non-Gaussian states. By extension, operations mapping Gaussian states to Gaussian states are also called Gaussian. These Gaussian operations and states are the ones implementable with linear optics and quadratic non-linearities [4], and are hence relatively easy to construct experimentally. However, it is well known that for many important applications, Gaussian operations and Gaussian states are not sufficient. This takes the forms of no-go theorems for distillation and error correction [10, 12, 20], and the fact that all Gaussian computations can be simulated efficiently classically [2]. Furthermore, it is not possible to demonstrate non-locality or contextuality – which are increasingly understood to be important resources in quantum information – in the Gaussian regime.

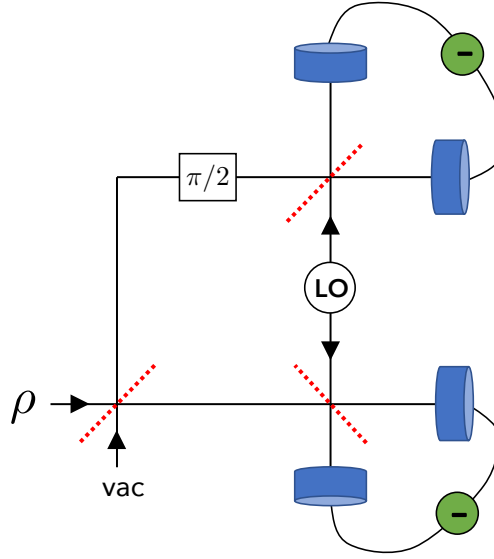
For continuous variable quantum devices, checking that the output state is close to a target state may be done with linear optics using optical homodyne tomography [19]. This method allows to reconstruct the Wigner function of a generic state using only Gaussian measurements, namely homodyne detection. Because of the continuous character of its outcomes, one must proceed to a discrete binning of the sample space, in order to build probability histograms. Then, the state representation in phase space is determined by a mathematical reconstruction.

For cases where we have a specific target state, more efficient options are possible. For multimode Gaussian states, more efficient certification methods have been derived with Gaussian measurements [1]. These methods involve the computation of a fidelity witness, i.e., a lower bound on the fidelity, from the measured samples. The cubic phase state certification protocol of [17] also introduces a fidelity witness and is an example of certification of a specific non-Gaussian state with Gaussian measurements, which assumes an i.i.d. state preparation. The verification protocol for Gaussian continuous variable weighted hypergraph states of [25] removes this assumption, again for this specific family of states.

2 Results

In this work we address two main issues. *Firstly*, existing continuous variable state tomography methods are not reliable in the sense of [7], because errors coming from the reconstruction procedure are indistinguishable from errors coming from the data. *Secondly*, to the best of our knowledge there is no Gaussian verification protocol for non-Gaussian states without i.i.d. assumption (a possible route using Serfling’s bound was mentioned in Ref. [17] for removing the i.i.d. assumption for their protocol).

We thus introduce a general *receive-and-measure* protocol for building trust for continuous variable quantum states, using solely Gaussian measurements, namely heterodyne detection [11, 26]. This protocol allows to perform reliable continuous variable quantum state tomography based on heterodyne detection, with analytical confidence intervals, which we



■ **Figure 1** A schematic representation of heterodyne measurement of a state ρ . The dashed red lines represent balanced beamsplitters. *LO* stands for local oscillator, i.e., strong coherent state, and *vac* for vacuum state. The blue circles are photodiode detectors.

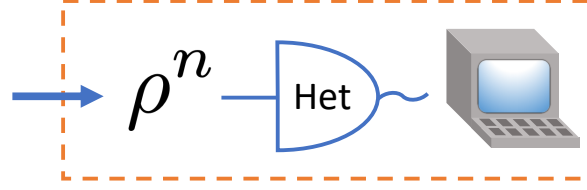
refer to as *heterodyne tomography* in what follows. This tomography technique only requires a single fixed measurement setting, compared to homodyne tomography. This protocol also provides a means for certifying continuous variable quantum states with an energy test, under the i.i.d. assumption. Finally, the same protocol also allows to verify continuous variable states, without the i.i.d. assumption. For these three applications, the measurements performed are the same. It is only the number of subsystems to be measured and the classical post-processing performed that differ from one application to another.

We detail the structure of the protocol in the following. We give an estimator for the expected value of any operator acting on a state with bounded support over the Fock basis (Theorem 1) by deriving an approximate version of the optical equivalence theorem for antinormal ordering [5]. The estimate is expressed as an expected value under heterodyne detection. Similar estimates have been obtained in the context of imperfect heterodyne detection [21, 22]. We go beyond these works in different respects: using this result, we introduce a reliable heterodyne tomography method and compute analytical bounds on its efficiency (Theorem 3). We then derive a *receive-and-measure* certification protocol (against i.i.d. adversary) for continuous variable quantum states, with Gaussian measurements (Theorem 4). We further promote this certification technique to a verification protocol against fully malicious adversary (Theorem 5), using a de Finetti reduction for infinite-dimensional systems [24].

3 Description of the protocol

Continuous variable quantum states live in an infinite-dimensional Hilbert space \mathcal{H} , spanned by the Fock basis $\{|n\rangle\}_{n \in \mathbb{N}}$, and are equivalently represented in phase space by their Husimi Q function [5], a smoother relative of the Wigner function. Given a single-mode state ρ , its Q function is defined as:

$$Q_\rho(\alpha) = \frac{1}{\pi} \text{Tr}(|\alpha\rangle\langle\alpha| \rho) = \text{Tr}(\Pi_\alpha \rho), \quad (1)$$



■ **Figure 2** A schematic representation of the protocol. The tester (within the dashed rectangle) receives a continuous variable quantum state ρ^n over n subsystems. This state could be for example the outcome of n successive runs of a physical experiment, the output of a commercial quantum device, or directly sent by some untrusted quantum server. The tester measures with heterodyne detection some of the subsystems of ρ^n , and uses the samples and efficient classical post-processing to deduce information about the remaining subsystems.

for all $\alpha \in \mathbb{C}$, where $|\alpha\rangle$ is a coherent state and where $\{\Pi_\alpha\}_{\alpha \in \mathbb{C}} = \{\frac{1}{\pi} |\alpha\rangle\langle\alpha|\}_{\alpha \in \mathbb{C}}$ is the Positive Operator Valued Measure for heterodyne detection.

This detection, also called double homodyne or eight-port homodyne [11], consists in splitting the measured state with a beamsplitter, and measuring both ends with homodyne detection (Fig. 1). This corresponds to a joint noisy measurement of quadratures q and p . This is a Gaussian measurement, which yields two real outcomes, corresponding to the real and imaginary parts of α . The Q function of a single-mode state thus is a probability density function over \mathbb{C} and measuring a state with heterodyne detection amounts to sampling from its Q function.

Using this detection, one may acquire knowledge about an unknown continuous variable quantum state. More precisely, we define the following *receive-and-measure* protocol, depicted in Fig. 2: given a quantum state ρ^n over n subsystems, measure some of the subsystems with heterodyne detection. Then, post-process the samples obtained to retrieve information about the remaining subsystems. The number subsystems to be measured and the post-processing performed depend on the application considered.

We show in the following sections how this protocol may be used to perform reliable tomography, certification and verification of continuous variable quantum states, and we detail the corresponding choices of subsystems and the classical post-processing for each task.

4 Heterodyne estimator

This section contains our main technical result, an estimator for the expected value of an operator acting on a state with bounded support over the Fock basis, from samples of heterodyne detection of the state. From this result, we derive various protocols in the following sections, ranging from tomography to state verification.

We denote by $\mathbb{E}_{\alpha \leftarrow D}[f(\alpha)]$ the expected value of a function f for samples drawn from a distribution D . Let us introduce for $k, l \geq 0$ the polynomials

$$\mathcal{L}_{k,l}(z) = e^{zz^*} \frac{(-1)^{k+l}}{\sqrt{k!}\sqrt{l!}} \frac{\partial^{k+l}}{\partial z^k \partial z^{*l}} e^{-zz^*}, \quad (2)$$

for $z \in \mathbb{C}$, which are, up to a normalisation, the Laguerre 2D polynomials, appearing in particular in the expressions of Wigner function of Fock states [29]. For any operator

3:6 Building Trust for Continuous Variable Quantum States

$A = \sum_{k,l=0}^{+\infty} A_{kl} |k\rangle\langle l|$ and all $E \in \mathbb{N}$, we define with these polynomials the function

$$f_A(z, \eta) = \frac{1}{\eta} e^{(1-\frac{1}{\eta})zz^*} \sum_{k,l=0}^E \frac{A_{kl}}{\sqrt{\eta^{k+l}}} \mathcal{L}_{k,l} \left(\frac{z}{\sqrt{\eta}} \right), \quad (3)$$

for all $z \in \mathbb{C}$, and all $0 < \eta < 1$. We omit the dependency in E for brevity. The function $z \mapsto f_A(z, \eta)$, being a polynomial multiplied by a converging Gaussian function, is bounded over \mathbb{C} . With the same notations, we also define the following constant:

$$K_A = \sum_{k,l=0}^E |A_{kl}| \sqrt{(k+1)(l+1)}. \quad (4)$$

The optical equivalence theorem for antinormal ordering [5] gives an equivalence between the expectation value of an operator in Hilbert space and the expectation value of its Glauber-Sudarshan P function. The P function is however highly singular in general and our results are based instead on the following approximate version of this equivalence when the P function is replaced by the bounded function f :

► **Theorem 1.** *Let $E \in \mathbb{N}$ and let $0 < \eta < \frac{2}{E}$. Let also $A = \sum_{k,l=0}^{+\infty} A_{kl} |k\rangle\langle l|$ be an operator and let $\rho = \sum_{k,l=0}^E \rho_{kl} |k\rangle\langle l|$ be a density operator with bounded support. Then,*

$$\left| \text{Tr}(A\rho) - \mathbb{E}_{\alpha \leftarrow Q_\rho} [f_A(\alpha, \eta)] \right| \leq \eta K_A, \quad (5)$$

where the function f and the constant K are defined in Eqs. (3) and (4).

For all theorems, the proof techniques are given in appendix A and the detailed proofs may be found in [6]. This result provides an estimator for the expected value of any operator A acting on a continuous variable state ρ with bounded support over the Fock basis. This estimator is the expected value of a bounded function f_A over samples drawn from the probability density corresponding to a Gaussian measurement of ρ , namely heterodyne detection. The optical equivalence theorem for antinormal ordering corresponds to the limit $\eta \rightarrow 0$. The right hand side of Eq. (5) is an energy bound, which depends on the operator A , the value E and the precision parameter η .

When the operator A is the density matrix of a continuous variable pure state $|\Psi\rangle$, the previous estimator approximates the fidelity $F(\Psi, \rho) = \langle \Psi | \rho | \Psi \rangle$ between $|\Psi\rangle\langle\Psi|$ and ρ . With the same notations:

► **Corollary 2.** *Let $E \in \mathbb{N}$ and let $0 < \eta < \frac{2}{E}$. Let also $|\Psi\rangle\langle\Psi| = \sum_{k,l=0}^{+\infty} \psi_k \psi_l^* |k\rangle\langle l|$ be a normalised pure state and let $\rho = \sum_{k,l=0}^E \rho_{kl} |k\rangle\langle l|$ be a density operator with bounded support. Then,*

$$\left| F(\Psi, \rho) - \mathbb{E}_{\alpha \leftarrow Q_\rho} [f_\Psi(\alpha, \eta)] \right| \leq \eta K_\Psi \leq \frac{\eta}{2} (E+1)(E+2), \quad (6)$$

where the function f_A and the constant K_A are defined in Eqs. (3) and (4), for $A = |\Psi\rangle\langle\Psi|$.

This result provides an estimator for the fidelity between any target pure state $|\Psi\rangle$ and any continuous variable (mixed) state ρ with bounded support over the Fock basis. This estimator is the expected value of a bounded function f_Ψ over samples drawn from the probability density corresponding to a Gaussian measurement of ρ , namely heterodyne detection. The right hand side of Eq. (6) is an energy bound, which may be refined depending on the

expression of $|\Psi\rangle$. In particular, the second bound is independent of the target state $|\Psi\rangle$. The assumption of bounded support makes sense for tomography, but not necessarily in an adversarial setting. We will relax this condition for the certification and verification protocols in the following, and indeed estimate the energy bound from the heterodyne measurements. Errors in this estimation are taken into account in the confidence statements.

Given these results, one may choose a target pure state $|\Psi\rangle$, and measure with heterodyne detection various copies of the output (mixed) state ρ of a quantum device with bounded support over the Fock basis. Then, using the samples obtained, one may estimate the expected value of f_Ψ , thus obtaining an estimate of the fidelity between the states $|\Psi\rangle\langle\Psi|$ and ρ . Using this result, we introduce a reliable method for performing continuous variable quantum state tomography using heterodyne detection.

5 Reliable continuous variable state tomography

Continuous variable quantum state tomography methods usually make two assumptions: firstly that the measured states are independent identical copies (i.i.d. assumption, for *independently and identically distributed*), and secondly that the measured states have a bounded support over the Fock basis [19]. With the same assumptions, we present a reliable method for state tomography with heterodyne detection which has the advantage of providing analytical confidence intervals. Our method directly provides estimates of the elements of the state density matrix, phase included. As such, neither mathematical reconstruction of the phase, nor binning of the sample space is needed, since the samples are used only to compute expected values of bounded functions. Moreover, only a single fixed Gaussian measurement setting is needed, namely heterodyne detection (Fig. 1).

For tomographic application, all copies of the state are measured. For $n \geq 1$, let $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ be samples from heterodyne detection of n copies of a quantum state ρ . For $\epsilon > 0$ and $k, l \in \mathbb{N}$, we define

$$\rho_{kl}^\epsilon = \frac{1}{n} \sum_{i=1}^n f_{|l\rangle\langle k|}(\alpha_i, \epsilon/K_{|l\rangle\langle k|}), \quad (7)$$

where the function f_A and the constant K_A are defined in Eqs. (3) and (4), for $A = |l\rangle\langle k|$, and where $\epsilon > 0$ is a free parameter. The quantity ρ_{kl}^ϵ is the average of the function $f_{|l\rangle\langle k|}$ over the samples $\alpha_1, \dots, \alpha_n$. The next result shows that this estimator approximates the matrix element k, l of this state with high probability. We use the notations of Theorem 1.

► **Theorem 3** (Reliable heterodyne tomography). *Let $\epsilon, \epsilon' > 0$, $n \geq 1$ and $\alpha_1, \dots, \alpha_n$ be samples obtained by measuring with heterodyne detection n copies of a state $\rho = \sum_{k,l=0}^E \rho_{kl} |k\rangle\langle l|$ with bounded support, for $E \in \mathbb{N}$. Then*

$$|\rho_{kl} - \rho_{kl}^\epsilon| \leq \epsilon + \epsilon', \quad (8)$$

for all $0 \leq k, l \leq E$, with probability greater than

$$1 - 4 \sum_{0 \leq k \leq l \leq E} \exp \left[-\frac{n\epsilon^{2+k+l}\epsilon'^2}{4C_{kl}} \right], \quad (9)$$

where the estimate ρ_{kl}^ϵ is defined in Eq. (7) and where

$$C_{kl} = [(k+1)(l+1)]^{1+\frac{k+l}{2}} 2^{|l-k|} \binom{\max(k,l)}{\min(k,l)} \quad (10)$$

is a constant independent of ρ .

In light of this result, the principle for performing reliable heterodyne tomography is straightforward and as follows: n identical copies $\rho^{\otimes n}$ of the output quantum state of a physical experiment or quantum device are measured with heterodyne detection, yielding the values $\alpha_1, \dots, \alpha_n$. These values are used to compute the estimates ρ_{kl}^ϵ , defined in Eq. (7), for all k, l in the range of energy of the experiment. Then, Theorem 3 directly provides confidence intervals for all these estimates of ρ_{kl} , the matrix elements of the density operator ρ , without the need for a binning of the sample space or any additional data reconstruction, using a single measurement setting. For a desired precision ϵ and a failure probability δ , the number of samples needed scales as $n = \text{poly}(1/\epsilon, \log(1/\delta))$.

Both homodyne and heterodyne quantum state tomography assume a bounded support over the Fock basis for the output state considered, i.e., that all matrix elements are equal to zero beyond a certain value, and that the output quantum states are i.i.d., i.e., that all measured output states are independent and identical. While these assumptions are natural when looking at the output of a physical experiment, corresponding to a noisy partially trusted quantum device with bounded energy, they may be questionable in the context of untrusted devices. We remove these assumptions in what follows: we first drop the bounded support assumption, deriving a certification protocol for continuous variable quantum states of an i.i.d. device with heterodyne detection ; then, we drop both assumptions, deriving a general verification protocol for continuous variable quantum states against an adversary who can potentially be fully malicious.

6 State certification with Gaussian measurements

Given an untrusted source of quantum states, the purpose of state certification and state verification protocols is to check whether if its output state is close to a given target state, or far from it. To achieve this, a verifier tests the output state of the source. Ideally, one would like to obtain an upper bound on the probability that the state is not close from the target state, given that it passed a test. However, this is known to be impossible without prior knowledge of the tested state distribution [14]. Indeed, writing this conditional probability

$$\Pr[\text{incorrect}|\text{accept}] = \frac{\Pr[\text{incorrect} \cap \text{accept}]}{\Pr[\text{accept}]}, \quad (11)$$

in a situation where the device always produces a bad output state, it is rejected by the verifier's test most of the time, so the acceptance probability is very small and the conditional probability is equal to 1. Therefore, the quantity that will always be bounded in certification and verification protocols, in which one does not have prior knowledge of the device, is the joint probability that the tested state is not close to the target state *and* that it passes the test. Equivalently, we obtain lower bounds on the probability that the tested state is close to the target state or that it fails the test.

We first consider the certification of the output of an i.i.d. quantum device, i.e., which output state is the same at each round. However, we do not assume that the output states of the device have bounded support over the Fock basis anymore. This is instead ensured probabilistically using the samples from heterodyne detection.

Our continuous variable quantum state certification protocol is then as follows: let $|\Psi\rangle$ be a target pure state, of which one wants to certify m copies. The values s and E are free parameters of the protocol. One instructs the i.i.d. device to prepare $n + m$ copies of $|\Psi\rangle$, and the device outputs an i.i.d. (mixed) state $\rho^{\otimes(n+m)}$. One keeps m copies $\rho^{\otimes m}$, and measures the n others with heterodyne detection, obtaining the samples $\alpha_1, \dots, \alpha_n$. One records the

number r of samples such that $|\alpha_i|^2 > E$. We refer to this step as *support estimation*. For a given $\epsilon > 0$, one also computes with the same samples the estimate

$$F_\Psi(\rho) = \left[\frac{1}{n} \sum_{i=1}^n f_\Psi(\alpha_i, \epsilon/(mK_\Psi)) \right]^m, \quad (12)$$

where the function f_A and the constant K_A are defined in Eqs. (3) and (4), for $A = |\Psi\rangle\langle\Psi|$, and where $\epsilon > 0$ is a free parameter. The next result quantifies how close this estimate is from the fidelity between the remaining m copies of the output state $\rho^{\otimes m}$ of the tested device and m copies of the target state $|\Psi\rangle\langle\Psi|^{\otimes m}$.

► **Theorem 4** (Gaussian certification of continuous variable quantum states). *Let $\epsilon, \epsilon' > 0$, let $s \leq n$, and let $\alpha_1, \dots, \alpha_n$ be samples obtained by measuring with heterodyne detection n copies of a state ρ . Let E in \mathbb{N} , and let r be the number of samples such that $|\alpha_i|^2 > E$. Let also $|\Psi\rangle$ be a pure state. Then for all $m \in \mathbb{N}^*$,*

$$|F(\Psi^{\otimes m}, \rho^{\otimes m}) - F_\Psi(\rho)| \leq \epsilon + \epsilon', \quad (13)$$

or $r > s$, with probability greater than

$$1 - (P_{\text{Support}}^{\text{iid}} + P_{\text{Hoeffding}}^{\text{iid}}), \quad (14)$$

where

$$P_{\text{Support}}^{\text{iid}} = \frac{(s+1)^{3/2}}{n} \exp \left[\frac{(s+1)^2}{n+1} \right], \quad (15)$$

$$P_{\text{Hoeffding}}^{\text{iid}} = 2 \exp \left[-\frac{n\epsilon^{2+2E}\epsilon'^2}{2m^{4+2E}C_\Psi^2} \right], \quad (16)$$

where the estimate $F_\Psi(\rho)$ is defined in Eq. (12), and where

$$C_\Psi = \sum_{k,l=0}^E |\psi_k \psi_l| \left(\frac{\epsilon}{m} \right)^{E - \frac{k+l}{2}} K_\psi^{1 + \frac{k+l}{2}} \sqrt{2^{|l-k|} \binom{\max(k,l)}{\min(k,l)}} \quad (17)$$

is a constant independent of ρ , with the constant K defined in Eq. (4).

This results implies that the quantity $F_\Psi(\rho)$ is a good estimate of the fidelity $F(\Psi^{\otimes m}, \rho^{\otimes m})$, or the score at the support estimation step is higher than s , with high probability. The values of the energy parameters E and s should be chosen to guarantee completeness, i.e., that if the correct state $|\Psi\rangle$ is sent, then $r \leq s$ with high probability.

This theorem is valid for all continuous variable target pure states $|\Psi\rangle$, and the failure probability may be greatly reduced depending on the expression of $|\Psi\rangle$. The number of samples needed for certifying a given number of copies m with a precision ϵ and a failure probability δ scales as $n = \text{poly}(m, 1/\epsilon, 1/\delta)$. Note that the same protocol may be used to obtain reliable estimates of $\text{Tr}(A\rho)$ for any operator A under the i.i.d. assumption, by setting $m = 1$ and replacing Ψ by A in Eq. (12).

This certification protocol is promoted to a verification protocol in the following section, by removing the i.i.d. assumption.

7 State verification with Gaussian measurements

We now consider an adversarial setting, where a verifier delegates the preparation of a continuous variable quantum state to a potentially malicious party, called the *prover*. One could see the verifier as the experimentalist in the laboratory and the prover as the noisy device, where we aim not to make any assumptions about its correct functionality or noise model. Given the absence of any direct error correction mechanism that permits a fault tolerant run of the device, the aim of verification is to ensure that a wrong outcome is not being accepted. In the context of state verification, this amounts to making sure that the output state of the tested device is close to an ideal target state.

The prover is not supposed to have i.i.d. behaviour. In particular, when asked for various copies of the same state, the prover may actually send a large state entangled over all subsystems, possibly also entangled with a quantum system on his side. In that case, the certification protocol derived in the previous section is not reliable. With usual tomography measurements, the number of samples needed for a given precision of the fidelity estimate scales exponentially in the number of copies to verify. This is an essential limitation of quantum tomography techniques, because they check all possible correlations between the different subsystems.

However we prove that, because of the symmetry of the protocol, the verifier can assume that the prover is sending permutation-invariant states, i.e., states that are invariant under any permutation of their subsystems. With a specific support estimation step, reduced states of permutation-invariant states are close to mixture almost-i.i.d. states, i.e., states that are i.i.d. on almost all subsystems. At the heart of this reduction is the de Finetti theorem for infinite-dimensional systems of [24], which allows restricting to an almost-i.i.d. prover.

Our verification protocol is then as follows: the verifier wants to verify m copies of a target pure state $|\Psi\rangle$. The values n, k, q, s and E are free parameters of the protocol. The prover is instructed to prepare $n + k$ copies of $|\Psi\rangle$ and send them to the verifier. The verifier picks k subsystems at random and measures them with heterodyne detection, obtaining the samples β_1, \dots, β_k , and records the number r of values $|\beta_i|^2 > E$. The verifier discards $4q$ subsystems at random and measures all the others but m chosen at random with heterodyne detection, obtaining the samples $\alpha_1, \dots, \alpha_{n-4q-m}$. Finally, the verifier computes with these samples the estimate

$$F_\Psi(\rho) = \left[\frac{1}{n-4q-m} \sum_{i=1}^{n-4q-m} f_\Psi(\alpha_i, \epsilon/(mK_\Psi)) \right]^m, \quad (18)$$

where the function f_A and the constant K_A are defined in Eqs. (3) and (4), for $A = |\Psi\rangle\langle\Psi|$ and where $\epsilon > 0$ is a free parameter. Note that this estimate is identical to the one defined in Eq. (12), replacing n by $n - 4q - m$.

► **Theorem 5** (Gaussian verification of continuous variable quantum states). *Let $n \geq 1$, let $s \leq k$, and let ρ^{n+k} be a state over $n + k$ subsystems. Let β_1, \dots, β_k be samples obtained by measuring k subsystems at random with heterodyne detection and let ρ^n be the remaining state after the measurement. Let E in \mathbb{N} , and let r be the number of samples such that $|\beta_i|^2 > E$. Let also $q \geq m$, and let ρ^m be the state remaining after discarding $4q$ subsystems of ρ^n at random, and measuring $n - 4q - m$ other subsystems at random with heterodyne detection, yielding the samples $\alpha_1, \dots, \alpha_{n-4q-m}$. Let $\epsilon, \epsilon' > 0$ and let $\epsilon'' = \sqrt{\frac{m(4q+m-1)}{n-4q}}$. Let $|\Psi\rangle$ be a target pure state. Then,*

$$|F(\Psi^{\otimes m}, \rho^m) - F_\Psi(\rho)| \leq \epsilon + \epsilon' + \epsilon'' + P_{deFinetti}, \quad (19)$$

or $r > s$, with probability greater than

$$1 - (P_{\text{support}} + P_{\text{deFinetti}} + P_{\text{Hoeffding}}), \quad (20)$$

where

$$P_{\text{support}} = 8k^{3/2} \exp \left[-\frac{k}{9} \left(\frac{q}{n} - \frac{2s}{k} \right)^2 \right], \quad (21)$$

$$P_{\text{deFinetti}} = q^{(E+1)^2/2} \exp \left[-\frac{2q(q+1)}{n} \right], \quad (22)$$

$$P_{\text{Hoeffding}} = 2 \binom{n-4q}{4q} \exp \left[-\frac{n-8q}{2m^{4+2E}} \left(\frac{\epsilon^{1+E} \epsilon'}{C_\Psi} - \frac{8qm^{2+E}}{n-4q-m} \right)^2 \right], \quad (23)$$

where the estimate $F_\Psi(\rho)$ is defined in Eq. (18), and where C_Ψ is a constant independent of ρ defined in Eq. (17).

This result implies that the quantity $F_\Psi(\rho)$ is a good estimate of the fidelity $F(\Psi^{\otimes m}, \rho^m)$, or the score at the support estimation step is higher than s , with high probability. Like for the certification protocol, the values of the energy parameters E and s should be chosen by the verifier to guarantee completeness, i.e., that if the prover sends the correct state $|\Psi\rangle$, then $r \leq s$ with high probability.

For specific choices of the free parameters of the protocol either the estimate $F_\Psi(\rho)$ is polynomially precise in m , or $r > s$, with exponential probability in m , with $n, k, q = \text{poly}(m)$. In particular, the efficiency of the protocol may be greatly refined by taking into account the expression of $|\Psi\rangle$ in the Fock basis, and optimizing over the free parameters.

This verification protocol let the verifier gain confidence about the precision of the estimate of the fidelity in Eq. (18). If the value of the estimate is close enough to 1, the verifier may decide to use the state to run a computation. Indeed, statements on the fidelity of a state allow inferring the correctness of any trusted computation done afterwards using this state. Let $\beta > 0$, and let \mathcal{O} be the observable corresponding to the result of the trusted computation performed on ρ^m , the reduced state over m subsystems instead of $|\Psi\rangle^{\otimes m}$, m copies of the target state $|\Psi\rangle$. In other words, \mathcal{O} encodes the resources which the verifier can perform perfectly (ancillary states, evolution and measurements), the imperfections being encoded in ρ . Then, $F(\Psi^{\otimes m}, \rho^m) \geq 1 - \beta$ implies the following bound on the total variation distance between the probability densities of the computation output of the actual and the target computations:

$$\|P_{\Psi^{\otimes m}}^{\mathcal{O}} - P_{\rho^m}^{\mathcal{O}}\|_{\text{tvd}} \leq D(\Psi^{\otimes m}, \rho^m) \leq \sqrt{\beta}, \quad (24)$$

by standard properties of the trace distance D [13]. What this means is that the distribution of outcomes for the state ρ^m sent by the prover is almost indistinguishable from the distribution of outcomes for m copies of the ideal state $|\Psi\rangle$, when the fidelity is close enough to one.

8 Discussion

Determining an unknown continuous variable quantum state is especially difficult since it is described by possibly infinitely many complex parameters. Existing methods like homodyne quantum state tomography require many different measurement settings, and

heavy classical post-processing. For that purpose, we have introduced a reliable method for heterodyne quantum state tomography, which uses heterodyne detection as a single Gaussian measurement setting, and allows the retrieval of the density matrix of an unknown quantum state without the need for data reconstruction nor binning of the sample space. For data reconstruction methods such as Maximum Likelihood, errors from the reconstruction procedure are usually indistinguishable from errors coming from the tested quantum device. For that reason, such methods do not extend well to the task of verification, unlike our method.

Building on these tomography techniques, and with the addition of cryptographic techniques such as the de Finetti theorem, we have derived a protocol for verifying various copies of a continuous variable quantum state, without i.i.d. assumption, with Gaussian measurements. This protocol is robust, as it directly gives a confidence interval on an estimate of the fidelity between the tested state and the target pure state. We emphasize that, while the target state is pure, the tested state is not required to be pure.

Our verification protocol is complementary to the approach of [25], in which a measurement-only verifier performs continuous variable quantum computing by delegating the preparation of Gaussian cluster states to a prover, and has to perform non-Gaussian measurements. In our approach, the measurement-only verifier may perform continuous variable quantum computing by delegating the preparation of non-Gaussian states to the prover, and has to perform Gaussian measurement, which are much easier to perform experimentally.

Our protocol may be tailored to different uses and assumptions, from tomography to verification, simply by changing the classical post-processing. We expect this protocol to be useful for the validation of continuous variable quantum devices in the NISQ [23] era and onwards.

In particular, an interesting perspective would be fine-tuning the various parameters of the protocol for specific target states in order to optimise its efficiency, thus reducing the number of samples needed for a given confidence interval. Another interesting prospect would be extending our main technical result, Theorem 1, which applies to operators, to quantum maps. Also, in the case where the operator is the density matrix of a target pure state, our result provide an estimate for the fidelity, and it would be interesting to extend this to target mixed states.

References

- 1 Leandro Aolita, Christian Gogolin, Martin Kliesch, and Jens Eisert. Reliable quantum certification of photonic state preparations. *Nature communications*, 6:8498, 2015. doi:10.1038/ncomms9498.
- 2 Stephen D. Bartlett, Barry C. Sanders, Samuel L. Braunstein, and Kae Nemoto. Efficient classical simulation of continuous variable quantum information processes. *Phys. Rev. Lett.*, 88:097904, February 2002. doi:10.1103/PhysRevLett.88.097904.
- 3 C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proc. IEEE International Conference on Computers, Systems and Signal Processing*, volume 175, page 8, Bangalore, December 1984. doi:10.1016/j.tcs.2011.08.039.
- 4 Samuel L. Braunstein and Peter van Loock. Quantum information with continuous variables. *Rev. Mod. Phys.*, 77:513–577, June 2005. doi:10.1103/RevModPhys.77.513.
- 5 Kevin E Cahill and Roy J Glauber. Density operators and quasiprobability distributions. *Physical Review*, 177(5):1882, 1969. doi:10.1103/PhysRev.177.1882.
- 6 Ulysse Chabaud, Tom Douce, Frédéric Grosshans, Elham Kashefi, and Damian Markham. Building trust for continuous variable quantum states. *quant-ph*, 2019. arXiv:1905.12700.

- 7 Matthias Christandl and Renato Renner. Reliable quantum state tomography. *Physical Review Letters*, 109(12):120403, 2012. doi:10.1103/PhysRevLett.109.120403.
- 8 G. M. D’Ariano and H. P. Yuen. Impossibility of measuring the wave function of a single quantum system. *Physical review letters*, 76(16):2832, 1996. doi:10.1103/PhysRevLett.76.2832.
- 9 G Mauro D’Ariano, Matteo GA Paris, and Massimiliano F Sacchi. Quantum tomography. *Advances in Imaging and Electron Physics*, 128:206–309, 2003. arXiv:quant-ph/0302028.
- 10 Jens Eisert, Stefan Scheel, and Martin B Plenio. Distilling gaussian states with gaussian operations is impossible. *Physical review letters*, 89(13):137903, 2002. doi:10.1103/PhysRevLett.89.137903.
- 11 Alessandro Ferraro, Stefano Olivares, and Matteo GA Paris. Gaussian states in continuous variable quantum information. *quant-ph*, 2005. arXiv:quant-ph/0503237.
- 12 Jaromír Fiurášek. Gaussian transformations and distillation of entangled gaussian states. *Physical review letters*, 89(13):137904, 2002. doi:10.1103/PhysRevLett.89.137904.
- 13 Christopher A Fuchs and Jeroen Van De Graaf. Cryptographic distinguishability measures for quantum-mechanical states. *IEEE Transactions on Information Theory*, 45(4):1216–1227, 1999.
- 14 Alexandru Gheorghiu, Theodoros Kapourniotis, and Elham Kashefi. Verification of quantum computation: An overview of existing approaches. *Theory of Computing Systems*, 4:715–808, 2019. doi:10.1007/s00224-018-9872-3.
- 15 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American statistical association*, 58(301):13–30, 1963. doi:10.1080/01621459.1963.10500830.
- 16 Anthony Leverrier, Raúl García-Patrón, Renato Renner, and Nicolas J Cerf. Security of continuous-variable quantum key distribution against general attacks. *Physical review letters*, 110(3):030502, 2013. doi:10.1103/PhysRevLett.110.030502.
- 17 Nana Liu, Tommaso F Demarie, Si-Hui Tan, Leandro Aolita, and Joseph F Fitzsimons. Client-friendly continuous-variable blind and verifiable quantum computing. *Physical Review A*, 100(6):062309, 2019. doi:PhysRevA.100.062309.
- 18 Seth Lloyd and Samuel L Braunstein. Quantum computation over continuous variables. In *Quantum Information with Continuous Variables*, pages 9–17. Springer, 1999. doi:10.1103/PhysRevLett.82.1784.
- 19 Alexander I Lvovsky and Michael G Raymer. Continuous-variable optical quantum-state tomography. *Reviews of Modern Physics*, 81(1):299, 2009. doi:10.1103/RevModPhys.81.299.
- 20 Julien Niset, Jaromír Fiurášek, and Nicolas J Cerf. No-go theorem for gaussian quantum error correction. *Physical review letters*, 102(12):120501, 2009. doi:10.1103/PhysRevLett.102.120501.
- 21 Matteo GA Paris. On density matrix reconstruction from measured distributions. *Optics communications*, 124(3-4):277–282, 1996. doi:10.1016/0030-4018(96)00019-3.
- 22 Matteo GA Paris. Quantum state measurement by realistic heterodyne detection. *Physical Review A*, 53(4):2658, 1996. doi:10.1103/PhysRevA.53.2658.
- 23 John Preskill. Quantum computing in the nisq era and beyond. *Quantum*, 2:79, 2018. doi:10.22331/q-2018-08-06-79.
- 24 Renato Renner and J Ignacio Cirac. de finetti representation theorem for infinite-dimensional quantum systems and applications to quantum cryptography. *Physical review letters*, 102(11):110504, 2009. doi:10.1103/PhysRevLett.102.110504.
- 25 Yuki Takeuchi, Atul Mantri, Tomoyuki Morimae, Akihiro Mizutani, and Joseph F Fitzsimons. Resource-efficient verification of quantum computing using serfling’s bound. *npj Quantum Information*, 5:27, 2019. doi:10.1038/s41534-019-0142-2.
- 26 Yong Siah Teo, Christian R Muller, Hyunseok Jeong, Zdenek Hradil, Jaroslav Rehacek, and Luis L Sanchez-Soto. When heterodyning beats homodyning: an assessment with quadrature moments. *quant-ph*, 2017. arXiv:1701.07539.

- 27 Thomas Vidick. http://users.cms.caltech.edu/~vidick/verification_bulletin.pdf, 2018.
- 28 Eugene Paul Wigner. On the quantum correction for thermodynamic equilibrium. In *Part I: Physical Chemistry. Part II: Solid State Physics*, pages 110–120. Springer, 1997. doi:10.1103/PhysRev.40.749.
- 29 Alfred Wünsche. Laguerre 2d-functions and their application in quantum optics. *Journal of Physics A: Mathematical and General*, 31(40):8267, 1998. doi:10.1088/0305-4470/31/40/017.
- 30 Shota Yokoyama, Ryuji Ukai, Seiji C Armstrong, Chanond Sornphiphatphong, Toshiyuki Kaji, Shigenari Suzuki, Jun-ichi Yoshikawa, Hidehiro Yonezawa, Nicolas C Menicucci, and Akira Furusawa. Ultra-large-scale continuous-variable cluster states multiplexed in the time domain. *Nature Photonics*, 7(12):982, 2013. doi:10.1038/nphoton.2013.287.

A Proof techniques

This section details the primary mathematical tools used in the proofs of the theorems, along with some intuition. The full technical proofs can be found in [6].

The function $z \mapsto f_A(z, \eta)$ defined in Eq. (3) for $\eta > 0$ is a bounded approximation of the Glauber-Sudarshan function P_A of the operator A . This approximation is parametrised by a precision η , and a cutoff value E . The optical equivalence theorem for antinormal ordering [5] reads

$$\text{Tr}(A\rho) = \int Q_\rho(\alpha) P_A(\alpha) d^2\alpha. \quad (25)$$

Given that

$$\mathbb{E}_{\alpha \leftarrow Q_\rho}[f_A(\alpha, \eta)] = \int Q_\rho(\alpha) f_A(\alpha, \eta) d^2\alpha, \quad (26)$$

we can expect that $\mathbb{E}_{\alpha \leftarrow Q_\rho}[f_A(\alpha, \eta)]$ is an approximation of $\text{Tr}(A\rho)$ parametrised by η and E . Theorem 1 formalises this statement.

The proof of Theorem 3 combines Theorem 1 with Hoeffding inequality [15], which quantifies the speed of convergence of the sample mean towards the expected value of a bounded i.i.d. random variable:

► **Lemma 6 (Hoeffding).** *Let $\lambda > 0$, let $n \geq 1$, let z_1, \dots, z_n be i.i.d. complex random variables from a probability density D over \mathbb{R} , and let $f : \mathbb{C} \mapsto \mathbb{R}$ such that $|f(z)| \leq M$, for $M > 0$ and all $z \in \mathbb{C}$. Then*

$$\Pr \left[\left| \frac{1}{n} \sum_{i=1}^n f(z_i) - \mathbb{E}_{z \leftarrow D}[f(z)] \right| \geq \lambda \right] \leq 2 \exp \left[-\frac{n\lambda^2}{2M^2} \right]. \quad (27)$$

The proof then follows by applying this inequality for $D = Q_\rho$, and $f = f_{|k\rangle\langle l|}$, for all values of k, l between 0 and E , together with the union bound.

Theorem 4 removes the bounded support assumption and its proof is similar to the one of Theorem 3, with the addition of a support estimation step, using samples from heterodyne detection. The main result utilised here is the fact that for all E [16]

$$1 - \Pi_{\leq E} = \sum_{n=E+1}^{+\infty} |n\rangle\langle n| \leq \frac{2}{\pi} \int_{|\alpha|^2 \geq E} |\alpha\rangle\langle \alpha| d^2\alpha, \quad (28)$$

where $\Pi_{\leq E}$ is the projector onto the space of states of support bounded by E . This result allows to bound the probability of having a large support and obtaining a low score at the support estimation step.

The proof of Theorem 5 is the most technical. This proof combines three main ingredients: a support estimation step for permutation-invariant states using samples from heterodyne detection, the de Finetti reduction from [24] and a refined version of Hoeffding inequality for superpositions of almost-i.i.d. states under a product measurement. The three terms appearing in the expression of the probability in the theorem correspond to these three ingredients, respectively.

Uncloneable Quantum Encryption via Oracles

Anne Broadbent 

Department of Mathematics and Statistics, University of Ottawa, Canada
abroadbe@uottawa.ca

Sébastien Lord 

Department of Mathematics and Statistics, University of Ottawa, Canada
slord050@uottawa.ca

Abstract

Quantum information is well known to achieve cryptographic feats that are unattainable using classical information alone. Here, we add to this repertoire by introducing a new cryptographic functionality called *uncloneable encryption*. This functionality allows the encryption of a classical message such that two collaborating but isolated adversaries are prevented from simultaneously recovering the message, even when the encryption key is revealed. Clearly, such functionality is unattainable using classical information alone.

We formally define uncloneable encryption, and show how to achieve it using Wiesner’s conjugate coding, combined with a quantum-secure pseudorandom function (qPRF). Modelling the qPRF as an oracle, we show security by adapting techniques from the quantum one-way-to-hiding lemma, as well as using bounds from quantum monogamy-of-entanglement games.

2012 ACM Subject Classification Theory of computation → Quantum computation theory; Theory of computation → Cryptographic primitives; Security and privacy → Symmetric cryptography and hash functions

Keywords and phrases Quantum Cryptography, Symmetric Key, Monogamy-of-Entanglement

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.4

Related Version A full version of the paper is available at <https://arxiv.org/abs/1903.00130>.

Funding This material is based upon work supported by the U.S. Air Force Office of Scientific Research under award number FA9550-17-1-0083, Canada’s NSERC, an Ontario ERA, and the University of Ottawa’s Research Chairs program.

1 Introduction

A key distinction between classical and quantum information is given by the *no-cloning principle*: unlike bits, arbitrary qubits cannot be perfectly copied [11, 18, 26]. This principle is the basis of many of the feats of quantum cryptography, including quantum money [25] and quantum key distribution (QKD) [6] (for a survey on quantum cryptography, see [9]).

In QKD, two parties establish a shared secret key, using public quantum communication combined with an authentic classical channel. The quantum communication allows to *detect* eavesdropping: when the parties detect only a small amount of eavesdropping, they can produce a shared string that is essentially guaranteed to be private. Gottesman [15] studied *quantum tamper-detection* in the case of *encryption schemes*: in this work, a classical message is encrypted into a quantum ciphertext such that, at decryption time, the receiver will *detect* if an adversary could have information about the plaintext when the key is revealed. We note that classical information alone cannot produce such encryption schemes, since it is always possible to perfectly *copy* ciphertexts.

Notably, Gottesman left open the question of an encryption scheme that would *prevent* the *splitting* of a ciphertext. In other words, would it be possible to encrypt a classical message into a quantum ciphertext, such that no attack at the ciphertext level would be



© Anne Broadbent and Sébastien Lord;

licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 4; pp. 4:1–4:22

Leibniz International Proceedings in Informatics



Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

significantly successful in producing *two* quantum registers, each of which, when combined with the decryption key, could be used to reconstruct the plaintext?

In this work, we define, construct and prove security for a scheme that answers Gottesman’s question in the positive. We call this *uncloneable encryption*. The core technical aspects of this work were first presented in one of the author’s M.Sc. thesis [16].

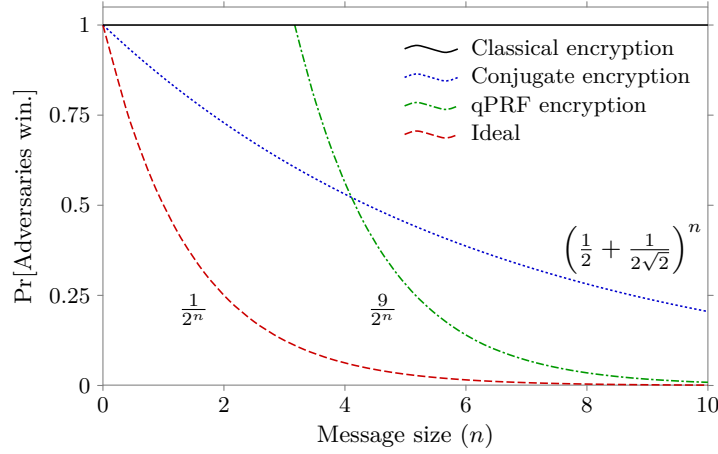
1.1 Summary of Contributions

We consider encryption schemes that encode classical plaintexts into quantum ciphertexts, which we formalize in Definition 4. For simplicity, in this work, we consider only the one-time, symmetric-key case. Next, we define uncloneable encryption (Definition 8). Informally, this can be thought of as a game, played between the honest sender (Alice) and two malicious recipients (Bob and Charlie). First, Alice picks a message $m \in \{0, 1\}^n$ and a key $k \in \{0, 1\}^{\kappa(\lambda)}$ (κ is a polynomial in some security parameter, λ). She encrypts her message into a quantum ciphertext register R . Initially, Bob and Charlie are physically together, and they receive R . They apply a quantum map to produce two registers: Bob keeps register B and Charlie keeps register C . Bob and Charlie are then isolated. In the next phase, Alice reveals k to both parties. Using k and their quantum register, Bob and Charlie produce m_B and m_C respectively. Bob and Charlie *win* if and only if $m_B = m_C = m$. The scheme is *t-uncloneable secure* if their winning probability is upper bounded by $2^{-n+t} + \eta(\lambda)$ for a negligible η .

Assuming that Alice picks her message uniformly at random, our results are summarized in Figure 1, where we plot upper bounds for the winning probability of Bob and Charlie against various types of encodings, according to the length of m . First of all, if the encoding is classical, then Bob and Charlie can each keep a copy of the ciphertext. Combined with the key k , each party decrypts to obtain m . This gives the horizontal line at $\Pr[\text{Adversaries win}] = 1$. Next, a lower bound on the winning probability for *any* encryption scheme is $\frac{1}{2^n}$ (corresponding to the parties coordinating a random guess). This is the *ideal* curve. Our goal is therefore to produce an encryption scheme that matches the ideal curve as close as possible.

It may seem that asking that Alice sample her message uniformly at random would be particularly restrictive, but this is not the case – we show in Theorem 9 that security in the case of uniformly sampled messages implies security in the case of non-uniformly sampled messages, if the message size does not grow with the security parameter. Specifically, if Bob and Charlie can win with probability at most $2^{-n+t} + \eta(\lambda)$ when the message is sampled uniformly at random, for some t and some negligible function η , then they can win with probability at most $2^{-h+t} + \eta'(\lambda)$ if the message m is sampled from a distribution with a min-entropy of h where η' is a negligible function which is larger than η .

Our first attempt at realizing uncloneable encryption (Appendix A) shows that the well-known Wiesner conjugate coding [25] already achieves a security bound that is better than any classical scheme. For any two bit strings $x, \theta \in \{0, 1\}^n$, define the Wiesner state $|x^\theta\rangle = H^{\theta_1}|x_1\rangle \otimes \dots \otimes H^{\theta_n}|x_n\rangle$. The encryption uses a random key $r, \theta \in \{0, 1\}^n$ and maps a classical message m into the quantum state $\rho = |(m \oplus r)^\theta\rangle\langle(m \oplus r)^\theta|$; given (r, θ) , decryption consists in measuring in the basis determined by θ to obtain x and then computing $x \oplus r$. We sketch a proof that this satisfies a notion of security for encryption schemes. The question of uncloneability then boils down to: “How well can an adversary *split* ρ into *two* registers, each of which, combined with (θ, r) can reconstruct m ?” This question is answered in prior work on *monogamy-of-entanglement games* [20]: an optimal strategy wins with probability $\left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^n$. This is again illustrated in Figure 1.



■ **Figure 1** Upper-bounds on winning probabilities for various types of encodings (up to negligible functions of λ) for messages sampled uniformly at random.

In order to improve this bound, we use a quantum-secure pseudorandom function (qPRF, see Definition 3) $f_\lambda : \{0, 1\}^\lambda \times \{0, 1\}^\lambda \rightarrow \{0, 1\}^n$. The encryption (see Section 4.1) consists of a quantum state $\rho = |r^\theta\rangle\langle r^\theta|$ for random $r, \theta \in \{0, 1\}^\lambda$, together with a classical string $c = m \oplus f_\lambda(s, r)$ for a random s . The key k consists in θ and s . Once again, it can be shown that this is an encryption scheme in a more usual sense and we sketch this argument in Section 4.1. Intuitively, the use of f_λ affords us a gain in uncloneable security, because an adversary who wants to output m would need to know the pre-image of m under $f_\lambda(s, \cdot)$. Reaching a formal proof along these lines, however, is tricky. First, we model the qPRF using a quantum random oracle [8]; this limits the adversaries' interaction with the qPRF to be black-box quantum queries (we refer to Section 4.3 for further details on this modelling). Next, the quantum random oracle model is notoriously tricky to use and many of the techniques in the classical literature are not directly applicable. Fortunately, we can adapt techniques from Unruh's quantum one-way-to-hiding lemma [22] to the two-player setting, which enables us to recover a precise statement along the lines of the intuition above. We thus complete the proof of our main Theorem 16, obtaining the bound $9 \cdot \frac{1}{2^n} + \text{negl}(\lambda)$. This is the fourth and final curve in Figure 1.

In addition to the above, we formally define a different type of uncloneable security: inspired by more standard security definitions of *indistinguishability*, we define *uncloneable-indistinguishability* (Definition 11). This security definition bounds the advantage that the adversaries have at *simultaneously* distinguishing between an encryption of 0^n and an encryption of a plaintext of length n , as prepared by the adversaries. In a series of results (Theorems 12 and 17 and Corollary 18), we show that our main protocol achieves this security notion against adversaries that use *unentangled strategies* and as long as the message size does not grow with λ . As discussed in Section 1.2, there are interesting uses cases where we can assume that the adversaries do not share entanglement.

We note that our protocols (both Definition 19 and Definition 13) have the desirable property of being *prepare-and-measure* schemes. This means that the quantum technology for the honest users is limited to the preparation of single-qubit pure states, as well as to single-qubit measurements; these quantum technologies are mature and commercially available. (Note, however, that quantum storage remains a major challenge at the implementation level).

1.2 Applications

While our focus is on the conceptual contribution of defining and proving a new primitive, we believe that uncloneable encryption could have many applications. We give two such examples.

1.2.1 Quantum Money

As it captures the idea of “uncloneable classical information” in a very generic manner, uncloneable encryption can be used as a tool to build other primitives which leverage the uncloneability of quantum states. Such constructions help us understand the landscape of quantum cryptography. As an example, any uncloneable secure encryption scheme naturally yields a private-key quantum money scheme [2, 25].

To obtain quantum money from an uncloneable encryption scheme, we identify the notion of “simultaneously passing the bank’s verification” with the notion of “simultaneously obtaining the correct plaintext”. To generate a banknote, the bank samples a message m , a key k , a serial number s and produces as output $(s, \text{Enc}(k, m))$, where $\text{Enc}(k, m)$ is the uncloneable encryption of m with the key k . When the bank is asked to verify a banknote, it verifies the serial number in its database to retrieve k , decrypts the ciphertext and verifies if the message obtained is indeed m .

The uncloneable security guarantee implies that the probability of a malicious party producing two banknotes which pass this test is negligible. If this were not the case, we could use the attack which counterfeits the banknote to essentially copy the ciphertext in the underlying uncloneable encryption scheme. The adversaries tasked with obtaining the message once the key is revealed then simply decrypt as if they were the honest receivers.

1.2.2 Preventing Storage Attacks by Classical Adversaries

Indistinguishable-uncloneable encryption prevents a single eavesdropping adversary with no quantum memory from collecting ciphertexts exchanged by two honest parties in the hope of later learning the key. We sketch an argument for this fact.

Suppose such an adversary obtains a ciphertext from an uncloneable-indistinguishable encryption scheme. We claim that they cannot correctly determine if the ciphertext corresponds to the encryption of 0^n or of some known message m with non-negligible advantage, even if the decryption key becomes known after their measurement of the ciphertext. If such an adversary existed, it could be used to break the uncloneable-indistinguishable security of the encryption scheme. Indeed, the almost-classical eavesdropper could create two copies of their classical memory and distribute it to the two adversaries who attempt to obtain the message once the key is revealed.¹

Note that the adversaries in this attack do not share any entanglement and so we can apply Corollary 18 which states that our encryption scheme is uncloneable-indistinguishable secure under this condition.

Our work is currently in the private-key setting, but can be extended in a straightforward way to the public-key setting. In this scenario, we can still guarantee the secrecy of the message even if the eavesdropper is later able to determine the decryption key from the publicly-known encryption key. In other words, an eavesdropping adversary with no quantum memory would need to attack the ciphertext during transmission. This is known as *long-term* security or *everlasting* security [21].

¹ We thank an anonymous reviewer for this suggestion.

1.3 More on Related Work

Starting with the foundational work of Wiesner [25], a rich body of literature has considered the encoding of classical information into quantum states in order to take advantage of quantum properties for cryptography.

- **Quantum Key Recycling.** The concept of quantum key recycling is a precursor to the QKD protocol, developed by Bennett, Brassard, and Breidbart [7] (the manuscript was prepared in 1982 but only published recently). According to this protocol, it is possible to encrypt a classical message into a quantum state, such that information-theoretic security is assured, but in addition, a tamper detection mechanism would allow the one-time pad key to be re-used in the case that no eavesdropping is detected. Quantum key recycling has been the object of recent related work [10, 13].
- **Tamper-Evident Encryption.** We referred above to tamper-detection in the case of encryption, which we will also call *tamper-evident encryption*. However, we emphasize that the author originally called this contribution *uncloneable encryption* [15]. We justify this choice of re-labelling in quoting the conclusion of the work:

One difficulty with such generalizations is that it is unclear to what extent the name “uncloneable encryption” is really deserved. I have not shown that a message protected by uncloneable encryption cannot be copied – only that Eve cannot copy it without being detected. Is it possible for Eve to create two states, (...), which can each be used (in conjunction with the secret key) to extract a good deal of information about the message? Or can one instead prove bounds, for instance, on the sum of the information content of the various purported copies? [15]

Since our work addresses this question, we have appropriately re-labeled prior work according to a seemingly more accurate name. To the best of our knowledge, the precise relationship between quantum key-recycling, tamper-evident encryption, and uncloneable encryption is unknown (see Section 1.4).

- **Quantum Copy-Protection.** Further related work includes the study of *quantum copy-protection*, as initiated by Aaronson [1]. Informally, this is a means to encode a function (from a given family of functions) into a quantum program state, such that an honest party can evaluate the function given the program state, but it would be impossible to somehow *split* the quantum program state so as to enable *two* parties to simultaneously evaluate the function. Aaronson gave protocols for quantum copy-protection in an oracle model, but left wide open the question of quantum copy-protection in the plain model. In a way, uncloneable encryption is a first step towards quantum copy-protection, since it prevents copying of *data*, which can be seen as a unit of information that is even simpler than a function.

1.4 Outlook and Future Work

In this work, we challenge one of the tacit assumptions of encryptions, namely that adversaries can always copy ciphertexts. We believe that this has the potential to significantly change the landscape of cryptography, for instance in terms of techniques for *key management* [5]. Furthermore, our techniques could become building blocks for a theory of uncloneable cryptography.

Our work leads to many follow-up questions, broadly classified according to the following themes:

- **Improvements.** There are many possible improvements to the current work. For instance: Could our scheme be made resilient to errors? Can we remove the reliance on the oracle, and/or on the qPRF? Could an encryption scheme simultaneously be uncloneable *and* provide *tamper detection*? Would achieving uncloneable-indistinguishable security be possible, without any restrictions on the adversary's strategy?
- **Links with related work.** What are the links, if any, between uncloneable encryption, tamper-evident encryption [15], and quantum encryption with key recycling [7, 10, 13]? We note that both uncloneable encryption and quantum encryption with key recycling [13] make use of theorems developed in the context of one-sided device-independent QKD [20]. Can we make more formal links between these primitives?
- **More uncloneability.** Finally, our work paves the way for the study of more complex uncloneable primitives. Could this lead to uncloneable programs [1]? What about in complexity theory, could we define and realize uncloneable *proofs* [1]?

1.5 Outline

The remainder of the paper is structured as follows. In Section 2, we introduce some basic notation and useful results from the literature. In Section 3, we formally define uncloneable encryption schemes and their security. Our main scheme is described in Section 4 (with a toy scheme based on Wiesner conjugate coding being described in Appendix A). Due to lack of space, most proofs are relegated to Appendix B and the remainder can be found in the full version.

2 Preliminaries

In this section, we present our notation and techniques from prior works used in this paper.

2.1 Notation and Basics of Quantum Information

We denote the set of all functions of the form $f : \{0, 1\}^n \rightarrow \{0, 1\}^m$ by $\text{Bool}(n, m)$. We denote the set of strictly positive natural numbers by \mathbb{N}^+ . All Hilbert spaces are finite dimensional. We overload the expectation symbol \mathbb{E} in the following way: If X is a finite set, \mathcal{X} a random variable on X , and $f : X \rightarrow \mathbb{R}$ some function, we define $\mathbb{E}_{x \leftarrow \mathcal{X}} f(x)$ to be $\sum_{x \in X} \Pr[\mathcal{X} = x] f(x)$. If we omit the random variable then we assume a uniform distribution, i.e.: $\mathbb{E}_x f(x) = \frac{1}{|X|} \sum_{x \in X} f(x)$. If \mathcal{X} is a random variable distributed over a finite set X , then its min-entropy is given by $-\max_{x \in X} \Pr[\mathcal{X} = x]$. A function $\eta : \mathbb{N} \rightarrow \mathbb{R}$ is said to be negligible if for all $n \in \mathbb{N}$ there exists an $x_n > 0$ such that $x > x_n$ implies that $|\eta(x)| < x^{-n}$.

A comprehensive introduction to quantum information and quantum computing may be found in [17, 24]. We fix some notation in the following paragraphs.

Let $\mathcal{Q} = \mathbb{C}^2$ be the state space of a single qubit. In particular, \mathcal{Q} is a two-dimensional complex Hilbert space spanned by the orthonormal set $\{|0\rangle, |1\rangle\}$. For any $n \in \mathbb{N}^+$, we write $\mathcal{Q}(n) = \mathcal{Q}^{\otimes n}$ and note that $\{|s\rangle = |s_1\rangle \otimes |s_2\rangle \otimes \dots \otimes |s_n\rangle\}_{s \in \{0,1\}^n}$ forms an orthonormal basis of $\mathcal{Q}(n)$.

Let \mathcal{H} be a Hilbert space. The set of all unitary and density operators on \mathcal{H} are denoted, respectively, by $\mathcal{U}(\mathcal{H})$ and $\mathcal{D}(\mathcal{H})$. We recall that the operator norm of a linear operator $A : \mathcal{H} \rightarrow \mathcal{H}'$ between finite dimensional Hilbert spaces is given by $\|A\| = \max_{v \in \mathcal{H}, \|v\|=1} \|Av\|$ and satisfies the property that $\|Av\| \leq \|A\| \cdot \|v\|$. If A is either a projector or a unitary operator, then $\|A\| = 1$.

We use the term “quantum state” to refer to both unit vectors $|\psi\rangle \in \mathcal{H}$ and to density operators $\rho \in \mathcal{D}(\mathcal{H})$ on some Hilbert space.

Let $H \in \mathcal{U}(\mathcal{Q})$ be the Hadamard operator defined by $|0\rangle \mapsto \frac{|0\rangle+|1\rangle}{\sqrt{2}}$ and $|1\rangle \mapsto \frac{|0\rangle-|1\rangle}{\sqrt{2}}$. For any strings $x, \theta \in \{0,1\}^n$, we define the state $|x^\theta\rangle = H^{\theta_1}|x_1\rangle \otimes H^{\theta_2}|x_2\rangle \otimes \dots \otimes H^{\theta_n}|x_n\rangle$. Note that the set $\{|s^\theta\rangle\}_{s \in \{0,1\}^n}$ forms an orthonormal basis of $\mathcal{Q}(n)$. Following their use in [25], we call states of the form $|x^\theta\rangle$ Wiesner states and we call $\{|s^\theta\rangle\}_{s \in \{0,1\}^n}$ a Wiesner basis. For any $n \in \mathbb{N}^+$, the Einstein-Podolski-Rosen (EPR) [12] state is given by $|\text{EPR}_n\rangle = \frac{1}{\sqrt{2^n}} \sum_{x \in \{0,1\}^n} |x\rangle \otimes |x\rangle$.

A positive operator-valued measurement (POVM) on a Hilbert space \mathcal{H} is a finite collection of positive semidefinite operators $\{E_i\}_{i \in I}$ on \mathcal{H} which sum to the identity. A projective measurement is a POVM composed of projectors.

We also recall that physically permissible transformation of a quantum system precisely coincide with the set of completely positive trace preserving (CPTP) maps. In particular, CPTP map will map density operators to density operators.

A polynomial-time uniform family of circuits $\mathbf{C} = \{\mathbf{C}_\lambda\}_{\lambda \in \mathbb{N}^+}$ is a collection of quantum circuits indexed by \mathbb{N}^+ such that there exists a polynomial-time deterministic Turing machine T which, on input 1^λ , produces a description of \mathbf{C}_λ . We refer to such families as efficient circuits. Each circuit \mathbf{C}_λ defines and implements a certain CPTP map $C_\lambda : \mathcal{D}(\mathcal{H}_{\text{In},\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{\text{Out},\lambda})$ where the Hilbert spaces $\mathcal{H}_{\text{In},\lambda}$ and $\mathcal{H}_{\text{Out},\lambda}$ are implicitly defined by the circuit. Note that we consider general, i.e.: possibly non-unitary, circuits. These were introduced in [3]. It is worth noting that a universal gate set for general quantum circuits exists which is composed of only unitary gates, implementing maps of the form $\rho \mapsto U\rho U^\dagger$ for some unitary operator U , and two non-unitary maps which are the single qubit partial trace map $\text{Tr} : \mathcal{D}(\mathcal{Q}) \rightarrow \mathcal{D}(\mathbb{C})$ and the state preparation map $\text{Aux} : \mathcal{D}(\mathbb{C}) \rightarrow \mathcal{D}(\mathcal{Q})$ defined by $1 \mapsto |0\rangle\langle 0|$. Further information on this circuit model can be found in [23].

2.2 Monogamy-of-Entanglement Games

Monogamy-of-entanglement games were introduced and studied in [20]. In short, such a game is played by Alice against cooperating Bob and Charlie. Alice describes to Bob and Charlie a collection of different POVMs which she could use to measure a quantum state on a Hilbert space \mathcal{H}_A . These POVMs are indexed by a finite set Θ and each reports a measurement result taken from a finite set X . Bob and Charlie then produce a tripartite quantum state $\rho \in \mathcal{D}(\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, giving the A register to Alice, the B register to Bob and the C register to Charlie. Alice then picks a $\theta \in \Theta$, measures her subsystem with the corresponding POVM and obtains some result $x \in X$. She then announces θ to Bob and Charlie who are now isolated. Bob and Charlie win if and only if they can both simultaneously guess the result x .

Upper bounds on the winning probability of Bob and Charlie in such games was the primary subject of study in [20]. One of their main results, corresponding to a game where Alice measures in a random Wiesner basis, is as follows.

► **Theorem 1** ([20]). *Let $\lambda \in \mathbb{N}^+$. For any Hilbert spaces \mathcal{H}_B and \mathcal{H}_C , any collections of POVMs*

$$\left\{ \{B_x^\theta\}_{x \in \{0,1\}^\lambda} \right\}_{\theta \in \{0,1\}^n} \quad \text{and} \quad \left\{ \{C_x^\theta\}_{x \in \{0,1\}^\lambda} \right\}_{\theta \in \{0,1\}^n} \quad (1)$$

on these Hilbert spaces, and any state $\rho \in \mathcal{D}(\mathcal{Q}(\lambda) \otimes \mathcal{H}_B \otimes \mathcal{H}_C)$, we have that

$$\mathbb{E}_{\theta} \sum_{x \in \{0,1\}^{\lambda}} \text{Tr} [(|x^{\theta}\rangle\langle x^{\theta}| \otimes B_x^{\theta} \otimes C_x^{\theta}) \rho] \leq \left(\frac{1}{2} + \frac{1}{2\sqrt{2}} \right)^{\lambda}. \quad (2)$$

Using standard techniques, we recast this theorem in a context where Alice sends to Bob and Charlie a random Wiesner state and they split this state among themselves via a CPTP map Φ .

► **Corollary 2.** *Let $\lambda \in \mathbb{N}^+$. For any Hilbert spaces \mathcal{H}_B and \mathcal{H}_C , any collections of POVMs*

$$\left\{ \{B_x^{\theta}\}_{x \in \{0,1\}^{\lambda}} \right\}_{\theta \in \{0,1\}^{\lambda}} \quad \text{and} \quad \left\{ \{C_x^{\theta}\}_{x \in \{0,1\}^{\lambda}} \right\}_{\theta \in \{0,1\}^{\lambda}} \quad (3)$$

on these Hilbert spaces, and any CPTP map $\Phi : \mathcal{D}(\mathcal{Q}(\lambda)) \rightarrow \mathcal{D}(\mathcal{H}_B \otimes \mathcal{H}_C)$, we have that

$$\mathbb{E}_{\theta} \mathbb{E}_x \text{Tr} [(B_x^{\theta} \otimes C_x^{\theta}) \Phi(|x^{\theta}\rangle\langle x^{\theta}|)] \leq \left(\frac{1}{2} + \frac{1}{2\sqrt{2}} \right)^{\lambda}. \quad (4)$$

The proof can be found in the full version, but conceptually follows from a two-step argument. First, we only consider states of the form $(\mathbb{1} \otimes \Phi) |\text{EPR}_{\lambda}\rangle\langle \text{EPR}_{\lambda}|$ for some CPTP map Φ and where Alice keeps the intact subsystems from the EPR pairs. Then, we apply the correspondence between Alice measuring her half of an EPR pair in a random Wiesner basis and her sending a random Wiesner state. This correspondence is similar to the one used in the Shor-Preiskill proof of security for the BB84 QKD protocol [19].

Corollary 2 can be seen as the source of “uncloneability” for our upcoming protocols. When Alice sends a state $|x^{\theta}\rangle\langle x^{\theta}|$, picked uniformly at random, to Bob and Charlie, she has a guarantee that it is unlikely for both of them to learn x even if she later divulges θ . It is worth noting that Theorem 1 and Corollary 2 have no computational or hardness assumptions.

2.3 Oracles and Quantum-Secure Pseudorandom Functions

A quantum-secure pseudorandom function is a keyed function which appears random to an efficient quantum adversary who only sees its input/output behaviour and is ignorant of the particular key being used. We formally define this notion with the help of oracles. Quantum accessible oracles have been previously studied in the literature, for example in [8, 22].

For a function $H \in \text{Bool}(n, m)$, a circuit \mathcal{C} is said to have oracle access to H , denoted \mathcal{C}^H , if we add to its gate set a gate implementing the unitary operator $O^H \in \mathcal{U}(\mathcal{Q}(n)_Q \otimes \mathcal{Q}(m)_R)$ defined on computational basis states by

$$|x\rangle_Q \otimes |y\rangle_R \mapsto |x\rangle_Q \otimes |y \oplus H(x)\rangle_R. \quad (5)$$

Colloquially, we are giving \mathcal{C} a “black box” which computes the function H . Note that for any two functions $H, H' \in \text{Bool}(n, m)$, we can obtain the circuit $\mathcal{C}^{H'}$ from \mathcal{C}^H by replacing every instance of the O^H gate by the $O^{H'}$ gate.

Our definition of a quantum-secure pseudorandom function, inspired by [27], is as follows.

► **Definition 3** (Quantum-Secure Pseudorandom Function). *A quantum-secure pseudorandom function \mathcal{F} is a collection of functions*

$$\mathcal{F} = \left\{ f_{\lambda} : \{0,1\}^{\lambda} \times \{0,1\}^{\ell_{In}(\lambda)} \rightarrow \{0,1\}^{\ell_{Out}(\lambda)} \right\}_{\lambda \in \mathbb{N}^+} \quad (6)$$

where $\ell_{In}, \ell_{Out} : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and such that:

1. There is an efficient quantum circuit $F = \{F_\lambda\}_{\lambda \in \mathbb{N}^+}$ such that F_λ implements the CPTP map F_λ given by $\rho \mapsto U_\lambda \rho U_\lambda^\dagger$ where $U_\lambda \in \mathcal{U}(\mathcal{Q}(\lambda + \ell_{\text{In}}(\lambda) + \ell_{\text{Out}}(\lambda)))$ is defined by

$$U_\lambda(|k\rangle \otimes |a\rangle \otimes |b\rangle) = |k\rangle \otimes |a\rangle \otimes |b \oplus f_\lambda(k, a)\rangle. \quad (7)$$

2. For all efficient quantum circuits $D = \{D_\lambda^H\}_{\lambda \in \mathbb{N}^+}$ having oracle access to a function of the form $H \in \text{Bool}(\ell_{\text{In}}(\lambda), \ell_{\text{Out}}(\lambda))$, each implementing a CPTP map of the form $D_\lambda^H : \mathcal{D}(\mathbb{C}) \rightarrow \mathcal{D}(\mathcal{Q})$, there is a negligible function η such that:

$$\left| \mathbb{E}_k \text{Tr} [|0\rangle\langle 0| D_\lambda^{f_\lambda(k, \cdot)}(1)] - \mathbb{E}_H \text{Tr} [|0\rangle\langle 0| D_\lambda^H(1)] \right| \leq \eta(\lambda). \quad (8)$$

We should think of D as a circuit which attempts to distinguish two different cases: was it given oracle access to the pseudorandom function $f(k, \cdot) : \{0, 1\}^{\ell_{\text{In}}(\lambda)} \rightarrow \{0, 1\}^{\ell_{\text{Out}}(\lambda)}$ for a randomly sampled $k \in \{0, 1\}^\lambda$? Or to a function $H \in \text{Bool}(\ell_{\text{In}}(\lambda), \ell_{\text{Out}}(\lambda))$ sampled truly at random? The circuit takes no input and produces a single bit of output, via measuring a single qubit in the computational basis. The bound given in the definition ensures that the probability distribution of the output does not change by much in both scenarios.

In his work on quantum-secure pseudorandom functions [27], Zhandry showed that certain common constructions of pseudorandom functions are secure against quantum adversaries.

3 Uncloneable Encryption

The encryption of classical plaintexts into classical ciphertexts has been extensively studied. The study of encrypting quantum plaintexts into quantum ciphertexts has also received some attention, for example in [4]. Uncloneable encryption is a security notion for classical plaintexts which is impossible to achieve in any meaningful way with classical ciphertexts. Thus, we formally define a notion of quantum encryptions for classical messages in Section 3.1 and then give our security definitions in Section 3.2.

3.1 Quantum Encryptions of Classical Messages

A quantum encryption of classical messages scheme is a procedure which takes as input a plaintext and a key, in the form of classical bit strings, and produces a ciphertext in the form of a quantum state. We model these schemes as efficient quantum circuits and CPTP maps where classical bit strings are identified with computational basis states: $s \leftrightarrow |s\rangle\langle s|$. Our schemes are parametrized by a security parameter λ . In general, the message size $n = n(\lambda)$, the key size $\kappa = \kappa(\lambda)$, and the size of the ciphertext $\ell = \ell(\lambda)$ may depend on λ . This is formalized in Definition 4.

► **Definition 4** (Quantum Encryption of Classical Messages). A quantum encryption of classical messages (QECM) scheme is a triplet of efficient quantum circuits $\mathcal{S} = (\text{Key}, \text{Enc}, \text{Dec})$ implementing CPTP maps of the form

- $\text{Key}_\lambda : \mathcal{D}(\mathbb{C}) \rightarrow \mathcal{D}(\mathcal{H}_{K,\lambda})$,
- $\text{Enc}_\lambda : \mathcal{D}(\mathcal{H}_{K,\lambda} \otimes \mathcal{H}_{M,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{T,\lambda})$, and
- $\text{Dec}_\lambda : \mathcal{D}(\mathcal{H}_{K,\lambda} \otimes \mathcal{H}_{T,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{M,\lambda})$

where, for functions $n, \ell, \kappa : \mathbb{N}^+ \rightarrow \mathbb{N}^+$, the plaintext space is given by $\mathcal{H}_{M,\lambda} = \mathcal{Q}(n(\lambda))$, the ciphertext space is given by $\mathcal{H}_{T,\lambda} = \mathcal{Q}(\ell(\lambda))$, and the keyspace is given by $\mathcal{H}_{K,\lambda} = \mathcal{Q}(\kappa(\lambda))$.

For all $\lambda \in \mathbb{N}^+$, $k \in \{0, 1\}^{\kappa(\lambda)}$, and $m \in \{0, 1\}^{n(\lambda)}$, the maps must satisfy

$$\text{Tr}[|k\rangle\langle k| \text{Key}(1)] > 0 \implies \text{Tr}[|m\rangle\langle m| \text{Dec}_k \circ \text{Enc}_k(|m\rangle\langle m|)] = 1 \quad (9)$$

where λ is implicit, Enc_k is the CPTP map defined by $\rho \mapsto \text{Enc}(|k\rangle\langle k| \otimes \rho)$, and we define Dec_k analogously.

A short discussion on the key generation circuit, Key , is in order. First, note that Key takes no input. Indeed, the domain of Key_λ is $\mathcal{D}(\mathbb{C})$ and \mathbb{C} is the state space of zero qubits. In particular, there is a single valid quantum state on \mathbb{C} : $\mathcal{D}(\mathbb{C}) = \{1\}$. To generate a classical key to be used by the encryption and decryption circuits Enc_λ and Dec_λ , a party runs the circuit Key_λ and obtains the quantum state $\text{Key}_\lambda(1)$. This quantum state is then measured in the computational basis and the result of this measurement is used as the key. We then see that Equation (9) is a correctness condition which imposes that, for all keys that may be generated, a valid ciphertext is always correctly decrypted.

3.2 Security Notions

Now that we have formal definition for QECM schemes, we can define security notions for these schemes. We define three such notions:

1. Indistinguishable security. Conceptually inspired by the original security notion of indistinguishable encryptions [14], which considers classical plaintexts and classical ciphertexts, and similar in details to an analogue definition in [4] which considers quantum plaintexts and quantum ciphertexts, this security notion considers classical plaintexts and quantum ciphertexts. It is formally stated in Definition 6.
2. Uncloneable security. This security notion is novel to this work and captures, in the broadest sense, what we mean by an “uncloneable encryption scheme”. This security notion is defined in Definition 8 and is parametrized by a real value $0 \leq t \leq n$, where n is the message size. The case where $t = 0$ is ideal and $t = n$ is trivial. In particular, no encryption scheme with classical ciphertexts may achieve t -uncloneable security for $t < n$.
3. Uncloneable-indistinguishable security. This security notion is also novel to this work. It can be seen as a combination of indistinguishable and uncloneable security. It is formally defined in Definition 11.

Each of these security notions is defined in two steps. First, we define a type of attack (Definitions 5, 7 and 10). Then, we say that the QECM scheme achieves the given security notion if all admissible attacks have their winning probability appropriately bounded (Definitions 6, 8 and 11). The definitions for uncloneable security and uncloneable-indistinguishable security will formalize the games which we described in Section 1.1.

Note that many classical encryption schemes which are secure against quantum adversaries, such as the one-time pad, are indistinguishable secure but satisfy neither uncloneable security notions as their ciphertexts can always be perfectly copied. We also discuss in Appendix A a scheme which offers non-trivial uncloneable security but is not in any way uncloneable-indistinguishable secure.

We first define our notion of indistinguishable security.

► **Definition 5** (Distinguishing Attack). *Let \mathcal{S} be a QECM scheme. A distinguishing attack against \mathcal{S} is a pair of efficient quantum circuits $\mathcal{A} = (\mathcal{G}, \mathcal{A})$ implementing CPTP maps of the form*

- $G_\lambda : \mathcal{D}(\mathbb{C}) \rightarrow \mathcal{D}(\mathcal{H}_{S,\lambda} \otimes \mathcal{H}_{M,\lambda})$ and
- $A_\lambda : \mathcal{D}(\mathcal{H}_{S,\lambda} \otimes \mathcal{H}_{T,\lambda}) \rightarrow \mathcal{D}(\mathcal{Q})$

where $\mathcal{H}_{S,\lambda} = \mathcal{Q}(s(\lambda))$ for a function $s : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and $\mathcal{H}_{M,\lambda}$ and $\mathcal{H}_{T,\lambda}$ are as defined by \mathcal{S} .

► **Definition 6** (Indistinguishable Security). *Let \mathcal{S} be a QECM scheme. For a fixed and implicit value of λ , we define the CPTP map $\text{Enc}_k^1 : \mathcal{D}(\mathcal{H}_{M,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{T,\lambda})$ by*

$$\rho \mapsto \sum_{m \in \{0,1\}^n} \text{Tr}[|m\rangle\langle m| \rho] \cdot \text{Enc}_k(|m\rangle\langle m|) \quad (10)$$

and the CPTP map $Enc_k^0 : \mathcal{D}(\mathcal{H}_{M,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{T,\lambda})$ by

$$\rho \mapsto Enc_k(|0^n\rangle\langle 0^n|) \quad (11)$$

where $0^n \in \{0,1\}^n$ is the all zero bit string.

Then, we say that \mathcal{S} is indistinguishable secure if for all distinguishing attacks \mathcal{A} against \mathcal{S} , there exists a negligible function η such that

$$\mathbb{E}_b \mathbb{E}_{k \leftarrow \mathcal{K}} \text{Tr} [|b\rangle\langle b| A_\lambda \circ (\mathbb{1}_S \otimes Enc_k^b) \circ G(1)] \leq \frac{1}{2} + \eta(\lambda) \quad (12)$$

where λ is implicit on the left-hand side, $b \in \{0,1\}$, and \mathcal{K}_λ is the random variable distributed on the set $\{0,1\}^{\kappa(\lambda)}$ such that $\Pr[\mathcal{K}_\lambda = k] = \text{Tr}[|k\rangle\langle k| Key_\lambda(1)]$.

In Definition 6, the map Enc_k^0 should be seen as discarding whatever plaintext was given and producing the encryption of the all zero bit string. On the other hand, Enc_k^1 is the map which first measures the state given in the computational basis, to ensure that the plaintext is indeed a classical message, and then encrypts this message. We say that a QECM scheme has indistinguishable security if no efficient adversary can distinguish between both of these scenarios (by trying to determine the value of b) with more than a negligible advantage. This security notion allows us to show that the schemes we define do offer a level of security as encryption schemes.

Next, we formalize the intuitive definition for uncloneable security as given by the game described in Section 1.1. In Figure 2, we sketch out the relation between the various CPTP maps and the underlying Hilbert spaces considered in this definition.

► **Definition 7 (Cloning Attack).** Let \mathcal{S} be a QECM scheme. A cloning attack against \mathcal{S} is a triplet of efficient quantum circuits $\mathcal{A} = (A, B, C)$ implementing CPTP maps of the form

- $A_\lambda : \mathcal{D}(\mathcal{H}_{T,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{B,\lambda} \otimes \mathcal{H}_{C,\lambda})$,
- $B_\lambda : \mathcal{D}(\mathcal{H}_{K,\lambda} \otimes \mathcal{H}_{B,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{M,\lambda})$, and
- $C_\lambda : \mathcal{D}(\mathcal{H}_{K,\lambda} \otimes \mathcal{H}_{C,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{M,\lambda})$

where $\mathcal{H}_{B,\lambda} = \mathcal{Q}(\beta(\lambda))$ and $\mathcal{H}_{C,\lambda} = \mathcal{Q}(\gamma(\lambda))$ for some functions $\beta, \gamma : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and $\mathcal{H}_{K,\lambda}$, $\mathcal{H}_{M,\lambda}$, and $\mathcal{H}_{T,\lambda}$ are as defined by \mathcal{S} .

► **Definition 8 (Uncloneable Security).** A QECM scheme \mathcal{S} is $t(\lambda)$ -uncloneable secure if for all cloning attacks \mathcal{A} against \mathcal{S} there exists a negligible function η such that

$$\mathbb{E}_{m \leftarrow \mathcal{K}} \mathbb{E}_{k \leftarrow \mathcal{K}} \text{Tr} [(|m\rangle\langle m| \otimes |m\rangle\langle m|) (B_k \otimes C_k) \circ A \circ Enc_k(|m\rangle\langle m|)] \leq 2^{-n+t(\lambda)} + \eta(\lambda) \quad (13)$$

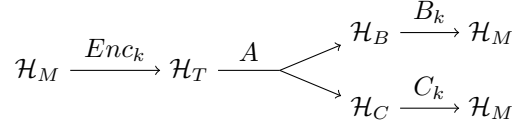
where λ is implicit on the left-hand side, \mathcal{K}_λ is a random variable distributed on $\{0,1\}^{\kappa(\lambda)}$ such that $\Pr[\mathcal{K}_\lambda = k] = \text{Tr}[|k\rangle\langle k| Key_\lambda(1)]$ and B_k is the CPTP map defined by $\rho \mapsto B(|k\rangle\langle k| \otimes \rho)$ and similarly for C_k .

If \mathcal{S} is 0-uncloneable secure, we simply say that it is uncloneable secure.

The left-hand side of Equation (13) is the probability, averaged over all messages and all keys, that both adversaries can correctly output the encrypted message.

We note that any encryption which produces classical ciphertexts cannot be t -uncloneable secure for any $t < n$. Indeed, an attack \mathcal{A} where A copies the classical ciphertext and where $B = C = \text{Dec}$ succeeds with probability 1.

Our definition of uncloneable security is with respect to messages sampled uniformly at random. However, if the length of the message is fixed, t -uncloneable security implies a similar security notion for messages sampled according to other distributions. We formalize this in the next theorem whose proof can be found in Appendix B.1.



■ **Figure 2** Schematic representation of the maps considered in a cloning attack (Definition 7). The k subscript indicates which maps have access to the encryption key.

► **Theorem 9.** *Let \mathcal{S} be a QECM scheme which is t -uncloneable secure and whose message size is constant, i.e.: $n(\lambda) = n$. Let \mathcal{M} be a random variable distributed over $\{0, 1\}^n$ with min-entropy h . Then, for any cloning attack \mathcal{A} on \mathcal{S} there is a negligible function η such that*

$$\mathbb{E}_{m \leftarrow \mathcal{M}} \mathbb{E}_{k \leftarrow \mathcal{K}} \text{Tr} [(|m\rangle\langle m| \otimes |m\rangle\langle m|) (B_k \otimes C_k) \circ A \circ \text{Enc}_k |m\rangle\langle m|] \leq 2^{-h+t(\lambda)} + \eta(\lambda) \quad (14)$$

where λ is implicit on the left-hand side.

Finally, we formalize the notion of uncloneable-indistinguishable security (see Section 1.1 for a description in terms of a game, and Figure 3 for the relation between the various CPTP maps and the underlying Hilbert spaces).

► **Definition 10** (Cloning-Distinguishing Attack). *Let \mathcal{S} be a QECM scheme. A cloning-distinguishing attack against \mathcal{S} is a tuple $\mathcal{A} = (\mathcal{G}, A, B, C)$ of efficient quantum circuits implementing CPTP maps of the form*

- $G_\lambda : \mathcal{D}(\mathbb{C}) \rightarrow \mathcal{D}(\mathcal{H}_{S,\lambda} \otimes \mathcal{H}_{M,\lambda})$,
- $A_\lambda : \mathcal{D}(\mathcal{H}_{S,\lambda} \otimes \mathcal{H}_{T,\lambda}) \rightarrow \mathcal{D}(\mathcal{H}_{B,\lambda} \otimes \mathcal{H}_{C,\lambda})$,
- $B_\lambda : \mathcal{D}(\mathcal{H}_{K,\lambda} \otimes \mathcal{H}_{B,\lambda}) \rightarrow \mathcal{D}(\mathcal{Q})$, and
- $C_\lambda : \mathcal{D}(\mathcal{H}_{K,\lambda} \otimes \mathcal{H}_{C,\lambda}) \rightarrow \mathcal{D}(\mathcal{Q})$

where $\mathcal{H}_{S,\lambda} = \mathcal{Q}(s(\lambda))$, $\mathcal{H}_{B,\lambda} = \mathcal{Q}(\beta(\lambda))$, and $\mathcal{H}_{C,\lambda} = \mathcal{Q}(\gamma(\lambda))$ for $s, \beta, \gamma : \mathbb{N}^+ \rightarrow \mathbb{N}^+$ and all other Hilbert spaces are as defined by \mathcal{S} .

► **Definition 11** (Uncloneable-Indistinguishable Security). *Let \mathcal{S} be a QECM scheme and define Enc_k^0 and Enc_k^1 as in Definition 6.*

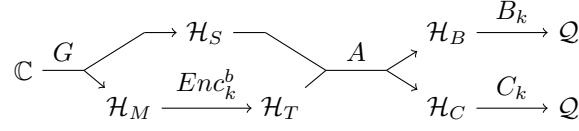
We say that \mathcal{S} is uncloneable-indistinguishable secure if for all cloning-distinguishing attacks \mathcal{A} there exists a negligible function η such that

$$\mathbb{E}_b \mathbb{E}_{k \leftarrow \mathcal{K}} \text{Tr} [(|b\rangle\langle b| \otimes |b\rangle\langle b|) (B_k \otimes C_k) \circ A \circ (\mathbb{1}_S \otimes \text{Enc}_k^b) \circ G(1)] \leq \frac{1}{2} + \eta(\lambda) \quad (15)$$

where λ is implicit on the left-hand side, \mathcal{K}_λ is the random variable distributed on $\{0, 1\}^{\kappa(\lambda)}$ such that $\Pr[\mathcal{K} = k] = \text{Tr}[|k\rangle\langle k| K(1)]$, B_k is the CPTP map defined by $\rho \mapsto B(|k\rangle\langle k| \otimes \rho)$, and similarly for C_k .

The left-hand side of Equation (15) is the probability, averaged all keys, that both adversaries can correctly determine if their submitted message (generated by \mathcal{G}) or the all 0 bit string was encrypted.

It is trivial to see, but worth noting, that uncloneable-indistinguishable security implies indistinguishable security. Indeed, if a scheme is not indistinguishable secure, then an adversary can determine which message was encrypted (with non-negligible advantage) without having to wait for the key to be divulged. Thus, instead of trying to split the ciphertext, the A circuit in an uncloneable-indistinguishable attack should attempt to determine which message was encrypted and simply pass on the result to the B and C circuits.



■ **Figure 3** Schematic representation of the maps considered in a cloning-distinguishing attack (Definition 10). The k subscript indicates which maps have access to the encryption key.

Finally, it can also be shown that any 0-uncloneable secure QECM scheme \mathcal{S} with constant message length is uncloneable-indistinguishable secure. The proof can be found in the full version and proceeds by using any cloning-distinguishing attack to construct a cloning attack. We then show that security against the constructed cloning attack implies security against the original distinguishing-cloning attack.

► **Theorem 12.** *Let \mathcal{S} be an 0-uncloneable secure QECM with constant message size, i.e.: $n(\lambda)$ is the constant function $n(\lambda) = n$, then \mathcal{S} is also uncloneable-indistinguishable secure.*

4 An Uncloneable Encryption Scheme

A first scheme which attempts to achieve a notion of uncloneable encryption is presented in Appendix A. It is based on a simple use of Wiesner states and illustrates the basic principle, but it is in many respects insufficient.

In Section 4.1, we present a refinement of the Appendix A protocol which uses quantum secure pseudorandom functions. The proof of the uncloneable security of this protocol relies on technical lemmas presented in Section 4.2. We give our final main results in Section 4.3.

4.1 Our qPRF Scheme

As discussed in Section 1.1, the motivation for this scheme is to use quantum-secure pseudorandom functions to attempt to “distill” the uncloneability found in the Wiesner state.

► **Definition 13** (\mathcal{F} -Conjugate Encryption). *Let $\mathcal{F} = \{f_\lambda : \{0, 1\}^\lambda \times \{0, 1\}^\lambda \rightarrow \{0, 1\}^{n(\lambda)}\}_{\lambda \in \mathbb{N}^+}$ be a quantum-secure pseudorandom function for a function $n : \mathbb{N}^+ \rightarrow \mathbb{N}^+$. We define the \mathcal{F} -conjugate encryption QECM scheme by the circuits implementing the following algorithms which are implicitly parametrized by λ . Note that the message size is the output size of the qPRF, $n(\lambda)$, the key size is $\kappa(\lambda) = 2\lambda$, and the ciphertext size is $\ell(\lambda) = \lambda + n(\lambda)$.*

■ **Algorithm 1** Key generation circuit, **Key**.

Input : None.

Output : A state $\rho \in \mathcal{D}(\mathcal{Q}(\kappa(\lambda)))$.

- 1 Sample $s \leftarrow \{0, 1\}^\lambda$ uniformly at random.
- 2 Sample $\theta \leftarrow \{0, 1\}^\lambda$ uniformly at random.
- 3 Output $\rho = |s\rangle\langle s| \otimes |\theta\rangle\langle \theta|$.

Algorithm 2 Encryption circuit, *Enc*.

Input : A plaintext $m \in \{0, 1\}^n$ and a key $(s, \theta) \in \{0, 1\}^\kappa$.
Output : A ciphertext $\rho \in \mathcal{D}(\mathcal{Q}(\ell(\lambda)))$.
1 Sample $x \leftarrow \{0, 1\}^\lambda$ uniformly at random.
2 Compute $c = m \oplus f_\lambda(s, x)$.
3 Output $\rho = |c\rangle\langle c| \otimes |x^\theta\rangle\langle x^\theta|$.

Algorithm 3 Decryption circuit, *Dec*.

Input : A ciphertext $|c\rangle\langle c| \otimes \rho \in \mathcal{D}(\mathcal{Q}(\ell))$ and a key $(s, \theta) \in \{0, 1\}^\kappa$.
Output : A plaintext $m \in \{0, 1\}^n$.
1 Compute $\rho' = H^\theta \rho H^\theta$.
2 Measure ρ' in the computational basis. Call the result r .
3 Output $m = c \oplus f_\lambda(s, r)$.

It is trivial to see that this scheme is correct. It is also straightforward to see that this scheme is indistinguishable secure (Definition 6). Indeed, if we replace the qPRF with a function chosen uniformly at random from $\text{Bool}(\lambda, n(\lambda))$, then the ciphertext, averaged over all keys, is independent of the plaintext. Security then follows from the fact that efficient adversaries cannot distinguish with non-negligible advantage between the qPRF and a function chosen randomly from $\text{Bool}(\lambda, n(\lambda))$.

4.2 Technical Lemmas

The following two lemmas form the core of the upcoming proofs of uncloneable security and they can be seen as extending Unruh's one-way-to-hiding lemma [22] to a two player setting. They are interpreted as follows.

We consider two adversaries who have oracle access to a function $H \in \text{Bool}(\lambda, n)$ which is chosen uniformly at random. Their goal is to simultaneously guess the value $H(x)$ for some value of x . The adversaries share a quantum state representing all the information they initially have on x . The lemmas relate the probability of both parties simultaneously guessing $H(x)$ to their probability of being able to both simultaneously guess x .

The first of these lemmas, Lemma 14, considers this problem in a setting where the adversaries do not share any entanglement. The second, Lemma 15, imposes no such restriction.

We show that the probability that both adversaries correctly guess $H(x)$ is upper bounded by $\frac{1}{2^n} + Q \cdot G$ or $\frac{9}{2^n} + Q' \cdot G'$ where Q and Q' are polynomial functions of the number of queries the adversaries make to the oracle and G and G' quantify their probability of guessing x with a particular strategy. The factor of 9 is present only if we allow the adversaries to share entanglement.

We can interpret G and G' in a manner very similar to its analogous quantity in Unruh's one-way-to-hiding lemma [22]. The adversaries, instead of continuing until the end of their computation, will stop immediately before a certain (randomly chosen) query to the oracle and measure their query register in the computational basis. Then, G is related to the probability that this procedure succeeds at letting both adversaries simultaneously obtain x , averaged over the possible stopping points and possible functions implemented by the oracle.

The key idea in the proof of these lemmas is that we can decompose the unitary operator representing each of the adversaries' computations into two "parts" (see Equation (26)). One

of these “parts” will never query the oracle on x and the other could query the oracle on x . This idea was present in the proof of Unruh’s one-way-to-hiding lemma [22].

Recall from Section 2.3 that we model queries to an oracle implementing a function H as a unitary operator O^H acting on a query and a response register with Hilbert spaces \mathcal{H}_Q and \mathcal{H}_R respectively. The action of this unitary operator on the computational basis states is given by $|x\rangle_Q \otimes |y\rangle_R \mapsto |x\rangle_Q \otimes |y \oplus H(x)\rangle_R$. A party having access to an oracle may also have some other register with Hilbert space \mathcal{H}_S with which they perform other computations. In general, their computation can then be modeled by an operator of the form $(UO^H)^q$ where U is a unitary operator on $\mathcal{H}_Q \otimes \mathcal{H}_R \otimes \mathcal{H}_S$ and q is the number of queries made to the oracle [8, 22].

The proof of Lemma 15 can be found in Appendix B.2. The proof of Lemma 14, which uses very similar ideas to those found in the proof of Lemma 15, can be found in the full version.

► **Lemma 14.** *Let $\lambda, n \in \mathbb{N}^+$. For $L \in \{B, C\}$, we let $s_L, q_L \in \mathbb{N}^+$, $\mathcal{H}_{L_Q} = \mathcal{Q}(\lambda)$, $\mathcal{H}_{L_R} = \mathcal{Q}(n)$, $\mathcal{H}_{L_S} = \mathcal{Q}(s_L)$, $U_L \in \mathcal{U}(\mathcal{H}_{L_Q} \otimes \mathcal{H}_{L_R} \otimes \mathcal{H}_{L_S})$, and $\{\pi_L^y\}_{y \in \{0,1\}^n}$ be a projective measurement on $\mathcal{H}_{L_Q} \otimes \mathcal{H}_{L_R} \otimes \mathcal{H}_{L_S}$.*

Finally, let $|\psi\rangle = |\psi_B\rangle \otimes |\psi_C\rangle$ be a separable unit vector with $|\psi_L\rangle \in \mathcal{Q}(n + \lambda + s_L)$ for $L \in \{B, C\}$ and $x \in \{0,1\}^\lambda$. Then, we have

$$\mathbb{E}_H \left\| \Pi^{H(x)} \left((U_B O_B^H)^{q_B} \otimes (U_C O_C^H)^{q_C} \right) |\psi\rangle \right\|^2 \leq \frac{1}{2^n} + (3q + 2)q\sqrt{M} \quad (16)$$

where $\Pi^{H(x)} = \pi_B^{H(x)} \otimes \pi_C^{H(x)}$, $q = q_B + q_C$ and

$$M = \mathbb{E}_k \mathbb{E}_\ell \mathbb{E}_H \mathbb{E}_{H'} \left\| \left(|x\rangle\langle x|_{B_Q} \otimes |x\rangle\langle x|_{C_Q} \right) \left((U_B O_B^H)^k \otimes (U_C O_C^{H'})^\ell \right) |\psi\rangle \right\|^2 \quad (17)$$

with $k \in \{0, \dots, q_B - 1\}$, $\ell \in \{0, \dots, q_C - 1\}$, and $H, H' \in \text{Bool}(\lambda, n)$.

► **Lemma 15.** *Let $\lambda, n \in \mathbb{N}^+$. For $L \in \{B, C\}$, we let $s_L, q_L \in \mathbb{N}^+$, $\mathcal{H}_{L_Q} = \mathcal{Q}(\lambda)$, $\mathcal{H}_{L_R} = \mathcal{Q}(n)$, $\mathcal{H}_{L_S} = \mathcal{Q}(s_L)$, $U_L \in \mathcal{U}(\mathcal{H}_{L_Q} \otimes \mathcal{H}_{L_R} \otimes \mathcal{H}_{L_S})$, and $\{\pi_L^y\}_{y \in \{0,1\}^n}$ be a projective measurement on $\mathcal{H}_{L_Q} \otimes \mathcal{H}_{L_R} \otimes \mathcal{H}_{L_S}$.*

Finally, let $|\psi\rangle \in \mathcal{Q}(2(\lambda + n) + s_B + s_C)$ be a unit vector and $x \in \{0,1\}^\lambda$. Then, we have

$$\mathbb{E}_H \left\| \Pi^{H(x)} \left((U_B O_B^H)^{q_B} \otimes (U_C O_C^H)^{q_C} \right) |\psi\rangle \right\|^2 \leq \frac{9}{2^n} + (3q_B q_C + 2)q_B q_C \sqrt{M} \quad (18)$$

where $\Pi^{H(x)} = \pi_B^{H(x)} \otimes \pi_C^{H(x)}$ and

$$M = \mathbb{E}_k \mathbb{E}_\ell \mathbb{E}_H \left\| \left(|x\rangle\langle x|_{B_Q} \otimes |x\rangle\langle x|_{C_Q} \right) \left((U_B O_B^H)^k \otimes (U_C O_C^H)^\ell \right) |\psi\rangle \right\|^2 \quad (19)$$

with $k \in \{0, \dots, q_B - 1\}$, $\ell \in \{0, \dots, q_C - 1\}$, and $H \in \text{Bool}(\lambda, n)$.

4.3 Main Results

We now have all the necessary tools to state our main results.

► **Theorem 16.** *Let \mathcal{S} be the QECM scheme defined in Definition 13. If the qPRF is modeled by a quantum oracle, then \mathcal{S} is $\log_2(9)$ -uncloneable secure.*

Our main results (Theorem 16) holds under the following assumptions:

1. The family of functions used in the encryption is indistinguishable from truly random functions for efficient adversaries (i.e.: it satisfies the indistinguishable property of a pseudorandom function).

2. The adversarial circuit A (the one which attempts to split the ciphertext) does not know precisely which function was used. This models the idea that the A circuit does not know the encryption key.
3. The adversarial circuits B and C (the ones attempting to guess the plaintext) may only interact with the function as a “black box”.

One way to model these assumptions is to use the quantum random oracle model, where in addition we specify that the A circuit cannot query the oracle. This captures the idea that all circuits, except the A circuit, are given the encryption key.

The above explains our design choice of presenting the scheme with a qPRF, which is modelled as an oracle in the proof. This allows us to assume that the B and C circuits only use the key k to query $f(k, \cdot)$ as a black box. By definition of a qPRF, and since all adversaries are efficient, this scenario is indistinguishable from the random oracle scenario discussed above.

The proof of Theorem 16 can be found in Appendix B.2. It essentially argues that Lemma 15 can be applied with a bound of $M \leq \left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^\lambda$, which is negligible in λ , due to Corollary 2.

We can strengthen this result if the adversaries do not share any entanglement (see Section 1.2 for an application).

► **Theorem 17.** *Let \mathcal{S} be the QECM scheme given in Definition 13. If the qPRF is modeled by a quantum oracle and the adversaries cannot share any entanglement, then \mathcal{S} is 0-uncloneable secure.*

Proof (Sketch). Follow the proof of Theorem 16 using Lemma 14 instead of Lemma 15. ◀

► **Corollary 18.** *Let \mathcal{S} be the QECM scheme given in Definition 13 with constant message size, i.e.: $n(\lambda) = n$. If the qPRF is modeled by a quantum oracle and the adversaries cannot share any entanglement, then \mathcal{S} is indistinguishable-uncloneable secure.*

Proof (Sketch). Use Theorem 17 with Theorem 12. ◀

References

- 1 Scott Aaronson. Quantum copy-protection and quantum money. In *24th Annual Conference on Computational Complexity—CCC 2009*, pages 229–242, 2009. doi:10.1109/CCC.2009.42.
- 2 Scott Aaronson and Paul Christiano. Quantum money from hidden subspaces. In *44th Annual ACM Symposium on Theory of Computing—STOC 2012*, pages 41–60, 2012. doi:10.1145/2213977.2213983.
- 3 Dorit Aharonov, Alexei Kitaev, and Noam Nisan. Quantum circuits with mixed states. In *30th Annual ACM Symposium on Theory of Computing—STOC 1998*, pages 20–30, 1998. doi:10.1145/276698.276708.
- 4 Gorjan Alagic, Anne Broadbent, Bill Fefferman, Tommaso Gagliardini, Christian Schaffner, and Michael St. Jules. Computational security of quantum encryption. In *Information Theoretic Security: 9th International Conference—ICITS 2016*, pages 47–71, 2016. doi:10.1007/978-3-319-49175-2_3.
- 5 Elaine Barker. Recommendation for key management part 1: General (revision 4). Technical Report SP 800-57, National Institute of Standards and Technology, 2016. doi:10.6028/NIST.SP.800-57pt1r4.
- 6 Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. In *International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984. arXiv:2003.06557.

- 7 Charles H. Bennett, Gilles Brassard, and Seth Breidbart. Quantum cryptography II: How to re-use a one-time pad safely even if $P=NP$. *Natural Computing*, 13(4):453–458, 2014. doi:10.1007/s11047-014-9453-6.
- 8 Dan Boneh, Özgür Dagdelen, Marc Fischlin, Anja Lehmann, Christian Schaffner, and Mark Zhandry. Random oracles in a quantum world. In *Advances in Cryptology—ASIACRYPT 2011*, pages 41–69, 2011. doi:10.1007/978-3-642-25385-0_3.
- 9 Anne Broadbent and Christian Schaffner. Quantum cryptography beyond quantum key distribution. *Designs, Codes and Cryptography*, 78(1):351–382, 2016. doi:10.1007/s10623-015-0157-4.
- 10 Ivan Damgård, Thomas Brochmann Pedersen, and Louis Salvail. A quantum cipher with near optimal key-recycling. In *Advances in Cryptology—CRYPTO 2005*, pages 494–510, 2005. doi:10.1007/11535218_30.
- 11 D. Dieks. Communication by EPR devices. *Physics Letters A*, 92(6):271–272, 1982. doi:10.1016/0375-9601(82)90084-6.
- 12 A. Einstein, B. Podolsky, and N. Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review Letters*, 47(10):777–780, 1935. doi:10.1103/physrev.47.777.
- 13 Serge Fehr and Louis Salvail. Quantum authentication and encryption with key recycling. In *Advances in Cryptology—EUROCRYPT 2017*, volume 3, pages 311–338, 2017. doi:10.1007/978-3-319-56617-7_11.
- 14 Shafi Goldwasser and Silvio Micali. Probabilistic encryption. *Journal of Computer and System Sciences*, 28(2):270–299, 1984. doi:10.1016/0022-0000(84)90070-9.
- 15 Daniel Gottesman. Unccloneable encryption. *Quantum Information & Computation*, 3(6):581–602, 2003. arXiv:quant-ph/0210062.
- 16 Sébastien Lord. Unccloneable quantum encryption via random oracles. Master’s thesis, University of Ottawa, 2019. doi:10.20381/ruor-23107.
- 17 Michael A. Nielsen and Issac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- 18 James L. Park. The concept of transition in quantum mechanics. *Foundations of Physics*, 1(1):23–33, 1970. doi:10.1007/BF00708652.
- 19 Peter W. Shor and John Preskill. Simple proof of security of the BB84 quantum key distribution protocol. *Physical Review Letters*, 85(2):441–444, 2000. doi:10.1103/physrevlett.85.441.
- 20 Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New Journal of Physics*, 15(10):103002, 2013. doi:10.1088/1367-2630/15/10/103002.
- 21 Dominique Unruh. Everlasting multi-party computation. In *Advances in Cryptology—CRYPTO 2013*, volume 2, pages 380–397, 2013. doi:10.1007/978-3-642-40084-1_22.
- 22 Dominique Unruh. Revocable quantum timed-release encryption. *Journal of the ACM*, 62(6):49, 2015. doi:10.1145/2817206.
- 23 John Watrous. Quantum computational complexity. In *Encyclopedia of complexity and systems science*, pages 7174–7201. Springer, 2009. doi:10.1007/978-3-642-27737-5_428-3.
- 24 John Watrous. *The Theory of Quantum Information*. Cambridge University Press, 1st edition, 2018.
- 25 Stephen Wiesner. Conjugate coding. *ACM SIGACT News*, 15(1):78–88, 1983. doi:10.1145/1008908.1008920.
- 26 W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982. doi:10.1038/299802a0.
- 27 Mark Zhandry. How to construct quantum random functions. In *53rd Annual Symposium on Foundations of Computer Science—FOCS 2012*, pages 679–687, 2012. doi:10.1109/FOCS.2012.37.

A

 Conjugate Encryption

Our first QECM scheme is a one-time pad encoded into Wiesner states. We emphasize that this scheme will not offer much in terms of uncloneable security but it remains an instructive example.

► **Definition 19** (Conjugate Encryption). *We define the conjugate encryption QECM scheme by the circuits implementing the following algorithms, each implicitly parametrized by λ . Note that the message size is $n(\lambda) = \lambda$, the key size is $\kappa(\lambda) = 2\lambda$ and the ciphertext size is $\ell(\lambda) = \lambda$.*

■ **Algorithm 4** The key generation circuit **Key**.

Input : None.
Output : A state $\rho \in \mathcal{D}(\mathcal{Q}(\kappa))$.
 1 Sample $r \leftarrow \{0, 1\}^n$ uniformly at random.
 2 Sample $\theta \leftarrow \{0, 1\}^n$ uniformly at random.
 3 Output $\rho = |r\rangle\langle r| \otimes |\theta\rangle\langle \theta|$.

■ **Algorithm 5** The encryption circuit **Enc**.

Input : A plaintext $m \in \{0, 1\}^n$ and a key $(r, \theta) \in \{0, 1\}^\kappa$.
Output : A ciphertext $\rho \in \mathcal{D}(\mathcal{Q}(n))$.
 1 Output $\rho = |(m \oplus r)^\theta\rangle\langle (m \oplus r)^\theta|$.

■ **Algorithm 6** The decryption circuit **Dec**.

Input : A ciphertext $\rho \in \mathcal{D}(\mathcal{Q}(n))$ and a key $(r, \theta) \in \{0, 1\}^\kappa$.
Output : A plaintext $m \in \{0, 1\}^n$.
 1 Compute $\rho' = H^\theta \rho H^\theta$.
 2 Measure ρ' in the computational basis. Call the result c . Output $c \oplus r$.

The correctness of this scheme is trivial to verify and it is indistinguishable secure. The indistinguishable security follows from the fact that if $Enc_{r,\theta}^0$ and $Enc_{r,\theta}^1$ are as defined in Definition 6, then for any state $\rho \in \mathcal{D}(\mathcal{H}_S \otimes \mathcal{Q}(n))$ we have that

$$\mathbb{E}_r \mathbb{E}_\theta \left(\mathbb{1}_S \otimes Enc_{(r,\theta)}^1 \right) (\rho) = \mathbb{E}_r \mathbb{E}_\theta \left(\mathbb{1}_S \otimes Enc_{(r,\theta)}^0 \right) (\rho). \quad (20)$$

We will need one small technical lemma before proceeding to the proof of uncloneable security for this scheme.

► **Lemma 20.** *Let $n \in \mathbb{N}^+$, $f : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \mathbb{R}$ be a function and $s \in \{0, 1\}^n$ be a string. Then, $\mathbb{E}_x f(x, x \oplus s) = \mathbb{E}_x f(x \oplus s, x)$.*

The proof of Lemma 20 may be found in the full version.

► **Theorem 21.** *The scheme in Definition 19 is $\lambda \log_2 \left(1 + \frac{1}{\sqrt{2}} \right)$ -uncloneable secure.*

Proof. It suffices to show that for any cloning attack \mathcal{A} the quantity

$$\mathbb{E}_m \mathbb{E}_r \mathbb{E}_\theta \text{Tr} [(|m\rangle\langle m| \otimes |m\rangle\langle m|) (B_{(r,\theta)} \otimes C_{(r,\theta)}) \circ A (|(m \oplus r)^\theta\rangle\langle (m \oplus r)^\theta|)] \quad (21)$$

is upper bounded by $\left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^\lambda$. By applying Lemma 20 with respect to the expectation over m , this quantity is the same as

$$\mathbb{E}_m \mathbb{E}_r \mathbb{E}_\theta \text{Tr} [(|m \oplus r\rangle\langle m \oplus r| \otimes |m \oplus r\rangle\langle m \oplus r|) (B_{(r,\theta)} \otimes C_{(r,\theta)}) \circ A (|m^\theta\rangle\langle m^\theta|)]. \quad (22)$$

We then see that for any fixed r , we can apply Corollary 2 to bound the expectation of the trace over m and θ by $\left(\frac{1}{2} + \frac{1}{2\sqrt{2}}\right)^\lambda$. Setting this quantity to be equal to 2^{-n+t} , recalling that $n = \lambda$, and solving for t completes the proof. \blacktriangleleft

Finally, note that this scheme cannot be uncloneable-indistinguishable secure if $n \geq 2$. Indeed, the adversaries could submit the all 1 plaintext to be encrypted and split the ciphertext such that each adversary gets half of the qubits. Once the key is revealed, the adversaries can then each obtain half of the message with probability 1. This is sufficient to distinguish between the two possible messages.

B Proofs

B.1 From Section 3

Proof of Theorem 9. For all $k \in \{0, 1\}^{\kappa(\lambda)}$ and $m \in \{0, 1\}^n$, define

$$p(k, m) = \text{Tr} [(|m\rangle\langle m| \otimes |m\rangle\langle m|) (B_k \otimes C_k) \circ A \circ \text{Enc}_k (|m\rangle\langle m|)]. \quad (23)$$

Recalling the min-entropy of \mathcal{M} and that \mathcal{S} is t -uncloneable, we may write

$$\begin{aligned} & \mathbb{E}_{m \leftarrow \mathcal{M}} \mathbb{E}_{k \leftarrow \mathcal{K}} \text{Tr} [(|m\rangle\langle m| \otimes |m\rangle\langle m|) (B_k \otimes C_k) \circ A \circ \text{Enc}_k (|m\rangle\langle m|)] \\ &= \sum_{m \in \{0,1\}^n} \Pr[\mathcal{M} = m] \mathbb{E}_{k \leftarrow \mathcal{K}} p(k, m) \leq 2^{-h} \cdot 2^n \mathbb{E}_{m \leftarrow \mathcal{M}} \mathbb{E}_{k \leftarrow \mathcal{K}} p(k, m) \leq 2^{-h} (2^t + 2^n \eta(\lambda)). \end{aligned} \quad (24)$$

Noting that $\lambda \mapsto 2^{-h+n}\eta(\lambda)$ is a negligible function concludes the proof. \blacktriangleleft

B.2 From Section 4

Before giving the proofs of Lemmas 14 and 15, we need the following three small lemmas. The first two, Lemma 22 and Lemma 23, have straightforward proofs which may be found in the full version. The third, Lemma 24, implicitly appears in [22].

► **Lemma 22.** *Let R be a ring with $a, b \in R$ and $c = a + b$. Then, for all $n \in \mathbb{N}^+$, we have that $c^n = a^n + \sum_{k=0}^{n-1} a^{n-k-1} b c^k$.*

► **Lemma 23.** *Let \mathcal{H} be a Hilbert space, $n \in \mathbb{N}^+$, and $\{v_0, v_1, \dots, v_n\}$ be $n+1$ vectors in \mathcal{H} such that $\|v_i\| \leq 1$ for all $i \in \{1, \dots, n\}$ and $\|\sum_{i=0}^n v_i\| \leq 1$. Then, we have that $\|\sum_{i=0}^n v_i\|^2 \leq \|v_0\|^2 + (3n+2) \sum_{i=1}^n \|v_i\|$.*

► **Lemma 24.** *Let $f : \text{Bool}(n, m) \rightarrow \mathbb{R}$ be a function and $x \in \{0, 1\}^n$ be a string. For any $H \in \text{Bool}(n, m)$ and $y \in \{0, 1\}^m$, define $H_{x,y} \in \text{Bool}(n, m)$ by*

$$s \mapsto \begin{cases} H(s) & \text{if } s \neq x, \\ y & \text{if } s = x. \end{cases} \quad (25)$$

Then, $\mathbb{E}_H f(H) = \mathbb{E}_H \mathbb{E}_y f(H_{x,y})$.

We can now give the proofs of our main technical lemma from Section 4.2.

Proof of Lemma 15. For $L \in \{B, C\}$, we define $P_L = |x\rangle\langle x|_{L_Q}$. Using Lemma 22 and the fact that we may write $U_L O_L^H = U_L O_L^H P_L + U_L O_L^H (\mathbb{1} - P_L)$, we have that

$$(U_L O_L^H)^{q_L} = \overbrace{(U_L O_L^H (\mathbb{1} - P_L))^{q_L}}^{=V_L^H} + \sum_{k=0}^{q_L-1} \overbrace{(U_L O_L^H (\mathbb{1} - P_L))^{q_L-k-1} U_L O_L^H P_L (U_L O_L^H)^k}^{=W_L^{H,k}} \quad (26)$$

and we define $W_L^H = \sum_{k=0}^{q_L-1} W_L^{H,k}$. This implies that

$$\begin{aligned} & \left\| \Pi^{H(x)} \left((U_B O_B^H)^{q_B} \otimes (U_C O_C^H)^{q_C} \right) |\psi\rangle \right\|^2 \\ &= \left\| \Pi^{H(x)} \left((O_B O_B^H)^{q_B} \otimes V_C^H + V_B^H \otimes W_C^H + W_B^H \otimes W_C^H \right) |\psi\rangle \right\|^2. \end{aligned} \quad (27)$$

We now claim that the contribution from the $W_B^H \otimes W_C^H$ operator corresponds to the M in the upper bound provided in the statement. Indeed, using Lemma 23, the definition of the various W operators, and properties of the operator norm on projectors and unitary operators, we have that

$$\begin{aligned} & \left\| \Pi^{H(x)} \left((O_B O_B^H)^{q_B} \otimes V_C^H + V_B^H \otimes W_C^H + W_B^H \otimes W_C^H \right) |\psi\rangle \right\|^2 \\ & \leq \left\| \Pi^{H(x)} \left((O_B O_B^H)^{q_B} \otimes V_C^H + V_B^H \otimes W_C^H \right) |\psi\rangle \right\|^2 \\ & \quad + (3q_B q_C + 2)q_B q_C \mathbb{E}_k \mathbb{E}_\ell \left\| (P_B \otimes P_C) \left((U_B O_B^H)^k \otimes (U_C O_C^H)^\ell \right) |\psi\rangle \right\|. \end{aligned} \quad (28)$$

Using Jensen's inequality, the above inequality and the definition of M , we have that

$$\begin{aligned} & \mathbb{E}_H \left\| \Pi^{H(x)} \left((O_B O_B^H)^{q_B} \otimes V_C^H + V_B^H \otimes W_C^H + W_B^H \otimes W_C^H \right) |\psi\rangle \right\|^2 \\ & \leq \mathbb{E}_H \left\| \Pi^{H(x)} \left((O_B O_B^H)^{q_B} \otimes V_C^H + V_B^H \otimes W_C^H \right) |\psi\rangle \right\|^2 + (3q_B q_C + 2)q_B q_C \sqrt{M}. \end{aligned} \quad (29)$$

It now suffices to show that

$$\mathbb{E}_H \left\| \Pi^{H(x)} \left((O_B O_B^H)^{q_B} \otimes V_C^H + V_B^H \otimes W_C^H \right) |\psi\rangle \right\|^2 \leq \frac{9}{2^n}. \quad (30)$$

By Lemma 24, this is equivalent to showing that

$$\mathbb{E}_H \mathbb{E}_y \left\| \Pi^y \left((U_B O_B^{H_{x,y}})^{q_B} \otimes V_C^{H_{x,y}} + V_B^{H_{x,y}} \otimes W_C^{H_{x,y}} \right) |\psi\rangle \right\|^2 \leq \frac{9}{2^n}. \quad (31)$$

In fact, it will be sufficient to show that for any particular H , the expectation over y is bounded by $9 \cdot 2^{-n}$. If, for any H , we define

$$\alpha = \mathbb{E}_y \left\| \Pi^y \left((U_B O_B^{H_{x,y}})^{q_B} \otimes V_C^{H_{x,y}} \right) |\psi\rangle \right\|^2 \quad \text{and} \quad \beta = \mathbb{E}_y \left\| \Pi^y \left(V_B^{H_{x,y}} \otimes W_C^{H_{x,y}} \right) |\psi\rangle \right\|^2 \quad (32)$$

then, using the triangle inequality and the fact that the operators in $\{\Pi^y\}_{y \in \{0,1\}^n}$ project on mutually orthogonal subspaces, we have that

$$\mathbb{E}_y \left\| \Pi^y \left((O_B O_B^{H_{x,y}})^{q_B} \otimes V_C^{H_{x,y}} + V_B^{H_{x,y}} \otimes W_C^{H_{x,y}} \right) |\psi\rangle \right\|^2 \leq \alpha + \beta + 2\sqrt{\alpha\beta}. \quad (33)$$

Now, noting that $V_B^{H_{x,y}}$ and $V_C^{H_{x,y}}$ do not depend on the value of y , as they always project on a subspace which does not query the oracle H on x , and using properties of the operator norm, we have that

$$\begin{aligned}
 \alpha &= \mathbb{E}_y \left\| \Pi^y \left(\left(U_B O_B^{H_{x,y}} \right)^{q_B} \otimes V_C^{H_{x,y}} \right) |\psi\rangle \right\|^2 \\
 &\leq \mathbb{E}_y \left\| \pi_B^y \otimes \mathbb{1}_C \right\|^2 \cdot \left\| \left(U_B O_B^{H_{x,y}} \right)^{q_B} \otimes \mathbb{1}_C \right\|^2 \cdot \left\| (\mathbb{1}_B \otimes \pi_C^y) (\mathbb{1}_B \otimes V_C^{H_{x,y}}) |\psi\rangle \right\|^2 \\
 &\leq \mathbb{E}_y \left\| (\mathbb{1}_B \otimes \pi_C^y) (\mathbb{1}_B \otimes V_C^{H_{x,y}}) |\psi\rangle \right\|^2 \\
 &= \frac{1}{2^n} \left\| (\mathbb{1}_B \otimes V_C^H) |\psi\rangle \right\|^2 \leq \frac{1}{2^n}.
 \end{aligned} \tag{34}$$

A similar reasoning yields that $\beta \leq 4 \cdot 2^{-n}$, where the 4 is a result of squaring the upper bound

$$\left\| W_C^{H_{x,y}} \right\| \leq \left\| \left(U_C O_C^{H_{x,y}} \right)^{q_C} \right\| + \left\| V_C^{H_{x,y}} \right\| \leq 2. \tag{35}$$

Finally, noting that $\alpha + \beta + 2\sqrt{\alpha\beta} \leq 9 \cdot 2^{-n}$ finishes the proof. \blacktriangleleft

Finally, we can give the proof of our main result.

Proof of Theorem 16. Let $\mathcal{A} = (\mathbf{A}, \mathbf{B}, \mathbf{C})$ be a cloning attack against \mathcal{S} as described in Definition 7. We need to show that the probability that the adversaries can simultaneously guess a message chosen uniformly at random is upper bounded by $9 \cdot 2^{-n} + \eta(\lambda)$ for a negligible function η . Furthermore, since the adversaries treat the qPRF as an oracle, it suffices to show that their winning probability is upper bounded by $9 \cdot 2^{-n} + \eta(\lambda)$ when averaged over all functions in $\text{Bool}(\lambda, n)$ and not only the functions $\{f_\lambda(s, \cdot)\}_{s \in \{0,1\}^\lambda}$. Indeed, by definition of a qPRF, their winning probability in both cases can differ by at most a negligible function of λ .

The remainder of the proof is an application of Lemma 15 followed by an application of Corollary 2.

Accounting for the randomness of the encryption and for a fixed and implicit λ , the quantity we wish to bound is given by

$$\omega = \mathbb{E}_H \mathbb{E}_\theta \mathbb{E}_x \mathbb{E}_m \text{Tr} \left[P^m (B_\theta^H \otimes C_\theta^H) \circ A(|m \oplus H(x)\rangle\langle m \oplus H(x)| \otimes |x^\theta\rangle\langle x^\theta|) \right] \tag{36}$$

where $P^m = |m\rangle\langle m| \otimes |m\rangle\langle m|$ and $H \in \text{Bool}(\lambda, n)$. Then, by using Lemma 20 with respect to the expectation over m to move the dependence on the string $H(x)$ from the state to the projector, we have that

$$\omega = \mathbb{E}_H \mathbb{E}_\theta \mathbb{E}_x \mathbb{E}_m \text{Tr} \left[P^{m \oplus H(x)} (B_\theta^H \otimes C_\theta^H) \circ A(|m\rangle\langle m| \otimes |x^\theta\rangle\langle x^\theta|) \right]. \tag{37}$$

Using standard purification arguments, we add auxiliary states $|\text{aux-B}\rangle\langle\text{aux-B}|$ and $|\text{aux-C}\rangle\langle\text{aux-C}|$ to the state $A(|m\rangle\langle m| \otimes |x^\theta\rangle\langle x^\theta|)$, replace the CPTP maps B_θ^H and C_θ^H by unitary operators on the resulting larger Hilbert spaces and replace the projectors $|m\rangle\langle m|$ by projectors $\{\pi_B^m\}_{m \in \{0,1\}^n}$ and $\{\pi_C^m\}_{m \in \{0,1\}^n}$ on these larger Hilbert spaces.

Following [8], these purified unitary operators will be of the form $(U_L^\theta O_L^H)^{q_L}$, acting on a Hilbert space of the form $\mathcal{Q}(\lambda)_{L_Q} \otimes \mathcal{Q}(n)_{L_R} \otimes \mathcal{Q}(s_L)_{L_S}$ for some $q_L, s_L \in \mathbb{N}^+$ as they model

4:22 Uncloneable Quantum Encryption via Oracles

oracle computations. In particular, we note that q_L represents the number of queries made to the oracle by that particular party. We also assume that

$$\begin{aligned} \rho^{m,x,\theta} &= A(|m\rangle\langle m| \otimes |x^\theta\rangle\langle x^\theta|) \otimes |\text{aux-B}\rangle\langle \text{aux-B}| \otimes |\text{aux-C}\rangle\langle \text{aux-C}| \\ &\in \mathcal{D}(\mathcal{Q}(\lambda)_{B_Q} \otimes \mathcal{Q}(n)_{B_R} \otimes \mathcal{Q}(s_B)_{B_S} \otimes \mathcal{Q}(\lambda)_{C_Q} \otimes \mathcal{Q}(n)_{C_R} \otimes \mathcal{Q}(s_C)_{C_S}). \end{aligned} \quad (38)$$

Next, we can write $\rho^{m,x,\theta}$ as an ensemble of pure states, which is to say that

$$\rho^{m,x,\theta} = \sum_{i \in I^{m,x,\theta}} p_i \left| \psi_i^{m,x,\theta} \right\rangle \left\langle \psi_i^{m,x,\theta} \right| \quad (39)$$

for some index set $I^{m,x,\theta}$, some non-zero p_i which sum to 1, and some unit vectors $\left| \psi_i^{m,x,\theta} \right\rangle$. It then follows that ω can be expressed as

$$\mathbb{E}_m \mathbb{E}_\theta \mathbb{E}_x \mathbb{E}_H \sum_{i \in I^{m,x,\theta}} p_i \left\| \left(\pi_B^{m \oplus H(x)} \otimes \pi_C^{m \oplus H(x)} \right) \left((U_B^\theta O_B^H)^{q_B} \otimes (U_C^\theta O_C^H)^{q_C} \right) \left| \psi_i^{m,x,\theta} \right\rangle \right\|^2. \quad (40)$$

Noting that we can bring the expectation with respect to H into the summation, we can then use Lemma 15 to upper bound ω by

$$\frac{9}{2^n} + q \mathbb{E}_m \mathbb{E}_\theta \mathbb{E}_x \sum_{i \in I^{m,x,\theta}} p_i \sqrt{\mathbb{E}_H \mathbb{E}_k \mathbb{E}_\ell \left\| Q_x \left((U_B O_B^H)^{q_B} \otimes (U_C O_C^H)^{q_C} \right) \left| \psi_i^{m,x,\theta} \right\rangle \right\|^2} \quad (41)$$

where $q = (3q_B q_C + 2)q_B q_C$ and $Q_x = |x\rangle\langle x|_{Q_B} \otimes |x\rangle\langle x|_{Q_C}$. Defining

$$\beta_x^{\theta,H,k} = \left((U_B^\theta O_B^H)^{q_B} \right)^\dagger |x\rangle\langle x|_{Q_B} \left((U_B^\theta O_B^H)^{q_B} \right), \quad (42)$$

and similarly for $\gamma_x^{\theta,H,\ell}$ by replacing every instance of B with C , we use Jensen's lemma to bring the remaining expectations and sums into the square root and obtain

$$\omega = \frac{9}{2^n} + q \sqrt{\mathbb{E}_m \mathbb{E}_\theta \mathbb{E}_x \mathbb{E}_H \mathbb{E}_k \mathbb{E}_\ell \text{Tr} \left[\left(\beta_x^{\theta,H,k} \otimes \gamma_x^{\theta,H,\ell} \right) \rho^{m,x,\theta} \right]}. \quad (43)$$

Letting Φ_m to be the CPTP map defined by

$$\rho \mapsto A(|m\rangle\langle m| \otimes \rho) \otimes |\text{aux-B}\rangle\langle \text{aux-B}| \otimes |\text{aux-C}\rangle\langle \text{aux-C}| \quad (44)$$

we see that, for any fixed H , k , ℓ , and m , Corollary 2 implies that

$$\mathbb{E}_x \mathbb{E}_\theta \text{Tr} \left[\left(\beta_x^{\theta,H,k} \otimes \gamma_x^{\theta,H,\ell} \right) \rho^{m,x,\theta} \right] \leq \left(\frac{1}{2} + \frac{1}{2\sqrt{2}} \right)^\lambda \quad (45)$$

since $\rho^{m,x,\theta} = \Phi_m(|x^\theta\rangle\langle x^\theta|)$. Thus,

$$\omega \leq \frac{9}{2^n} + q \left(\sqrt{\frac{1}{2} + \frac{1}{2\sqrt{2}}} \right)^\lambda. \quad (46)$$

Finally, since B and C are efficient quantum circuits, they may query the oracle a number of time which grows at most polynomially in λ . Thus, $q \leq p(\lambda)$ for some polynomial p . Noting that the function $\lambda \mapsto p(\lambda) \cdot \left(\sqrt{\frac{1}{2} + \frac{1}{2\sqrt{2}}} \right)^\lambda$ is a negligible function completes the proof. \blacktriangleleft

Quasirandom Quantum Channels

Tom Bannink

CWI, 1098 XG Amsterdam, Netherlands
QuSoft, Science Park 123, 1098 XG Amsterdam, Netherlands
tombannink@gmail.com

Jop Briët

CWI, 1098 XG Amsterdam, Netherlands
QuSoft, Science Park 123, 1098 XG Amsterdam, Netherlands
j.briet@cwi.nl

Farrokh Labib

CWI, 1098 XG Amsterdam, Netherlands
QuSoft, Science Park 123, 1098 XG Amsterdam, Netherlands
labib@cwi.nl

Hans Maassen

QuSoft, Science Park 123, 1098 XG Amsterdam, Netherlands
Korteweg-de Vries Institute for Mathematics, Radboud University, Nijmegen, Netherlands
H.Maassen@math.ru.nl

Abstract

Mixing (or quasirandom) properties of the natural transition matrix associated to a graph can be quantified by its distance to the complete graph. Different mixing properties correspond to different norms to measure this distance. For dense graphs, two such properties known as spectral expansion and uniformity were shown to be equivalent in seminal 1989 work of Chung, Graham and Wilson. Recently, Conlon and Zhao extended this equivalence to the case of sparse vertex transitive graphs using the famous Grothendieck inequality.

Here we generalize these results to the non-commutative, or “quantum”, case, where a transition matrix becomes a quantum channel. In particular, we show that for irreducibly covariant quantum channels, expansion is equivalent to a natural analog of uniformity for graphs, generalizing the result of Conlon and Zhao. Moreover, we show that in these results, the non-commutative and commutative (resp.) Grothendieck inequalities yield the best-possible constants.

2012 ACM Subject Classification Theory of computation → Quantum information theory

Keywords and phrases Quantum channels, quantum expanders, quasirandomness

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.5

Funding *Tom Bannink*: Supported by the Gravitation-grant NETWORKS-024.002.003 from the Dutch Research Council (NWO).

Jop Briët: Supported by the Gravitation-grant NETWORKS-024.002.003 from the Dutch Research Council (NWO). Additionally supported by an NWO VENI grant.

Farrokh Labib: Supported by the Gravitation-grant NETWORKS-024.002.003 from the Dutch Research Council (NWO).

Acknowledgements We would like to thank Māris Ozols, Michael Walter and Freek Witteveen for fruitful discussions.

1 Introduction

In a seminal work [8], Chung, Graham and Wilson – building on work of Thomason [33, 34] – proved that several seemingly distinct notions of quasirandomness for graphs are equivalent. In particular, they identified seven properties found in random graphs with high probability, that always coexist simultaneously in any large dense graph. Two of these properties are



© Tom Bannink, Jop Briët, Farrokh Labib, and Hans Maassen;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 5; pp. 5:1–5:20



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

spectral expansion and *uniformity* (defined below). A question of Chung and Graham [7] on the equivalence of these two properties in *sparse* graphs resulted in a line of research culminating in recent work of Conlon and Zhao [9], which introduced a surprising new item to the armory of combinatorics: the famous Grothendieck inequality [13]. In this paper, we draw a parallel line in the context of quantum information theory, where quantum channels take the place of graphs. In addition, we give a streamlined proof of the main result of [9] and show that the use of Grothendieck's inequality yields an optimal constant. Similarly, we show that the non-commutative Grothendieck inequality gives an optimal constant in the quantum setting.

Spectral expansion and uniformity

Spectral expansion is a linear-algebraic property given in terms of the transition matrix of a graph. This transition matrix is the normalized adjacency matrix, which for a d -regular graph $G = (V, E)$ is given by $A_{uv} = e(\{u\}, \{v\})/d$, where $e(S, T)$ denotes the number of edges connecting subsets $S, T \subseteq V$. We say that the graph G is an (n, d, λ) graph if $|V| = n$, it is d -regular and all but the largest eigenvalue of A , which is always 1, have modulus at most λ . The smallest value of λ for which this holds is denoted by $\lambda(G)$. Spectral expansion then refers to the property that $\lambda(G)$ is much smaller than 1, in which case G is referred to as a (spectral) expander. Expanders have many important applications in mathematics and computer science (we refer to [23] for an extensive survey). One such application is in randomized algorithms, which can exploit the fact that a random walk on an expander rapidly mixes (i.e. quickly converges to its limit distribution) to significantly reduce the amount of randomness needed.

Uniformity is a combinatorial property of the configuration of the edges. An n -vertex d -regular graph $G = (V, E)$ is ϵ -uniform if for all $S, T \subseteq V$,

$$\left| e(S, T) - \frac{d}{n} |S| |T| \right| \leq \epsilon dn \quad (1)$$

and $\epsilon(G)$ denotes the smallest value of ϵ for which this holds. Uniformity then refers to the property that this parameter is much smaller than 1; trivially any graph is 1-uniform. Intuitively, this says that for any two vertex subsets, the number of edges between those sets is close to the expected number of edges in a random graph with the same edge density.

A basic result known as the Expander Mixing Lemma [23] shows that for any regular graph G we have $\epsilon(G) \leq \lambda(G)$, which is to say that spectral expansion implies uniformity. A sequence G_n of d_n -regular graphs is called *dense* if $d_n \geq \Omega(n)$, and *sparse* if $d_n/n \rightarrow 0$. It was shown in [8] that in the dense case, a converse to the Expander Mixing Lemma $\epsilon(G_n) \leq o(1) \Rightarrow \lambda(G_n) \leq o(1)$ also holds. In contrast, Krivelevich and Sudakov [25] showed that this is false for sparse graphs, thereby answering the question posed in [7]. Their counterexample is not regular, however (and a later one from [4] is not connected). But in [9] it was shown that even regular sparse graphs (where $d_n \leq o(n)$) can simultaneously satisfy $\epsilon(G_n) \leq o(1)$ and $\lambda(G_n) \geq \Omega(1)$. Surprisingly, Kohayakawa, Rödl, and Schacht [24] showed that Cayley graphs over abelian groups, including sparse ones, do again admit such a converse. Cayley graphs are an important class of regular graphs that include for instance the famous Ramanujan graphs of Margulis [27] and Lubotzky, Phillips and Sarnak [26]. Conlon and Zhao [9] generalized this to all Cayley graphs and showed that this implies the same for all vertex-transitive graphs in general, for which they showed that $\lambda(G) \leq 4K_G \epsilon(G)$, where $1.6769 \dots \leq K_G < 1.7822 \dots$ is the famous *Grothendieck constant*, whose exact value is currently unknown; the bounds shown here are the best known and were shown by Davie

and Reeds (independently) in [11, 30] and Braverman et al. in [5], respectively. Spectral expansion and uniformity are thus equivalent notions of quasirandomness for dense graphs and vertex-transitive graphs.

Quasirandomness in quantum information theory

A transition matrix, such as the normalized adjacency matrix of a graph, maps probability vectors¹ to probability vectors. A natural non-commutative generalization of a transition matrix is a *quantum channel*, a completely positive trace preserving linear map $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$; see Section 2 for formal definitions. Quantum channels are the most general operations on quantum systems that are physically realizable. They encapsulate the “classical” transition matrices by restricting them to diagonal matrices whose diagonals form probability vectors; we discuss this in more detail in Section 3. In quantum information theory, general linear maps from $M_n(\mathbb{C})$ to itself are referred to as *superoperators*. Since superoperators are in one-to-one correspondence with bilinear forms on $M_n(\mathbb{C}) \times M_n(\mathbb{C})$, they also appear in the context of (generalizations of) Bell inequalities from physics in the form of quantum XOR games [31, 10], as well as in combinatorial optimization [28].

The graph-theoretic concepts mentioned above have natural analogues for superoperators, which we discuss next.

In independent work, Hastings [18] and Ben-Aroya, Schwartz and Ta-Schma [3] introduced *quantum expanders* as a special class of quantum channels defined analogously to spectral expanders. For a unital² quantum channel Φ , the expansion parameter is given by

$$\lambda(\Phi) = \|\Phi - \Pi\|_{S_2 \rightarrow S_2} = \sup \{ \|(\Phi - \Pi)(X)\|_{S_2} : \|X\|_{S_2} \leq 1 \}, \quad (2)$$

where $\Pi : X \mapsto \frac{1}{n} \text{Tr}(X) \text{Id}$ is the projection onto the identity, $\|X\|_{S_2} = \sqrt{\langle X, X \rangle}$ is the Frobenius (or Schatten-2) norm and $\langle X, Y \rangle = \frac{1}{n} \text{Tr}(Y^* X)$ is the normalized trace inner product. A quantum channel is an expander if $\lambda(\Phi)$ is much smaller than 1. Also quantum expanders found many applications, one of which is again randomness reduction, where randomness takes on the form of random unitary matrices. Since a k -qubit unitary requires 4^k real parameters, sampling one from the uniform distribution (Haar probability measure) is very expensive. A 1-design is a fixed collection of unitaries U_1, \dots, U_m such that the superoperator $\Phi(X) = \frac{1}{m} \sum_{i=1}^m U_i X U_i^*$ exactly effects the projection Π , thus mimicking in a finite way the Haar measure on $U(n)$. Quantum expanders can be used to construct *approximate* 1-designs, meaning that $\Phi(X)$ and $\Pi(X)$ are close in trace distance³ instead of precisely equal. Another application is in cryptography where Ambainis and Smith [1] used quantum expanders to construct short quantum one-time pads. It was shown in [18] that truly random quantum channels (given by independent Haar-uniform U_i as described above) are quantum expanders with high probability, supporting the idea that this is a notion of quasirandomness.

In this work we introduce a natural notion of uniformity for superoperators, informally given by how well they mimic the action of Π on projectors on subspaces, which may be thought of as generalizations of vertex subsets in graphs. This is similar to Hasting’s notion of edge expansion for quantum channels [18]. In particular, we say that Φ is ϵ -uniform if for any two subspaces $V, W \subseteq \mathbb{C}^n$ with associated projections P_V, P_W , it holds that

$$|\langle P_V, (\Phi - \Pi)(P_W) \rangle| \leq \epsilon. \quad (3)$$

¹ We use the convention of writing probability vectors as *column* vectors instead of row vectors.

² This is the superoperator analogue of regularity for graphs, defined in Section 2.

³ The trace distance is the distance induced by the Schatten-1 norm, defined in Section 2.

Let $\epsilon(\Phi)$ denote the smallest ϵ for which this holds. As we show in Section 3.3, the parameters $\lambda(\Phi)$ and $\epsilon(\Phi)$ reduce to their graphical analogs under a suitable embedding of graphs into quantum channels.

Finally, also symmetry, which in the graph-theoretic context takes the form of vertex transitivity, is an important property of quantum channels. In particular, *irreducibly covariant* quantum channels, which turn out to generalize vertex-transitive graphs (see Section 3), play an important role in questions about the capacity of quantum channels as noisy transmitters of quantum information [22]. A now famous result of Hastings [19] shows that the minimum output capacity in general does not have the intuitively natural property of being sub-additive under tensor products. However, it was shown earlier by Holevo [21], that the capacity is additive for the subclass of irreducibly covariant quantum channels.

Summary of our results

In this work we make a first step in the study of the equivalence of quasirandom properties for quantum channels, or superoperators in general, and show optimality in the case of vertex-transitive graphs and covariant quantum channels.

- (Section 3.2) Our main result shows that under irreducible covariance, expansion and uniformity are equivalent for superoperators. In particular, while a simple analogue of the classical Expander Mixing Lemma implies that $\epsilon(\Phi) \leq \lambda(\Phi)$ in general, we show using a non-commutative version of Grothendieck's inequality due to Haagerup [14], that for this class of superoperators, also $\lambda(\Phi) \leq 2\pi^2\epsilon(\Phi)$ always holds. This implies the same result for vertex-transitive graphs with \mathbb{C} -weighted edges, essentially proved in [9] with the factor 2 replaced by the *complex* Grothendieck constant $1.3380\dots \leq K_G^{\mathbb{C}} \leq 1.4049\dots$
- (Section 3.3) We show that a construction of sparse regular graphs from [9] can be embedded to give a sequence of quantum channels Φ_n that are not irreducibly covariant and for which it holds that $\epsilon(\Phi_n) \leq o(1)$ and $\lambda(\Phi_n) \geq \Omega(1)$.
- (Section 3.4) We show that for *randomizing* channels, a notion introduced in [2], the two notions of quasirandomness are also equivalent. This can be interpreted as a generalization of the same statement for dense graphs proved in [8].
- (Section 4.1) We show that the result of [9] cannot be improved in the sense that the factors $4K_G$ and $\pi^2 K_G^{\mathbb{C}}$ are optimal in the case of vertex-transitive graphs with \mathbb{R} -weighted and \mathbb{C} -weighted edges, respectively.
- (Section 4.2) Our work leaves open whether the factor $2\pi^2$ in our main result is optimal. However, our proof consists of two steps, the first of which gives a factor 2 and the second a factor π^2 , and we show these steps are individually optimal. We prove that the first step is optimal by showing that an example of Haagerup and Ito [16] for the non-commutative Grothendieck inequality is irreducibly covariant, which uses some representation theory of $SO(n)$. The optimality of the second step follows directly from a result of [9].

2 Preliminaries

Write $[n] = \{1, \dots, n\}$. For a finite set S , write $\mathbb{E}_{s \in S}$ for $\frac{1}{|S|} \sum_{s \in S}$. For a compact set S , write $C(S)$ for the set of continuous functions from S to \mathbb{C} . For a compact group Γ , write $\mathbb{E}_{g \in \Gamma}$ for the integral with respect to the (unique) Haar probability measure on Γ .

Write $M_n(\mathbb{C})$ for the set of complex $n \times n$ matrices and let $U(n) = \{X \in M_n(\mathbb{C}) : X^*X = \text{Id}\}$ be the set of unitary matrices. Here, all maps of the form $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ are linear, and we refer to these as superoperators. A superoperator Φ is *unital* if $\Phi(\text{Id}) = \text{Id}$

and it is *completely positive* if for all $k \in \mathbb{N}$ the superoperator $\text{Id} \otimes \Phi : M_k \otimes M_n \rightarrow M_k \otimes M_n$ maps positive semidefinite matrices to positive semidefinite matrices. Completely positive superoperators that are trace preserving are called *quantum channels*.

We normalize inner products so that for $x, y \in \mathbb{C}^n$ we define $\langle y, x \rangle = \mathbb{E}_{i \in [n]} \bar{y}_i x_i$ and for matrices $X, Y \in M_n(\mathbb{C})$ we have $\langle Y, X \rangle = \frac{1}{n} \text{Tr}[Y^* X]$.

Norms

For $p \in [1, \infty)$, $x \in \mathbb{C}^n$ and $X \in M_n(\mathbb{C})$, the L_p norm and (normalized) Schatten- p norm are defined by

$$\|x\|_{L_p} = \left(\mathbb{E}_{i \in [n]} |x_i|^p \right)^{1/p} \quad \text{and} \quad \|X\|_{S_p} = \left(\frac{1}{n} \text{Tr}[(X^* X)^{p/2}] \right)^{1/p}$$

and $\|x\|_{L_\infty} = \max_i |x_i|$ and $\|X\|_{S_\infty} = \sup\{|\langle Xx, y \rangle| : \|x\|_{L_2}, \|y\|_{L_2} \leq 1\}$. Note that for the identity matrix $\text{Id} \in M_n$ we have $\|\text{Id}\|_{S_p} = 1$ for all $p \in [1, \infty]$.

► **Proposition 1.** *Let $p \geq 1$ and let $X \in M_n(\mathbb{C})$. Then $\|X\|_{S_p} \geq \|(X_{11}, \dots, X_{nn})\|_{L_p}$.*

Proof. For a vector $x \in \mathbb{C}^n$, denote by $\text{Diag}(x)$ the $n \times n$ matrix with x on the diagonal and for a matrix X denote by $\text{diag}(X)$ the matrix where we set the off-diagonal elements to 0. A small computation shows that

$$\mathbb{E}_{s \in \{\pm 1\}^n} \text{Diag}(s) X \text{Diag}(s) = \text{diag}(X).$$

Since the Schatten- p norms are invariant under conjugation with a unitary matrix, applying the above with the triangle inequality gives

$$\|(X_{11}, \dots, X_{nn})\|_{L_p} = \|\text{diag}(X)\|_{S_p} \leq \mathbb{E}_{s \in \{\pm 1\}^n} \|\text{Diag}(s) X \text{Diag}(s)\|_{S_p} = \|X\|_{S_p}. \quad \blacktriangleleft$$

For $q \in [1, \infty]$, define $q' \in [1, \infty]$ to be its dual given by $\frac{1}{q} + \frac{1}{q'} = 1$. For $p, q \in [1, \infty]$, a matrix $A \in M_n(\mathbb{C})$ and a superoperator $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$, define

$$\begin{aligned} \|A\|_{L_p \rightarrow L_q} &= \sup\{|\langle y, Ax \rangle| : \|x\|_{L_p} \leq 1, \|y\|_{L_{q'}} \leq 1\} \\ \|\Phi\|_{S_p \rightarrow S_q} &= \sup\{|\langle Y, \Phi(X) \rangle| : \|X\|_{S_p} \leq 1, \|Y\|_{S_{q'}} \leq 1\}. \end{aligned}$$

Also define the *cut norms* by

$$\begin{aligned} \|A\|_{\text{cut}} &= \max\{|\langle y, Ax \rangle| : x, y \in \{0, 1\}^n\} \\ \|\Phi\|_{\text{cut}} &= \sup\{|\langle Y, \Phi(X) \rangle| : X, Y \text{ projectors}\}. \end{aligned}$$

It is then not hard to see that if G is a d -regular graph with normalized adjacency matrix A , then $\epsilon(G) = \|A - \frac{1}{n} J\|_{\text{cut}}$, where J is the all-ones matrix. Similarly, we have $\epsilon(\Phi) = \|\Phi - \Pi\|_{\text{cut}}$.

We have the following relation between these norms, the proof of which is a simple generalization of the same result from [9] for matrices.

► **Lemma 2.** *For any superoperator Φ , we have $\|\Phi\|_{\text{cut}} \leq \|\Phi\|_{S_\infty \rightarrow S_1} \leq \pi^2 \|\Phi\|_{\text{cut}}$ and π^2 is the best possible constant.*

Proof. First note that the cut norm as defined above can also be written as

$$\|\Phi\|_{\text{cut}} = \sup\{|\langle Y, \Phi(X) \rangle| : X, Y \succeq 0, \|X\|_{S_\infty}, \|Y\|_{S_\infty} \leq 1\}, \quad (4)$$

because the set $\{X : X \succeq 0, \|X\|_{S_\infty} \leq 1\}$ is the convex hull of the set of projectors. Hence, by linearity the supremum in (4) will always be attained by projectors.

The first inequality of the lemma follows by dropping the positive semidefinite constraint. For the second inequality, let z be a complex number of norm 1, and w a uniform random complex number of norm 1. Then

$$z = \pi \mathbb{E}_w [w \mathbf{1}_{\{\Re(z\bar{w}) \geq 0\}}].$$

Note that $\mathbb{E}_w[f(w)] = \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) d\theta$, hence the equality follows by using $\int_{-\pi/2}^{\pi/2} \cos(\theta) d\theta = 2$. We have $\|\Phi\|_{S_\infty \rightarrow S_1} = \sup\{|\langle Y, \Phi(X) \rangle| : \|X\|_{S_\infty}, \|Y\|_{S_\infty} \leq 1\}$. The set of matrices X such that $\|X\|_{S_\infty} \leq 1$ is the convex hull of the set of unitary matrices, so by linearity we can assume that the supremum in $\|\Phi\|_{S_\infty \rightarrow S_1}$ is obtained by unitary X, Y . Unitary matrices are diagonalizable, so write $X = UAU^*$ and $Y = VBV^*$ with U, V unitary and A, B diagonal. Let $u, w \in \mathbb{C}$, $|u| = |w| = 1$ be uniform random complex numbers and define diagonal matrices A', B' as $A'_{ii}(w) = \mathbf{1}_{\{\Re(A_{ii}w) \geq 0\}}$ and $B'_{ii}(u) = \mathbf{1}_{\{\Re(B_{ii}\bar{u}) \geq 0\}}$. By the above we have $A = \pi \mathbb{E}_w[wA'(w)]$ and similar for B , so we have $X = \pi \mathbb{E}_w[wUA'(w)U^*]$ and $Y = \pi \mathbb{E}_u[uVB'(u)V^*]$. Now, $UA'(w)U^*$ and $VB'(u)V^*$ are projections for all values of w and u , as required in the definition of the cut norm. Therefore

$$\begin{aligned} \|\Phi\|_{S_\infty \rightarrow S_1} &= |\langle Y, \Phi(X) \rangle| = \pi^2 |\mathbb{E}_{u,w} \bar{u} w \langle VB'(u)V^*, \Phi(UA'(w)U^*) \rangle| \\ &\leq \pi^2 \mathbb{E}_{u,w} |\langle VB'(u)V^*, \Phi(UA'(w)U^*) \rangle| \\ &\leq \pi^2 \mathbb{E}_{u,w} \|\Phi\|_{\text{cut}} \\ &= \pi^2 \|\Phi\|_{\text{cut}}, \end{aligned}$$

completing the first part of the proof. Conlon and Zhao show that π^2 is the best possible constant in the commutative case, using the matrix $A \in M_n(\mathbb{C})$ given by $A_{st} = e^{2\pi i(s-t)/n}$. This matrix satisfies $\|A\|_{L_\infty \rightarrow L_1} = n$ and one can show $\|A\|_{\text{cut}} = (\pi^{-2} + o(1))n$. By Proposition 10 in Section 3.3, their example can be embedded into a superoperator with the same norms so π^2 is also the best possible constant here. \blacktriangleleft

Define the *Grothendieck norm* of a matrix $A \in M_n(\mathbb{C})$ by

$$\|A\|_G := \sup \left\{ \left| \frac{1}{n} \sum_{i,j=1}^n A_{ij} \langle x_i, y_j \rangle \right| : d \in \mathbb{N}, x_i, y_j \in \mathbb{C}^d, \|x_i\|_{L_2} \leq 1, \|y_j\|_{L_2} \leq 1 \right\}.$$

Then, the *complex Grothendieck constant* is given by

$$K_G^{\mathbb{C}} := \sup \left\{ \frac{\|A\|_G}{\|A\|_{L_\infty \rightarrow L_1}} : n \in \mathbb{N}, A \in M_n(\mathbb{C}) \right\}.$$

The current best upper and lower bounds on $K_G^{\mathbb{C}}$ are 1.4049 [15] and 1.338 [11], respectively. The real version of the Grothendieck constant, denoted by K_G and mentioned in the introduction, is obtained by replacing the underlying field in the above quantities by the reals.

Some basic group theory

Given a graph $G = (V, E)$, a permutation $\pi : V \rightarrow V$ is an *automorphism* of G if for all $u, v \in V$, we have $\{\pi(u), \pi(v)\} \in E \Leftrightarrow \{u, v\} \in E$. The automorphisms of G form a group under composition, which we call $\text{Aut}(G)$. Then, G is said to be *vertex transitive* if for every $u, v \in V$, there is a $\pi \in \text{Aut}(G)$ such that $\pi(u) = v$. For superoperators, we have the following

analogous definitions. A unitary representation of a group Γ on \mathbb{C}^n is a homomorphism from Γ to $U(n)$ and it is irreducible if the only subspaces of \mathbb{C}^n that are left invariant by the group action are the zero-dimensional subspace and \mathbb{C}^n itself.

► **Definition 3** (Irreducible covariance). *A superoperator $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ is irreducibly covariant if there exist a compact group Γ and continuous irreducible unitary representations $U, V : \Gamma \rightarrow U(n)$ such that for all $g \in \Gamma$ and $X \in M_n(\mathbb{C})$, we have*

$$\Phi(U(g)XU^*(g)) = V(g)\Phi(X)V^*(g).$$

3 Converse expander mixing lemmas

In this section, we prove the “converse expander mixing lemmas” announced in the first and third bullet in the introduction. As a warm-up, we start with a proof of the commutative case due to Conlon and Zhao, which we reprove in a slightly different manner analogous to how we will prove the non-commutative case.

3.1 Commutative case

In the following, let S be a compact set and Γ be a compact group acting continuously and transitively on S . The Haar probability measure on Γ induces a measure on S (by pullback) according to which the L_p -norm (for $p \in [1, \infty)$) and inner product of $f, g \in C(S)$ are given by

$$\|f\|_{L_p} = \left(\mathbb{E}_{\pi \in \Gamma} |f(\pi(s_0))|^p \right)^{\frac{1}{p}} \quad \text{and} \quad \langle f, g \rangle = \mathbb{E}_{\pi \in \Gamma} \overline{f(\pi(s_0))} g(\pi(s_0)), \quad (5)$$

where (by transitivity) s_0 can be taken to be some arbitrary but fixed element of S . We lift the action of Γ on S to an action on $C(S)$ by precomposition, that is, for any function $f \in C(S)$ and element $\pi \in \Gamma$, define the function f^π by $f^\pi(s) := f(\pi(s))$. Furthermore, for a linear map $A : C(S) \rightarrow C(S)$ define A^π by $A^\pi f := (Af)^\pi$ and say that A is transitive covariant with respect to Γ if for any $\pi \in \Gamma$ we have $A^\pi = A$.⁴ We sometimes omit the group and simply say A is *transitive covariant* if such a group Γ exists.

In [9], the following result is proved (over the real numbers) for the case $S = [n]$, in which case transitive covariant linear maps A are simply $n \times n$ matrices which commute with the permutation matrices of a transitive subgroup Γ of S_n . However, their proof easily implies the more general version below.

► **Theorem 4** (Conlon–Zhao). *Let S be as above and let $A : C(S) \rightarrow C(S)$ be a linear map that is transitive covariant with respect to Γ . Then,*

$$\|A\|_{L_2 \rightarrow L_2} \leq K_G^{\mathbb{C}} \|A\|_{L_\infty \rightarrow L_1}.$$

Here we give a somewhat more streamlined proof of this result based on a well-known factorization version of Grothendieck’s inequality [13] (see also [29]), which will serve as a stepping stone to the proof of the non-commutative case.⁵ In our setting the inequality asserts the following

⁴ In general one says A is *covariant* with respect to Γ , but we say *transitive* to emphasize that we require Γ to act transitively on S .

⁵ The main difference is that in [9], the result is first proved for weighted Cayley graphs, after which it is shown that this implies the result for transitive covariant matrices.

► **Theorem 5** (Commutative Grothendieck inequality (factorization)). *Let S be as above and let $A : C(S) \rightarrow C(S)$ be a linear map. Then, there exist probability measures λ, ν on S such that for all $f, g \in C(S)$, we have*

$$|\langle g, Af \rangle| \leq K_G^{\mathbb{C}} \|A\|_{L_\infty \rightarrow L_1} \left(\int_S |f(s)|^2 d\lambda(s) \right)^{1/2} \left(\int_S |g(s)|^2 d\nu(s) \right)^{1/2}.$$

Proof of Theorem 4. It follows from the triangle inequality and transitivity that

$$|\langle g, Af \rangle| \leq \mathbb{E}_{\pi \in \Gamma} |\langle g, A^\pi f \rangle| = \mathbb{E}_{\pi \in \Gamma} |\langle g^\pi, Af^\pi \rangle|.$$

By Theorem 5 and the AM-GM inequality there are probability measures λ, ν on S such that the above right-hand side is at most

$$\frac{K_G^{\mathbb{C}} \|A\|_{L_\infty \rightarrow L_1}}{2} \mathbb{E}_{\pi \in \Gamma} \left(\int_S |f^\pi(s)|^2 d\lambda(s) + \int_S |g^\pi(s)|^2 d\nu(s) \right) = \frac{K_G^{\mathbb{C}} \|A\|_{L_\infty \rightarrow L_1}}{2} (\|f\|_{L_2}^2 + \|g\|_{L_2}^2),$$

where we switched the order of the integrals (using Tonelli's theorem) and the expression (5) for the L_2 norm. For $\|f\|_{L_2} = \|g\|_{L_2} = 1$ this shows $\|A\|_{L_2 \rightarrow L_2} \leq K_G^{\mathbb{C}} \|A\|_{L_\infty \rightarrow L_1}$. ◀

3.2 Non-commutative case

Our main technical result is as follows.

► **Theorem 6.** *Let $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ be an irreducibly covariant superoperator. Then, $\|\Phi\|_{S_\infty \rightarrow S_1} \leq \|\Phi\|_{S_2 \rightarrow S_2} \leq 2\|\Phi\|_{S_\infty \rightarrow S_1}$.*

Since the supremum in $\|\Phi\|_{S_\infty \rightarrow S_1}$ is taken over X, Y with S_∞ -norm equal to 1, the first inequality of the theorem follows from the fact that $\|X\|_{S_2} \leq \|X\|_{S_\infty}$. As projectors have Schatten- ∞ norm 1, the first inequality also easily implies the analogue of the Expander Mixing Lemma, that is, $\epsilon(\Phi) \leq \lambda(\Phi)$, where $\lambda(\Phi)$ and $\epsilon(\Phi)$ are as in (2) and (3), respectively; note that when Φ is irreducibly covariant, so is $\Phi - \Pi$. The second inequality is proved at the end of this section and in Section 4.2 we show that the factor 2 in the theorem is optimal. With Lemma 2, which relates the uniformity parameter $\epsilon(\Phi)$ to $\|\Phi - \Pi\|_{S_\infty \rightarrow S_1}$, Theorem 6 then immediately gives the following result stated in the introduction.

► **Corollary 7** (Converse Quantum Expander Mixing Lemma). *Let $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ be an irreducibly covariant superoperator. Then, $\lambda(\Phi) \leq 2\pi^2 \epsilon(\Phi)$.*

In this non-commutative setting we use the following analog of Theorem 5 (a factorization version of the non-commutative Grothendieck inequality), proved by Haagerup in [14]; see also [29]. A density matrix is a positive semidefinite matrix with trace equal to 1.

► **Theorem 8** (Haagerup). *Let $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ be a superoperator. Then, there exist density matrices $\rho_1, \rho_2, \sigma_1, \sigma_2$ such that for any $X, Y \in M_n(\mathbb{C})$, we have*

$$|\langle Y, \Phi(X) \rangle| \leq \|\Phi\|_{S_\infty \rightarrow S_1} (\text{Tr}[\rho_1 X^* X] + \text{Tr}[\rho_2 X X^*])^{1/2} (\text{Tr}[\sigma_1 Y^* Y] + \text{Tr}[\sigma_2 Y Y^*])^{1/2}. \quad (6)$$

We also use the following lemma.

► **Lemma 9.** *Let Γ be a compact group. Then, a unitary representation $U : \Gamma \rightarrow U(n)$ is irreducible if and only if for any $X \in M_n(\mathbb{C})$, we have*

$$\mathbb{E}_{g \in \Gamma} U(g) X U(g)^* = \text{Tr}(X) \frac{1}{n} \text{Id}.$$

Proof. By Schur's lemma, if U is an irreducible representation, then for $T \in M_n(\mathbb{C})$

$$\left[\forall g \in \Gamma \quad U(g)TU(g)^* = T \right] \iff \left[\exists \lambda \in \mathbb{C} \quad T = \lambda \text{Id} \right].$$

Let $T_X = \mathbb{E}_{g \in \Gamma} U(g)XU(g)^*$, then by the group structure we have $U(g)T_XU(g)^* = T_X$ for all $g \in \Gamma$. Therefore, if U is irreducible then $T_X = \lambda_X \text{Id}$. By taking the trace, it follows that $\lambda_X = \text{Tr}(X)/n$. In the other direction, if U is reducible then there exists a projector P onto an irreducible subspace that is left invariant, i.e. $U(g)PU(g)^* = P$ for all $g \in \Gamma$, so $T_P \neq \lambda \text{Id}$. \blacktriangleleft

Proof of Theorem 6. Denote by Γ and $U, V: \Gamma \rightarrow U(n)$ the group and irreducible representations such that Φ is irreducibly covariant with respect to Γ (see Definition 3). For any $X, Y \in M_n(\mathbb{C})$ write $X_g = U(g)XU(g)^*$ and $Y_g = V(g)YV(g)^*$, then we have

$$|\langle Y, \Phi(X) \rangle| = \mathbb{E}_{g \in \Gamma} |\langle Y_g, \Phi(X_g) \rangle|.$$

By Theorem 8 and the AM-GM inequality, there exist density matrices $\rho_1, \rho_2, \sigma_1, \sigma_2$ such that the right hand side is bounded from above by

$$\frac{1}{2} \|\Phi\|_{S_\infty \rightarrow S_1} \mathbb{E}_{g \in \Gamma} \left(\text{Tr}[\rho_1 X_g^* X_g] + \text{Tr}[\rho_2 X_g X_g^*] + \text{Tr}[\sigma_1 Y_g^* Y_g] + \text{Tr}[\sigma_2 Y_g Y_g^*] \right).$$

By Lemma 9 we have $\mathbb{E}_{g \in \Gamma} X_g^* X_g = \mathbb{E}_{g \in \Gamma} U(g)X^* X U(g)^* = \frac{1}{n} \text{Tr}[X^* X] \text{Id} = \|X\|_{S_2}^2 \text{Id}$. Let ρ be a density matrix, then $\mathbb{E}_{g \in \Gamma} \text{Tr}[\rho X_g^* X_g] = \|X\|_{S_2}^2$. The same holds for $\mathbb{E}_{g \in \Gamma} \text{Tr}[\rho X_g X_g^*]$ but with U , and for Y with V , so we see that the above quantity is equal to

$$\|\Phi\|_{S_\infty \rightarrow S_1} (\|X\|_{S_2}^2 + \|Y\|_{S_2}^2).$$

If $\|X\|_{S_2} = \|Y\|_{S_2} = 1$ we obtain $\|\Phi\|_{S_2 \rightarrow S_2} \leq 2\|\Phi\|_{S_\infty \rightarrow S_1}$. \blacktriangleleft

3.3 Embedding graphs into quantum channels

In this subsection, we elucidate the claim that quantum channels generalize graphs and prove the result stated in the second bullet in the introduction, namely that there are non-irreducible quantum channels for which a converse expander mixing lemma does not hold.

We consider the following embeddings. For $A \in M_n(\mathbb{C})$, define $\Phi_A: M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ as

$$\Phi_A(X) = \sum_{i,j} A_{ij} X_{jj} E_{ii}, \quad (7)$$

where E_{ij} is the matrix with a single 1 at position (i, j) . When A is a transition matrix, i.e., its column sums are 1, then it is not hard to see that Φ_A is completely positive and trace preserving and that $\Phi_{\frac{1}{n}J} = \Pi$. Several other ways exist to create quantum expanders from expander graphs, see for example [20] and [17], but as we show below, our embedding given above carries over all relevant properties of the graph we consider here.

Conlon and Zhao [9] give an infinite sequence of d -regular graphs G_n that are $o(1)$ -uniform but for which $\lambda(G_n) \geq 1/2$. Combined with the following proposition, this immediately gives the result stated in the second bullet in the introduction.

► **Proposition 10.** *Let $A \in M_n(\mathbb{C})$ and $p, q \in [1, \infty]$. Then, for Φ_A as in (7), we have*

$$\|\Phi_A - \Pi\|_{S_p \rightarrow S_q} = \|A - \frac{1}{n}J\|_{L_p \rightarrow L_q} \quad \text{and} \quad \|\Phi_A - \Pi\|_{\text{cut}} = \|A - \frac{1}{n}J\|_{\text{cut}}.$$

Proof. Let $B = A - \frac{1}{n}J$, then $\Phi_A - \Pi = \Phi_B$. By compactness and definition of $\|\cdot\|_{S_p \rightarrow S_q}$ we can assume there is an $X \in M_n(\mathbb{C})$ such that $\|\Phi_B\|_{S_p \rightarrow S_q} = \|\Phi_B(X)\|_{S_q} / \|X\|_{S_p}$. Write $X = \text{diag}(x) + X_{\text{other}}$ where $x \in \mathbb{C}^n$ is the diagonal of X , and X_{other} are the off-diagonal entries. Note that by definition of Φ_B we have $\Phi_B(X) = \Phi_B(\text{diag}(x)) = \text{diag}(Bx)$. By definition of Schatten norms, $\|\text{diag}(x)\|_{S_p} = \|x\|_{L_p}$ and by Proposition 1 we have $\|X\|_{S_p} \geq \|x\|_{L_p}$. We have

$$\|B\|_{L_p \rightarrow L_q} \geq \frac{\|Bx\|_{L_q}}{\|x\|_{L_p}} \geq \frac{\|\text{diag}(Bx)\|_{S_q}}{\|X\|_{S_p}} = \frac{\|\Phi_B(X)\|_{S_q}}{\|X\|_{S_p}} = \|\Phi_B\|_{S_p \rightarrow S_q}$$

Now let $y \in \mathbb{C}^n$ be such that $\|B\|_{L_p \rightarrow L_q} = \|By\|_{L_q} / \|y\|_{L_p}$. Then

$$\|\Phi_B\|_{S_p \rightarrow S_q} \geq \frac{\|\Phi_B(\text{diag}(y))\|_{S_q}}{\|\text{diag}(y)\|_{S_p}} = \frac{\|\text{diag}(By)\|_{S_q}}{\|y\|_{L_p}} = \frac{\|By\|_{L_q}}{\|y\|_{L_p}} = \|B\|_{L_p \rightarrow L_q}.$$

This proves the first part.

The cut norm of a matrix takes the supremum over $x, y \in \{0, 1\}^n$. Instead we can relax this to $x, y \in [0, 1]^n$, since by linearity the supremum will always be attained by the extreme points. Similarly, for the superoperator case, we use Equation (4). Then, there exist $x, y \in [0, 1]^n$ such that $\|B\|_{\text{cut}} = |\langle Bx, y \rangle|$. We have $\text{diag}(x), \text{diag}(y) \succeq 0$ and $\|\text{diag}(x)\|_{S_\infty}, \|\text{diag}(y)\|_{S_\infty} \leq 1$. Therefore

$$\|\Phi_B\|_{\text{cut}} \geq |\langle \text{diag}(y), \Phi_B(\text{diag}(x)) \rangle| = |\langle \text{diag}(y), \text{diag}(Bx) \rangle| = |\langle y, Bx \rangle| = \|B\|_{\text{cut}}.$$

In the other direction, let $X, Y \in M_n(\mathbb{C})$ such that $X, Y \succeq 0$ and $\|X\|_{S_\infty}, \|Y\|_{S_\infty} \leq 1$. Define x, y to be the diagonals of X, Y , i.e. $x_i = X_{ii}$ and $y_i = Y_{ii}$. By Proposition 1 we have $\|x\|_{L_\infty}, \|y\|_{L_\infty} \leq 1$. Since $X, Y \succeq 0$ we know all diagonal entries of X and Y are real and non-negative, so we have $x, y \in [0, 1]^n$. We conclude

$$\|B\|_{\text{cut}} \geq |\langle y, Bx \rangle| = |\langle \text{diag}(y), \text{diag}(Bx) \rangle| = |\langle Y, \Phi_B(X) \rangle| = \|\Phi_B\|_{\text{cut}},$$

completing the proof. ◀

Note that $\|A - \frac{1}{n}J\|_{L_2 \rightarrow L_2}$ is the second largest eigenvalue in absolute value of the matrix A , so spectral expansion is preserved under the embedding of graphs into quantum channels. Also, uniformity is preserved since the cut-norm does not change.

The following proposition shows that the embedding (7) preserves transitivity. This shows that our Theorem 6 generalizes the main result of [9], albeit with a slightly worse constant.

► **Proposition 11.** *For any $A \in M_n(\mathbb{C})$, A is vertex transitive if and only if Φ_A is irreducibly covariant.*

Proof. Suppose A is vertex transitive. Let $\pi \in \text{Aut}(A)$ be a permutation and $P_\pi \in M_n(\mathbb{C})$ be the associated permutation matrix, so that $P_\pi A P_\pi^* = A$. Then,

$$\begin{aligned} \Phi_A(P_\pi X P_\pi^*) &= \sum_{i,j} A_{ij} (P_\pi X P_\pi^*)_{jj} E_{ii} \\ &= \sum_{i,j} A_{ij} X_{\pi^{-1}(j)\pi^{-1}(j)} E_{ii} \\ &= \sum_{i,j} A_{i\pi(j)} X_{jj} E_{ii} \\ &= \sum_{i,j} A_{\pi(i)\pi(j)} X_{jj} E_{\pi(i)\pi(i)} \\ &= \sum_{i,j} A_{\pi(i)\pi(j)} X_{jj} (P_\pi E_{ii} P_\pi^*) = P_\pi \Phi_A(X) P_\pi^*. \end{aligned}$$

This shows that for all $\pi \in \text{Aut}(A)$ we have $\Phi_A(P_\pi X P_\pi^*) = P_\pi \Phi_A(X) P_\pi^*$.

Let $\mathbb{T} = \{c \in \mathbb{C} : |c| = 1\}$ be the complex unit circle. For $\alpha \in \mathbb{T}^n$, define $U_\alpha := \text{diag}(\alpha)$. We have $U_\alpha E_{ii} U_\alpha^* = |\alpha_i|^2 E_{ii} = E_{ii}$ and $(U_\alpha X U_\alpha^*)_{ii} = |\alpha_i|^2 X_{ii} = X_{ii}$. Therefore

$$\Phi_A(U_\alpha X U_\alpha^*) = \sum_{i,j} A_{ij} (U_\alpha X U_\alpha^*)_{jj} E_{ii} = \sum_{i,j} A_{ij} X_{jj} U_\alpha E_{ii} U_\alpha^* = U_\alpha \Phi_A(X) U_\alpha^*.$$

We combine these two observations as follows. First we have that

$$\left(\mathbb{E}_{\alpha \in \mathbb{T}^n} U_\alpha X U_\alpha^* \right)_{ij} = \mathbb{E}_{\alpha \in \mathbb{T}^n} \alpha_i X_{ij} \overline{\alpha_j} = \int_0^{2\pi} \int_0^{2\pi} \alpha_i X_{ij} \overline{\alpha_j} d\alpha_i d\alpha_j = X_{ii} \delta_{ij}$$

If A is vertex transitive then for all $x \in \mathbb{C}^n$ we have $\mathbb{E}_{\pi \in \text{Aut}(A)} P_\pi \text{diag}(x) P_\pi^* = (\mathbb{E}_i x_i) \text{Id}$. Therefore

$$\mathbb{E}_{\substack{\pi \in \text{Aut}(A) \\ \alpha \in \mathbb{T}^n}} (P_\pi U_\alpha X (P_\pi U_\alpha)^*) = \mathbb{E}_{\pi \in \text{Aut}(A)} P_\pi \left(\mathbb{E}_{\alpha \in \mathbb{T}^n} U_\alpha X U_\alpha^* \right) P_\pi^* = \frac{\text{Tr}(X)}{n} \text{Id}.$$

Letting $G \subset M_n(\mathbb{C})$ be the subgroup generated by the U_α and P_π for $\pi \in \text{Aut}(A)$, we see that for any $g \in G$

$$\Phi_A(g X g^*) = g \Phi_A(X) g^*$$

and by the previous equation and Lemma 9, G acts irreducibly on \mathbb{C}^n (and it is unitary). This proves Φ is irreducibly covariant with respect to the group G with equal representations.

For the other direction, let $U : G \rightarrow U(n)$ be the irreducible representation such that Φ_A is irreducibly covariant, i.e. $\Phi_A(U(g) X U(g)^*) = U(g) \Phi_A(X) U(g)^*$ for all $g \in G$. Define $P_g \in M_n(\mathbb{C})$ as $(P_g)_{ij} = |U(g)_{ij}|^2$ so that $(U(g) E_{jj} U(g)^*)_{ii} = (P_g)_{ij}$. Then

$$\begin{aligned} A_{kl} &= \text{Tr}[E_{kk} \Phi_A(E_{ll})] = \text{Tr}[U(g) E_{kk} U(g)^* \Phi_A(U(g) E_{ll} U(g)^*)] \\ &= \sum_{ij} A_{ij} (P_g)_{jl} (P_g)_{ik} = (P_g^T A P_g)_{kl}, \end{aligned}$$

showing $P_g^T A P_g = A$. Since $U(g)$ is unitary, P_g is doubly stochastic so by Birkhoff's Theorem P_g is a convex combination of permutation matrices, i.e., $P_g = \mathbb{E}_i \Pi_i$ for some (not necessarily uniform) probability distribution and where Π_i is a permutation matrix. We have

$$A_{kl} = (P_g^T A P_g)_{kl} = \mathbb{E}_i \mathbb{E}_j (\Pi_i^T A \Pi_j)_{kl} = \mathbb{E}_i \mathbb{E}_j A_{\pi_i(k) \pi_j(l)}.$$

Since A is $\{0, 1\}$ -valued, it follows that if $A_{kl} = 1$ then all elements of the convex combination on the right-hand side must be 1, and if $A_{kl} = 0$ then all elements of the right hand side must be 0. Therefore, for all i we have $\Pi_i^T A \Pi_i = A$. By irreducibility, we have for all k, l that

$$\frac{1}{n} = \frac{\text{Tr}[E_{kk}]}{n} \text{Id}_l = \left(\mathbb{E}_{g \in G} U(g) E_{kk} U^*(g) \right)_{ll} = \mathbb{E}_{g \in G} |U(g)_{lk}|^2,$$

showing $\mathbb{E}_{g \in G} (P_g)_{lk} = 1/n$. It follows that there is a $g \in G$ such that $(P_g)_{lk} > 0$. Decomposing P_g into permutation matrices shows there is a $\Pi \in \text{Aut}(A)$ such that $\Pi_{lk} = 1$. This holds for all k, l , proving the lemma. \blacktriangleleft

3.4 Randomizing superoperators

We prove the following analogue of one of the results from [8] showing that for any d -regular graph G , it holds that $\lambda(G) \leq (2\epsilon(G)/\delta^2)^{1/4}$, where $\delta = d/n$ is the edge density. This in particular establishes a tight relation between spectral expansion and uniformity for sequences of graphs with $\delta_n \geq \Omega(1)$. For $A \in M_n(\mathbb{C})$, we have $\|A\|_{L_1 \rightarrow L_\infty} = n \sup_{ij} |A_{ij}|$, and for an n -vertex d -regular graph with normalized adjacency matrix A we have $\sup_{ij} |A_{ij}| = \frac{1}{d}$ so $\|A\|_{L_1 \rightarrow L_\infty} = \frac{1}{\delta}$. Therefore, a sequence of graphs with normalized adjacency matrices A_n is dense exactly when $\|A_n\|_{L_1 \rightarrow L_\infty} \leq \mathcal{O}(1)$.

A superoperator Φ is said to be η -randomizing if $\|\Phi\|_{S_1 \rightarrow S_\infty} \leq \eta$, which when $\eta \leq \mathcal{O}(1)$, may thus be seen as an analogue of density. Note that by Proposition 10 the embedding of any dense graph is $\mathcal{O}(1)$ -randomizing.

► **Proposition 12.** *Let $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ be a unital superoperator that is $\mathcal{O}(1)$ -randomizing. Then, $\lambda(\Phi) \leq \mathcal{O}(\epsilon(\Phi)^{1/4})$.*

To prove Proposition 12, we require the following lemma.

► **Lemma 13.** *Let $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ be a superoperator and let $C = \|\Phi\|_{S_1 \rightarrow S_\infty}$. Then we have $\|\Phi\|_{S_2 \rightarrow S_2} \leq \left(C^3 \|\Phi\|_{S_\infty \rightarrow S_1} \right)^{1/4}$.*

Proof. Note that by definition of C we have $|\langle Q, \Phi(P) \rangle| \leq C \|Q\|_{S_1} \|P\|_{S_1}$. Let $X, Y \in M_n(\mathbb{C})$ be such that $\langle Y, \Phi(X) \rangle = \|\Phi\|_{S_2 \rightarrow S_2}$ with $\|X\|_{S_2} = \|Y\|_{S_2} = 1$. Write $X = \frac{1}{n} \sum_{i=1}^n \lambda_i P_i$ and $Y = \frac{1}{n} \sum_{i=1}^n \mu_i Q_i$ with P_i, Q_i rank-1 matrices with $\|Q_i\|_{S_1} = \|P_i\|_{S_1} = 1$. We have $\|\lambda\|_{L_2} = \|\mu\|_{L_2} = 1$ and by applying Cauchy-Schwarz twice,

$$\begin{aligned} |\langle Y, \Phi(X) \rangle|^4 &= \left| \mathbb{E}_{ij} \lambda_i \mu_j \langle Q_j, \Phi(P_i) \rangle \right|^4 \\ &\leq \left(\mathbb{E}_i \lambda_i^2 \right)^2 \left(\mathbb{E}_j \left| \mathbb{E}_i \mu_j \langle Q_j, \Phi(P_i) \rangle \right|^2 \right)^2 \\ &= \left(\mathbb{E}_{i,j,j'} \mu_j \mu_{j'} \langle Q_j, \Phi(P_i) \rangle \langle P_i, \Phi^*(Q_{j'}) \rangle \right)^2 \\ &\leq \left(\mathbb{E}_{j,j'} \mu_j^2 \mu_{j'}^2 \right) \left(\mathbb{E}_{j,j'} \left| \mathbb{E}_i \langle Q_j, \Phi(P_i) \rangle \langle P_i, \Phi^*(Q_{j'}) \rangle \right|^2 \right) \\ &= \mathbb{E}_{i,i',j,j'} \langle Q_j, \Phi(P_i) \rangle \langle P_i, \Phi^*(Q_{j'}) \rangle \langle Q_{j'}, \Phi(P_{i'}) \rangle \langle P_{i'}, \Phi^*(Q_j) \rangle, \end{aligned}$$

where all indices are averaged from 1 to n . Now we see

$$\begin{aligned}
| \langle Y, \Phi(X) \rangle |^4 &\leq \mathbb{E}_{i,j} \langle Q_j, \Phi(P_i) \rangle \left\langle \mathbb{E}_{j'} \langle Q_{j'}, \Phi(P_i) \rangle Q_{j'}, \Phi \left(\mathbb{E}_{i'} \langle P_{i'}, \Phi^*(Q_j) \rangle P_{i'} \right) \right\rangle \\
&\leq \mathbb{E}_{i,j} | \langle Q_j, \Phi(P_i) \rangle | \|\Phi\|_{S_\infty \rightarrow S_1} \|\mathbb{E}_{j'} \langle Q_{j'}, \Phi(P_i) \rangle Q_{j'}\|_{S_\infty} \|\mathbb{E}_{i'} \langle P_{i'}, \Phi^*(Q_j) \rangle P_{i'}\|_{S_\infty} \\
&\leq \mathbb{E}_{i,j} | \langle Q_j, \Phi(P_i) \rangle | \|\Phi\|_{S_\infty \rightarrow S_1} \max_{j'} | \langle Q_{j'}, \Phi(P_i) \rangle | \max_{i'} | \langle Q_j, \Phi(P_{i'}) \rangle | \\
&\leq C^3 \|\Phi\|_{S_\infty \rightarrow S_1}.
\end{aligned}$$

Proof of Proposition 12. Let $\Pi(X) = \frac{1}{n} \text{Tr}[X] \text{Id}$ and $\mathcal{E} = \Phi - \Pi$, then $\|\mathcal{E}\|_{\text{cut}} \leq \epsilon$ by assumption. Define $C = \|\Phi\|_{S_1 \rightarrow S_\infty}$. We have $\|\Pi\|_{S_1 \rightarrow S_\infty} = 1$ so by the triangle inequality, $\|\mathcal{E}\|_{S_1 \rightarrow S_\infty} \leq C + 1$. Using Lemma 2 and Lemma 13 applied to \mathcal{E} we find $\|\mathcal{E}\|_{S_2 \rightarrow S_2} \leq ((C + 1)^3 \pi^2 \epsilon)^{1/4}$.

4 Optimality of constants

4.1 Commutative case

In this section we prove the fourth bullet point in our introduction. Theorem 4 shows that $K_G^{\mathbb{C}}$ bounds the ratio of the $L_2 \rightarrow L_2$ and $L_\infty \rightarrow L_1$ norms, and Lemma 2 (the matrix version) shows that π^2 bounds the ratio of the $L_\infty \rightarrow L_1$ norm and the cut norm. We now prove the optimality of the combined inequality.

Let $S^{m-1} = \{x \in \mathbb{C}^m : \|x\|_{L_2} = 1\}$ denote the $(m-1)$ -dimensional unit sphere endowed with its Haar probability measure μ .

► **Theorem 14.** *For any $\epsilon > 0$ there exist positive integers m, k and a transitive covariant linear map $M : C(S^{m-1} \times [k]) \rightarrow C(S^{m-1} \times [k])$ such that $\|M\|_{L_2 \rightarrow L_2} \geq (\pi^2 K_G^{\mathbb{C}} - \epsilon) \|M\|_{\text{cut}}$.*

The optimality of π^2 between the $L_\infty \rightarrow L_1$ norm and the cut norm is already covered in Lemma 2. We show that $K_G^{\mathbb{C}}$ is optimal in the sense that Theorem 4 cannot be improved (despite the fact that the exact value of the Grothendieck constant $K_G^{\mathbb{C}}$ is unknown). We do this in Lemma 15 below. Then in Lemma 16 we show that any map can be lifted to one on a bigger space with appropriately bounded cut norm. The combination of these lemmas proves our theorem.

In the introduction we also mentioned the optimal constant $4K_G$ in the case where the field is \mathbb{R} instead of \mathbb{C} . The proofs below still apply in this case, with only small modifications.

► **Lemma 15.** *For any $\epsilon > 0$ there exists a positive integer m and a transitive covariant linear map $B : C(S^{m-1}) \rightarrow C(S^{m-1})$ such that $\|B\|_{L_2 \rightarrow L_2} \geq (K_G^{\mathbb{C}} - \epsilon) \|B\|_{L_\infty \rightarrow L_1}$.*

Proof. By definition of the Grothendieck constant, for any $\epsilon > 0$ there exists an $n \in \mathbb{N}$ and a linear map $A \in M_n(\mathbb{C})$ such that $\|A\|_G \geq (K_G^{\mathbb{C}} - \epsilon) \|A\|_{L_\infty \rightarrow L_1}$. This map A might not be transitive covariant, so from it we will now construct a transitive covariant linear map $B : C(S^{2n-1}) \rightarrow C(S^{2n-1})$ such that $\|B\|_{L_\infty \rightarrow L_1} \leq \|A\|_{L_\infty \rightarrow L_1}$ and $\|B\|_{L_2 \rightarrow L_2} \geq \|A\|_G$. This idea is based on a lemma found in [6].

Let $x^i, y^j \in S^{2n-1}$ be the vectors that attain the Grothendieck norm for A , which can always be assumed to be $2n$ -dimensional since there are only $2n$ of them, so

$$\|A\|_G = \left| \frac{1}{n} \sum_{i,j} A_{ij} \langle x^i, y^j \rangle \right|.$$

Define the map B by

$$\langle f, B(g) \rangle = \frac{1}{n} \sum_{i,j} A_{ij} \int_{U(2n)} f(Ux^i) g(Uy^j) dU.$$

To bound $\|B\|_{L_\infty \rightarrow L_1}$ we have to bound $|\langle f, B(g) \rangle|$ for $f, g : S^{2n-1} \rightarrow [-1, 1]$. By the triangle inequality,

$$|\langle f, B(g) \rangle| \leq \int_{U(2n)} \left| \frac{1}{n} \sum_{i,j} A_{ij} f(Ux^i) g(Uy^j) \right| dU \leq \int_{U(2n)} \|A\|_{L_\infty \rightarrow L_1} dU \leq \|A\|_{L_\infty \rightarrow L_1}.$$

Now for each $i \in [2n]$ let $f_i \in C(S^{2n-1})$ be given by $f_i(x) = x_i$ (i.e. the i -th coordinate). Then,

$$\begin{aligned} \frac{1}{2n} \sum_{i=1}^{2n} \langle f_i, B(f_i) \rangle &\leq \frac{1}{2n} \sum_{i=1}^{2n} \|B\|_{L_2 \rightarrow L_2} \|f_i\|_{L_2}^2 \\ &= \|B\|_{L_2 \rightarrow L_2} \int_{S^{2n-1}} \frac{1}{2n} \sum_{i=1}^{2n} x_i^2 d\mu(x) \\ &= \|B\|_{L_2 \rightarrow L_2}. \end{aligned}$$

On the other hand,

$$\frac{1}{2n} \sum_{i=1}^{2n} \langle f_i, B(f_i) \rangle = \frac{1}{n} \sum_{i,j} A_{ij} \int_{U(2n)} \langle Ux^i, Uy^j \rangle dU = \frac{1}{n} \sum_{i,j} A_{ij} \langle x^i, y^j \rangle = \|A\|_G,$$

so we conclude $\|B\|_{L_2 \rightarrow L_2} \geq \|A\|_G$. We will show B is transitive covariant with respect to $\Gamma = U(2n)$. To show B is invariant, we have to prove that for all $V \in U(2n)$ we have $\langle f^V, B(g^V) \rangle = \langle f, B(g) \rangle$. Indeed,

$$\begin{aligned} \langle f^V, B(g^V) \rangle &= \frac{1}{n} \sum_{i,j} A_{ij} \int_{U(2n)} f(VUx^i) g(VUy^j) dU \\ &= \frac{1}{n} \sum_{i,j} A_{ij} \int_{U(2n)} f(U'x^i) g(U'y^j) dU' = \langle f, B(g) \rangle, \end{aligned}$$

which completes the proof. ◀

► **Lemma 16.** *Let S be any compact set and let $B : C(S) \rightarrow C(S)$ be a linear map. For any $\epsilon > 0$ there exists a $k \in \mathbb{N}$ and a linear map $M : C(S \times [k]) \rightarrow C(S \times [k])$ such that*

$$\frac{\|M\|_{\text{cut}}}{\|M\|_{L_2 \rightarrow L_2}} \leq \left(\frac{1}{\pi^2} + \epsilon \right) \frac{\|B\|_{L_\infty \rightarrow L_1}}{\|B\|_{L_2 \rightarrow L_2}}$$

and if B is transitive covariant then so is M .

Proof. We will choose k large enough, to be determined later. For any $f, g \in C(S \times [k])$ define $f^i \in C(S)$ as $f^i(s) := f(s, i)$, and similar for g^i . Define $\omega = e^{2\pi i/k}$. Define a linear map $M : C(S \times [k]) \rightarrow C(S \times [k])$ as

$$(M(f))(t, j) := \frac{1}{k} \sum_{i=1}^k \omega^{i-j} B(f^i)(t), \quad \text{for } t \in S \text{ and } j \in [k].$$

We then have

$$\langle g, M(f) \rangle_{S \times [k]} = \frac{1}{k^2} \left\langle \sum_i \omega^i g^i, B \left(\sum_j \omega^j f^j \right) \right\rangle_S$$

where one factor of $\frac{1}{k}$ comes from our normalization of the inner product. This implies

$$|\langle g, M(f) \rangle_{S \times [k]}| \leq \|B\|_{L_\infty \rightarrow L_1} \left\| \frac{1}{k} \sum_{i=1}^k \omega^i g^i \right\|_{L_\infty} \left\| \frac{1}{k} \sum_{j=1}^k \omega^j f^j \right\|_{L_\infty}. \quad (8)$$

If $f, g \in C(S \times [k])$ are the $[0, 1]$ -valued functions that attain the cut norm of M , then by (8)

$$\|M\|_{\text{cut}} \leq \left(\frac{1}{\pi^2} + \epsilon \right) \|B\|_{L_\infty \rightarrow L_1},$$

where we used Lemma 17 to bound $\left\| \frac{1}{k} \sum_{i=1}^k \omega^i g^i \right\|_{L_\infty}$.

Let $u, v \in C(S)$ with $\|u\|_{L_2} = \|v\|_{L_2} = 1$ be such that $\|B\|_{L_2 \rightarrow L_2} = \langle v, B(u) \rangle_S$. Now define $f_{(u)}, g_{(v)} \in C(S \times [k])$ as $f_{(u)}(s, i) := \omega^{-i} u(s)$ and $g_{(v)}(s, i) := \omega^{-i} v(s)$, which also have L_2 -norm equal to 1. We then see

$$\|M\|_{L_2 \rightarrow L_2} \geq \langle g_{(v)}, M(f_{(u)}) \rangle_{S \times [k]} = \langle v, B(u) \rangle_S = \|B\|_{L_2 \rightarrow L_2}.$$

The combination of these observations completes the first part of the proof. Now assume B is transitive covariant with respect to Γ , so $B(f^\pi)(\pi^{-1}(s)) = B(f)(s)$ for all $s \in S$ and $\pi \in \Gamma$. Define a new group Γ' as the cartesian product $\Gamma' = \Gamma \times \mathbb{Z}_k$. For $(\pi, m) \in \Gamma'$ define the action $(\pi, m) : S \times [k] \rightarrow S \times [k]$ as $(\pi, m)(s, i) = (\pi(s), i + m)$. By entering $f^{(\pi, m)}$ into the definition of M it follows that $M^{(\pi, m)} = M$, so M is transitive covariant with respect to Γ' , completing the proof. \blacktriangleleft

► **Lemma 17.** *Let $\epsilon > 0$, then there exists a $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$ and $x \in [0, 1]^k$ we have*

$$\left| \frac{1}{k} \sum_{j=1}^k e^{2\pi i j/k} x_j \right| \leq \frac{1}{\pi} + \epsilon.$$

Proof. First let k_0 be arbitrary, to be determined later and $k \geq k_0$. Define $y \in [-1, 1]^k$ as $y_i = 2x_i - 1$, then

$$\left| \frac{1}{k} \sum_{j=1}^k e^{2\pi i j/k} x_j \right| = \frac{1}{2} \left| \frac{1}{k} \sum_{j=1}^k e^{2\pi i j/k} y_j \right| = \frac{1}{2} e^{2\pi i \phi} \frac{1}{k} \sum_{j=1}^k e^{2\pi i j/k} y_j.$$

In the first equality we used that $\sum_{j=1}^k e^{2\pi i j/k} = 0$. In the second equality we used that there exists a ϕ such that the full expression becomes real and positive. Since $e^{i\theta} = \cos(\theta) + i \sin(\theta)$ and the full expression is real, we know the sin component vanishes and therefore

$$\frac{1}{2} \frac{1}{k} \sum_{j=1}^k e^{2\pi i(\phi + j/k)} y_j = \frac{1}{2} \frac{1}{k} \sum_{j=1}^k \cos(2\pi(\phi + j/k)) y_j.$$

Now note that $\cos(2\pi(\phi + j/k)) y_j \leq |\cos(2\pi(\phi + j/k))|$ and hence

$$\frac{1}{2} \frac{1}{k} \sum_{j=1}^k |\cos(2\pi(\phi + j/k))| \xrightarrow{k \rightarrow \infty} \frac{1}{2} \int_0^1 |\cos(2\pi(\phi + x))| dx = \frac{1}{\pi}.$$

This completes the proof. \blacktriangleleft

4.2 Non-commutative case

In the non-commutative case we show optimality of Theorem 6. By Lemma 2, the factor π^2 between the cut-norm and $S_\infty \rightarrow S_1$ -norm is also optimal. In contrast with the commutative case, our work leaves the optimality of the combined inequality in Corollary 7 as an open problem. Straightforward analogues of the techniques employed in Lemma 16 did not follow through in the non-commutative case.

► **Proposition 18.** *For any $\epsilon > 0$, there exists a positive integer n and an irreducibly covariant superoperator $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ such that $\|\Phi\|_{S_2 \rightarrow S_2} \geq (2 - \epsilon)\|\Phi\|_{S_\infty \rightarrow S_1}$.*

One of the forms of the non-commutative Grothendieck inequality, equivalent to Theorem 8, is the following [29]. Let $\Phi : M_n(\mathbb{C}) \rightarrow M_n(\mathbb{C})$ be a linear map and $x_i, y_j \in M_n(\mathbb{C})$ finite sets of matrices. Then,

$$\left| \sum_i \langle x_i, \Phi(y_i) \rangle \right| \leq K'_G \|\Phi\|_{S_\infty \rightarrow S_1} \left(\frac{\|\sum_i x_i^* x_i\| + \|\sum_i x_i x_i^*\|}{2} \cdot \frac{\|\sum_i y_i^* y_i\| + \|\sum_i y_i y_i^*\|}{2} \right)^{1/2} \quad (9)$$

where $K'_G \leq 2$ and the norms on the right hand side are operator norms $\|\cdot\|_{S_\infty}$. To show tightness, i.e. $K'_G \geq 2$, Haagerup and Itoh [16] (see [29] for a survey) gave an explicit family of operators for which (9) gives a lower bound of K'_G approaching 2. We will show that slight modifications of these operators are irreducibly covariant, which proves Proposition 18. It is instructive to repeat their construction. The proof uses techniques familiar in the context of the antisymmetric Fock space, but our proof is self contained.

► **Lemma 19** ([16]). *For each $n \in \mathbb{N}$ there exists a $d \in \mathbb{N}$ and a linear map $\Phi : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ with sets of matrices $\{x_i\}, \{y_i\}$ such that (9) yields $K'_G \geq (2n + 1)/(n + 1)$.*

Proof. Let $H = \mathbb{C}^{2n+1}$ and consider the antisymmetric k -fold tensor product $H^{\wedge k}$ which is a linear subspace of the k -fold tensor product $H^{\otimes k}$. A basis of $H^{\wedge k}$ is formed by vectors $e_{i_1} \wedge e_{i_2} \wedge \cdots \wedge e_{i_k}$ with $i_1 < \cdots < i_k$ where the e_i are standard basis vectors of H . Here \wedge is the wedge product or exterior product, which has the property $x \wedge y = -y \wedge x$ and is given by $x \wedge y = x \otimes y - y \otimes x$, for $x, y \in H$. We will consider $k = n$ and $k = n + 1$ so that the dimension of $H^{\wedge k}$ is $d = \binom{2n+1}{n}$ for both $k = n$ and $k = n + 1$.

For $1 \leq i \leq (2n + 1)$, define $c_i : H^{\wedge n} \rightarrow H^{\wedge(n+1)}$ as $c_i(x) := e_i \wedge x$, which physicists call the fermionic creation operator. Its adjoint $c_i^* : H^{\wedge(n+1)} \rightarrow H^{\wedge n}$ is known as the annihilation operator. By the antisymmetric property, $c_i(x) = 0$ whenever e_i was present in x , i.e., when $x = e_i \wedge x'$. The operator $c_i c_i^*$, also known as the number operator, is a projector onto the space spanned by basis vectors in which e_i is present. The operator $c_i^* c_i$ is a projector onto the space where e_i is *not* present. Since there are always $(n + 1)$ vectors present in $H^{\wedge(n+1)}$ and $(n + 1)$ vectors *not* present in $H^{\wedge n}$, we have

$$\sum_{i=1}^{2n+1} c_i c_i^* = (n + 1) \text{Id}_{H^{\wedge(n+1)}} \quad \text{and} \quad \sum_{i=1}^{2n+1} c_i^* c_i = (n + 1) \text{Id}_{H^{\wedge n}}.$$

We will now argue that

$$\langle c_i, c_j \rangle := \frac{1}{d} \text{Tr}(c_i^* c_j) = \delta_{i,j} \frac{n + 1}{2n + 1}, \quad (10)$$

$$\left\| \sum_{i=1}^{2n+1} \alpha_i c_i \right\|_{S_1} = \|\alpha\|_{L_2} \frac{n + 1}{\sqrt{2n + 1}} \quad \text{for } \alpha \in \mathbb{C}^{2n+1}. \quad (11)$$

The $\delta_{i,j}$ in (10) follows because $\langle x, c_i^* c_j x \rangle = 0$ for any $x = e_{k_1} \wedge \dots \wedge e_{k_n}$ when $i \neq j$. The factor $\frac{n+1}{2n+1}$ follows by taking the trace of one of the sums above and noting that by symmetry in i , every term of the sum must have the same trace. To prove (11), first note that for any unitary $U \in U(2n+1)$ we have

$$U^{\otimes(n+1)} \cdot c_i \cdot (U^{\otimes n})^{-1} = \sum_j U_{ji} c_j, \quad (12)$$

which can be shown by proving it for all basis states:

$$\begin{aligned} U^{\otimes(n+1)} c_i (U^{\otimes n})^{-1} (e_{k_1} \wedge \dots \wedge e_{k_n}) &= U^{\otimes(n+1)} c_i (U^{-1} e_{k_1} \wedge \dots \wedge U^{-1} e_{k_n}) \\ &= U^{\otimes(n+1)} (e_i \wedge U^{-1} e_{k_1} \wedge \dots \wedge U^{-1} e_{k_n}) \\ &= (U e_i \wedge e_{k_1} \wedge \dots \wedge e_{k_n}) \\ &= \left(\sum_j U_{ji} e_j \wedge e_{k_1} \wedge \dots \wedge e_{k_n} \right) \\ &= \sum_j U_{ji} c_j (e_{k_1} \wedge \dots \wedge e_{k_n}). \end{aligned}$$

The trace-norm is unitarily invariant, so (12) implies $\|c_i\|_{S_1} = \|\sum_j U_{ji} c_j\|_{S_1}$. Since $c_i^* c_i$ is a projector, we have $\sqrt{c_i^* c_i} = c_i^* c_i$ and hence $\|c_i\|_{S_1} = \frac{1}{d} \text{Tr}(c_i^* c_i)$. Now let $\alpha \in \mathbb{C}^{2n+1}$ with $\sum_i |\alpha_i|^2 = 1$, then there is a unitary $U \in U(2n+1)$ such that the i -th row of U is α . Note that $\|\alpha\|_{L_2} = 1/\sqrt{2n+1}$ since we use normalized L_2 -norms, which implies (11).

Since the dimensions of $H^{\wedge n}$ and $H^{\wedge(n+1)}$ are equal, we can identify the space of linear maps $L(H^{\wedge n}, H^{\wedge(n+1)})$ with $M_d(\mathbb{C})$ (by choosing bases for $H^{\wedge n}$ and $H^{\wedge(n+1)}$), and define the following operator $\Phi : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$,

$$\Phi(x) = \sum_{i=1}^{2n+1} \langle c_i, x \rangle c_i.$$

Consider (9) for Φ with $x_i = y_i = c_i$. For the left hand side, note that by (10) we have

$$\left| \sum_{j=1}^{2n+1} \langle c_j, \Phi(c_j) \rangle \right| = \left| \sum_{i,j=1}^{2n+1} \langle c_i, c_j \rangle \langle c_j, c_i \rangle \right| = \frac{(n+1)^2}{2n+1}.$$

For the right-hand side of (9), we require $\|\Phi\|_{S_\infty \rightarrow S_1} = \sup_{\|x\|_{S_\infty}=1} \|\Phi(x)\|_{S_1}$. For any $x \in M_d(\mathbb{C})$, define $v^{(x)} \in \mathbb{C}^{2n+1}$ as $v_i^{(x)} = \langle c_i, x \rangle$. Note that $\|v\|_{L_2} = \sup_{\|\alpha\|_{L_2}=1} |\langle v, \alpha \rangle|$. First apply (11) to obtain

$$\|\Phi(x)\|_{S_1} = \left\| \sum_{i=1}^{2n+1} \langle c_i, x \rangle c_i \right\|_{S_1} = \|v^{(x)}\|_{L_2} \frac{n+1}{\sqrt{2n+1}} = \sup_{\|\alpha\|_{L_2}=1} |\langle v^{(x)}, \alpha \rangle| \frac{n+1}{\sqrt{2n+1}}.$$

Using (11) again, we compute $\sup_{\|x\|_{S_\infty}=1} |\langle v^{(x)}, \alpha \rangle|$ for arbitrary α with $\|\alpha\|_{L_2} = 1$,

$$\sup_{\|x\|_{S_\infty}=1} |\langle v^{(x)}, \alpha \rangle| = \sup_{\|x\|_{S_\infty}=1} \frac{1}{2n+1} \left| \langle x, \sum_i \alpha_i c_i \rangle \right| = \frac{1}{2n+1} \left\| \sum_i \alpha_i c_i \right\|_{S_1} = \frac{n+1}{(2n+1)\sqrt{2n+1}}.$$

We obtain $\|\Phi\|_{S_\infty \rightarrow S_1} = (n+1)^2/(2n+1)^2$. Now (9) yields $\frac{(n+1)^2}{2n+1} \leq K'_G \frac{(n+1)^2}{(2n+1)^2} \cdot (n+1)$ and therefore $\frac{2n+1}{n+1} \leq K'_G$. \blacktriangleleft

We use the following fact from [12, Theorem 19.14], about the representations of the odd dimensional complex special orthogonal groups on wedge products of *complex* vector spaces.

► **Lemma 20.** *Let $n, k \in \mathbb{N}$, $N := 2n + 1$ and let $R_k : \mathrm{SO}(N, \mathbb{C}) \rightarrow \mathrm{GL}((\mathbb{C}^N)^{\wedge k})$ be given by $A \mapsto A^{\otimes k}$. This representation is irreducible.*

Below, we actually need that the *real* special orthogonal group $\mathrm{SO}(N, \mathbb{R})$ acts irreducibly on the same anti-symmetric space. Fortunately, this is implied by Lemma 20; see [12, pp. 439]. We will also use the fact that R_k and R_{N-k} are *unitarily* equivalent to each other. This is the content of the following proposition [32, Proposition IX.10.4].

► **Proposition 21.** *For positive integer n and $N = 2n + 1$ and $k \in \{1, \dots, N\}$, let R_k be the representation as in lemma 20. Then, there exists an isometry $V_k : (\mathbb{C}^N)^{\wedge k} \rightarrow (\mathbb{C}^N)^{\wedge(N-k)}$ such that*

$$V_k R_k(A) = R_{N-k}(A) V_k, \quad \forall A \in \mathrm{SO}(N, \mathbb{R}).$$

of Proposition 18. Let d be the dimension of $(\mathbb{C}^N)^{\wedge n}$ and let $\Phi : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ be as in the proof of Lemma 19. For each $k \in \mathbb{N}$, let $R_k : \mathrm{SO}(N, \mathbb{R}) \rightarrow \mathrm{GL}(H^{\wedge k})$ be the representation $A \mapsto A^{\otimes k}$, which is irreducible by Lemma 20. Define, for notational convenience, $\pi := R_{n+1}$ and $\rho := R_n$. We first show that for all $A \in \mathrm{SO}(N, \mathbb{R})$, we have

$$\Phi(\pi(A)x\rho^*(A)) = \pi(A)\Phi(x)\rho^*(A). \quad (13)$$

For the left-hand side, note that

$$\begin{aligned} \Phi(\pi(A)x\rho^*(A)) &= \sum_i \langle c_i, \pi(A)x\rho^*(A) \rangle c_i \\ &= \sum_i \langle \pi(A)^* c_i \rho(A), x \rangle c_i \\ &= \sum_i \left\langle \sum_j A_{ij} c_j, x \right\rangle c_i \\ &= \sum_{ij} A_{ij} \langle c_j, x \rangle c_i, \end{aligned}$$

where we used (12) from the proof of Lemma 19 and noting that $\mathrm{SO}(N, \mathbb{R}) \subset \mathrm{U}(N)$ is a subgroup. Using (12) again for the right-hand side, we have

$$\begin{aligned} \pi(A)\Phi(x)\rho^*(A) &= \sum_i \langle c_i, x \rangle \pi(A)c_i \rho^*(A) \\ &= \sum_i \langle c_i, x \rangle \sum_j A_{ji} c_j \\ &= \sum_{ij} A_{ij} \langle c_j, x \rangle c_i. \end{aligned}$$

which proves (13).

Define a new superoperator $\Phi' : M_d(\mathbb{C}) \rightarrow M_d(\mathbb{C})$ by

$$\Phi'(x) = \Phi(xV^*)V,$$

where $V := V_{n+1}$ is the isometry as in Lemma 21 (we view V as a matrix in $M_d(\mathbb{C})$ by choosing basis). We first note that this Φ' might also be used in Lemma 19 to show that the non-commutative Grothendieck constant is 2, since Schatten-norms are unitarily invariant.

Hence, if we show that Φ' is irreducibly covariant, we are done. This follows from the following computation, where we use (13) and the fact that $V\pi(A) = \rho(A)V$ for all $A \in \text{SO}(N, \mathbb{R})$:

$$\begin{aligned}\Phi'(\pi(A)x\pi(A)^*) &= \Phi(\pi(A)x\pi(A)^*V^*)V \\ &= \Phi(\pi(A)xV^*\rho(A)^*)V \\ &\stackrel{(13)}{=} \pi(A)\Phi(xV^*)\rho(A)^*V \\ &= \pi(A)\Phi(xV^*)V\pi(A)^* \\ &= \pi(A)\Phi'(x)\pi^*(A),\end{aligned}$$

where the second-last line follows since $\rho(A)^* = V\pi(A)^*V^*$. Hence, Φ' is irreducibly covariant with respect to the irreducible representation π of $\text{SO}(N, \mathbb{R})$. ◀

References

- 1 Andris Ambainis and Adam Smith. Small pseudo-random families of matrices: Derandomizing approximate quantum encryption. In Klaus Jansen, Sanjeev Khanna, José D. P. Rolim, and Dana Ron, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 249–260, Berlin, Heidelberg, 2004. Springer Berlin Heidelberg.
- 2 Guillaume Aubrun. On almost randomizing channels with a short Kraus decomposition. *Comm. Math. Phys.*, 288(3):1103–1116, 2009. doi:10.1007/s00220-008-0695-y.
- 3 Avraham Ben-Aroya, Oded Schwartz, and Amnon Ta-Shma. Quantum expanders: Motivation and construction. *Theory of Computing*, 6(1):47–79, 2010. doi:10.4086/toc.2010.v006a003.
- 4 Béla Bollobás and Vladimir Nikiforov. Hermitian matrices and graphs: singular values and discrepancy. *Discrete Math.*, 285(1-3):17–32, 2004. doi:10.1016/j.disc.2004.05.006.
- 5 M. Braverman, K. Makarychev, Y. Makarychev, and A. Naor. The Grothendieck constant is strictly smaller than Krivine’s bound. *Forum Math. Pi*, 1:453–462, 2013. Preliminary version in FOCS’11. <https://arxiv.org/abs/1103.6161>.
- 6 Jop Briët. *Grothendieck inequalities, nonlocal games and optimization*. PhD thesis, Institute for Logic, Language and Computation, 2011.
- 7 Fan Chung and Ronald Graham. Sparse quasi-random graphs. *Combinatorica*, 22(2):217–244, 2002. Special issue: Paul Erdős and his mathematics. doi:10.1007/s004930200010.
- 8 Fan R. K. Chung, Ronald L. Graham, and Richard M. Wilson. Quasi-random graphs. *Combinatorica*, 9(4):345–362, 1989. doi:10.1007/BF02125347.
- 9 David Conlon and Yufei Zhao. Quasirandom Cayley graphs. *Discrete Anal.*, pages Paper No. 6, 14, 2017.
- 10 Tom Cooney, Marius Junge, Carlos Palazuelos, and David Pérez-García. Rank-one quantum games. *computational complexity*, 24(1):133–196, 2015.
- 11 A. Davie. Lower bound for K_G . Unpublished, 1984.
- 12 William Fulton and Joe Harris. *Representation theory: a first course*, volume 129. Springer Science & Business Media, 2013.
- 13 A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Bol. Soc. Mat. São Paulo*, 8:1–79, 1953.
- 14 Uffe Haagerup. The Grothendieck inequality for bilinear forms on C^* -algebras. *Adv. in Math.*, 56(2):93–116, 1985. doi:10.1016/0001-8708(85)90026-X.
- 15 Uffe Haagerup. A new upper bound for the complex Grothendieck constant. *Israel J. Math.*, 60(2):199–224, 1987. doi:10.1007/BF02790792.
- 16 Uffe Haagerup and Takashi Itoh. Grothendieck type norms for bilinear forms on C^* -algebras. *J. Operator Theory*, 34(2):263–283, 1995.
- 17 Aram Wettroth Harrow. Quantum expanders from any classical cayley graph expander. *Quantum Information & Computation*, 8(8):715–721, 2008. URL: <http://www.rintonpress.com/xxqic8/qic-8-89/0715-0721.pdf>.

- 18 M. B. Hastings. Random unitaries give quantum expanders. *Phys. Rev. A* (3), 76(3):032315, 11, 2007. doi:10.1103/PhysRevA.76.032315.
- 19 Matthew B Hastings. Superadditivity of communication capacity using entangled inputs. *Nature Physics*, 5(4):255, 2009.
- 20 Matthew B. Hastings and Aram Wettroth Harrow. Classical and quantum tensor product expanders. *Quantum Information & Computation*, 9(3):336–360, 2009. URL: <http://www.rintonpress.com/xxqic9/qic-9-34/0336-0360.pdf>.
- 21 Alexander S Holevo. Remarks on the classical capacity of quantum channel. *arXiv preprint quant-ph/0212025*, 2002.
- 22 Alexander S. Holevo. The additivity problem in quantum information theory. In *International Congress of Mathematicians. Vol. III*, pages 999–1018. Eur. Math. Soc., Zürich, 2006.
- 23 Shlomo Hoory, Nathan Linial, and Avi Wigderson. Expander graphs and their applications. *Bull. Amer. Math. Soc.*, 43:439–561, 2006.
- 24 Yoshiharu Kohayakawa, Vojtěch Rödl, and Mathias Schacht. Discrepancy and eigenvalues of Cayley graphs. *Czechoslovak Math. J.*, 66(141)(3):941–954, 2016. doi:10.1007/s10587-016-0302-x.
- 25 M. Krivelevich and B. Sudakov. Pseudo-random graphs. In *More sets, graphs and numbers*, volume 15 of *Bolyai Soc. Math. Stud.*, pages 199–262. Springer, Berlin, 2006. doi:10.1007/978-3-540-32439-3_10.
- 26 Alexander Lubotzky, Ralph Phillips, and Peter Sarnak. Ramanujan graphs. *Combinatorica*, 8(3):261–277, 1988.
- 27 Grigorii Aleksandrovich Margulis. Explicit group-theoretical constructions of combinatorial schemes and their application to the design of expanders and concentrators. *Problems of Information Transmission*, 24(1):39–46, 1988.
- 28 Assaf Naor, Oded Regev, and Thomas Vidick. Efficient rounding for the noncommutative Grothendieck inequality. *Theory Comput.*, 10(11):257–295, 2014. Earlier version in STOC’13.
- 29 Gilles Pisier. Grothendieck’s theorem, past and present. *Bull. Amer. Math. Soc. (N.S.)*, 49(2):237–323, 2012. doi:10.1090/S0273-0979-2011-01348-9.
- 30 J. Reeds. A new lower bound on the real Grothendieck constant. Manuscript (<http://www.dtc.umn.edu/~reedsj/bound2.dvi>), 1991.
- 31 Oded Regev and Thomas Vidick. Quantum XOR games. *ACM Trans. Comput. Theory*, 7(4):Art. 15, 43, 2015. doi:10.1145/2799560.
- 32 Barry Simon. *Representations of finite and compact groups*. Number 10. American Mathematical Soc., 1996.
- 33 Andrew Thomason. Pseudorandom graphs. In *Random graphs ’85 (Poznań, 1985)*, volume 144 of *North-Holland Math. Stud.*, pages 307–331. North-Holland, Amsterdam, 1987.
- 34 Andrew Thomason. Random graphs, strongly regular graphs and pseudorandom graphs. In *Surveys in combinatorics 1987 (New Cross, 1987)*, volume 123 of *London Math. Soc. Lecture Note Ser.*, pages 173–195. Cambridge Univ. Press, Cambridge, 1987.

Towards Quantum One-Time Memories from Stateless Hardware

Anne Broadbent

Department of Mathematics and Statistics, University of Ottawa, Canada
abroadbe@uottawa.ca

Sevag Gharibian

Department of Computer Science, Paderborn University, Germany
Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
sevag.gharibian@upb.de

Hong-Sheng Zhou

Department of Computer Science, Virginia Commonwealth University, Richmond, VA, USA
hszhou@vcu.edu

Abstract

A central tenet of theoretical cryptography is the study of the minimal assumptions required to implement a given cryptographic primitive. One such primitive is the one-time memory (OTM), introduced by Goldwasser, Kalai, and Rothblum [CRYPTO 2008], which is a classical functionality modeled after a non-interactive 1-out-of-2 oblivious transfer, and which is complete for one-time classical and quantum programs. It is known that secure OTMs do not exist in the standard model in both the classical and quantum settings. Here, we propose a scheme for using quantum information, together with the assumption of stateless (i.e., reusable) hardware tokens, to build statistically secure OTMs. Via the semidefinite programming-based quantum games framework of Gutoski and Watrous [STOC 2007], we prove security for a malicious receiver, against a linear number of adaptive queries to the token, in the quantum universal composability framework, but leave open the question of security against a polynomial amount of queries. Compared to alternative schemes derived from the literature on quantum money, our scheme is technologically simple since it is of the “prepare-and-measure” type. We also show our scheme is “tight” according to two scenarios.

2012 ACM Subject Classification Theory of computation → Cryptographic protocols

Keywords and phrases quantum cryptography, one-time memories, semi-definite programming

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.6

Related Version A full version of the paper is available at <https://arxiv.org/abs/1810.05226>.

Funding *Anne Broadbent*: U.S. Air Force Office of Scientific Research under award number FA9550-17-1-0083, Canada’s NSERC, Ontario ERA, and the University of Ottawa’s Research Chairs program.
Sevag Gharibian: NSF grants CCF-1526189 and CCF-1617710.

Acknowledgements We thank referees for pointing out the impossibility result against quantum queries applies only if we model the token as a *reversible* process, as well as for finding an error in a prior version of this work. We thank Kai-Min Chung and Jamie Sikora for related discussions.

1 Introduction

Theoretical cryptography centers around building cryptographic primitives secure against adversarial attacks. In order to allow a broader set of such primitives to be implemented, one often considers restricting the power of the adversary. For example, one can limit the *computing* power of adversaries to be polynomial bounded [68, 7], restrict the *storage* of adversaries to be bounded or noisy [49, 11, 22], or make *trusted setups* available to honest players [39, 6, 14, 16, 36, 55, 42, 46, 47, 48, 41, 40], to name a few. One well-known trusted



© Anne Broadbent, Sevag Gharibian, and Hong-Sheng Zhou;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 6; pp. 6:1–6:25



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

setup is *tamper-proof hardware* [38, 30], which is assumed to provide a specific input-output functionality, and which can only be accessed in a “black box” fashion. The hardware can maintain a state (i.e., is *stateful*) and possibly carry out complex functionality, but presumably may be difficult or expensive to implement or manufacture. This leads to an interesting research direction: Building cryptography primitives using the *simplest* (and hence easiest and cheapest to manufacture) hardware.

In this respect, two distinct simplified notions of hardware have captured considerable interest. The first is the notion of a *one-time memory (OTM)* [30], which is arguably the simplest possible notion of *stateful* hardware. An OTM, modeled after a non-interactive 1-out-of-2 oblivious transfer, behaves as follows: first, a player (called the *sender*) embeds two values s_0 and s_1 into the OTM, and then gives the OTM to another player (called the *receiver*). The receiver can now read his choice of precisely one of s_0 or s_1 ; after this “use” of the OTM, however, the unread bit is lost forever. Interestingly, OTMs are complete for implementing *one-time* use programs (OTPs): given access to OTMs, one can implement statistically secure OTPs for any efficiently computable program in the universal composability (UC) framework [32]. (OTPs, in turn, have applications in software protection and one-time proofs [30].) In the quantum UC model, OTMs enable *quantum* one-time programs [9]. (This situation is analogous to the case of *oblivious transfer* being complete for two-party secure function evaluation [39, 36].) Unfortunately, OTMs are inherently *stateful*, and thus represent a very strong cryptographic assumption – any physical implementation of such a device must somehow maintain internal knowledge between activations, i.e., it must completely “self-destruct” after a single use.

This brings us to a second important simplified notion of hardware known as a *stateless* token [17], which keeps no record of previous interactions. On the positive side, such hardware is presumably easier to implement. On the negative side, an adversary can run an experiment with stateless hardware as many times as desired, and each time the hardware is essentially “reset”. (Despite this, stateless hardware has been useful in achieving *computationally secure* multi-party computation [17, 32, 19], and *statistically secure* commitments [23].) It thus seems impossible for stateless tokens to be helpful in implementing any sort of “self-destruct” mechanism. Indeed, classically stateful tokens are trivially more powerful than stateless ones, as observed in, e.g., [32]. This raises the question:

Can quantum information, together with a classical stateless token, be used to simulate “self destruction” of a hardware token?

In particular, a natural question along these lines is whether quantum information can help implement an OTM. Unfortunately, it is known that quantum information *alone* cannot implement an OTM (or, more generally, any one-time program) [9]; see also Section 4 below. We thus ask the question: What are the minimal cryptographic assumptions required in a quantum world to implement an OTM?

1.1 Contributions and summary of techniques

We propose what is, to our knowledge, the first prepare-and-measure quantum protocol that constructs OTMs from stateless hardware tokens. For this protocol, we are able to rigorously prove information theoretic security against an adversary making a *linear* (in n , the security parameter) number of adaptive queries to the token. While we conjecture that security holds also for *polynomially* many queries, note that already in this setting of linearly many adaptive queries, our protocol achieves something impossible classically (i.e., classically, obtaining security against a linear number of queries is impossible). We also show stand-alone security against a malicious sender.

Historical Note. We proposed the concept that quantum information could provide a “stateless to stateful” transformation in a preliminary version of this work [8]; however, that work claimed security against a *polynomial* number of token queries, obtained via a reduction from the interactive to the non-interactive setting. We thank an anonymous referee for catching a subtle, but important bug which ruled out the proof approach of [8]. The current paper employs a different proof approach, which models interaction with the token as a “quantum game” via semidefinite programming. Since our original paper was posted, recent work [20] has shown an alternate quantum “stateful to stateless” transformation via quantum money constructions [3]. Specifically, in [20], security against a polynomial number of queries is achieved, albeit with respect to a new definition of “OTMs relative to an oracle” (while the security results of the present paper are with respect to the well-established simulation-based definition of [32, 38]). Furthermore, [20] directly applies known quantum money constructions, which require difficult-to-prepare highly entangled states. Our focus here, in contrast, is to take a “first-principles” approach and build a technologically simple-to-implement scheme which requires no entanglement, but rather the preparation of just one of four single qubit states, $|0\rangle, |1\rangle, |+\rangle, |-\rangle$. Indeed, the two works are arguably complementary in that the former focuses primarily on *applications* of “stateful” single-use tokens, while our focus is on the most technologically simple way to *implement* such “stateful” tokens.

Construction. Our construction is inspired by Wiesner’s *conjugate coding* [65]: the quantum portion of the protocols consists in n quantum states chosen uniformly at random from $\{|0\rangle, |1\rangle, |+\rangle, |-\rangle\}$ (note this encoding is independent of the classical bits of the OTM functionality). We then couple this n -qubit quantum state, $|\psi\rangle$ (the *quantum key*) with a *classical* stateless hardware token, which takes as inputs a choice bit b , together with an n -bit string y . If $b = 0$, the hardware token verifies that the bits of y that correspond to *rectilinear* ($|0\rangle$ or $|1\rangle$, i.e., Z basis) encoded qubits of $|\psi\rangle$ are consistent with the measurement of $|\psi\rangle$ in the computational basis, in which case the bit s_0 is returned. If $b = 1$, the hardware token verifies that the bits of y that correspond to *diagonal* ($|+\rangle$ or $|-\rangle$, i.e., X basis) encoded qubits of $|\psi\rangle$ are consistent with the measurement of $|\psi\rangle$ in the diagonal basis, in which case the bit s_1 is returned.¹ The honest use of the OTM is thus intuitive: for choice bit $b = 0$, the user measures each qubit of the quantum key in the rectilinear basis to obtain an n -bit string y , and inputs (b, y) into the hardware token. If $b = 1$, the same process is applied, but with measurements in the diagonal basis.

Assumption. Crucially, we assume the hardware token accepts *classical* input only (alternatively and equivalently, the token immediately measures its quantum input in the standard basis), i.e., it cannot be queried in superposition. Although this may seem a strong assumption, in Section 4 we show that any token which can be queried in superposition in a reversible way, cannot be used to construct a secure OTM (with respect to our setting in which the adversary is allowed to apply arbitrary quantum operations). Similar classical-input hardware has previously been considered in, e.g., [60, 9].

Security and intuition. Stand-alone security against a malicious sender is relatively simple to establish, since the protocol consists in a single message from the sender to the receiver, and stand-alone security only requires simulation of the *local* view of the adversary.

¹ We note that a simple modification using a classical one-time pad could be used to make *both* the quantum state and hardware token independent of s_0 and s_1 : the token would output one of two uniformly random bits r_0 and r_1 , which could each be used to decrypt a single bit, s_0 or s_1 .

The intuition underlying security against a malicious receiver is clear: in order for a receiver to extract a bit s_b as encoded in the OTM, she must perform a complete measurement of the qubits of $|\psi\rangle$ in order to obtain a classical key for s_b (since, otherwise, she would likely fail the test as imposed by the hardware token). But such a measurement would invalidate the receiver’s chance of extracting the bit s_{1-b} ! This is exactly the “self-destruct”-like property we require in order to implement an OTM. This intuitive notion of security was present in Wiesner’s proposal for quantum money [65], and is often given a physical explanation in terms of the no-cloning theorem [67] or Heisenberg uncertainty relation [35].

Formally, we work in the statistical (i.e., information-theoretic) setting of the quantum *Universal Composability* (UC) framework [59], which allows us to make strong security statements that address the *composability* of our protocol within others. As a proof technique, we describe a simulator, such that for any “quantum environment” wishing to interact with the OTM, the environment statistically cannot tell whether it is interacting with the *ideal* OTM functionality or the *real* OTM instance provided by our scheme. The security of this simulator requires a statement of the following form: Given access to a (randomly chosen) “quantum key” $|\psi_k\rangle$ and corresponding stateless token V_k , it is highly unlikely for an adversary to successfully extract keys for *both* the secret bits s_0 and s_1 held by V_k . We are able to show this statement for any adversary which makes a linear number of queries, by which we mean an adversary making m queries succeeds with probability at most $O(2^{2m-0.228n})$ (for n the number of quantum key bits in $|\psi_k\rangle$). In other words, if the adversary makes at most $m = cn$ queries with $c < 0.114$, then its probability of cheating successfully is exponentially small in n . We conjecture, however, that a similar statement holds for any $m \in \text{poly}(n)$, i.e., that the protocol is secure against polynomially many queries.

To show security against linearly many queries, we exploit the semidefinite programming-based quantum games framework of Gutoski and Watrous (GW) [33] to model interaction with the token. Intuitively, GW is useful for our setting, since it is general enough to model multiple rounds of adaptive queries to the token, even when the receiver holds quantum “side information” in the form of $|\psi\rangle$. We describe this technique in Sections 2.1 and 3.4, and provide formal details in the full version. Summarizing, we show the following.

► **Main Theorem (informal).** *There exists a protocol Π , which together with a classical stateless token and the ability to randomly prepare single qubits in one of four pure states, implements the OTM functionality with statistical security in the UC framework against a corrupted receiver making a linear number of adaptive queries.*

As stated above, we conjecture that our protocol is actually secure against polynomially many adaptive queries. However, we are unable to show this claim using our present proof techniques, and hence leave this question open. Related to this, we make the following comments: (1) As far as we are aware, the Main Theorem above is the only known formal proof of any type of security for conjugate coding in the interactive setting with $\Omega(1)$ queries. Moreover, as stated earlier, classically security against $\Omega(1)$ queries is trivially impossible. (2) Our proof introduces the GW semidefinite programming framework from quantum interactive proofs to the study of conjugate coding-based schemes. This framework allows handling multiple challenges in a unified fashion: arbitrary quantum operations by the user, classical queries to the token, and the highly non-trivial assumption of quantum side information for the user (the “quantum key” state sent to the user.)

Towards security against polynomially many queries. Regarding the prospects of proving security against polynomially many adaptive queries, we generally believe it requires a significant new insight into how to design a “good” feasible solution to the primal semidefinite program (SDP) obtained via GW. However, in addition to our proof for linear security, in the

full version we give evidence potentially supporting our conjecture for polynomial security. Namely, we first simplify the SDPs obtained from GW, and derive the corresponding dual SDPs. These derivations apply for any instantiation of the GW framework, i.e. they are not specific to our setting, and hence may prove useful elsewhere. We then give a feasible solution Y to the dual SDP. While Y is simple to state, it is somewhat involved to analyze. A heuristic analysis suggests Y 's dual objective function value has precisely the behavior needed to show security, i.e. the value scales as $m/\sqrt{2^n}$, for m queries and n key bits. If Y were to be the *optimal* solution to the dual SDP, this would strongly suggest the optimal cheating probability is essentially $m/\sqrt{2^n}$. However, we explicitly show Y is not optimal, and so $m/\sqrt{2^n}$ is only a *lower bound* on the optimal cheating probability. Nevertheless, we conjecture that while Y is not optimal, it is *approximately* optimal; this would imply the desired polynomial security claim. Unfortunately, the only techniques we are aware of to show approximate optimality require a better primal SDP solution, which appears challenging.

Further related work. Our work contributes to the growing list of functionalities achievable with quantum information, yet unachievable classically. This includes: unconditionally secure key expansion [4], physically uncloneable money [65, 51, 53], a reduction from oblivious transfer to bit commitment [5, 21] and to other primitives such as “cut-and choose” functionality [27], and revocable time-release quantum encryption [61]. Importantly, these protocols all make use of the technique of conjugate coding [65], which is also an important technique used in protocols for OT in the bounded quantum storage and noisy quantum storage models [22, 63] (see [10] for a survey).

Various proof techniques have been developed in the context of conjugate coding, including entropic uncertainty relations [64]. In the context of QKD, another technique is the use of de Finetti reductions [58] (which exploit the symmetry of the scheme in order to simplify the analysis). Recently, semidefinite programming (SDP) approaches have been applied to analyze security of conjugate coding [51] for quantum money, in the setting of one round of interaction with a “stateful” bank. SDPs are also the technical tool we adopt for our proof (Section 3.4), though here we require the more advanced quantum games SDP framework of Gutoski and Watrous [33] to deal with multiple adaptive interactions with stateless tokens. Reference [53] has also made use of Gavinsky’s [28] quantum retrieval games framework.

Somewhat similar to [53], Aaronson and Christiano [1] have studied quantum money schemes in which one interacts with a verifier. They introduce an “inner product adversary method” to lower bound the number of queries required to break their scheme.

We remark that [53] and [51] have studied schemes based on conjugate coding similar to ours, but in the context of quantum money. In contrast to our setting, the schemes of [53] and [51] (for example) involve dynamically chosen random challenges from a verifier to the holder of a “quantum banknote”, whereas in our work here the “challenges” are fixed (i.e., measure all qubits in the Z or X basis to obtain secret bit s_0 or s_1 , respectively), and the verifier is replaced by a stateless token. Thus, [51], for example, may be viewed as using a “stateful” verifier, whereas our focus here is on a “stateless” verifier (i.e., a token).

Also, prior work has achieved oblivious transfer using quantum information, together with some assumption (e.g., bit commitment [5], bounded quantum storage [22]). These protocols typically use an interaction phase similar to the “commit-and-open” protocol of [5]; because we are working in the non-interactive setting, these techniques appear to be inapplicable.

Finally, Liu [43, 44, 45] has given stand-alone secure OTMs using quantum information in the *isolated-qubit model*. Liu’s approach is nice in that it avoids the use of trusted setups. In return, however, Liu must use the isolated-qubit model, which restricts the adversary to

perform only single-qubit operations (no entangling gates are permitted); this restriction is, in some sense, necessary if one wants to avoid trusted setups, as a secure OTM in the plain quantum model cannot exist (see Section 4). In contrast, in the current work we allow unbounded and unrestricted quantum adversaries, but as a result require a trusted setup. In addition, we remark the security notion of OTMs of [43, 44, 45] is weaker than the simulation-based notion studied in this paper, and it remains an interesting open question whether the type of OTM in [43, 44, 45] is secure under composition (in the current work, the UC framework gives us security under composition for free).

Significance. Our results show a strong separation between the classical and quantum settings, since classically, stateless tokens cannot be used to securely implement OTMs. To the best of our knowledge, our work is the first to combine conjugate coding with *stateless* hardware tokens. Moreover, while our protocol shares similarities with prior work in the setting of quantum money, building OTMs appears to be a new focus here ².

Our protocol has a simple implementation, fitting into the single-qubit prepare-and-measure paradigm, which is widely used as the “benchmark” for a “physically feasible” quantum protocol (in this model, one needs only the ability to prepare single-qubit states $|0\rangle, |1\rangle, |+\rangle, |-\rangle$, and to perform single-qubit projective measurements. In particular, no entangled states are required, and in principle no quantum memory is required, since qubits can be measured one-by-one as they arrive). In addition, from a theoretical cryptographic perspective, our protocol is attractive in that its implementation requires an assumption of a stateless hardware token, which is easier and cheaper to mass produce than a stateful token.

In terms of security guarantees, we allow *arbitrary* operations on behalf of a malicious quantum receiver in our protocol (i.e., all operations allowed by quantum mechanics), with the adversary restricted in that the stateless token is assumed only usable as a black box. The security we obtain is statistical, with the only computational assumption being on the number of *queries* made to the token (recall we show security for a linear number of queries, and conjecture security for polynomially many queries). Finally, our security analysis is in the quantum UC framework against a corrupted receiver; this means our protocol can be easily composed with many others; for example, combining our results with [9]’s protocol immediately yields UC-secure quantum OTPs against a dishonest receiver.

Finally, our scheme is “tight” with respect to two impossibility results (Section 4), both of which assume the adversary has black-box access to both the token and its inverse operation³. First, the assumption that the token be queried only in the computational basis cannot be relaxed: If the token can be queried in superposition, then an adversary can easily break an OTM scheme. Second, our scheme has the property that corresponding to each secret bit s_i held by the token, there are exponentially many valid keys one can input to the token to extract s_i . We show that for any “measure-and-access” OTM (i.e., an OTM in which one measures a given quantum key and uses the classical measurement result to access a token to extract data, of which our protocol is an example), a polynomial number of keys implies the ability to break the scheme with inverse polynomial probability (more generally, Δ keys allows probability at least $1/\Delta^2$ of breaking the scheme).

² We remark, however, that a reminiscent concept of single usage of quantum “tickets” in the context of quantum money is very briefly mentioned in Appendix S.4.1 of [53].

³ This is common in the oracle model of quantum computation, where a function $f : \{0, 1\}^n \mapsto \{0, 1\}$ is implemented via the (self-inverse) unitary mapping $U_f|x\rangle|y\rangle = |x\rangle|y \oplus f(x)\rangle$.

Open Questions. While our work shows the fundamental advantage that quantum information yields in a stateful to stateless reduction, it does leave a number of open questions:

1. **Security against polynomially many queries.** Can our security proof be strengthened to show information theoretic security against a polynomial number of queries to the token? We conjecture this to be the case, but finding a formal proof has been elusive.
2. **Composable security against a malicious sender.** While we show composable security against a malicious receiver, our protocol can achieve standalone security against a malicious sender. Could an adaptation of our protocol ensure composable security against a malicious sender as well?⁴
3. **Non-reversible token.** Our impossibility result for quantum one-time memories with *quantum* queries (Section 4) assumes the adversary has access to reversible tokens; can a similar impossibility result be shown for non-reversible tokens? In Section 4, we briefly discuss why it may be difficult to extend the techniques of our impossibility results straightforwardly when the adversary does *not* have access to the inverse of the token.
4. **Imperfect devices.** While our prepare-and-measure scheme is technologically simple, it is still unrealizable with current technology, due to the requirement of perfect quantum measurements. We leave open the question of tolerance to a small amount of noise.

Organization. Section 2 covers preliminaries, including ideal functionalities for an OTM and stateless token, background on quantum channels, semidefinite programming, and the Gutoski-Watrous (GW) framework for quantum games. Section 3 gives our construction for an OTM based on a stateless hardware token; the proof ideas for security are also provided. Section 4 discusses “tightness” of our construction by showing two impossibility results for “relaxations” of our scheme. In the Appendix, we discuss classical UC and quantum UC (Appendix A); Appendix B establishes notation required in the definition of stand-alone security against a malicious sender. Due to space constraints, our formal security proof against a linear number of queries to the token (used to finish the security proof in Section 3) is deferred to the full version, along with simplifications of the GW SDP, derivation of its dual, and a dual feasible solution which we conjecture to be approximately optimal.

2 Preliminaries

Notation. Two binary distributions \mathbf{X} and \mathbf{Y} are *indistinguishable*, denoted $\mathbf{X} \approx \mathbf{Y}$, if $|\Pr(X_n = 1) - \Pr(Y_n = 1)| \leq \text{negl}(n)$. We define single-qubit $|0\rangle_+ = |0\rangle$ and $|1\rangle_+ = |1\rangle$, so that $\{|0\rangle_+, |1\rangle_+\}$ form the *rectilinear basis*. We define $|0\rangle_\times = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)$ and $|1\rangle_\times = \frac{1}{\sqrt{2}}(|0\rangle - |1\rangle)$, so that $\{|0\rangle_\times, |1\rangle_\times\}$ form the *diagonal basis*. For strings $x = x_1, x_2, \dots, x_n \in \{0, 1\}^n$ and $\theta = \theta_1, \theta_2, \dots, \theta_n \in \{+, \times\}^n$, define $|x\rangle_\theta = \bigotimes_{i=1}^n |x_i\rangle_{\theta_i}$. For \mathcal{X} a finite dimensional complex Hilbert space, $\mathcal{L}(\mathcal{X})$, $\text{Herm}(\mathcal{X})$, $\text{Pos}(\mathcal{X})$, and $\mathcal{D}(\mathcal{X})$ denote the sets of linear, Hermitian, positive semidefinite, and density operators acting on \mathcal{X} , respectively. Notation $A \succeq B$ means $A - B$ is positive semidefinite.

Quantum universal composition (UC) framework. We study simulation-based security in this paper. In particular, we prove security of our construction against a malicious receiver in the quantum universal composition (UC) framework [59]. See Appendix A for a description of classical UC [14] and quantum UC [59]. In the next two paragraphs, we introduce the ideal functionalities of one-time memory and stateless hardware token.

⁴ We note that this would require a different protocol, since in our current construction, a cheating sender could program the token to abort based on the user’s input.

One-time memory (OTM). The one-time memory (OTM) functionality \mathcal{F}_{OTM} involves two parties, the sender and the receiver, and consists of two phases, “Create” and “Execute”. Please see Functionality 1 below for details; for the sake of simplicity, we have omitted the session/party identifiers as they should be implicitly clear from the context. We sometimes refer to this functionality \mathcal{F}_{OTM} as an *OTM token*.

Functionality 1 Ideal functionality \mathcal{F}_{OTM} .

1. **Create:** Upon input (s_0, s_1) from the sender, with $s_0, s_1 \in \{0, 1\}$, send **create** to the receiver and store (s_0, s_1) .
 2. **Execute:** Upon input $b \in \{0, 1\}$ from the receiver, send s_b to receiver. Delete any trace of this instance.
-

Stateless hardware. The original work of Katz [38] introduces the ideal functionality $\mathcal{F}_{\text{wrap}}$ to model stateful tokens in the UC-framework. In the ideal model, a party that wants to create a token, sends the Turing machine to $\mathcal{F}_{\text{wrap}}$. $\mathcal{F}_{\text{wrap}}$ will then run the machine (keeping the state), when the designated party will ask for it. The same functionality can be adapted to model stateless tokens. It is sufficient that the functionality does not keep the state between two executions. A simplified version of the $\mathcal{F}_{\text{wrap}}$ functionality as shown in [17] (that is very similar to the $\mathcal{F}_{\text{wrap}}$ of [38]) is described below. Note that, again for the sake of simplicity, we have omitted the session/party identifiers as they should be implicitly clear from the context. Although the environment and adversary are unbounded, we specify that stateless

Functionality 2 Ideal functionality $\mathcal{F}_{\text{wrap}}$.

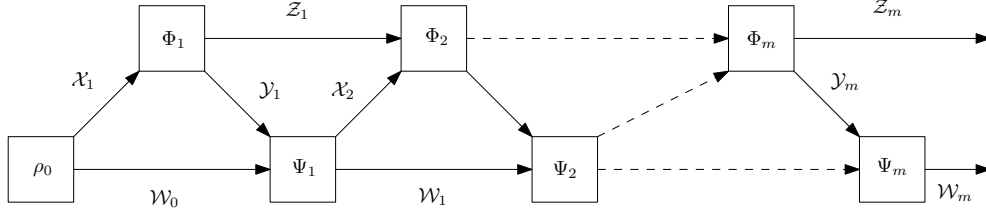
The functionality is parameterized by a polynomial $p(\cdot)$, and implicit security parameter n .

1. **Create:** Upon input (create, M) from the sender, where M is a Turing machine, send **create** to the receiver and store M .
 2. **Execute:** Upon input (run, msg) from the receiver, execute $M(\text{msg})$ for at most $p(n)$ steps, and let out be the response. Let $\text{out} := \perp$ if M does not halt in $p(n)$ steps. Send out to the receiver.
-

hardware can be queried only a polynomial number of times. This is necessary; otherwise the hardware token model is vacuous (with unbounded queries, the entire input-output behavior of stateless hardware can be extracted).

Quantum channels. A linear map $\Phi : \mathcal{L}(\mathcal{X}) \mapsto \mathcal{L}(\mathcal{Y})$ is a *quantum channel* if Φ is trace-preserving and completely positive (TPCP). Such maps take density operators to density operators. A useful representation of linear maps (or “superoperators”) $\Phi : \mathcal{L}(\mathcal{X}) \mapsto \mathcal{L}(\mathcal{Y})$ is the Choi-Jamiołkowski representation, $J(\Phi) \in \mathcal{L}(\mathcal{Y} \otimes \mathcal{X})$. The latter is defined (with respect to some choice of orthonormal basis $\{|i\rangle\}$ for \mathcal{X}) as $J(\Phi) = \sum_{i,j} \Phi(|i\rangle\langle j|) \otimes |i\rangle\langle j|$. The following properties of $J(\Phi)$ hold [18, 37]: (1) Φ is completely positive if and only if $J(\Phi) \succeq 0$, and (2) Φ is trace-preserving if and only if $\text{Tr}_{\mathcal{Y}}(J(\Phi)) = I_{\mathcal{X}}$. In a nutshell, the Gutoski-Watrous (GW) framework generalizes this definition to *interacting* strategies [33].

Semidefinite programs. We review semidefinite programs (SDPs) from the perspective of quantum information, as done e.g., in the notes of Watrous [62] or [51]. Given any 3-tuple (A, B, Φ) for operators $A \in \text{Herm}(\mathcal{X})$ and $B \in \text{Herm}(\mathcal{Y})$, and Hermiticity-preserving linear map $\Phi : \mathcal{L}(\mathcal{X}) \mapsto \mathcal{L}(\mathcal{Y})$, one can state a *primal* and *dual* semidefinite program:



■ **Figure 1** A general interaction between two quantum parties.

Primal problem (P)	Dual problem (D)
$\sup \quad \text{Tr}(AX)$	$\inf \quad \text{Tr}(BY)$
s.t. $\Phi(X) = B,$	s.t. $\Phi^*(Y) \succeq A$
$X \in \text{Pos}(\mathcal{X}),$	$Y \in \text{Herm}(\mathcal{Y}),$

where Φ^* denotes the *adjoint* of Φ , which is the unique map satisfying $\text{Tr}(A^\dagger \Phi(B)) = \text{Tr}((\Phi^*(A))^\dagger B)$ for all $A \in \mathcal{L}(\mathcal{Y})$ and $B \in \mathcal{L}(\mathcal{X})$. Not all SDPs have feasible solutions (i.e. a solution satisfying all constraints); in this case, optimal values are $-\infty$ for P and ∞ for D.

2.1 The Gutoski-Watrous framework for quantum games

We now recall the Gutoski-Watrous (GW) framework for quantum games [33], which can be used to model quantum interactions between spatially separated parties. The setup most relevant to our protocol here is depicted in Figure 1. Here, we imagine one party, A , prepares an initial state $\rho_0 \in \mathcal{D}(\mathcal{X}_1 \otimes \mathcal{W}_0)$. Register \mathcal{X}_1 is then sent to the second party (\mathcal{W}_0 is kept as private memory), B , who applies some quantum channel $\Phi_i : \mathcal{L}(\mathcal{X}_1) \mapsto \mathcal{L}(\mathcal{Y}_1 \otimes \mathcal{Z}_1)$. B keeps register \mathcal{Z}_1 as private memory, and sends \mathcal{Y}_1 back to A , who applies channel $\Psi_1 : \mathcal{L}(\mathcal{W}_0 \otimes \mathcal{Y}_1) \mapsto \mathcal{L}(\mathcal{X}_2 \otimes \mathcal{W}_1)$, and sends \mathcal{X}_2 to B . The protocol continues for m messages back and forth, until the final operation $\Psi_m : \mathcal{L}(\mathcal{W}_m \otimes \mathcal{Y}_m) \mapsto \mathbb{C}$, in which A performs a two-outcome measurement (specifically, a POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$, meaning $\Lambda_0, \Lambda_1 \succeq 0$, $\Lambda_0 + \Lambda_1 = I$) in order to decide whether to reject (Λ_0) or accept (Λ_1). As done in [33], without loss of generality (by the Stinespring dilation theorem) all channels are given by linear isometries A_k , i.e. $\Phi_k(X) = A_k X A_k^\dagger$. Reference [33] refers to (Φ_1, \dots, Φ_m) as a *strategy* and $(\rho_0, \Psi_1, \dots, \Psi_m)$ as a *co-strategy*. In our setting, the former is “non-measuring”, meaning it makes no final measurement after Φ_m is applied, whereas the latter is “measuring”, since we will apply a final measurement on space \mathcal{W}_m (not depicted in Figure 1).

Intuitively, since our protocol (Section 3.1) begins with the token sending the user a quantum key $|x\rangle_\theta$, we will model the token as a *measuring co-strategy*, and the user as a *strategy*. The advantage to doing so is that the GW framework allows one to (recursively) characterize any such strategy (resp., co-strategy) via a set of linear (in)equalities and positive semi-definite constraints. (In this sense, the GW framework generalizes the Choi-Jamiołkowski representation for channels to a “Choi-Jamiołkowski” representation for strategies/co-strategies.) To state these constraints, we first write down the Choi-Jamiołkowski (CJ) representation of a strategy (resp., measuring co-strategy) from [33].

CJ representation of (non-measuring) strategy. The CJ representation of a strategy (A_1, \dots, A_m) is given by matrix [33]

$$\text{Tr}_{\mathcal{Z}_m}(\text{vec}(A) \text{vec}(A)^\dagger), \quad (1)$$

where $A \in \mathcal{L}(\mathcal{X}_1 \otimes \cdots \otimes \mathcal{X}_m, \mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m \otimes \mathcal{Z}_m)$ is the product of the isometries A_i ,

$$A := (I_{\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_{m-1}} \otimes A_m) \cdots (A_1 \otimes I_{\mathcal{X}_2 \otimes \cdots \otimes \mathcal{X}_m}), \quad (2)$$

and the $\text{vec} : \mathcal{L}(\mathcal{S}, \mathcal{T}) \mapsto \mathcal{T} \otimes \mathcal{S}$ mapping is the linear extension of the map $|i\rangle\langle j| \mapsto |i\rangle|j\rangle$ defined on all standard basis states $|i\rangle, |j\rangle$.

CJ representation of (measuring) co-strategy. Let $P := \{\Lambda_0, \Lambda_1\}$ denote a POVM with reject and accept measurement operators Λ_0 and Λ_1 , respectively. A measuring strategy which ends with a measurement via POVM Λ replaces, for $\Lambda_a \in \Lambda$, Equation (1) with [33]

$$Q_a := \text{Tr}_{\mathcal{Z}_m}((\Lambda_a \otimes I_{\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m}) \text{vec}(A) \text{vec}(A)^\dagger) = \text{Tr}_{\mathcal{Z}_m}(\text{vec}(B_a) \text{vec}(B_a)^\dagger), \quad (3)$$

for $B_a := (\sqrt{\Lambda_a} \otimes I_{\mathcal{Y}_1 \otimes \cdots \otimes \mathcal{Y}_m})A$. To convert this to a *co*-strategy, one takes the transpose of the operators defined above (with respect to the standard basis).

Optimization characterization over strategies and co-strategies. With CJ representations for strategies and co-strategies in hand, one can formulate [33] the optimal probability with which a strategy can force a corresponding co-strategy to output a desired result as follows. Fix any Q_a from a measuring co-strategy $\{Q_0, Q_1\}$, as in Equation (3). Then, Corollary 7 and Theorem 9 of [33] show that the maximum probability with which a (non-measuring) strategy can force the co-strategy to output result a is given by

$$\min: \quad p \quad (4)$$

$$\text{subject to: } Q_a \preceq p R_m \quad (5)$$

$$R_k = P_k \otimes I_{\mathcal{Y}_k} \quad \text{for } 1 \leq k \leq m \quad (6)$$

$$\text{Tr}_{\mathcal{X}_k}(P_k) = R_{k-1} \quad \text{for } 1 \leq k \leq m \quad (7)$$

$$R_0 = 1 \quad (8)$$

$$R_k \in \text{Pos}(\mathcal{Y}_{1,\dots,k} \otimes \mathcal{X}_{1,\dots,k}) \quad \text{for } 1 \leq k \leq m \quad (9)$$

$$P_k \in \text{Pos}(\mathcal{Y}_{1,\dots,k-1} \otimes \mathcal{X}_{1,\dots,k}) \quad \text{for } 1 \leq k \leq m \quad (10)$$

$$p \in [0, 1] \quad (11)$$

Intuition. The minimum p denotes the optimal “success” probability, meaning the optimal probability of forcing the co-strategy to output a (Theorem 9 of [33]). The variables above, in addition to p , are $\{R_i\}$ and $\{P_i\}$, where the optimization is happening over all m -round co-strategies R_m satisfying Equation (5). How do we enforce that R_m encodes such an m -round co-strategy? This is given by the (recursive) Equations (6)-(10). Specifically, Corollary 7 of [33] states that R_m is a valid m -round co-strategy if and only if all of the following hold: (1) $R_m \succeq 0$, (2) $R_m = P_m \otimes I_{\mathcal{Y}_m}$ for $P_m \succeq 0$ and \mathcal{Y}_m the last incoming message register to the co-strategy, (3) $\text{Tr}_{\mathcal{X}_m}(P_m)$ is a valid $m-1$ round co-strategy (this is the recursive part of the definition). An intuitive sense as to why conditions (2) and (3) should hold is as follows: For any m -round co-strategy R_m , let R_{m-1} denote R_m restricted to the first $m-1$ rounds. Then, to operationally obtain R_{m-1} from R_m , the co-strategy first ignores the last incoming message in register \mathcal{Y}_m . This is formalized via a partial trace over \mathcal{Y}_m , which (once pushed through the CJ formalism⁵) translates into the $\otimes I_{\mathcal{Y}_k}$ term

⁵ Recall that the CJ representation of the trace map is the identity matrix (up to scaling).

in Equation (6). Since the co-strategy is now ignoring the last *incoming* message \mathcal{Y}_m , any measurement it makes after $m - 1$ rounds is independent of the last *outgoing* message \mathcal{X}_m . Thus, we can trace out \mathcal{X}_m as well, obtaining a co-strategy R_{m-1} on just the first $m - 1$ rounds; this is captured by Equation (7).

3 Feasibility of Quantum OTMs using Stateless Hardware

In this section, we present a *quantum* construction for one-time memories by using stateless hardware (Section 3.1). We also state our main theorem (Theorem 1). In Section 3.3, we describe the Simulator and prove Theorem 1 using the technical results of the full version. The intuition and techniques behind the proofs in the full version are sketched in Section 3.4.

3.1 Construction

We now present the OTM protocol Π in the $\mathcal{F}_{\text{wrap}}$ hybrid model, between a sender P_s and a receiver P_r . Here the security parameter is n .

- Upon receiving input (s_0, s_1) from the environment where $s_0, s_1 \in \{0, 1\}$, the sender:
 - The sender chooses uniformly random $x \in_R \{0, 1\}^n$ and $\theta \in_R \{+, \times\}^n$, and prepares $|x\rangle_\theta$. Based on (s_0, s_1, x, θ) , the sender prepares program M as in **Program 1**.

Program 1 Program for hardware token.

Hardcoded values: $s_0, s_1 \in \{0, 1\}$, $x \in \{0, 1\}^n$, and $\theta \in \{+, \times\}^n$

Inputs: $y \in \{0, 1\}^n$ and $b \in \{0, 1\}$, where y is a claimed measured value for the quantum register, and b the evaluator's choice bit

1. If $b = 0$, check that the $\theta = +$ positions return the correct bits in y according to x . If Accept, output s_0 . Otherwise output \perp .
 2. If $b = 1$, check that the $\theta = \times$ positions return the correct bits in y according to x . If Accept, output s_1 . Otherwise output \perp .
-

- The sender sends $|x\rangle_\theta$ to the receiver.
- The sender sends (create, M) to functionality $\mathcal{F}_{\text{wrap}}$, and the functionality sends **create** to notify the receiver.
- The receiver P_r operates as follows:

Upon input b from the environment, and $|x\rangle_\theta$ from the receiver, and **create** notification from $\mathcal{F}_{\text{wrap}}$,

 - If $b = 0$, measure $|x\rangle_\theta$ in computational basis to get y . Input $(\text{run}, (y, b))$ into $\mathcal{F}_{\text{wrap}}$.
 - If $b = 1$, apply $H^{\otimes n}$ to $|x\rangle_\theta$, then measure in computational basis to get y . Input $(\text{run}, (y, b))$ into $\mathcal{F}_{\text{wrap}}$.

Return the output of $\mathcal{F}_{\text{wrap}}$ to the environment.

It is easy to see that the output of $\mathcal{F}_{\text{wrap}}$ is s_b for both $b = 0$ and $b = 1$.

Note again that the hardware token, as defined in **Program 1**, accepts only classical input (i.e., it cannot be queried in superposition). As mentioned earlier, relaxing this assumption yields impossibility of a secure OTM implementation (assuming the receiver also has access to the token's inverse operation), as shown in Section 4.

3.2 Stand-Alone Security Against a Malicious Sender

We note that in protocol Π of Section 3.1, once the sender prepares and sends the token, she is no longer involved (and in particular, the sender does not receive any further communication from the receiver). We call such a protocol a *one-way* protocol. Because of this simple structure, and because the ideal functionality $\mathcal{F}_{\text{wrap}}$ also does not return any message to the sender, we can easily establish stand-alone security against a malicious sender (Appendix B).

3.3 UC-Security against a corrupt receiver

Our main theorem, which establishes security against a corrupt receiver is now stated.

► **Theorem 1.** *Construction Π above quantum-UC-realizes \mathcal{F}_{OTM} in the $\mathcal{F}_{\text{wrap}}$ hybrid model with statistical security against an actively-corrupted receiver making at most cn number of adaptive queries to the token, for any fixed constant $c < 0.114$.*

To prove Theorem 1, we now construct and analyze an appropriate simulator.

3.3.1 The simulator

In order to prove Theorem 1, for an adversary \mathcal{A} that corrupts the receiver, we build a simulator \mathcal{S} (having access to the OTM functionality \mathcal{F}_{OTM}), such that for any unbounded environment \mathcal{Z} , the executions in the real model and that in simulation are statistically indistinguishable. Our simulator \mathcal{S} is given below:

- The simulator emulates an internal copy of the adversary \mathcal{A} who corrupts the receiver. The simulator emulates the communication between \mathcal{A} and the external environment \mathcal{Z} by forwarding the communication messages between \mathcal{A} and \mathcal{Z} .
- The simulator \mathcal{S} needs to emulate the whole view for the adversary \mathcal{A} . First, \mathcal{S} picks dummy inputs $\tilde{s}_0 = 0$ and $\tilde{s}_1 = 0$, and randomly chooses $x \in \{0, 1\}^n$, and $\theta \in \{+, \times\}^n$, and generates program \tilde{M} . Then the simulator plays the role of the sender to send $|x\rangle_\theta$ to the adversary \mathcal{A} (who controls the corrupted receiver). The simulator also emulates $\mathcal{F}_{\text{wrap}}$ to notify \mathcal{A} by sending `create` to indicate the hardware is ready for queries.
- For each query $(\text{run}, (b, y))$ to $\mathcal{F}_{\text{wrap}}$ from the adversary \mathcal{A} , the simulator evaluates program \tilde{M} (created based on $\tilde{s}_0, \tilde{s}_1, x, \theta$) as in the construction, and then acts as follows:
 1. If this is a rejecting input, output \perp .
 2. If this is the first accepting input, call the external \mathcal{F}_{OTM} with input b , and learn the output s_b from \mathcal{F}_{OTM} . Output s_b .
 3. If this is a subsequent accepting input, output s_b (as above).

3.3.2 Analysis

We now show that the simulation and the real model execution are statistically indistinguishable. There are two cases in an execution of the simulation which we must consider:

- *Case 1: In all its queries to $\mathcal{F}_{\text{wrap}}$, the accepting inputs of \mathcal{A} have the same choice bit b .* In this case, the simulation is perfectly indistinguishable.
- *Case 2: In its queries to $\mathcal{F}_{\text{wrap}}$, \mathcal{A} produces accepting inputs for both $b = 0$ and $b = 1$.* In this case, it is possible that the simulation fails (the environment can distinguish the real model from the ideal model), since the simulator is only able to retrieve a single bit from the external OTM functionality \mathcal{F}_{OTM} (either corresponding to $b = 0$ or $b = 1$).

Thus, whereas in Case 1 the simulator behaves perfectly, in Case 2 it is in trouble. Fortunately, in Theorem 2 we show that the probability that Case 2 occurs is exponentially small in n , the number of qubits comprising $|x\rangle_\theta$, provided the number of queries to the token is at most cn for any $c < 0.114$. Specifically, we show that for an arbitrary m -query strategy (i.e., any quantum strategy allowed by quantum mechanics, whether efficiently implementable or not, which queries the token at most m times), the probability of Case 2 occurring is at most $O(2^{2m-0.228n})$. This concludes the proof.

3.4 Security analysis for the token: Intuition

Our simulation proof showing statistical security of our Quantum OTM construction of Section 3.1 relies crucially on Theorem 2, stated below. For this, we now introduce notation in line with the formal analysis of the full version.

With respect to the construction of Section 3.1, let us replace each two-tuple $(x, \theta) \in \{0, 1\}^n \times \{+, \times\}^n$ by a single string $z \in \{0, 1\}^{2n}$, which we denote the *secret key*. Bits $2i$ and $2i + 1$ of z specify the basis and value of conjugate coding qubit i for $i \in \{1, \dots, n\}$ (i.e., $z_{2i} = \theta_i$ and $z_{2i+1} = x_i$). Also, rename the “quantum key” (or conjugate coding key) $|\psi_z\rangle := |x\rangle_\theta \in (\mathbb{C}^2)^{\otimes n}$. Thus, the protocol begins by having the sender pick a *secret key* $z \in \{0, 1\}^{2n}$ uniformly at random, and preparing a joint state

$$|\psi\rangle = \frac{1}{2^n} \sum_{z \in \{0, 1\}^{2n}} |\psi_z\rangle_R |z\rangle_T. \quad (12)$$

The first register, R , is sent to the receiver, while the second register, T , is kept by the token. (Thus, the token knows the secret key z , and hence also which $|\psi_z\rangle$ the receiver possesses.) The mixed state describing the receiver’s state of knowledge at this point is given by

$$\rho_R := \frac{1}{2^{2n}} \sum_{z \in \{0, 1\}^{2n}} |\psi_z\rangle\langle\psi_z|.$$

► **Theorem 2.** *Given a single copy of ρ_R , and the ability to make m (adaptive) queries to the hardware token, the probability that an unbounded quantum adversary can force the token to output both bits s_0 and s_1 scales as $O(2^{2m-0.228n})$.*

Thus, the probability of an unbounded adversary (i.e., which applies arbitrary trace-preserving completely positive (TPCP) maps, which are not necessarily efficiently implementable) to successfully cheat using $m = cn$ for $c < 0.114$ queries is exponentially small in the quantum key size, n . The proof of Theorem 2 is in the full version; here, we give intuition.

Proof intuition. The challenge in analyzing security of the protocol is the fact that the receiver (a.k.a. the user) is not only given adaptive query access to the token, but also a copy of the quantum “resource state” ρ_R , which it may arbitrarily tamper with (in any manner allowed by quantum mechanics) while making queries. Luckily, the GW framework [33] (Section 2.1) is general enough to model such “queries with quantum side information”. The framework outputs an SDP, Γ (Equation (13)), the optimal value of which will encode the optimal cheating probability for a cheating user of our protocol. Giving a feasible solution for Γ will hence suffice to upper bound this cheating probability, yielding Theorem 2.

Coherently modeling quantum queries to the token. To model the interaction between the token and user, we first recall that all queries to the token must be classical by assumption. To model this process *coherently* in the GW framework, we hence imagine (solely for the purposes of the security analysis) that the token behaves as follows:

1. It first sends state ρ_R to the user.
2. When it receives as i th query a quantum state ρ_i from the user, it sends response string r_i to the user, and “copies” ρ_i via transversal CNOT gates to a private memory register \mathcal{W}_i , along with r_i . It does not access ρ_i again throughout the protocol, and only accesses r_i again in Step 3. For clarity, the token runs a classical circuit, and conditions each response r_i solely on the current incoming message, ρ_i .
3. After all communication, the token “measures” its responses (r_1, \dots, r_m) in the Z -basis to decide whether to accept (user successfully cheated) or reject (user failed to cheat).

The “copying” phase of Step 2 accomplishes two tasks: First, since the token will never read the “copies” of ρ_i again, the principle of deferred measurement [52] implies the transversal CNOT gates effectively simulate measuring ρ_i in the standard basis. In other words, without loss of generality the user is reduced to feeding a classical string \tilde{y} to the token. Second, we would like the entire security analysis to be done in a unified fashion in a single framework, the GW framework. To this end, we want the token itself to “decide” at the end of the protocol whether the user has successfully cheated (i.e. extracted both secret bits). Storing all responses r_i in Step 2 allows us to simulate such a final measurement in Step 3. We reiterate that, crucially, once the token “copies” ρ_i and r_i to W_i , it (1) never accesses (i.e. reads or writes to) ρ_i again and (2) only accesses r_i again in the final standard basis measurement of Step 3. Together, these ensure all responses r_i are independent, as required..

Formalization in GW framework. To place the discussion thus far into the formal GW framework, we return to Figure 1. The bottom “row” of Figure 1 will depict the token’s actions, and the top row the user’s actions. As outlined above, the protocol begins by imagining the token sends initial state $\rho_0 = \rho_R$ to the user via register \mathcal{X}_1 . The user then applies an arbitrary sequence of TPCP maps Φ_i to its private memory (modeled by register \mathcal{Z}_i in round i), each time sending a query \tilde{y}_i (which is, as discussed above a classical string without loss of generality) to the token via register \mathcal{Y}_i . Given any such query \tilde{y}_i in round i , the token applies its own TPCP map Ψ_i to determine how to respond to the query. In our protocol, the Ψ_i correspond to coherently applying a classical circuit, i.e. a sequence of unitary gates mapping the standard basis to itself. Specifically, their action is fully determined by Program 1, and in principle all Ψ_i are identical since the token is stateless (i.e., the action of the token in round i is unaffected by previous rounds $\{1, \dots, i-1\}$). (We use the term “in principle”, as recall from above that in the security analysis we model each Ψ_i as classically copying (\tilde{y}_i, r_i) to a distinct private register W_i .) Finally, after receiving the m th query \tilde{y}_m in register \mathcal{Y}_m , we imagine the token makes a measurement (not depicted in Fig. 1) based on the query responses (r_1, \dots, r_m) it returned; if the user managed to extract both s_0 and s_1 via queries, then the token “accepts”; otherwise it “rejects”. (Again, we are using the fact that in our security analysis, the token keeps a history of all its responses r_i , solely for the sake of this final measurement.)

With this high-level setup, the output of the GW framework is a semidefinite program, Γ :

$$\text{min: } p \tag{13}$$

$$\text{subject to: } Q_1 \preceq R_{m+1} \tag{14}$$

$$R_k = P_k \otimes I_{\mathcal{Y}_k} \quad \text{for } 1 \leq k \leq m+1 \tag{15}$$

$$\text{Tr}_{\mathcal{X}_k}(P_k) = R_{k-1} \quad \text{for } 1 \leq k \leq m+1 \tag{16}$$

$$R_0 = p \tag{17}$$

$$R_k \in \text{Pos}(\mathcal{Y}_{1,\dots,k} \otimes \mathcal{X}_{1,\dots,k}) \quad \text{for } 1 \leq k \leq m+1 \tag{18}$$

$$P_k \in \text{Pos}(\mathcal{Y}_{1,\dots,k-1} \otimes \mathcal{X}_{1,\dots,k}) \quad \text{for } 1 \leq k \leq m+1 \tag{19}$$

Above, Q_1 encodes the actions of the token, i.e. the co-strategy in the bottom row of Figure 1. The variable p denotes the “cheating probability” (i.e., the probability with which both s_0 and s_1 are extracted), subject to linear constraints (Equations (15)-(19)) which enforce that operator R_{m+1} encodes a valid co-strategy (see Section 2.1). Theorem 9 of [33] now says that the minimum p above encodes precisely the optimal cheating probability for a user which is constrained only by the laws of quantum mechanics. Since Γ is a minimization problem, to upper bound the cheating probability it hence suffices to give a feasible solution $(p, R_1, \dots, R_{m+1}, P_1, \dots, P_{m+1})$ for Γ , which will be our approach.

Intuition for Q_1 and an upper bound on p . It remains to give intuition as to how one derives Q_1 in Γ , and how an upper bound on the optimal p is obtained. Without loss of generality, one may assume that each of the token’s TPCP maps Ψ_i are given by *isometries* $A_i : \mathcal{Y}_i \otimes \mathcal{W}_{i-1} \mapsto \mathcal{X}_{i+1} \otimes \mathcal{W}_i$, meaning $A_i^\dagger A_i = I_{\mathcal{Y}_i \otimes \mathcal{W}_{i-1}}$ (due to the Stinespring dilation theorem). (We omit the first isometry which prepares state ρ_0 in our discussion here for simplicity.) Let us denote their sequential application by a single operator $A := A_m \cdots A_1$. Then, the Choi-Jamiołkowski representation of A is given by [33] (Section 2.1) $\text{Tr}_{\mathcal{Z}_m}(\text{vec}(A) \text{vec}(A)^\dagger)$, where we trace out the token’s private memory register \mathcal{Z}_m . However, since in our security analysis, we imagine the token also makes a final measurement via some POVM $\Lambda = \{\Lambda_0, \Lambda_1\}$, whereupon obtaining outcome Λ_1 the token “accepts”, and upon outcome Λ_0 the token rejects, we require a slightly more complicated setup. Letting $B_1 := \Lambda_1 A$, we define Q_1 as [33] $Q_1 = \text{Tr}_{\mathcal{Z}_m}(\text{vec}(B_1) \text{vec}(B_1)^\dagger)$.

The full derivation of Q_1 is deferred to the full version; here, we state Q_1 with intuition:

$$Q_1 = \frac{1}{4^n} \sum_{s \in T} |t_m s_{t_m}\rangle \langle t_m s_{t_m}|_{\mathcal{X}_{m+1}} \otimes \cdots \otimes |t_1 s_{t_1}\rangle \langle t_1 s_{t_1}|_{\mathcal{X}_2} \otimes \left(\sum_{(\tilde{y}, z) \in Y_t} |\tilde{y}_m\rangle \langle \tilde{y}_m|_{\mathcal{Y}_m} \otimes \cdots \otimes |\tilde{y}_1\rangle \langle \tilde{y}_1|_{\mathcal{Y}_1} \otimes |\psi_z\rangle \langle \psi_z|_{\mathcal{X}_1} \right).$$

Intuitively, each string $t_i s_{t_i} \in \{0, 1\}^3$ encodes the response r_i of the token given the i th query from the user; hence, the corresponding projectors in Q_1 act on spaces \mathcal{X}_2 through \mathcal{X}_{m+1} . Each string $\tilde{y}_i \in \{0, 1\}^{n+1}$ denotes the i th query sent from the user to the token, where each $\tilde{y}_i = b_i \circ y_i$ in the notation of Program 1, i.e. $b_i \in \{0, 1\}$ is the choice bit for each query. Each such message is passed via register \mathcal{Y}_i . The states $|\psi_z\rangle$ and strings z are defined as in the beginning of Section 3.4; recall $z \in \{0, 1\}^{2n}$ and $|\psi_z\rangle \in (\mathbb{C}^2)^{\otimes n}$ denote the secret key and corresponding quantum key, respectively. Finally, the relation Y_t encodes the constraint that for all $i \in \{1, \dots, m\}$, the tuple (\tilde{y}_i, z) (i.e. the i th message to the token, \tilde{y}_i , and secret key z) is consistent with the response returned by the token, t_i .

Upper bounding p . To now upper bound p , we give a feasible solution R_{m+1} satisfying the constraints of Γ . Note that giving even a solution which attains $p = 1$ for all n and m is *non-trivial* – such a solution is given in the full version. Here, we give a solution which attains $p \in O(2^{2m-0.228n})$, as claimed in Theorem 2. Namely, we set

$$R_{m+1} = \frac{1}{|T|} \sum_{t \in T} |t_m s_{t_m}\rangle \langle t_m s_{t_m}|_{\mathcal{X}_{m+1}} \otimes \cdots \otimes |t_1 s_{t_1}\rangle \langle t_1 s_{t_1}|_{\mathcal{X}_2} \otimes I_{Y_1 \otimes \cdots \otimes Y_m} \otimes \frac{I}{2^n} \chi_1.$$

This satisfies constraint (15) of Γ due to the identity term $I_{Y_1 \otimes \cdots \otimes Y_m}$. The renormalization factor $(|T| 2^n)^{-1}$ above ensures that tracing out all \mathcal{X}_i registers yields $R_0 = 1$ in constraint (17) of Γ . We are thus reduced to choosing the minimum p satisfying constraint (14).

Now, observe we have chosen R_{m+1} to align with the block-diagonal structure of Q_1 on registers $\mathcal{X}_2, \dots, \mathcal{X}_m$. Since registers $\mathcal{Y}_1 \otimes \dots \otimes \mathcal{Y}_m$ and \mathcal{X}_1 of R_{m+1} are proportional to the identity matrix, it thus suffices to characterize the largest eigenvalue of Q_1 , $\lambda_{\max}(Q_1)$. This is done in the full version, which shows $\lambda_{\max}(Q_1) = \frac{2}{4^n} \left(1 + \frac{1}{\sqrt{2}}\right)^n$. Combining this bound on $\lambda_{\max}(Q_1)$ with the parameters of R_{m+1} above now yields the desired claim that $p \in O(2^{2m-0.228n})$. For $m < 0.114n$ queries, this implies that the probability that a user of the token successfully cheats and thus that the simulation fails is exponentially small in the key size, n . Simplifications of the GW SDP, the derivation of its dual SDP, and a conjectured approximately optimal dual feasible solution are given in the full version.

4 Impossibility Results

We now discuss “tightness” of our protocol with respect to impossibility results. To begin, it is easy to argue that OTMs cannot exist in the plain model (i.e., without additional assumptions) in both the classical and quantum settings: in the classical setting, impossibility holds, since software can always be copied. Quantumly, this follows by a rewinding argument [9]. Here, we give two no-go results for the quantum setting which support the idea that our scheme is “tight” in terms of the minimality of the assumptions it uses. Both results assume the token is reversible, meaning the receiver can run both the token and its inverse operation. Note that if the receiver is *not* given access to the token’s inverse operation, it is unlikely for our no-go techniques to go through. This is because, in the most general case where the token is an arbitrary unitary U , which the receiver may apply as a black box, simulating $U^{-1} = U^\dagger$ appears difficult [26, 57]; see the full version for a discussion.

Result 1: Tokens which can be queried in superposition. In our construction, we require that all queries to the token be classical strings, i.e., no querying in superposition is allowed. It is easy to argue via a standard rewinding argument that relaxing this requirement yields impossibility of a secure OTM, as long as access to the token’s adjoint (inverse) operation is given, as we now show. Specifically, let M be a quantum OTM implemented using a hardware token. Since the token access is assumed to be reversible, we may model it as an oracle O_f realizing a function $f : \{0, 1\}^n \mapsto \{0, 1\}^m$ in the standard way, i.e., for all $y \in \{0, 1\}^n$ and $b \in \{0, 1\}^m$, $O_f|y\rangle|b\rangle = |y\rangle|b \oplus f(y)\rangle$. Now, suppose our OTM stores two secret bits s_0 and s_1 , and provides the receiver with an initial state $|\psi\rangle \in A \otimes B \otimes C$, where A , B , and C are the algorithm’s workspace, *query* (i.e., input to O_f), and *answer* (i.e., O_f ’s answers) registers, respectively. By definition, an honest receiver must be able to access precisely one of s_0 or s_1 with certainty, given $|\psi\rangle$. Thus, for any $i \in \{0, 1\}$, there exists a quantum query algorithm $A_i = U_m O_f \dots O_f U_2 O_f U_1$ for unitaries $U_i \in \mathcal{U}(A \otimes B \otimes C)$ such that $A_i|\psi\rangle = |\psi'\rangle_{AB}|s_i\rangle_C$. For any choice of i , however, this implies a malicious receiver can now classically copy s_i to an external register, and then “rewind” by applying A_i^\dagger to $|\psi'\rangle_{AB}|s_i\rangle_C$ to recover $|\psi\rangle$. Applying $A_{i'}$ for $i' \neq i$ to $|\psi\rangle$ now yields the second bit i' with certainty as well. We conclude that a quantum OTM which allows superposition queries to a reversible stateless token is insecure.

► **Remark 3.** Above, the OTM outputs s_i with certainty. A similar argument holds if s_i is output with probability at least $1 - \epsilon$ for small $\epsilon > 0$ via the Gentle Measurement Lemma [66].

Result 2: Tokens with a bounded number of keys. We observed superposition queries to the token prevent an OTM from being secure. One can also ask how simple a hardware token with classical queries can be, while still allowing a secure OTM. Below, we consider such a strengthening in which the token is forced to have a bounded number of keys.

To formalize this, we define the notion of a “measure-and-access (MA)” OTM, i.e., an OTM in which given an initial state $|\psi\rangle$, an honest receiver applies a prescribed measurement to $|\psi\rangle$, and feeds the resulting classical string (i.e., key) y into the token O_f to obtain s_i . Our construction is an example of a MA memory in which each bit s_i has an *exponential* number of valid keys y such that $f(y) = s_i$. Can the construction can be strengthened such that each s_i has a bounded number (e.g., a polynomial number) of keys? We now show that such a strengthening would preclude security, assuming the token is reversible.

► **Lemma 4.** *Let M be an MA memory with oracle O_f , such that O_f cannot be queried in superposition. If a secret bit s_i has at most Δ keys y_i such that $f(y_i) = s_i$, then given a single copy of $|\psi\rangle$, one can extract both s_0 and s_1 from M with probability at least $1/\Delta^2$.*

Thus, if a secret bit b_i has at most polynomially many keys, then any measure-and-access OTM can be broken with at least inverse polynomial probability. The proof is in the full version. In this sense, in the setting of measure-and-access memories, our construction is tight – in order to bound the adversary’s success probability of obtaining both secret bits by an inverse exponential, we require each secret bit to have exponentially many valid keys.

References

- 1 Scott Aaronson and Paul Christiano. Quantum money from hidden subspaces. In *Proc. 44th Symposium on Theory of Computing (STOC) 2012*, pages 41–60, 2012. Full version available as arXiv:1203.4740. doi:10.1145/2213977.2213983.
- 2 Donald Beaver. Secure multiparty protocols and zero-knowledge proof systems tolerating a faulty minority. *Journal of Cryptology*, 4(2):75–122, 1991.
- 3 Shalev Ben-David and Or Sattath. Quantum tokens for digital signatures. arXiv:1609.09047, 2018.
- 4 Charles H. Bennett and Gilles Brassard. Quantum cryptography: Public key distribution and coin tossing. *International Conference on Computers, Systems and Signal Processing*, pages 175–179, 1984.
- 5 Charles H. Bennett, Gilles Brassard, Claude Crépeau, and Marie-Hélène Skubiszewska. Practical quantum oblivious transfer. In Joan Feigenbaum, editor, *CRYPTO’91*, volume 576 of *LNCS*, pages 351–366. Springer, August 1992.
- 6 Manuel Blum, Paul Feldman, and Silvio Micali. Non-interactive zero-knowledge and its applications. In *20th ACM STOC*, pages 103–112. ACM Press, May 1988.
- 7 Manuel Blum and Silvio Micali. How to generate cryptographically strong sequences of pseudo random bits. In *23rd FOCS*, pages 112–117. IEEE Computer Society Press, November 1982.
- 8 Anne Broadbent, Sevag Gharibian, and Hong-Sheng Zhou. Quantum one-time memories from stateless hardware. arXiv:1511.01363, November 2015.
- 9 Anne Broadbent, Gus Gutoski, and Douglas Stebila. Quantum one-time programs. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part II*, volume 8043 of *LNCS*, pages 344–360. Springer, August 2013. doi:10.1007/978-3-642-40084-1_20.
- 10 Anne Broadbent and Christian Schaffner. Quantum cryptography beyond quantum key distribution. *Designs, Codes and Cryptography*, 78(1):351–382, 2016.
- 11 Christian Cachin and Ueli Maurer. Unconditional security against memory-bounded adversaries. In *Advances in Cryptology - CRYPTO 1997*, LNCS, pages 292–306. Springer, 1997. doi:10.1007/BFb0052243.
- 12 Ran Canetti. Security and composition of multiparty cryptographic protocols. *Journal of Cryptology*, 13(1):143–202, 2000.
- 13 Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. Cryptology ePrint Archive, Report 2000/067, 2000. URL: <http://eprint.iacr.org/2000/067>.
- 14 Ran Canetti. Universally composable security: A new paradigm for cryptographic protocols. In *42nd FOCS*, pages 136–145. IEEE Computer Society Press, October 2001.

- 15 Ran Canetti, Yevgeniy Dodis, Rafael Pass, and Shabsi Walfish. Universally composable security with global setup. In Salil P. Vadhan, editor, *TCC 2007*, volume 4392 of *LNCS*, pages 61–85. Springer, February 2007.
- 16 Ran Canetti, Yehuda Lindell, Rafail Ostrovsky, and Amit Sahai. Universally composable two-party and multi-party secure computation. In *34th ACM STOC*, pages 494–503. ACM Press, May 2002.
- 17 Nishanth Chandran, Vipul Goyal, and Amit Sahai. New constructions for UC secure computation using tamper-proof hardware. In Nigel P. Smart, editor, *EUROCRYPT 2008*, volume 4965 of *LNCS*, pages 545–562. Springer, April 2008.
- 18 Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Alg. Appl.*, 10:285, 1975.
- 19 Seung Geol Choi, Jonathan Katz, Dominique Schröder, Arkady Yerukhimovich, and Hong-Sheng Zhou. (efficient) universally composable oblivious transfer using a minimal number of stateless tokens. In Yehuda Lindell, editor, *TCC 2014*, volume 8349 of *LNCS*, pages 638–662. Springer, February 2014. doi:10.1007/978-3-642-54242-8_27.
- 20 Kai-Min Chung, Marios Georgiou, Ching-Yi Lai, and Vassilis Zikas. Cryptography with disposable backdoors. eprint:2018/352, 2018.
- 21 Ivan Damgård, Serge Fehr, Carolin Lunemann, Louis Salvail, and Christian Schaffner. Improving the security of quantum protocols via commit-and-open. In Shai Halevi, editor, *CRYPTO 2009*, volume 5677 of *LNCS*, pages 408–427. Springer, August 2009.
- 22 Ivan Damgård, Serge Fehr, Louis Salvail, and Christian Schaffner. Cryptography in the bounded quantum-storage model. In *Symposium on Foundations of Computer Science - FOCS 2005*, pages 449–458. IEEE, 2005. doi:10.1109/SFCS.2005.30.
- 23 Ivan Damgård and Alessandra Scafuro. Unconditionally secure and universally composable commitments from physical assumptions. In Kazue Sako and Palash Sarkar, editors, *ASIACRYPT 2013, Part II*, volume 8270 of *LNCS*, pages 100–119. Springer, December 2013. doi:10.1007/978-3-642-42045-0_6.
- 24 Frédéric Dupuis, Jesper Buus Nielsen, and Louis Salvail. Secure two-party quantum evaluation of unitaries against specious adversaries. In *Advances in Cryptology – Proc. CRYPTO 2010*, *LNCS*, pages 685–706. Springer, 2010.
- 25 Frédéric Dupuis, Jesper Buus Nielsen, and Louis Salvail. Actively secure two-party evaluation of any quantum operation. In *Advances in Cryptology – Proc. CRYPTO 2012*, volume 7417 of *LNCS*, pages 794–811. Springer, 2012. doi:10.1007/978-3-642-32009-5_46.
- 26 Bill Fefferman and Shelby Kimmel. Quantum vs. Classical Proofs and Subset Verification. In Igor Potapov, Paul Spirakis, and James Worrell, editors, *43rd International Symposium on Mathematical Foundations of Computer Science (MFCS 2018)*, volume 117 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 22:1–22:23, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.MFCS.2018.22.
- 27 Serge Fehr, Jonathan Katz, Fang Song, Hong-Sheng Zhou, and Vassilis Zikas. Feasibility and completeness of cryptographic tasks in the quantum world. In Amit Sahai, editor, *TCC 2013*, volume 7785 of *LNCS*, pages 281–296. Springer, March 2013. doi:10.1007/978-3-642-36594-2_16.
- 28 Dmitry Gavinsky. Quantum money with classical verification. In *Computational Complexity (CCC), 2012 IEEE 27th Annual Conference on*, pages 42–52, June 2012. doi:10.1109/CCC.2012.10.
- 29 Oded Goldreich, Silvio Micali, and Avi Wigderson. How to play any mental game or A completeness theorem for protocols with honest majority. In Alfred Aho, editor, *19th ACM STOC*, pages 218–229. ACM Press, May 1987.
- 30 Shafi Goldwasser, Yael Tauman Kalai, and Guy N. Rothblum. One-time programs. In David Wagner, editor, *CRYPTO 2008*, volume 5157 of *LNCS*, pages 39–56. Springer, August 2008.

- 31 Shafi Goldwasser and Leonid A. Levin. Fair computation of general functions in presence of immoral majority. In Alfred J. Menezes and Scott A. Vanstone, editors, *CRYPTO'90*, volume 537 of *LNCS*, pages 77–93. Springer, August 1991.
- 32 Vipul Goyal, Yuval Ishai, Amit Sahai, Ramarathnam Venkatesan, and Akshay Wadia. Founding cryptography on tamper-proof hardware tokens. In Daniele Micciancio, editor, *TCC 2010*, volume 5978 of *LNCS*, pages 308–326. Springer, February 2010.
- 33 Gus Gutoski and John Watrous. Toward a general theory of quantum games. In *Proceedings of the 39th ACM Symposium on Theory of Computing (STOC 2007)*, pages 565–574, 2007. arXiv:quant-ph/0611234v2.
- 34 Sean Hallgren, Adam Smith, and Fang Song. Classical cryptographic protocols in a quantum world. In Phillip Rogaway, editor, *CRYPTO 2011*, volume 6841 of *LNCS*, pages 411–428. Springer, August 2011.
- 35 Werner Heisenberg. Schwankungserscheinungen und quantenmechanik. *Zeitschrift fuer Physik*, 40(7):501–506, July 1927. doi:10.1007/BF01440827.
- 36 Yuval Ishai, Manoj Prabhakaran, and Amit Sahai. Founding cryptography on oblivious transfer - efficiently. In David Wagner, editor, *CRYPTO 2008*, volume 5157 of *LNCS*, pages 572–591. Springer, August 2008.
- 37 Andrzej Jamiolkowski. Linear transformations which preserve trace and positive semi-definiteness of operators. *Rep. Math. Phys.*, 3:275, 1972.
- 38 Jonathan Katz. Universally composable multi-party computation using tamper-proof hardware. In Moni Naor, editor, *EUROCRYPT 2007*, volume 4515 of *LNCS*, pages 115–128. Springer, May 2007.
- 39 Joe Kilian. Founding cryptography on oblivious transfer. In *20th ACM STOC*, pages 20–31. ACM Press, May 1988.
- 40 Daniel Kraschewski, Hemanta K. Maji, Manoj Prabhakaran, and Amit Sahai. A full characterization of completeness for two-party randomized function evaluation. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 659–676. Springer, May 2014. doi:10.1007/978-3-642-55220-5_36.
- 41 Daniel Kraschewski and Jörn Müller-Quade. Completeness theorems with constructive proofs for finite deterministic 2-party functions. In Yuval Ishai, editor, *TCC 2011*, volume 6597 of *LNCS*, pages 364–381. Springer, March 2011.
- 42 Huijia Lin, Rafael Pass, and Muthuramakrishnan Venkitasubramaniam. A unified framework for concurrent security: universal composability from stand-alone non-malleability. In Michael Mitzenmacher, editor, *41st ACM STOC*, pages 179–188. ACM Press, 2009.
- 43 Yi-Kai Liu. Building one-time memories from isolated qubits. In Moni Naor, editor, *ITCS 2014*, pages 269–286. ACM, January 2014.
- 44 Yi-Kai Liu. Single-shot security for one-time memories in the isolated qubits model. In Juan A. Garay and Rosario Gennaro, editors, *CRYPTO 2014, Part II*, volume 8617 of *LNCS*, pages 19–36. Springer, August 2014. doi:10.1007/978-3-662-44381-1_2.
- 45 Yi-Kai Liu. Privacy amplification in the isolated qubits model. In Elisabeth Oswald and Marc Fischlin, editors, *EUROCRYPT 2015, Part II*, volume 9057 of *LNCS*, pages 785–814. Springer, April 2015. doi:10.1007/978-3-662-46803-6_26.
- 46 Hemanta K. Maji, Manoj Prabhakaran, and Mike Rosulek. Complexity of multi-party computation problems: The case of 2-party symmetric secure function evaluation. In Omer Reingold, editor, *TCC 2009*, volume 5444 of *LNCS*, pages 256–273. Springer, March 2009.
- 47 Hemanta K. Maji, Manoj Prabhakaran, and Mike Rosulek. A zero-one law for cryptographic complexity with respect to computational UC security. In Tal Rabin, editor, *CRYPTO 2010*, volume 6223 of *LNCS*, pages 595–612. Springer, August 2010.
- 48 Ueli Maurer and Renato Renner. Abstract cryptography. In Bernard Chazelle, editor, *ICS 2011*, pages 1–21. Tsinghua University Press, January 2011.

- 49 Ueli M. Maurer. Protocols for secret key agreement by public discussion based on common information. In *Advances in Cryptology - CRYPTO 1992*, volume 740 of *LNCS*, pages 461–470. Springer, 1992. doi:10.1007/3-540-48071-4_32.
- 50 Silvio Micali and Phillip Rogaway. Secure computation (abstract). In Joan Feigenbaum, editor, *CRYPTO'91*, volume 576 of *LNCS*, pages 392–404. Springer, August 1992.
- 51 Abel Molina, Thomas Vidick, and John Watrous. Optimal counterfeiting attacks and generalizations for Wiesner's quantum money. In Kazuo Iwama, Yasuhito Kawano, and Mio Murao, editors, *Theory of Quantum Computation, Communication, and Cryptography*, volume 7582 of *Lecture Notes in Computer Science*, pages 45–64, 2013.
- 52 M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.
- 53 Fernando Pastawski, Norman Y Yao, Liang Jiang, Mikhail D Lukin, and J Ignacio Cirac. Unforgeable noise-tolerant quantum tokens. *Proceedings of the National Academy of Sciences*, 109(40):16079–16082, 2012.
- 54 Birgit Pfiztmann and Michael Waidner. A model for asynchronous reactive systems and its application to secure message transmission. In *Proc. 22nd IEEE Symposium on Security & Privacy (S&P) 2001*, pages 184–200. IEEE, 2001. Full version available at <http://eprint.iacr.org/2000/066>. doi:10.1109/SECPRI.2001.924298.
- 55 Manoj Prabhakaran and Mike Rosulek. Cryptographic complexity of multi-party computation problems: Classifications and separations. In David Wagner, editor, *CRYPTO 2008*, volume 5157 of *LNCS*, pages 262–279. Springer, August 2008.
- 56 Manoj Prabhakaran and Amit Sahai. New notions of security: Achieving universal composability without trusted setup. In László Babai, editor, *36th ACM STOC*, pages 242–251. ACM Press, June 2004.
- 57 Marco Túlio Quintino, Qingxiuxiong Dong, Atsushi Shimbo, Akihito Soeda, and Mio Murao. Reversing unknown quantum transformations: Universal quantum circuit for inverting general unitary operations. *Phys. Rev. Lett.*, 123:210502, November 2019. doi:10.1103/PhysRevLett.123.210502.
- 58 Renato Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zürich (Switzerland), September 2008. doi:10.1142/S0219749908003256.
- 59 Dominique Unruh. Universally composable quantum multi-party computation. In Henri Gilbert, editor, *EUROCRYPT 2010*, volume 6110 of *LNCS*, pages 486–505. Springer, May 2010.
- 60 Dominique Unruh. Everlasting multi-party computation. In Ran Canetti and Juan A. Garay, editors, *CRYPTO 2013, Part II*, volume 8043 of *LNCS*, pages 380–397. Springer, August 2013. doi:10.1007/978-3-642-40084-1_22.
- 61 Dominique Unruh. Revocable quantum timed-release encryption. In Phong Q. Nguyen and Elisabeth Oswald, editors, *EUROCRYPT 2014*, volume 8441 of *LNCS*, pages 129–146. Springer, May 2014. doi:10.1007/978-3-642-55220-5_8.
- 62 John Watrous. Lecture 7: Semidefinite programming, 2011. Latest version available at: <https://cs.uwaterloo.ca/~watrous/CS766/LectureNotes/07.pdf>.
- 63 Stephanie Wehner, Christian Schaffner, and Barbara M. Terhal. Cryptography from noisy storage. *Physical Review Letters*, 100(22):220502, June 2008. doi:10.1103/PhysRevLett.100.220502.
- 64 Stephanie Wehner and Andreas Winter. Entropic uncertainty relations—a survey. *New J. Phys.*, 12(2):025009, February 2010. doi:10.1088/1367-2630/12/2/025009.
- 65 Stephen Wiesner. Conjugate coding. *ACM SIGACT News*, 15(1):78–88, 1983. Original article written circa 1970. doi:10.1145/1008908.1008920.
- 66 Andreas Winter. Coding theorem and strong converse for quantum channels. *IEEE Transactions on Information Theory*, 45:2481–2485, 1999.
- 67 William K. Wootters and Wojciech H. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982. doi:10.1038/299802a0.
- 68 Andrew Chi-Chih Yao. Theory and applications of trapdoor functions. In *23rd FOCS*, pages 80–91. IEEE Computer Society Press, November 1982.

A Universal Composition (UC) Framework

We consider simulation-based security. The Universal Composability (UC) framework was proposed by Canetti [14, 13], culminating a long sequence of simulation-based security definitions (*c.f.* [29, 31, 50, 2, 12]); please see also [54, 56, 15, 42, 48] for alternative/extended frameworks. Recently Unruh [59] extend the UC framework to the quantum setting. Next, we provide a high-level description of the original classical UC model by Canetti [14, 13], and then the quantum UC model by Unruh [59].

A.1 Classical UC Model ([14, 13])

Machines. The basic entities involved in the UC model are players P_1, \dots, P_k where k is polynomial of security parameter n , an adversary \mathcal{A} , and an environment \mathcal{Z} . Each entity is modeled as a interactive Turing machine (ITM), where \mathcal{Z} could have an additional non-uniform string as advice. Each P_i has identity i assigned to it, while \mathcal{A} and \mathcal{Z} have special identities $id_{\mathcal{A}} := \text{adv}$ and $id_{\mathcal{Z}} := \text{env}$.

Protocol Execution. A protocol specifies the programs for each P_i , which we denote as $\pi = (\pi_1, \dots, \pi_k)$. The execution of a protocol is coordinated by the environment \mathcal{Z} . It starts by preparing inputs to all players, who then run their respective programs on the inputs and exchange messages of the form $(id_{\text{sender}}, id_{\text{receiver}}, \text{msg})$. \mathcal{A} can corrupt an arbitrary set of players and control them later on. In particular, \mathcal{A} can instruct a corrupted player sending messages to another player and also read messages that are sent to the corrupted players. During the course of execution, the environment \mathcal{Z} also interacts with \mathcal{A} in an arbitrary way. In the end, \mathcal{Z} receives outputs from all the other players and generates one bit output. We use $\text{EXEC}[\mathcal{Z}, \mathcal{A}, \pi]$ denote the distribution of the environment \mathcal{Z} 's (single-bit) output when executing protocol π with \mathcal{A} and the P_i 's.

Ideal Functionality and Dummy Protocol. Ideal functionality \mathcal{F} is a trusted party, modeled by an ITM again, that perfectly implements the desired multi-party computational task. We consider an “dummy protocol”, denoted $P^{\mathcal{F}}$, where each party has direct communication with \mathcal{F} , who accomplishes the desired task according to the messages received from the players. The execution of $P^{\mathcal{F}}$ with environment \mathcal{Z} and an adversary, usually called the simulator \mathcal{S} , is defined analogous as above, in particular, \mathcal{S} monitors the communication between corrupted parties and the ideal functionality \mathcal{F} . Similarly, we denote \mathcal{Z} 's output distribution as $\text{EXEC}[\mathcal{Z}, \mathcal{S}, P^{\mathcal{F}}]$.

► **Definition 5** (Classical UC-secure Emulation). *We say π (classically) UC-emulates π' if for any adversary \mathcal{A} , there exists a simulator \mathcal{S} such that for all environments \mathcal{Z} ,*

$$\text{EXEC}[\mathcal{Z}, \mathcal{A}, \pi] \approx \text{EXEC}[\mathcal{Z}, \mathcal{S}, \pi'] \quad (20)$$

We here consider that \mathcal{A} and \mathcal{Z} are computationally unbounded, and we call it statistical UC-security. We require the running time \mathcal{S} is polynomial in that of \mathcal{A} . We call this property Polynomial Simulation.

Let \mathcal{F} be a well-formed two party functionality. We say π (classically) UC-realizes \mathcal{F} if for all adversary \mathcal{A} , there exists a simulator \mathcal{S} such that for all environments \mathcal{Z} , $\text{EXEC}[\mathcal{Z}, \mathcal{A}, \pi] \approx \text{EXEC}[\mathcal{Z}, \mathcal{S}, P^{\mathcal{F}}]$. We also write $\text{EXEC}[\mathcal{Z}, \mathcal{A}, \pi] \approx \text{EXEC}[\mathcal{Z}, \mathcal{S}, \mathcal{F}]$ if the context is clear.

UC-secure protocols admit a general composition property, demonstrated in the following universal composition theorem.

► **Theorem 6** (UC Composition Theorem [13]). *Let π, π' and σ be n -party protocols. Assume that π UC-emulates π' . Then σ^π UC-emulates $\sigma^{\pi'}$.*

A.2 Quantum UC Model ([59])

Now, we give a high-level description of quantum UC model by Unruh [59].

Quantum Machine. In the quantum UC model, all players are modeled as quantum machines. A quantum machine is a sequence of quantum circuits $\{M^n\}_{n \in \mathbb{N}}$, for each security parameter n . M^n is a completely positive trace preserving operator on space $\mathcal{H}^{\text{state}} \otimes \mathcal{H}^{\text{class}} \otimes \mathcal{H}^{\text{quant}}$, where $\mathcal{H}^{\text{state}}$ represents the internal workspace of M^n and $\mathcal{H}^{\text{class}}$ and $\mathcal{H}^{\text{quant}}$ represent the spaces for communication, where for convenience we divide the messages into classical and quantum parts. We allow a non-uniform quantum advice⁶ to the machine of the environment \mathcal{Z} , while all other machines are uniformly generated.

Protocol Execution. In contrast to the communication policy in classical UC model, we consider a network \mathbf{N} which contains the space $\mathcal{H}_{\mathbf{N}} := \mathcal{H}^{\text{class}} \otimes \mathcal{H}^{\text{quant}} \otimes_i \mathcal{H}_i^{\text{state}}$. Namely, each machine maintains individual internal state space, but the communication space is shared among all. We assume $\mathcal{H}^{\text{class}}$ contains the message $(id_{\text{sender}}, id_{\text{receiver}}, \text{msg})$ which specifies the sender and receiver of the current message, and the receiver then processes the quantum state on $\mathcal{H}^{\text{quant}}$. Note that this communication model implicitly ensures authentication. In a protocol execution, \mathcal{Z} is activated first, and at each round, one player applies the operation defined by its machine M^n on $\mathcal{H}^{\text{class}} \otimes \mathcal{H}^{\text{quant}} \otimes \mathcal{H}^{\text{state}}$. In the end \mathcal{Z} generates a one-bit output. Denote $\text{EXEC}[\mathcal{Z}, \mathcal{A}, \Pi]$ the output distribution of \mathcal{Z} .

Ideal Functionality. All functionalities we consider in this work are classical, i.e., the inputs and outputs are classical, and its program can be implemented by an efficient classical Turing machine. Here in the quantum UC model, the ideal functionality \mathcal{F} is still modeled as a quantum machine for consistency, but it only applies classical operations. Namely, it measures any input message in the computational basis to get a classical bit-string, and implements the operations specified by the classical computational task.

We consider an “dummy protocol”, denoted $P^{\mathcal{F}}$, where each party has direct communication with \mathcal{F} , who accomplishes the desired task according to the messages received from the players. The execution of $P^{\mathcal{F}}$ with environment \mathcal{Z} and an adversary, usually called the simulator \mathcal{S} , is defined analogous as above, in particular, \mathcal{S} monitors the communication between corrupted parties and the ideal functionality \mathcal{F} . Similarly, we denote \mathcal{Z} ’s output distribution as $\text{EXEC}[\mathcal{Z}, \mathcal{S}, P^{\mathcal{F}}]$. For simplicity, we also write it as $\text{EXEC}[\mathcal{Z}, \mathcal{S}, \mathcal{F}]$.

► **Definition 7** (Quantum UC-secure Emulation). *We say Π quantum-UC-emulates Π' if for any quantum adversary \mathcal{A} , there exists a (quantum) simulator \mathcal{S} such that for all quantum environments \mathcal{Z} ,*

$$\text{EXEC}[\mathcal{Z}, \mathcal{A}, \Pi] \approx \text{EXEC}[\mathcal{Z}, \mathcal{S}, \Pi'] \quad (21)$$

⁶ Unruh’s model only allows classical advice, but we tend to take the most general model. It is easy to justify that almost all results remain unchanged, including the composition theorem. See [34, Section 5] for more discussion.

We consider here that \mathcal{A} and \mathcal{Z} are computationally unbounded, we call it (quantum) statistical UC-security. We require the running time \mathcal{S} is polynomial in that of \mathcal{A} . We call this property Polynomial Simulation.

Similarly, (quantum) computational UC-security can be defined. Let \mathcal{F} be a well-formed two party functionality. We say Π **quantum-UC-realizes** \mathcal{F} if for all quantum adversary \mathcal{A} , there exists a (quantum) simulator \mathcal{S} such that for all quantum environments \mathcal{Z} , $\text{EXEC}[\mathcal{Z}, \mathcal{A}, \Pi] \approx \text{EXEC}[\mathcal{Z}, \mathcal{S}, P^{\mathcal{F}}]$.

Quantum UC-secure protocols also admit general composition:

► **Theorem 8** (Quantum UC Composition Theorem [59, Theorem 11]). *Let Π, Π' and Σ be quantum-polynomial-time protocols. Assume that Π quantum UC-emulates Π' . Then Σ^Π quantum UC-emulates $\Sigma^{\Pi'}$.*

► **Remark 9.** Out of the two protocol parties (the sender and the receiver), we consider security only in the case of the receiver being a corrupted party. Note that we are only interested in cases where the same party is corrupted with respect to all composed protocol. Furthermore, we only consider static corruption.

B Stand-Alone Security in the case of a Malicious Sender

In order to define stand-alone security against a malicious sender (Definition 11), in our context, we closely follow definitions given in prior work [24], which we now recall. (Note that, instead of considering the *approximate* case for security, we are able to use the *exact* one.)

► **Definition 10.** An n -step quantum two-party protocol with oracle calls, denoted $\Pi^\mathcal{O} = (\mathcal{A}, \mathcal{B}, \mathcal{O}, n)$ consists of:

1. input space \mathcal{A}_0 and \mathcal{B}_0 for parties \mathcal{A} and \mathcal{B} respectively.
2. memory spaces $\mathcal{A}_1, \dots, \mathcal{A}_n$ and $\mathcal{B}_1, \dots, \mathcal{B}_n$ for \mathcal{A} and \mathcal{B} , respectively.
3. An n -tuple of quantum operations $(\mathcal{A}_1, \dots, \mathcal{A}_n)$ for \mathcal{A} , $\mathcal{A}_i : \mathcal{L}(\mathcal{A}_{i-1}) \mapsto \mathcal{L}(\mathcal{A}_i)$, $(1 \leq i \leq n)$.
4. An n -tuple of quantum operations $(\mathcal{B}_1, \dots, \mathcal{B}_n)$ for \mathcal{B} , $\mathcal{B}_i : \mathcal{L}(\mathcal{B}_{i-1}) \mapsto \mathcal{L}(\mathcal{B}_i)$, $(1 \leq i \leq n)$.
5. Memory spaces $\mathcal{A}_1, \dots, \mathcal{A}_n$ and $\mathcal{B}_1, \dots, \mathcal{B}_n$ can be written as $\mathcal{A}_i = \mathcal{A}_i^\mathcal{O} \otimes \mathcal{A}_i'$ and $\mathcal{B}_i = \mathcal{B}_i^\mathcal{O} \otimes \mathcal{B}_i'$, $(1 \leq i \leq n)$ and $\mathcal{O} = (\mathcal{O}_1, \dots, \mathcal{O}_n)$ is an n -tuple of quantum operations: $\mathcal{O}_i : \mathcal{L}(\mathcal{A}_i^\mathcal{O} \otimes \mathcal{B}_i^\mathcal{O}) \mapsto \mathcal{L}(\mathcal{A}_i^\mathcal{O} \otimes \mathcal{B}_i^\mathcal{O})$, $(1 \leq i \leq n)$.

If $\Pi^\mathcal{O} = (\mathcal{A}, \mathcal{B}, \mathcal{O}, n)$ is an n -turn two-party protocol, then the final state of the interaction upon input $\rho_{\text{in}} \in \mathcal{D}(\mathcal{A}_0 \otimes \mathcal{B}_0 \otimes \mathcal{R})$ where \mathcal{R} is a system of dimension $\dim \mathcal{A}_0 \dim \mathcal{B}_0$, is:

$$[\mathcal{A} \circledast \mathcal{B}](\rho_{\text{in}}) = (\mathbb{K}_{\mathcal{L}(\mathcal{A}'_n \otimes \mathcal{B}'_n \otimes \mathcal{R})} \otimes \mathcal{O}_n)(\mathcal{A}_n \otimes \mathcal{B}_n \otimes \mathbb{K}_{\mathcal{R}}) \dots (\mathbb{K}_{\mathcal{L}(\mathcal{A}'_1 \otimes \mathcal{B}'_1 \otimes \mathcal{R})} \otimes \mathcal{O}_1)(\mathcal{A}_1 \otimes \mathcal{B}_1 \otimes \mathbb{K}_{\mathcal{R}})(\rho_{\text{in}}). \quad (22)$$

As in [24], we specify that an oracle \mathcal{O} can be a communication oracle or an ideal functionality oracle.

An *adversary* $\tilde{\mathcal{A}}$ for an honest party \mathcal{A} in $\Pi^\mathcal{O} = (\mathcal{A}, \mathcal{B}, \mathcal{O}, n)$ is an n -tuple of quantum operations matching the input and outputs spaces of \mathcal{A} . A *simulator* for $\tilde{\mathcal{A}}$ is a sequence of quantum operations $(\mathcal{S}_i)_{i=1}^n$ where \mathcal{S}_i has the same input-output spaces as the maps of $\tilde{\mathcal{A}}$ at step i . In addition, \mathcal{S} has access to the ideal functionality for the protocol Π .

► **Definition 11.** An n -step quantum two-party protocol with oracle calls, $\Pi^\mathcal{O} = (\mathcal{A}, \mathcal{B}, \mathcal{O}, n)$ is statistically stand-alone secure against a corrupt \mathcal{A} if for every adversary $\tilde{\mathcal{A}}$ there exists a simulator \mathcal{S} such that for every input ρ_{in} ,

$$\text{Tr}_{\mathcal{B}_n \otimes \mathcal{R}}(\tilde{\mathcal{A}} \circledast \mathcal{B}) = \text{Tr}_{\mathcal{B}_n \otimes \mathcal{R}}(\mathcal{S} \circledast \mathcal{B}). \quad (23)$$

We note that Definition 11 is weaker than some other definitions for active security used in the literature, e.g., [25], because we ask only that the *local* view of the adversary be simulated.

Given the simple structure of our protocol and ideal functionality, the construction and proof of the simulator is straightforward as shown below.

► **Theorem 12.** *Protocol Π is statistically stand-alone secure against a corrupt sender.*

Proof. Since Π consists in a single message from the sender to the receiver (together with a call to the ideal functionality for the token), we have that $\mathcal{A} = (\mathcal{A}_1)$. Furthermore, since the ideal functionality $\mathcal{F}_{\text{wrap}}$ does not return anything to the sender, there is no need for our simulator \mathcal{S} to call an ideal functionality.

We thus build \mathcal{S} that runs \mathcal{A} on the input in register \mathcal{A}_0 . When \mathcal{A} calls the $\mathcal{F}_{\text{wrap}}$ ideal functionality, the simulator does nothing. Since Π is a one-way protocol, and since the ideal functionality also does not allow communication from the receiver to the sender,

$$\text{Tr}_{\mathcal{B}_n \otimes \mathcal{R}}(\tilde{\mathcal{A}} \circ \mathcal{B}) = \mathcal{A}(\text{Tr}_{\mathcal{B}_0 \otimes \mathcal{R}}(\rho_{\text{in}})) = \mathcal{S}(\text{Tr}_{\mathcal{B}_0 \otimes \mathcal{R}}(\rho_{\text{in}})). \quad (24)$$

This concludes the proof. ◀

C Proof of Lemma 4

For clarity, implicitly in our proof below, we model the oracle O_f as having three possible outputs: 0, 1, or 2, where 2 is output whenever O_f is fed an invalid key y . This is required for the notion of having “few” keys to make sense (i.e., there are 2^n candidate keys, and only two secret bits, each of which is supposed to have a bounded number of keys). Note that our construction indeed fits into this framework.

Proof. Observe first that an honest receiver Alice wishing to extract s_i acts as follows. She applies a unitary $U_i \in \mathcal{U}(A \otimes B)$ to get state

$$|\phi_1\rangle := U_i |\psi\rangle_{AB} |0\rangle_C. \quad (25)$$

She then measures B in the computational basis and postselects on result $y \in \{0, 1\}^n$, obtaining state

$$|\phi_2\rangle := |\phi_y\rangle_A |y\rangle_B |0\rangle_C. \quad (26)$$

She now treats y as a “key” for s_i , i.e., she applies O_f to $B \otimes C$ to obtain her desired bit s_i , i.e.,

$$|\phi_3\rangle := |\phi_y\rangle_A |y\rangle_B |s_i\rangle_C. \quad (27)$$

A malicious receiver Bob wishing to extract s_0 and s_1 now acts similarly to the rewinding strategy for superposition queries. Suppose without loss of generality that s_0 has at most Δ keys. Then, Bob first applies U_0 to prepare $|\phi_1\rangle$ from Equation (25), which we can express as

$$|\phi_1\rangle = \sum_{y \in \{0,1\}^n} \alpha_y |\psi_y\rangle_A |y\rangle_B |0\rangle_C. \quad (28)$$

for $\sum_y |\alpha_y|^2 = 1$. Since measuring B next would allow us to retrieve s_0 in register C with certainty, we have that all y appearing in the expansion above satisfy $f(y) = s_0$. Moreover,

since s_0 has at most Δ keys, there exists a key y' such that $|\alpha_{y'}|^2 \geq 1/\Delta$. Bob now measures B in the computational basis to obtain $|\phi_2\rangle$ from Equation (26), obtaining y' with probability at least $1/\Delta$. Feeding y' into O_f yields s_0 . Having obtained y' , we have that $|\langle\phi_1|\phi_2\rangle|^2 \geq 1/\Delta$, implying

$$\left| \langle\psi|U_0^\dagger|\phi_{y'}\rangle|y'\rangle \right|^2 \geq 1/\Delta, \quad (29)$$

i.e., Bob now applies U_0^\dagger to recover a state with “large” overlap with initial state $|\psi\rangle$.

To next recover s_1 , define $|\psi_{\text{good}}\rangle := U_1|\psi\rangle$ and $|\psi_{\text{approx}}\rangle := U_1U_0^\dagger|\phi_{y'}\rangle|y'\rangle$. Bob applies U_1 to obtain

$$|\psi_{\text{approx}}\rangle = \beta_1|\psi_{\text{good}}\rangle + \beta_2|\psi_{\text{good}}^\perp\rangle, \quad (30)$$

where $\sum_i |\beta_i|^2 = 1$, $\langle\psi_{\text{good}}|\psi_{\text{good}}^\perp\rangle = 0$, and $|\beta_1|^2 \geq 1/\Delta$. Define

$$\Pi_{\text{good}} := \sum_{y \in \{0,1\}^n \text{ s.t. } f(y)=s_1} |y\rangle\langle y|.$$

Then, the probability that measuring B in the computational basis now yields a valid key for s_1 is

$$\langle\psi_{\text{approx}}|\Pi_{\text{good}}|\psi_{\text{approx}}\rangle \geq |\beta_1|^2 \geq \frac{1}{\Delta}, \quad (31)$$

where we have used the fact that $\Pi_{\text{good}}|\psi_{\text{good}}\rangle = |\psi_{\text{good}}\rangle$ (since an honest receiver can extract s_1 with certainty). We conclude that Bob can extract both s_0 and s_1 with probability at least $1/\Delta^2$. \blacktriangleleft

Beyond Product State Approximations for a Quantum Analogue of Max Cut

Anurag Anshu 

Institute for Quantum Computing, University of Waterloo, Canada
Department of Combinatorics and Optimization, University of Waterloo, Canada
Perimeter Institute for Theoretical Physics, Waterloo, Canada
aanshu@uwaterloo.ca

David Gosset 

Institute for Quantum Computing, University of Waterloo, Canada
Department of Combinatorics and Optimization, University of Waterloo, Canada
dgosset@uwaterloo.ca

Karen Morenz 

Institute for Quantum Computing, University of Waterloo, Canada
Department of Combinatorics and Optimization, University of Waterloo, Canada
Department of Chemistry, University of Toronto, Canada
karen.morenz@mail.utoronto.ca

Abstract

We consider a computational problem where the goal is to approximate the maximum eigenvalue of a two-local Hamiltonian that describes Heisenberg interactions between qubits located at the vertices of a graph. Previous work has shed light on this problem's approximability by *product states*. For any instance of this problem the maximum energy attained by a product state is lower bounded by the Max Cut of the graph and upper bounded by the standard Goemans-Williamson semidefinite programming relaxation of it. Gharibian and Parekh described an efficient classical approximation algorithm for this problem which outputs a product state with energy at least 0.498 times the maximum eigenvalue in the worst case, and observe that there exist instances where the best product state has energy 1/2 of optimal. We investigate approximation algorithms with performance exceeding this limitation which are based on optimizing over tensor products of few-qubit states and shallow quantum circuits. We provide an efficient classical algorithm which achieves an approximation ratio of at least 0.53 in the worst case. We also show that for any instance defined by a 3 or 4-regular graph, there is an efficiently computable shallow quantum circuit that prepares a state with energy larger than the best product state (larger even than its semidefinite programming relaxation).

2012 ACM Subject Classification Theory of computation

Keywords and phrases Approximation algorithms, Quantum many-body systems

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.7

Related Version A preprint is available at <https://arxiv.org/abs/2003.14394>.

Funding *Anurag Anshu*: Supported by the Canadian Institute for Advanced Research, through funding provided to the Institute for Quantum Computing by the Government of Canada and the Province of Ontario. Perimeter Institute is also supported in part by the Government of Canada and the Province of Ontario.

David Gosset: Supported by the Natural Sciences and Engineering Research Council of Canada, IBM Research, and the Canadian Institute for Advanced Research.

Karen Morenz: Supported by Vanier Canada Graduate Scholarship from the Natural Sciences and Engineering Research Council of Canada, and by the Transformative Quantum Technologies Quantum Graduate Visitors Program.

Acknowledgements We thank Sevag Gharibian, Hosho Katsura, Eunou Lee, and Ojas Parekh for comments and helpful discussions.



© Anurag Anshu, David Gosset, and Karen Morenz;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 7; pp. 7:1–7:15



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

In this paper we continue a line of recent work which aims to understand the power and limitations of approximation algorithms for quantum constraint satisfaction problems. Consider an n -qubit local Hamiltonian of the form

$$H = \sum_{ij} h_{ij}. \quad (1)$$

Here each term h_{ij} is a Hermitian operator which acts nontrivially only on qubits i and j and we shall assume that $h_{ij} \geq 0$. Estimating the maximum energy $\|H\|$ is a quantum constraint satisfaction problem which is a special case of the well-studied 2-local Hamiltonian problem, and it is known that computing an estimate of $\|H\|$ within a given small additive error $\epsilon = 1/\text{poly}(n)$ is QMA-complete [16, 15]. Consequently, this sort of precise estimate is unlikely to admit efficient algorithms. An estimate λ is an r -approximation of $\|H\|$, or achieves approximation ratio r , if

$$r \leq \frac{\lambda}{\|H\|} \leq 1.$$

The classical PCP theorem places stringent bounds on the efficiency of good approximation algorithms for this problem even in the special case where H is diagonal in the computational basis. It states that there exists a constant $r < 1$ such that computing an r -approximation to $\|H\|$ is NP-hard [3]. A major open question in this area is whether or not the problem is in fact QMA-hard for some $r < 1$. Whereas the standard PCP theorem already implies hardness of approximation, the quantum PCP conjecture targets the more fine-grained question of whether or not such approximations can be checked efficiently given a concise classical witness. These considerations also motivate the study of efficient classical or quantum algorithms for such quantum approximation problems, as measured by the achievable approximation ratio.

A natural way to establish a lower bound $\|H\| \geq \alpha$ is to exhibit a state $|\phi\rangle$ satisfying $\langle\phi|H|\phi\rangle \geq \alpha$. Several previous works have bounded the approximation ratios that can be achieved by product states $\phi = \phi_1 \otimes \phi_2 \otimes \dots \otimes \phi_n$ [10, 5, 14, 6, 12]. Gharibian and Kempe have shown that there always exists a product state which achieves an approximation ratio $r = 0.5$ [10]. This is also easily seen to be the best possible approximation guarantee for product states, as there are simple examples which saturate this bound. It is not known if a product state achieving a ratio $1/2$ can be computed efficiently in the general case; the most recent progress is an efficient algorithm which outputs a product state that achieves a ratio of $r = 0.328$ [1]. On the other hand, it is known that efficient classical algorithms can achieve approximation ratios arbitrarily close to 1 if we are willing to specialize to certain families of 2-local Hamiltonians. Such algorithms are known if the graph which describes the nonzero interactions between qubits is either (a) a $d = O(1)$ dimensional lattice, (b) a planar graph [4, 5] or (c) dense graphs, in which the number of edges is close to maximal, i.e. $\Omega(n^2)$ [10, 5].

For completeness, we note that Ref. [6] considers a different approximation problem for Hamiltonians where the terms h_{ij} are traceless (rather than positive semidefinite) and describes an efficient $r = O(1/\log(n))$ approximation algorithm based on product states, generalizing the classical result of Charikar and Wirth [8]. A related work [14] considers a slightly different notion of approximation ratio, again achieved by product states in the traceless setting.

An n -qubit product state is an appealing generalization of a classical n -bit string, and has the desirable feature that it can be manipulated and stored efficiently by classical algorithms.

Moreover, some of the known approximation algorithms for classical constraint satisfaction problems which are based on semidefinite programming have a natural extension to product states [6, 12]. But how good are these algorithms, and can we hope to do better using efficient algorithms that are based on entangled states? Of course, most n -qubit states do not have concise classical descriptions and cannot even be prepared efficiently using a quantum computer. Since we are aiming for efficient algorithms, we shall restrict our attention to entangled quantum states prepared by polynomial size quantum circuits.

We shall focus our attention on a specific family of Hamiltonians studied previously in Ref. [12] which defines a quantum analogue of the Max Cut problem. Unless otherwise specified, throughout this paper we shall consider graphs $G = (V, E, w)$ with nonnegative edge weights $w : E \rightarrow \mathbb{R}_{\geq 0}$, and we write $n = |V|$. We shall also assume that the maximum edge weight is upper bounded by $O(n^c)$ for some $c = O(1)$.

For completeness we begin by reviewing facts about the classical Max Cut problem. Recall that the maximum cut of a weighted graph G is defined to be

$$\text{MC}(G) = \max_{z \in \{\pm 1\}^n} \text{Cut}_G(z) \quad \text{where} \quad \text{Cut}_G(z) = \sum_{\{i,j\} \in E} \frac{w_{ij}}{2} (I - z_i z_j). \quad (2)$$

An approximation algorithm for the Max Cut problem due to Goemans and Williamson [13] is based on the following semidefinite programming relaxation of Eq. (2):

$$\text{SDP}(G) = \max_{M \in \mathbb{R}^{n \times n} : M \geq 0, \text{diag}(M) = I} \sum_{\{i,j\} \in E} \frac{w_{ij}}{2} (I - M_{ij}). \quad (3)$$

A matrix M achieving the maximum SDP value $\text{SDP}(G)$ can be computed efficiently using standard classical algorithms. The Goemans-Williamson algorithm then uses a randomized procedure which maps M to a bit string z which is guaranteed to satisfy

$$\text{Cut}_G(z) \geq 0.8785 \cdot \text{SDP}(G) \quad (4)$$

for all graphs G [13].

The quantum Max Cut problem as considered in Ref. [12] is defined by a family of local Hamiltonians Eq. (1) where each term h_{ij} is proportional to the two-qubit singlet state $|s\rangle = \sqrt{2}^{-1}(|01\rangle - |10\rangle)$. In particular, given a graph $G = (V, E, w)$ we define

$$H_G = \sum_{\{i,j\} \in E} w_{ij} h_{ij} \quad h_{ij} = \frac{1}{2} (I - X_i X_j - Y_i Y_j - Z_i Z_j) = 2|s\rangle\langle s|_{ij}. \quad (5)$$

We are interested in approximating the maximum eigenvalue of H_G which we write as

$$\text{OPT}(G) = \|H_G\|.$$

Estimating this quantity can be viewed as a quantum analogue of the classical Max Cut problem. Indeed, a constraint $(I - z_i z_j)$ in the Max Cut problem Eq. (2) has maximal energy when the corresponding two entries disagree, i.e., $z_i \neq z_j$. Analogously, a constraint h_{ij} in the Hamiltonian Eq. (5) has maximal energy for a quantum state $|\psi\rangle$ when the two qubits are antisymmetric under swap, i.e., $\text{SWAP}_{ij}|\psi\rangle = -|\psi\rangle$. In this sense, the classical and quantum constraints represent two different notions of disagreement.

Piddock and Montanaro have shown that the problem of computing a precise estimate of $\text{OPT}(G)$ is QMA-complete [19], and recent work has focused on its approximability using product states [12]. Let us now see how the problem of optimizing the energy of Eq. (5)

7:4 Beyond Product State Approximations

over product states is directly related to the Max Cut problem Eq. (2) and its semidefinite relaxation Eq. (3). An n -qubit product state ϕ can be specified (up to a global phase) by n normalized vectors $v^{(j)} \in \mathbb{R}^3$:

$$|\phi\rangle\langle\phi| = \bigotimes_{j=1}^n \frac{1}{2} \left(I + v_1^{(j)} X + v_2^{(j)} Y + v_3^{(j)} Z \right) \quad \|v^{(j)}\| = 1,$$

and its energy is given by

$$\text{Tr} [|\phi\rangle\langle\phi| H_G] = \sum_{\{i,j\} \in E} \frac{w_{ij}}{2} (1 - v^{(i)} \cdot v^{(j)}). \quad (6)$$

Defining

$$\alpha(k) = \max_{\{v_i \in \mathbb{R}^k : \|v_i\|=1\}} \sum_{\{i,j\} \in E} \frac{w_{ij}}{2} (1 - v_i \cdot v_j). \quad (7)$$

we see that

$$\alpha(1) = \text{MC}(G) \leq \text{PROD}(G) = \alpha(3) \leq \text{SDP}(G) = \alpha(n).$$

The Goemans-Williamson algorithm for the Max Cut problem has been generalized by Briet, de Oliveira Filho, and Vallentin to obtain efficient algorithms for approximating $\alpha(k)$ for $1 < k < n$ [7]. The resulting approximation ratios obtained become larger as k increases towards $k = n$ where the optimal value can be computed efficiently and exactly by semidefinite programming. Their result for the case $k = 3$ at hand is summarized below.

► **Theorem 1** ([7]). *There exists an efficient randomized classical algorithm which computes an estimate μ such that*

$$0.956 \leq \frac{\mu}{\text{PROD}(G)} \leq 1.$$

This algorithm (and other randomized algorithms discussed in this paper) may fail with some small probability, say 0.01, in which case the output of the algorithm is a flag indicating failure.

Since $\text{PROD}(G) \geq 0.5 \cdot \text{OPT}(G)$ [10], the algorithm described in Theorem 1 can be used to approximate $\text{OPT}(G)$ with ratio at least $0.5 \cdot 0.956 = 0.478$, as observed in Ref. [11]. The recently proposed approximation algorithm of Gharibian and Parekh [12] is based on rounding a solution to a semidefinite program relaxation of $\text{OPT}(G)$, and obtains an even higher ratio of 0.498. The authors of Ref. [12] note that their algorithm is almost optimal (as far as product states are concerned), since there exists a very simple graph – just two vertices connected by a single weight one edge – for which the optimal product state is equal to $0.5 \cdot \text{OPT}(G)$.

Our first result shows that if all edge weights are equal then this limitation of product states only occurs in small graphs. That is, for sufficiently large connected graphs with uniform weights, it is always possible to efficiently find a product state with a strictly larger approximation ratio:

► **Theorem 2.** *Suppose $G = (V, E, w)$ is a connected and unweighted graph, i.e., $w_{ij} = 1$ for all $\{i, j\} \in E$. Then*

$$\frac{\text{PROD}(G)}{\text{OPT}(G)} \geq \frac{4}{7} - O(|E|^{-1}). \quad (8)$$

The efficient randomized algorithm from Theorem 1 computes an r -approximation to $\text{OPT}(G)$, where $r \geq 0.546 - O(|E|^{-1})$.

In the more general setting where the weights may not be uniform, one can of course construct examples of connected graphs where all weights are vanishingly small except for the weight of a single edge. In this limit we already know that it is impossible to beat a ratio of 0.5 using product states. Our next result shows that by considering tensor products of one- and two-qubit states it is possible to guarantee a strictly better approximation ratio.

► **Theorem 3.** *Let $G = (V, E, w)$ be a weighted graph. Then there is a tensor product $\phi = \phi_1 \otimes \phi_2 \otimes \dots \otimes \phi_L$ of 1- and 2-qubit states $\{\phi_j\}$ such that*

$$\frac{\langle \phi | H_G | \phi \rangle}{\text{OPT}(G)} \geq 0.55.$$

Moreover, there is an efficient randomized algorithm which outputs an r -approximation to $\text{OPT}(G)$, where $r \geq 0.53$.

Theorem 3 provides the best currently known efficient approximation algorithm for this problem, improving slightly on Ref. [12]. Moreover, it establishes that although there exist graphs where the best product state is only 1/2 of the optimal energy, efficient classical algorithms can go slightly beyond this ratio.

Our next result shows that, for a family of low-degree graphs it is possible to efficiently beat product states on *every* graph from the family. In particular, given any 3- or 4-regular graph G , we can efficiently compute a constant-depth quantum circuit which prepares a state with energy strictly larger than the best product state energy $\text{PROD}(G)$ (in fact, larger than its semidefinite relaxation $\text{SDP}(G)$).

► **Theorem 4.** *Suppose $G = (V, E, w)$ is a k -regular graph with $k \in \{3, 4\}$. There is a depth- $(k+1)$ quantum circuit $U(G)$ that can be efficiently computed by a randomized classical algorithm such that the state $|\phi\rangle = U(G)|0^n\rangle$ approximates $\text{OPT}(G)$ with a strictly larger ratio than that of any product state. Moreover,*

$$\frac{\langle \phi | H_G | \phi \rangle}{\text{PROD}(G)} \geq \frac{\langle \phi | H_G | \phi \rangle}{\text{SDP}(G)} > 1.001$$

The low depth quantum circuit used in Theorem 4 is inspired (and similar to) the quantum approximate optimization algorithm described by Farhi, Goldstone, and Gutmann [9]. The circuit is directly obtained from any computational basis state $z \in \{0, 1\}^n$ with a large enough cut value $\text{Cut}_G(z)$ ¹; in particular, it is sufficient to use a bit string satisfying Eq. (4) which can be computed efficiently using the Goemans-Williamson algorithm. The quantum computation starts from the computational basis state $|z\rangle$ and then applies a low-depth quantum circuit composed of a sequence of commuting two-qubit gates of the form

$$e^{i\theta P(j)P(k)}$$

where for each qubit v we choose a Pauli operator $P(v) \in \{X_v, Y_v\}$ depending only on the bit z_v . To prove the theorem we compute the energy of this state as a function of the variational parameter θ and then optimize.

In summary, we have shown that for the quantum Max Cut problem there are efficient algorithms which beat any approximation algorithm based on product states. A natural open question is whether this is also true for the more general problem of approximating the

¹ Note that we previously defined $\text{Cut}_G(z)$ for inputs $z \in \{-1, 1\}^n$. Here and below we extend this definition to bit string inputs $z \in \{0, 1\}^n$ by identifying each bit z_j with the ± 1 -valued variable $(-1)^{z_j}$.

maximum energy of a two-local Hamiltonian Eq. (1). One may also ask if the semidefinite programming method [13] can be used in some novel way to efficiently obtain approximation ratios which go beyond the limitations of product states. For the quantum Max Cut problem, Ref. [12] provides a semidefinite program which upper bounds the optimal energy $\text{OPT}(G)$. A central challenge here is that we do not (yet) know a randomized rounding scheme which maps an SDP solution to an entangled state.

2 Tensor products of few qubit states

In this Section, we prove Theorems 2 and 3. We shall use the following upper bound for the special case where G is a star graph. The lemma shows that the maximum energy for any star with at least 3 vertices is always less than the trivial upper bound $2 \sum_{e \in E} w_e$ which comes from the triangle inequality. This can be interpreted as a consequence of the monogamy of entanglement—the center spin cannot be maximally entangled with all of the points of the star. Along similar lines, Ref. [11] provides a different upper bound on $\|H_G\|$ using a monogamy of entanglement bound known as the Coffman-Kundu-Wootters inequality.

► **Lemma 5.** *Suppose $G = (V, E, w)$ is a star graph with nonnegative weights. Then*

$$\|H_G\| \leq \max_{e \in E} w_e + \sum_{e \in E} w_e. \quad (9)$$

Proof. Define the total spin operators

$$\vec{S} = \frac{1}{2} \left(\sum_{j \in V} X_j, \sum_{j \in V} Y_j, \sum_{j \in V} Z_j \right).$$

Let $S_x = \frac{1}{2} \sum_{j \in V} X_j$, $S_y = \frac{1}{2} \sum_{j \in V} Y_j$, $S_z = \frac{1}{2} \sum_{j \in V} Z_j$ and note that the Hamiltonian Eq. (5), $S^2 = S_x^2 + S_y^2 + S_z^2$, and S_z are mutually commuting. It is shown in Ref. [17] that the maximum eigenvalue of Eq. (5) on any (nonnegatively) weighted complete bipartite graph with bipartition $V = A \sqcup B$ is attained by an eigenvector ϕ which satisfies

$$S^2|\phi\rangle = s(s+1)|\phi\rangle \quad S_z|\phi\rangle = s|\phi\rangle \quad s = (|A| - |B|)/2.$$

A star graph is a complete bipartite graph with $|A| = |V| - 1$ and $|B| = 1$. Therefore the result of Lieb and Mattis implies that a maximum eigenvalue is attained by a state ϕ satisfying $S_z|\phi\rangle = |V|/2 - 1$. In particular, ϕ is equal to the maximum eigenvector of H_G restricted to the $|V|$ -dimensional subspace

$$\mathcal{Q} = \text{span}\{|100 \dots 0\rangle, |010 \dots 0\rangle, \dots, |00 \dots, 01\rangle\}$$

spanned by computational basis states with Hamming weight equal to 1. It is easily seen that the Hamiltonian restricted to this subspace is the Laplacian matrix of G . More precisely,

$$H_G|_{\mathcal{Q}} = L(G)$$

where $L(G)$ is the graph Laplacian of G , defined by

$$L(G)_{ij} = \begin{cases} \sum_{e \sim i} w_e & i = j \\ -w_e & e = \{i, j\} \in E \\ 0 & \text{otherwise.} \end{cases}$$

The lemma follows by upper bounding the norm of the Laplacian of a star graph

$$\|L(G)\| \leq \sum_{e \in E} w_e + \max_{e \in E} w_e. \quad (10)$$

The upper bound Eq. (10) is obtained using an argument from Ref. [18]. In particular, we note that

$$\|L(G)\| = \|W^{-1}L(G)W\|$$

where W is a diagonal matrix such that $W_{ii} = \sum_{e \sim i} w_e$, and then use Gershgorin's circle theorem to upper bound the right hand side. Computing the Gershgorin discs for a star graph we arrive at

$$\|W^{-1}L(G)W\| \leq \max \left\{ \sum_{e \in E} w_e + \left(\sum_{e \in E} w_e \right)^{-1} \sum_{e \in E} w_e^2, \max_{e \in E} w_e + \sum_{e \in E} w_e \right\} \quad (11)$$

$$= \max_{e \in E} w_e + \sum_{e \in E} w_e. \quad (12)$$

◀

We note that for a star graph with uniform weights the upper bound Eq. (9) becomes an equality, as can be seen using the rules for addition of angular momentum.

Next, we consider the case of uniform weights $w_{ij} = 1$ on an arbitrary connected graph. Using Lemma 5, we exhibit a product state with approximation ratio better than $\frac{1}{2}$.

► **Theorem 6.** *Suppose $G = (V, E, w)$ is a connected graph with uniform weights, i.e., $w_{ij} = 1$ for all $\{i, j\} \in E$. Then*

$$\frac{\text{PROD}(G)}{\text{OPT}(G)} \geq \frac{1}{3} + \frac{2}{3} \left(\frac{|E|}{2|E| + |V|} \right). \quad (13)$$

Moreover, there exists a computational basis state with energy satisfying the above inequality.

Proof. For any vertex $v \in V$ define a Hamiltonian h_v which has support only on qubit v and its neighbors:

$$h_v = \sum_{j: \{v, j\} \in E} \frac{1}{2} (I - X_v X_j - Y_v Y_j - Z_v Z_j).$$

Note that we may write $H_G = \frac{1}{2} \sum_{v \in V} h_v$, where the factor of $1/2$ compensates for the fact that the Hamiltonian term corresponding to each edge of the graph appears twice on the right hand side. Now using the triangle inequality we get

$$\text{OPT}(G) \leq \frac{1}{2} \sum_{v \in V} \|h_v\|. \quad (14)$$

Let us write d_v for the degree of vertex v . Then

$$\|h_v\| \leq d_v + 1,$$

where we used Lemma 5. Substituting in Eq. (14) gives

$$\text{OPT}(G) \leq \frac{1}{2} \sum_{v \in V} (d_v + 1) = |E| + |V|/2. \quad (15)$$

7:8 Beyond Product State Approximations

The upper bound Eq. (15) and its proof is similar to the one derived by Anderson in Ref. [2] for the special case of bipartite graphs. To see why Eq. (15) is nontrivial, note that since G is a connected graph on $|V|$ vertices, it satisfies $|E| \geq |V| - 1$ (the minimum is attained by a tree). Thus Eq. (15) implies

$$\text{OPT}(G) \leq \frac{1}{2}(3|E| + 1), \quad (16)$$

which improves upon the naive upper bound $\text{OPT}(G) \leq 2|E|$ which is obtained by applying the triangle inequality directly to Eq. (5).

We need only a little bit more to get the Theorem from Eq. (15). Let us write

$$H_G = \frac{|E|}{2} + H^X(G) + H^Y(G) + H^Z(G)$$

where

$$H^X(G) = -\frac{1}{2} \sum_{\{i,j\} \in E} X_i X_j \quad H^Y(G) = -\frac{1}{2} \sum_{\{i,j\} \in E} Y_i Y_j \quad H^Z(G) = -\frac{1}{2} \sum_{\{i,j\} \in E} Z_i Z_j. \quad (17)$$

We denote their largest eigenvalues as $\lambda_{\max}^P(G)$ with $P = X, Y, Z$. Note that these 3 quantities are all equal. Applying the triangle inequality and using this fact gives

$$\text{OPT}(G) \leq \frac{|E|}{2} + 3\lambda_{\max}^Z(G). \quad (18)$$

Also note that we can lower bound $\text{PROD}(G)$ by the maximum energy of a computational basis state:

$$\text{PROD}(G) \geq \frac{|E|}{2} + \lambda_{\max}^Z(G). \quad (19)$$

Now combining Eqs. (18, 19) gives

$$\text{PROD}(G) \geq \frac{|E|}{2} + \frac{\text{OPT}(G) - |E|/2}{3} = \frac{1}{3}\text{OPT}(G) + \frac{1}{3}|E|.$$

Therefore

$$\frac{\text{PROD}(G)}{\text{OPT}(G)} \geq \frac{1}{3} + \frac{1}{3} \frac{|E|}{\text{OPT}(G)}.$$

Finally, substituting Eq. (15) in the second term we arrive at Eq. (13) and complete the proof. \blacktriangleleft

Proof of Theorem 2

Proof. Let T be a spanning tree of G , which can be computed efficiently and has $|V| - 1$ edges. Let $s \in \{0, 1\}^n$ be a bit string corresponding to a 2-coloring of T , i.e., $s_i \neq s_j$ whenever $\{i, j\}$ is an edge of T (of course, s can also be computed efficiently). Then

$$\langle s | H_G | s \rangle = \text{Cut}_G(s) \geq |V| - 1,$$

and combining with Eq. (15) gives

$$\frac{\langle s | H_G | s \rangle}{\text{OPT}(G)} \geq \frac{2|V| - 2}{2|E| + |V|}.$$

Putting this together with Theorem 6 we arrive at

$$\frac{\text{PROD}(G)}{\text{OPT}(G)} \geq \max \left\{ \frac{2(|V| - 1)}{2|E| + |V|}, \frac{4|E| + |V|}{6|E| + 3|V|} \right\} \quad (20)$$

$$(21)$$

Now let $x = (|V| - 1)/|E|$ and note that $x \in [0, 1]$, and

$$\frac{\text{PROD}(G)}{\text{OPT}(G)} \geq \min_{0 \leq x \leq 1} \max \left\{ \frac{2x}{2+x}, \frac{4+x}{6+3x} \right\} - O(|E|^{-1}) \quad (22)$$

$$= 4/7 - O(|E|^{-1}). \quad (23)$$

The randomized approximation algorithm of 1 outputs an estimate which is an α -approximation to $\text{PROD}(G)$ with ratio $\alpha \geq 0.956$. Eq. (23) implies that this estimate is an r -approximation of $\text{OPT}(G)$ with $r \geq \alpha \cdot (4/7 - O(|E|^{-1})) = 0.546 - O(|E|^{-1})$. ◀

Proof of Theorem 3

Proof. Note that in the weighted case we may run through exactly the same arguments used to obtain Eq. (13). Eq. (15) is replaced by

$$\text{OPT}(G) \leq W + \frac{1}{2} \sum_{v \in V} \max_{e \sim v} w_e$$

where $W = \sum_{e \in E} w_e$, and correspondingly we have

$$\frac{\text{PROD}(G)}{\text{OPT}(G)} \geq \frac{1}{3} + \frac{2}{3} \left(\frac{W}{2W + \sum_{v \in V} \max_{e \sim v} w_e} \right). \quad (24)$$

Now let us focus on the expression

$$\sum_{v \in V} \max_{e \sim v} w_e.$$

We note that this quantity can be trivially upper bounded as $2W$ since each edge can appear at most twice in the sum (once for each of its incident vertices). This naive upper bound is not sufficient for our purposes, and so we perform a more careful analysis below.

► **Lemma 7.** *We may efficiently compute edge subsets $M, F \subseteq E$ such that M is a matching and F is a forest, and*

$$\sum_{v \in V} \max_{e \sim v} w_e = \sum_{e \in M} w_e + \sum_{e \in F} w_e.$$

Proof. Let us fix an ordering e_1, e_2, \dots, e_m of all the edges of G such that

$$w_{e_1} \leq w_{e_2} \leq \dots \leq w_{e_m}$$

(if all edge weights are distinct there is a unique such ordering, otherwise there is some freedom in the choice). Now for each vertex $v \in V$ we let $I(v) \in E$ be the (unique) edge incident to v which is maximal with respect to the above ordering. We define

$$F = \{I(v) : v \in V\}$$

$$M = \{e \in E : e = I(v) \text{ and } e = I(w) \text{ for two distinct vertices } v \neq w \in V\}.$$

7:10 Beyond Product State Approximations

At most one edge $e = I(v)$ incident to any given vertex v can appear in M , and hence M is a matching. To see that F is a forest, consider a graph $G' = (V, E, w')$ with the same vertex and edge sets as G , but where the edge weights are rescaled so that $w'(e_j) = -j$ for $1 \leq j \leq m$ (in particular, all edge weights are negative, distinct, and their magnitudes respect our chosen ordering). Then each edge of F is contained in any minimum spanning tree of G' , by the well-known cut property of minimum spanning trees. We infer that F does not contain any cycles, and is therefore a forest. \blacktriangleleft

Now let M, F be as in the lemma, and define the set of vertices $U \subseteq V$ which are not incident to an edge in M . Consider a random variable

$$|\phi_z\rangle = \left(\bigotimes_{e=\{i,j\} \in M} \frac{1}{\sqrt{2}} (|01\rangle - |10\rangle)_{ij} \right) \otimes |z\rangle_U$$

where $z \in \{0, 1\}^{|U|}$ is a uniformly random bit string. Then

$$\mathbb{E}_z[\langle \phi_z | H_G | \phi_z \rangle] = 2 \sum_{e \in M} w_e + \frac{1}{2} (W - \sum_{e \in M} w_e) = \frac{3}{2}m + \frac{1}{2}W \quad m \equiv \sum_{e \in M} w_e. \quad (25)$$

This shows that there exists a state ϕ_z with energy at least $\frac{3}{2}m + \frac{1}{2}W$.

Finally, since F is a forest, we may efficiently compute a computational basis state $s \in \{0, 1\}^n$ such that

$$\langle s | H_G | s \rangle \geq f \quad f \equiv \sum_{e \in F} w_e. \quad (26)$$

This follows since the Max Cut for a forest is achieved by an efficiently computable 2-coloring of the vertices. Putting together Eqs. (24,25,26) and Lemma 7, we see that there exists a tensor product ϕ of 1- and 2-qubit states such that

$$\frac{\langle \phi | H_G | \phi \rangle}{\text{OPT}(G)} \geq \max \left\{ \frac{2f}{2W + f + m}, \frac{3m + W}{2W + f + m}, \frac{1}{3} + \frac{2}{3} \left(\frac{W}{2W + f + m} \right) \right\} \quad (27)$$

$$\geq \min_{0 \leq x \leq y \leq 1} \frac{1}{2 + y + x} \max \left\{ 2y, 3x + 1, \frac{1}{3}(4 + x + y) \right\} \quad (28)$$

$$\geq 0.55, \quad (29)$$

where in the second line we set $x = m/W, y = f/W$, and in the last line we used a computer.

Now let us bound the approximation ratio achieved by an efficient randomized algorithm. First note that the state $|s\rangle$ in Eq. (26) can be computed efficiently. Moreover, using Eq. (25) and the fact that $\langle \phi_z | H_G | \phi_z \rangle$ is a random variable upper bounded by $2W$ we see that the probability of randomly sampling a bit string z with energy at least $\frac{3}{2}m + 0.49W$ is

$$\Pr \left[\langle \phi_z | H_G | \phi_z \rangle \geq \frac{3}{2}m + 0.49W \right] \geq \frac{0.01}{1.51}.$$

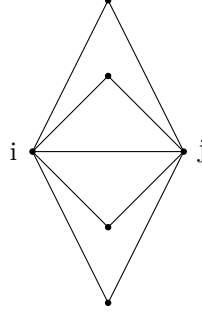
By randomly sampling $O(1)$ times, with high probability we will obtain a bit string with this energy. Finally, note that combining Theorem 1 with Eq. (24) we get a randomized algorithm that outputs a state with energy at least

$$0.956 \cdot \left(\frac{1}{3} + \frac{2}{3} \left(\frac{W}{2W + f + m} \right) \right).$$

Thus we may efficiently compute a state with approximation ratio at least

$$\min_{0 \leq x \leq y \leq 1} \frac{1}{2 + y + x} \max \left\{ 2y, 3x + 0.98, \frac{0.956}{3}(4 + x + y) \right\} \geq 0.53. \quad (30)$$

\blacktriangleleft



■ **Figure 1** An edge $\{i, j\}$ contained in 4 triangles.

3 Low degree regular graphs

In this section we consider the case of 3- or 4-regular graphs and we establish Theorem 4.

Given an n -vertex graph $G = (V, E, w)$, we shall consider the following algorithm. First, we use the classical Goemans-Williamson algorithm [13] to compute a bit string $z \in \{0, 1\}^n$ satisfying Eq. (4). This defines a partition of the edges into those which are satisfied and those which are not:

$$E_{\text{sat}} = \{\{u, v\} \in E : z_u = z_v\} \quad E_{\text{unsat}} = E \setminus E_{\text{sat}}. \quad (31)$$

Next we define a Pauli operator $P(j)$ for each qubit $1 \leq j \leq n$, which depends on the j -th bit of z :

$$P(j) = \begin{cases} X_j, & z_j = 1 \\ Y_j, & z_j = 0 \end{cases}.$$

Finally, we define a variational state

$$|\phi(\theta)\rangle = V(\theta)|z\rangle \quad \text{where} \quad V(\theta) = \exp\left(\sum_{\{j,k\} \in E} i\theta P(j)P(k)\right). \quad (32)$$

Here $\theta \in \mathbb{R}$ is a parameter that we will choose later. Note that $V(\theta)$ can be expressed as a product of *commuting* 2-qubit gates

$$V(\theta) = \prod_{\{j,k\} \in E} \exp(i\theta P(j)P(k)). \quad (33)$$

Moreover, if the graph G has maximum degree Δ then we may efficiently compute an edge coloring with $\Delta + 1$ colors such that no two edges with the same color share a vertex. If we order the gates Eq. (33) in $\Delta + 1$ layers according to this edge coloring we obtain a depth $\Delta + 1$ quantum circuit that implements $V(\theta)$.

The following lemma describes the energy of the variational state $\phi(\theta)$. Below we write d_j for the degree of vertex $j \in V$. We say that an edge $\{i, j\} \in E$ is contained in T triangles iff there are vertices k_1, k_2, \dots, k_T such that $\{i, k_1\}, \{i, k_2\}, \dots, \{i, k_T\} \in E$ and $\{j, k_1\}, \{j, k_2\}, \dots, \{j, k_T\} \in E$. This is depicted in Fig. 1.

7:12 Beyond Product State Approximations

► **Lemma 8.** *Let $G = (V, E, w)$ be a graph and let $\phi(\theta)$ be the variational state defined in Eq. (32). If $\{i, j\} \in E_{\text{sat}}$ is a satisfied edge contained in exactly T triangles then*

$$\langle \phi(\theta) | 2h_{ij} | \phi(\theta) \rangle = 1 + \sin(2\theta) \cos^{d_i-1}(2\theta) + \sin(2\theta) \cos^{d_j-1}(2\theta) + \frac{1 + \cos^T(4\theta)}{2} \cos^{d_i+d_j-2-2T}(2\theta). \quad (34)$$

On the other hand, if $\{i, j\} \in E_{\text{unsat}}$ is an unsatisfied edge contained in exactly T triangles, then

$$\langle \phi(\theta) | 2h_{ij} | \phi(\theta) \rangle = 1 - \cos^{d_i+d_j-2-2T}(2\theta). \quad (35)$$

We defer the proof of the Lemma until the end of this section. Let us now see how Lemma 8 can be used to lower bound the energy $\langle \phi(\theta) | H_G | \phi(\theta) \rangle$ when G is a 3- or 4-regular graph. In fact, we will only need Eq. (34) for the proof below; Eq. (35) is included only for completeness.

Proof of Theorem 4. In a d -regular graph, each edge may be contained in $T \leq d-1$ triangles. Note that the energy Eq. (34) of a satisfied edge is lower bounded by the same expression with $T = 0$ since the last term is monotonically increasing with T . Thus all satisfied edges in a d -regular graph have energy lower bounded as

$$\langle \phi(\theta) | 2h_{ij} | \phi(\theta) \rangle \geq 1 + 2 \cos^{d-1}(2\theta) \sin(2\theta) + \cos^{2d-2}(2\theta) \quad (36)$$

An unsatisfied edge in a d -regular graph always contributes a nonnegative energy since the Hamiltonian terms h_{ij} are positive semidefinite and the weights w_{ij} are nonnegative.

Thus for a d -regular graph G we have

$$\langle \phi(\theta) | H_G | \phi(\theta) \rangle \geq \frac{F(\theta, d)}{2} \sum_{\{i, j\} \in E_{\text{sat}}} w_{ij} = \frac{F(\theta, d)}{2} \text{Cut}_G(z), \quad (37)$$

where

$$F(\theta, d) = 1 + 2 \cos^{d-1}(2\theta) \sin(2\theta) + \cos^{2d-2}(2\theta).$$

For a fixed d we may compute $\theta^*(d) = \arg\max_{\theta} F(\theta, d)$ which maximizes the right hand side. Also note that since z is the output of the Goemans-Williamson approximation algorithm, it satisfies Eq. (4) and therefore

$$\frac{\langle \phi(\theta) | H_G | \phi(\theta) \rangle}{\text{SDP}(G)} \geq G(d) \equiv (0.8785) \cdot \frac{F(\theta^*(d), d)}{2}. \quad (38)$$

Using a computer we find $G(3) = 1.047\dots$ and $G(4) = 1.001\dots$, which completes the proof. ◀

Proof of Lemma 8. We shall compute

$$\langle \phi | 2h_{ij} | \phi \rangle = 1 - \langle \phi | X_i X_j | \phi \rangle - \langle \phi | Y_i Y_j | \phi \rangle - \langle \phi | Z_i Z_j | \phi \rangle \quad (39)$$

(here and below we write $\phi \equiv \phi(\theta)$ for ease of notation).

We treat the two cases separately: satisfied edges $\{i, j\} \in E_{\text{sat}}$ (i.e., $z_i \neq z_j$) and unsatisfied edges $\{i, j\} \in E_{\text{unsat}}$ (i.e., $z_i = z_j$).

Satisfied edge

Without loss of generality, assume that $z_i = 0$ and $z_j = 1$ (else we perform the same calculation with i and j interchanged). Using the standard commutation relations between Pauli operators, and the fact that the set of operators $\{P(k)P(\ell)\}_{(k,\ell) \in E}$ mutually commute we get

$$\begin{aligned}
& \langle \phi | X_i X_j | \phi \rangle \\
&= \langle z | \left(\prod_{\{k,\ell\} \in E} \exp(-i\theta P(k)P(\ell)) \right) X_i X_j \left(\prod_{\{k,\ell\} \in E} \exp(i\theta P(k)P(\ell)) \right) | z \rangle \\
&= \langle z | \left(\prod_{k: \{i,k\} \in E} \exp(-2i\theta Y_i P(k)) \right) X_i X_j | z \rangle \\
&= \langle z | \prod_{k: \{i,k\} \in E} (\cos(2\theta) - i \sin(2\theta) Y_i P(k)) X_i X_j | z \rangle \\
&= -i \cos^{d_i-1}(2\theta) \sin(2\theta) \langle z | Y_i P(j) X_i X_j | z \rangle = -\cos^{d_i-1}(2\theta) \sin(2\theta).
\end{aligned}$$

Here, the second last equality follows since $\langle z_k | P(k) | z_k \rangle = 0$ for all $k \neq i, j$. A similar calculation shows that

$$\langle \phi | Y_i Y_j | \phi \rangle = -\cos^{d_j-1}(2\theta) \sin(2\theta).$$

Finally, for the last term in Eq. (39), we will need to take the triangles into account:

$$\langle \phi | Z_i Z_j | \phi \rangle \tag{40}$$

$$\begin{aligned}
&= \langle z | \left(\prod_{\{k,\ell\} \in E} \exp(-i\theta P(k)P(\ell)) \right) Z_i Z_j \left(\prod_{\{k,\ell\} \in E} \exp(i\theta P(k)P(\ell)) \right) | z \rangle \\
&= \langle z | \left(\prod_{k: k \neq i, \{k,j\} \in E} \exp(-2i\theta P(k)X_j) \prod_{k: k \neq j, \{i,k\} \in E} \exp(-2i\theta Y_i P(k)) \right) Z_i Z_j | z \rangle \\
&= -\langle z | \prod_{k: k \neq i, \{k,j\} \in E} (\cos(2\theta) - i \sin(2\theta) P(k)X_j) \prod_{k \neq j: \{i,k\} \in E} (\cos(2\theta) - i \sin(2\theta) Y_i P(k)) | z \rangle
\end{aligned} \tag{41}$$

$$= -\sum_{a=0}^{\lfloor \frac{T}{2} \rfloor} \binom{T}{2a} \sin^{4a}(2\theta) \cos^{d_i+d_j-2-4a}(2\theta) \tag{42}$$

Where the last equality is obtained by noting that a pair of triangles $\{i, j, k\}$ and $\{i, j, l\}$ will give a term $(-i \sin(2\theta))^4 Y_i P(k) Y_i P(l) X_j P(k) X_j P(l) = \sin^4(2\theta) \cdot I$ inside the expectation value in Eq. (41). The summation in Eq. (42) runs over all even cardinality subsets of triangles. Thus, when $z_i \neq z_j$,

$$\begin{aligned}
& \langle \phi(\theta) | 2h_{ij} | \phi(\theta) \rangle = \\
& 1 + \sin(2\theta) \cos^{d_i-1}(2\theta) + \sin(2\theta) \cos^{d_j-1}(2\theta) + \sum_{a=0}^{\lfloor \frac{T}{2} \rfloor} \binom{T}{2a} \sin^{4a}(2\theta) \cos^{d_i+d_j-2-4a}(2\theta) \\
&= 1 + \sin(2\theta) \cos^{d_i-1}(2\theta) + \sin(2\theta) \cos^{d_j-1}(2\theta) + \frac{1}{2} \left(\frac{1}{\cos^{2T}(2\theta)} + \frac{\cos^T(4\theta)}{\cos^{2T}(2\theta)} \right) \cos^{d_i+d_j-2}(2\theta) \\
&= 1 + \sin(2\theta) \cos^{d_i-1}(2\theta) + \sin(2\theta) \cos^{d_j-1}(2\theta) + \frac{1 + \cos^T(4\theta)}{2} \cos^{d_i+d_j-2-2T}(2\theta).
\end{aligned}$$

Here, the second equality follows from the binomial expansion of $\frac{1}{2} ((1+x)^T + (1-x)^T)$ for $x = \frac{\sin^2(2\theta)}{\cos^2(2\theta)}$.

Unsatisfied edge

Suppose $z_i = z_j = 0$. Then $P(i) = Y_i$ and $P(j) = Y_j$ and so

$$\langle \phi | Y_i Y_j | \phi \rangle = \langle z | Y_i Y_j | z \rangle = 0.$$

On the other hand,

$$\begin{aligned} \langle \phi | X_i X_j | \phi \rangle &= \langle z | \left(\prod_{\{k, \ell\} \in E} \exp(-i\theta P(k)P(\ell)) \right) X_i X_j \left(\prod_{\{k, \ell\} \in E} \exp(i\theta P(k)P(\ell)) \right) | z \rangle \\ &= \langle z | \left(\prod_{k: k \neq j, \{i, k\} \in E} \exp(-2i\theta Y_i P(k)) \right) \left(\prod_{k: k \neq i, \{k, j\} \in E} \exp(-2i\theta P(k)Y_j) \right) X_i X_j | z \rangle \\ &= \langle z | \prod_{k: k \neq j, \{i, k\} \in E} (\cos(2\theta) - i \sin(2\theta) Y_i P(k)) \prod_{k: k \neq i, \{k, j\} \in E} (\cos(2\theta) - i \sin(2\theta) P(k) Y_j) X_i X_j | z \rangle \\ &= \sum_{a=1}^{\lfloor \frac{T+1}{2} \rfloor} \binom{T}{2a-1} \sin^{4a-2}(2\theta) \cos^{d_i+d_j-4a}(2\theta). \end{aligned} \quad (43)$$

The summation in Eq. (43) runs over all odd cardinality subsets of triangles. Finally,

$$\begin{aligned} \langle \phi | Z_i Z_j | \phi \rangle &= \langle z | \prod_{k: k \neq j, \{i, k\} \in E} (\cos(2\theta) - i \sin(2\theta) Y_i P(k)) \prod_{k: k \neq i, \{k, j\} \in E} (\cos(2\theta) - i \sin(2\theta) P(k) Y_j) Z_i Z_j | z \rangle \\ &= \sum_{a=0}^{\lfloor \frac{T}{2} \rfloor} \binom{T}{2a} \sin^{4a}(2\theta) \cos^{d_i+d_j-2-4a}(2\theta) \end{aligned}$$

Where again, we have used the fact that any pair of triangles will result in an identity term.

If $z_i = z_j = 1$, similar calculations show that the contributions from $\langle \phi | Y_i Y_j | \phi \rangle$ and $\langle \phi | X_i X_j | \phi \rangle$ are interchanged, but that their sum is unchanged. So for any unsatisfied edge, we have added lines to this equation:

$$\begin{aligned} \langle \phi | 2h_{ij} | \phi \rangle &= 1 - \sum_{a=1}^{\lfloor \frac{T+1}{2} \rfloor} \binom{T}{2a-1} \sin^{4a-2}(2\theta) \cos^{d_i+d_j-4a}(2\theta) \\ &\quad - \sum_{a=0}^{\lfloor \frac{T}{2} \rfloor} \binom{T}{2a} \sin^{4a}(2\theta) \cos^{d_i+d_j-2-4a}(2\theta) \\ &= 1 - \cos^{d_i+d_j-2}(2\theta) \left(\sum_{b=0}^T \binom{T}{b} \frac{\sin^{2b}(2\theta)}{\cos^{2b}(2\theta)} \right) = 1 - \cos^{d_i+d_j-2-2T}(2\theta), \end{aligned}$$

where we used the binomial expansion of $(1+x)^T$ for $x = \frac{\sin^2(2\theta)}{\cos^2(2\theta)}$. ◀

References

- 1 Hallgren, Lee, Parekh 2019. Announced in a contributed talk at QIP 2020 in Shenzhen, China.
- 2 PW Anderson. Limits on the energy of the antiferromagnetic ground state. *Physical Review*, 83(6):1260, 1951.
- 3 Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM (JACM)*, 45(3):501–555, 1998.

- 4 Nikhil Bansal, Sergey Bravyi, and Barbara M Terhal. Classical approximation schemes for the ground-state energy of quantum and classical ising spin hamiltonians on planar graphs. *arXiv preprint arXiv:0705.1115*, 2007.
- 5 Fernando GSL Brandao and Aram W Harrow. Product-state approximations to quantum states. *Communications in Mathematical Physics*, 342(1):47–80, 2016.
- 6 Sergey Bravyi, David Gosset, Robert König, and Kristan Temme. Approximation algorithms for quantum many-body problems. *Journal of Mathematical Physics*, 60(3):032203, 2019.
- 7 Jop Briët, Fernando Mário de Oliveira Filho, and Frank Vallentin. The positive semidefinite grothendieck problem with rank constraint. In *International Colloquium on Automata, Languages, and Programming*, pages 31–42. Springer, 2010.
- 8 Moses Charikar and Anthony Wirth. Maximizing quadratic programs: Extending grothendieck’s inequality. In *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 54–60. IEEE, 2004.
- 9 Edward Farhi, Jeffrey Goldstone, and Sam Gutmann. A quantum approximate optimization algorithm. *arXiv preprint arXiv:1411.4028*, 2014.
- 10 Sevag Gharibian and Julia Kempe. Approximation algorithms for qma-complete problems. *SIAM Journal on Computing*, 41(4):1028–1050, 2012.
- 11 Sevag Gharibian and Yi-Kai Liu. Approximation algorithms for the quantum heisenberg model. *Private communication*, 2020.
- 12 Sevag Gharibian and Ojas Parekh. Almost optimal classical approximation algorithms for a quantum generalization of max-cut. *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques (APPROX/RANDOM 2019)*, 145:31, 2019.
- 13 Michel X Goemans and David P Williamson. Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming. *Journal of the ACM (JACM)*, 42(6):1115–1145, 1995.
- 14 Aram W Harrow and Ashley Montanaro. Extremal eigenvalues of local hamiltonians. *Quantum*, 1:6, 2017.
- 15 Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem. In *International Conference on Foundations of Software Technology and Theoretical Computer Science*, pages 372–383. Springer, 2004.
- 16 Alexei Yu Kitaev, Alexander Shen, and Mikhail N Vyalyi. *Classical and quantum computation*. Number 47. American Mathematical Soc., 2002.
- 17 Elliott Lieb and Daniel Mattis. Ordering energy levels of interacting spin systems. *Journal of Mathematical Physics*, 3(4):749–751, 1962.
- 18 Russell Merris. A note on laplacian graph eigenvalues. *Linear algebra and its applications*, 285(1-3):33–35, 1998.
- 19 Stephen Piddock and Ashley Montanaro. The complexity of antiferromagnetic interactions and 2d lattices. *arXiv preprint arXiv:1506.04014*, 2015.

Simpler Proofs of Quantumness

Zvika Brakerski

Weizmann Institute of Science, Rehovot, Israel
zvika.brakerski@weizmann.ac.il

Venkata Koppula

Weizmann Institute of Science, Rehovot, Israel
venkata.koppula@weizmann.ac.il

Umesh Vazirani

University of California, Berkeley, CA, USA
vazirani@cs.berkeley.edu

Thomas Vidick

California Institute of Technology, Pasadena, CA, USA
vidick@cms.caltech.edu

Abstract

A proof of quantumness is a method for provably demonstrating (to a classical verifier) that a quantum device can perform computational tasks that a classical device with comparable resources cannot. Providing a proof of quantumness is the first step towards constructing a useful quantum computer.

There are currently three approaches for exhibiting proofs of quantumness: (i) Inverting a classically-hard one-way function (e.g. using Shor's algorithm). This seems technologically out of reach. (ii) Sampling from a classically-hard-to-sample distribution (e.g. BosonSampling). This may be within reach of near-term experiments, but for all such tasks known verification requires exponential time. (iii) Interactive protocols based on cryptographic assumptions. The use of a trapdoor scheme allows for efficient verification, and implementation seems to require much less resources than (i), yet still more than (ii).

In this work we propose a significant simplification to approach (iii) by employing the random oracle heuristic. (We note that we *do not* apply the Fiat-Shamir paradigm.)

We give a two-message (challenge-response) proof of quantumness based on any trapdoor claw-free function. In contrast to earlier proposals we do not need an adaptive hard-core bit property. This allows the use of smaller security parameters and more diverse computational assumptions (such as Ring Learning with Errors), significantly reducing the quantum computational effort required for a successful demonstration.

2012 ACM Subject Classification Theory of computation → Cryptographic protocols; Theory of computation → Quantum complexity theory

Keywords and phrases Proof of Quantumness, Random Oracle, Learning with Errors

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.8

Funding *Zvika Brakerski*: Supported by the Binational Science Foundation (Grant No. 2016726), and by the European Union Horizon 2020 Research and Innovation Program via ERC Project REACT (Grant 756482) and via Project PROMETHEUS (Grant 780701).

Venkata Koppula: Supported by the Binational Science Foundation (Grant No. 2016726), and by the European Union Horizon 2020 Research and Innovation Program via ERC Project REACT (Grant 756482) and via Project PROMETHEUS (Grant 780701).

Umesh Vazirani: Supported in part by ARO Grant W911NF-12-1-0541, NSF Grant CCF1410022, a Vannevar Bush faculty fellowship, and the Miller Institute at U.C. Berkeley through a Miller Professorship.

Thomas Vidick: Supported by NSF CAREER Grant CCF-1553477, AFOSR YIP award number FA9550-16-1-0495, a CIFAR Azrieli Global Scholar award, MURI Grant FA9550-18-1-0161, and the IQIM, an NSF Physics Frontiers Center (NSF Grant PHY-1125565).



© Zvika Brakerski, Venkata Koppula, Umesh Vazirani, and Thomas Vidick;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 8; pp. 8:1–8:14



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Quantum computing holds a promise of a qualitative leap in our ability to perform important computational tasks. These tasks include simulation of chemical and physical systems at the quantum level, generating true randomness, algorithmic tasks such as factoring large numbers, and more. However, constructing a quantum computer with capabilities beyond those of existing classical computers is technologically challenging. Indeed, whether it is possible or not remains to be proven; such a “proof” is the focus of the ongoing race to construct a useful quantum device, with records for device size and functionality set at an increasing rate by the likes of Google, IBM, and the increasing number of startups heavily invested in this race. This notion, known as “proof of quantumness”,¹ is generally viewed as a major milestone towards unlocking the powers of quantum computing. We can classify existing approaches towards proof of quantumness into three families:

1. There are tasks that are generally believed to be classically intractable, and for which quantum algorithms are known; most notably the factoring and discrete logarithm problems [16]. Constructing a quantum computer that can factor beyond our classical capabilities would constitute a valid proof of quantumness. Alas, in order to implement the factoring algorithm on relevant input sizes one requires fault-tolerant quantum computation, which seems technologically out of reach (see e.g. [8] for recent and highly optimized estimates ranging in the millions of qubits).
2. A different approach, introduced independently by Bremner, Jozsa and Shepperd [4] and by Aaronson and Arkhipov [1], is to use a quantum device to sample from distributions that are presumed to be hard to sample from classically. The intractability of classically achieving the task has not stood the same test of time as more established problems such as e.g. factoring, but can nonetheless be based on reasonable complexity-theoretic conjectures, at least for the problem of exact sampling. While quantum devices that can sample from these distributions appear to be “right around the corner”, the real challenges are in (i) showing hardness of approximate sampling – the quantum device will never be perfect – and (ii) the classical verification: verification for these methods generally requires investing exponential classical computational resources, and can thus only be performed for fairly small input lengths.
3. A new approach was recently proposed in [3]. They propose to use *post-quantum cryptography*, namely to rely on cryptographic assumptions that cannot be broken even by the quantum device. Rather than verifying that the quantum device has the ability to break the assumption, cryptography is used to compel the device to generate a quantum superposition in a way that can be efficiently verified using a secret key. This method is inherently interactive, unlike the previous two, and requires at least four rounds of communication. As a cryptographic building block it uses trapdoor claw-free function families (recall that claw-freeness was originally introduced in the context of digital signatures and constructed based on factoring [9]). In addition to claw freeness, the [3] approach also requires an additional adaptive hardcore bit property which appears to be hard to realize and is currently only known to be achievable based on the Learning with Errors (LWE) assumption [15].

The third approach is compelling in its ability to verify quantumness even of large quantum devices efficiently, but it still requires a large number of quantum operations.

¹ The term “quantum supremacy” is also used in the literature.

Furthermore, the interactive nature of the protocol requires the quantum device to retain a superposition while waiting for the verifier’s second message (a single random bit).

In this work we simplify the [3] approach and allow for it to be based on a more diverse set of computational assumptions. This marks a step towards a protocol that can be realistically implemented on an actual quantum device, and can be efficiently verified on a classical computer.

Our Results

We propose to use the *random oracle heuristic* as a tool to reduce the round complexity of the proof of quantumness protocol from [3], making it into a simple one-round message-response protocol. We note that it is unlikely that a similar result can be achieved *in the standard model* without introducing an additional hardness assumption. The reason is that a single-round message-response protocol in the standard model (i.e. without oracles) immediately implies that quantum samplers cannot be efficiently de-quantized (otherwise the protocol will have no soundness). Such a result therefore implies a (weak) separation between the BQP and BPP models. However, the LWE assumption does not appear to imply such a separation, and the current state of the art suggests that it is equally intractable in the quantum and classical settings.²

We show that using the random oracle heuristic, it is possible to implement the protocol in a single round while at the same time eliminating the need for an adaptive hard-core bit property, and thus relying on any family of claw-free functions. In particular, we propose a construction of trapdoor claw free functions which is analogous to that of [3] but relies on the Ring-LWE assumption [10, 11]. Ring-LWE based primitives are often regarded as more efficient than their LWE-based counterparts since they involve arithmetic over polynomial rings, which can be done more efficiently than over arbitrary linear spaces. Despite the similarity between LWE and Ring-LWE, proving an adaptive hard-core theorem for the latter appears to be a challenging task. This is since the LWE-based construction uses a so-called lossiness argument that is not known to be replicable in the Ring-LWE setting. We note that we can also instantiate our method using “pre-quantum” cryptography since soundness should hold only with respect to classical adversaries. Using a back-of-the-envelope calculation we estimate that it is possible to execute our protocol using superpositions over $\sim 8\lambda \log^2 \lambda$ qubits, for security parameter λ and the adversary would have advantage negligible in λ .

While we allow the use of trapdoor claw-free families based on arbitrary assumptions, which should allow for better security/efficiency trade-offs, our protocol still requires the quantum device to evaluate the random oracle on a quantum superposition, which could potentially create an additional burden. We point out that current and future heuristic instantiations of the random oracle model using explicit hash functions are assumed to enjoy efficient quantum implementation. Specifically, in evaluating the *post-quantum* security level of cryptographic constructions (e.g. for the NIST competition [14]), security is evaluated in the Quantum Random Oracle model where adversaries are assumed to evaluate hash functions on superpositions as efficiently as they do classically. Granted, this is just a model for an adversary, but it is customary to try to be as realistic as possible and not over-estimate the power of the adversary. We therefore consider the evaluation of the random hash function as a relatively lower-order addition to the cost of performing the quantumness test.

² This insight is due to a discussion with Omer Paneth.

Lastly, we compare our method to the most straightforward way to employ a random oracle for the purpose of round reduction, the Fiat-Shamir transform [7]. The basic protocol of [3] contains 4 messages, where the third message is simply a random bit. One can therefore do parallel repetition of the protocol (though the soundness of this transformation needs to be shown),³ and apply Fiat-Shamir to compress it into challenge-response form. Furthermore, for proofs of quantumness soundness is only required to hold against a classical adversary, so the standard security reduction for Fiat-Shamir should hold. This approach only requires to apply the random oracle to a classical input. However, it still requires the adaptive hard-core bit property and is therefore restricted to the LWE assumption. We believe that our protocol, being of a somewhat modified form compared to prior works, may be useful for future applications.

Our Technique

At a high level, a family of trapdoor claw free functions allows to sample a function $f : \{0, 1\} \times \{0, 1\}^n \rightarrow \{0, 1\}^n$ together with a trapdoor. The function has two branches $f(0, \cdot), f(1, \cdot)$ which are both injective, i.e. permutations (this is a simplified description, actual protocols use a relaxed “noisy” notion). It is guaranteed that it is computationally intractable to find a collision (“claw”) x_0, x_1 s.t. $f(0, x_0) = f(1, x_1)$, however given the trapdoor it is possible to find for all y the preimages x_0, x_1 s.t. $f(0, x_0) = f(1, x_1) = y$.

The [3] protocol sends a description of f to the quantum device, asks it to apply f on a uniform superposition of inputs and measure the image register, call the value obtained y . The quantum device is then left with a uniform superposition over the two preimages of y : $(0, x_0)$ and $(1, x_1)$. The value y is sent to the verifier who challenges the quantum device to measure the remaining superposition on inputs in either the standard or Hadamard basis. A classical adversary that can answer each query independently must also be able to answer both at the same time, which is ruled out by the adaptive hard core property.

We propose to enable the quantum device to generate a superposition over $(0, x_0, H(0, x_0))$ and $(1, x_1, H(1, x_1))$, where H is a one-bit hash function modeled as a random oracle. This can be done in a straightforward manner, similar to the previous method. The device is then asked to measure the resulting state in the Hadamard basis (always), and send the outcomes obtained to the verifier.⁴ Since the device makes a single measurement, there is no need for a challenge from the verifier, which effectively collapses the protocol to two messages. A quick calculation shows that the verifier receives a bit m and vector d s.t. in the case of a honest behavior the equation $m = d \cdot (x_0 \oplus x_1) \oplus H(0, x_0) \oplus H(1, x_1)$ holds. Finally, the verifier uses the trapdoor to recover x_0, x_1 from y and checks that the equation is satisfied. The crux of the security proof is that a classical adversary cannot query the oracle at both $(0, x_0)$ and $(1, x_1)$, otherwise it would have been able to find a claw and break the cryptographic assumption. Therefore at least one value out of $H(0, x_0)$ and $H(1, x_1)$ remains random, and thus the adversary cannot compute m, d that adhere to the required equation with probability greater than $1/2$. The proof thus follows from a simple extraction-style argument. In our main protocol, we use parallel repetition to argue that no prover can succeed with non-negligible probability.

³ Very recently, two concurrent works by Alagic et al. [2] and Chia et al. [6] showed that parallel repetition of Mahadev’s protocol indeed achieves negligible soundness error.

⁴ In fact we use a slight variant of this protocol, since measuring the H part in Hadamard basis has probability $1/2$ of erasing the information on that bit. Instead we append the H values directly to the phase. This is immaterial for the purpose of this exposition.

Discussion on the “Random Oracle” Heuristic

As discussed above, the Fiat-Shamir heuristic can be used for the quantum supremacy protocol of Brakerski et al. [3]. However, this would mean that the resulting scheme would require stronger assumptions (in particular, it would require noisy TCFs with the adaptive hardcore bit property). Secondly, starting with the work of Canetti et al. [5], many works have shown uninstantiability of the random oracle. These works show certain cryptographic primitives which are secure in the random oracle model, but are broken when instantiated by any concrete hash function. However, these constructions are very contrived, and in particular, do not apply to our protocol.

Efficiency of our Protocol, and Comparison to Previous Approaches

We would like to emphasize that at the current level of maturity of quantum technology, any estimate of “practical advantage” would be educated guesswork at best. The technology for any option is far from being available and it is hard to predict the direction that technology will take, and as a consequence the practical cost of implementing certain operations.

This state of affairs, we believe, highlights the importance of developing multiple approaches to tasks such as proof of quantumness. This way, an assortment of solutions will be ready to accommodate the different directions that technology may lead.

A second point that we wish to highlight before getting into technical calculations, is that our approach allows to use *any* family of trapdoor claw free permutations (and as we point out, for proofs of quantumness even “pre-quantum” candidates will suffice, e.g. if a candidate can be devised based on DDH in EC groups). This means that our back of the envelope calculation only refers to one specific way of using our scheme. Currently, we do not know any candidates for trapdoor claw free permutations based on such “pre-quantum” assumptions.

Our protocol can be executed using a quasi-linear number of qubits and, with the proper choice of candidate for the hash function, has quasi-linear computational complexity.

Comparison with [3]: Since we do not require the hardcore bit property, our input dimension n is smaller by a factor of at least $60 \log(\lambda)$. This follows due to Lemma 4.2 in [3]. Also, note that the parameter q must also grow, hence the overall number of qubits required to implement the protocol in [3] is $O(\lambda \log^3(\lambda))$, at least $100 \log(\lambda)$ times more. Secondly, since [3] is a four-round protocol, the prover must maintain its quantum state until it receives a challenge from the verifier.

Comparison to discrete log via Shor’s algorithm: Let n denote the number of bits required for representing the group elements. The current estimates for the number of qubits required for discrete log are $3n$, while the number of quantum gates required is $0.3n^3$ (see [8]). Similar to Shor’s algorithm for factoring/discrete log, our protocol is also a non-interactive one (that is, the verifier sends a challenge, and the prover responds with an answer).

Open Problems

Our work suggests a number of open problems in the context of utilizing random oracles in the regime of classical verification of quantum computation. Most desirably, whether it is possible to use the random oracle in order to eliminate the need for other assumptions, or at least the need for a trapdoor. Obtaining a publicly verifiable protocol is a highly desirable goal. We can also wonder whether our protocol can be used for the purposes of certifying randomness or verifying quantum computation. In the plain model, the adaptation of the proof of quantumness for these purposes was far from trivial and yet the protocol itself is almost identical. Improving the state of the art in certifying randomness and in verifiability using random oracles (or using other methods) is also an interesting open problem.

2 Preliminaries

2.1 Notations

For an integer n we write $[n]$ for the set $\{1, \dots, n\}$. For any finite set X , let $x \leftarrow \mathcal{X}$ denote a uniformly random element drawn from X . Similarly, for any distribution \mathcal{D} , let $x \leftarrow \mathcal{D}$ denote a sample from \mathcal{D} . For an element $x \in X$ we write $\text{BitDecomp}(x)$ for an arbitrarily chosen but canonical (depending only on the implicit set X) binary representation of x . For any density function f on domain X , let $\text{SUPP}(f)$ denote the support of f ; that is $\text{SUPP}(f) = \{x \in X : f(x) > 0\}$.

For density functions f_1, f_2 over the same finite domain X , the Hellinger distance between f_1 and f_2 is

$$H^2(f_1, f_2) = 1 - \sum_{x \in X} \sqrt{f_1(x)f_2(x)}.$$

The total variation distance between f_1 and f_2 is

$$\|f_1 - f_2\|_{\text{TV}} = \frac{1}{2} \sum_{x \in X} |f_1(x) - f_2(x)| \leq \sqrt{2H^2(f_1, f_2)}.$$

The following lemma relates the Hellinger distance and the trace distance of superpositions.

► **Lemma 1.** *Let X be a finite set and f_1, f_2 two density functions on X . Let*

$$|\psi_1\rangle = \sum_{x \in X} \sqrt{f_1(x)} |x\rangle, \text{ and } |\psi_2\rangle = \sum_{x \in X} \sqrt{f_2(x)} |x\rangle.$$

Then

$$\| |\psi_1\rangle - |\psi_2\rangle \|_{\text{tr}} \leq \sqrt{1 - (1 - H^2(f_1, f_2))^2}.$$

2.2 Ideal Lattices

In this section, we present some background on ideal lattices, the truncated discrete Gaussian distribution and the Ring Learning with Errors problem. For a positive integer B , modulus q , and dimension n , the truncated discrete Gaussian distribution is a distribution with support $\{x \in \mathbb{Z}_q^n : \|x\| \leq B\sqrt{n}\}$ defined as follows:

$$D_{\mathbb{Z}_q^n, B}(x) = \frac{\exp(-\pi\|x\|^2/B^2)}{\sum_{z \in \mathbb{Z}_q^n, \|z\| \leq B\sqrt{n}} \exp(-\pi\|z\|^2/B^2)}.$$

The Ring Learning with Errors (RLWE) assumption[11] is parameterized by a ring R , modulus $q \in \mathbb{N}$ and a noise distribution χ . Informally, the assumption states that given many samples of the form $(a, a \cdot s + e)$ where s is fixed for all samples, a is chosen uniformly at random and e is chosen from the error distribution χ for each sample, it is hard to compute s . The formal definition is given below. Here, we restrict ourselves to a special family of *cyclotomic rings*.

► **Assumption 1.** *Let n be a power of two, $f_n(X) = X^n + 1$ an irreducible polynomial over $\mathbb{Q}[X]$ and $R_n = \mathbb{Z}[X]/(f_n(X))$. Let $q = \{q_n\}_{n \in \mathbb{N}}$ be a family of moduli, $R_{n, q_n} = R_n/q_n R_n = \mathbb{Z}_{q_n}[X]/(f_n(X))$ the quotient space, and $\chi = \{\chi_n\}_{n \in \mathbb{N}}$ a family of error distributions, where χ_n is a distribution over R_{n, q_n} . For any secret s in R_{n, q_n} , let \mathcal{O}_s denote the oracle that, on*

each query, chooses $a \leftarrow R_{n,q_n}$, $e \leftarrow \chi_n$ and outputs $(a, a \cdot s + e \bmod q_n)$. The Ring Learning with Errors assumption $\text{RLWE}_{R,q,\chi}$, parameterized by the family of rings $\{R_n\}_{n=2^k, k \in \mathbb{N}}$, moduli family q and distribution family χ , states that for any PPT adversary \mathcal{A} , there exists a negligible function $\text{negl}(\cdot)$ such that for all security parameters $n = 2^k, k \in \mathbb{N}$,

$$\Pr \left[s \leftarrow \mathcal{A}^{O_s(\cdot)}(1^n) : s \leftarrow R_{n,q_n} \right] \leq \text{negl}(n).$$

Given many samples $\{a_i, a_i \cdot s + e_i\}_i$, one can efficiently find s using a *trapdoor* for the public elements $\{a_i\}_i$. There exists a sampling algorithm that can sample $\{a_i\}_i$ together with a trapdoor τ , and an inversion algorithm that uses τ to extract s from the set of evaluations $\{a_i \cdot s + e_i\}_i$. Without the trapdoor, the public elements $\{a_i\}_i$ look uniformly random.

► **Theorem 2** (Theorem 5.1 of [13] in the Ring setting). *Let n, m, q be such that n is a power of 2, $m = \Omega(\log q)$. There is an efficient randomized algorithm GENTRAP that takes as input $(1^n, 1^m, q)$, and returns $\mathbf{a} = (a_i)_i \in R_{n,q}^m$ and a trapdoor τ such that the distribution of \mathbf{a} is negligibly (in n) close to the uniform distribution over $R_{n,q}^m$. Moreover, there is an efficient algorithm INVERT and a universal constant C_T such that the following holds with overwhelming probability over the choice of $(\mathbf{a}, \tau) \leftarrow \text{GENTRAP}(1^n, 1^m, q)$:*

$$\text{for all } s \in R_{n,q}, \mathbf{e} \text{ such that } \|\mathbf{e}\| \leq \frac{q}{C_T \sqrt{n \log q}}, \text{INVERT}(\mathbf{a}, \tau, \mathbf{a} \cdot s + \mathbf{e}) = s.$$

2.3 Noisy Trapdoor Claw-Free Hash Functions

In this section we introduce the notion of noisy trapdoor claw-free functions (NTCFs). Let \mathcal{X}, \mathcal{Y} be finite sets and \mathcal{K} a set of keys. For each $k \in \mathcal{K}$ there should exist two (efficiently computable) injective functions $f_{k,0}, f_{k,1}$ that map \mathcal{X} to \mathcal{Y} , together with a trapdoor t_k that allows efficient inversion from $(b, y) \in \{0, 1\} \times \mathcal{Y}$ to $f_{k,b}^{-1}(y) \in \mathcal{X} \cup \{\perp\}$. For security, we require that for a randomly chosen key k , no polynomial time adversary can efficiently compute $x_0, x_1 \in \mathcal{X}$ such that $f_{k,0}(x_0) = f_{k,1}(x_1)$ (such a pair (x_0, x_1) is called a *claw*).

Unfortunately, we do not know how to construct such “clean” trapdoor claw-free functions. Hence, as in previous works [3, 12], we will use “noisy” version of the above notion. For each $k \in \mathcal{K}$, there exist two functions $f_{k,0}, f_{k,1}$ that map \mathcal{X} to a distribution over \mathcal{Y} .

The following definition is taken directly from [3].

► **Definition 3** (NTCF family). *Let λ be a security parameter. Let \mathcal{X} and \mathcal{Y} be finite sets. Let $\mathcal{K}_{\mathcal{F}}$ be a finite set of keys. A family of functions*

$$\mathcal{F} = \{f_{k,b} : \mathcal{X} \rightarrow \mathcal{D}_{\mathcal{Y}}\}_{k \in \mathcal{K}_{\mathcal{F}}, b \in \{0,1\}}$$

is called a noisy trapdoor claw free (NTCF) family if the following conditions hold:

1. **Efficient Function Generation.** *There exists an efficient probabilistic algorithm $\text{GEN}_{\mathcal{F}}$ which generates a key $k \in \mathcal{K}_{\mathcal{F}}$ together with a trapdoor t_k :*

$$(k, t_k) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda).$$

2. **Trapdoor Injective Pair.**

- a. *Trapdoor:* *There exists an efficient deterministic algorithm $\text{INV}_{\mathcal{F}}$ such that with overwhelming probability over the choice of $(k, t_k) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda)$, the following holds:*

$$\text{for all } b \in \{0, 1\}, x \in \mathcal{X} \text{ and } y \in \text{SUPP}(f_{k,b}(x)), \text{INV}_{\mathcal{F}}(t_k, b, y) = x.$$

- b. *Injective pair:* For all keys $k \in \mathcal{K}_{\mathcal{F}}$, there exists a perfect matching $\mathcal{R}_k \subseteq \mathcal{X} \times \mathcal{X}$ such that $f_{k,0}(x_0) = f_{k,1}(x_1)$ if and only if $(x_0, x_1) \in \mathcal{R}_k$.
- 3. **Efficient Range Superposition.** For all keys $k \in \mathcal{K}_{\mathcal{F}}$ and $b \in \{0, 1\}$ there exists a function $f'_{k,b} : \mathcal{X} \rightarrow \mathcal{D}_{\mathcal{Y}}$ such that the following hold.
 - a. For all $(x_0, x_1) \in \mathcal{R}_k$ and $y \in \text{SUPP}(f'_{k,b}(x_b))$, $\text{INV}_{\mathcal{F}}(t_k, b, y) = x_b$ and $\text{INV}_{\mathcal{F}}(t_k, b \oplus 1, y) = x_{b \oplus 1}$.
 - b. There exists an efficient deterministic procedure $\text{CHK}_{\mathcal{F}}$ that, on input $k, b \in \{0, 1\}$, $x \in \mathcal{X}$ and $y \in \mathcal{Y}$, returns 1 if $y \in \text{SUPP}(f'_{k,b}(x))$ and 0 otherwise. Note that $\text{CHK}_{\mathcal{F}}$ is not provided the trapdoor t_k .
 - c. For every k and $b \in \{0, 1\}$,

$$E_{x \leftarrow \mathcal{U}\mathcal{X}} [H^2(f_{k,b}(x), f'_{k,b}(x))] \leq 1/50.^5$$

Here H^2 is the Hellinger distance. Moreover, there exists an efficient procedure $\text{SAMP}_{\mathcal{F}}$ that on input k and $b \in \{0, 1\}$ prepares the state

$$\frac{1}{\sqrt{|\mathcal{X}|}} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}} \sqrt{(f'_{k,b}(x))(y)} |x\rangle |y\rangle. \quad (1)$$

- 4. **Claw-Free Property.** For any PPT adversary \mathcal{A} , there exists a negligible function $\text{negl}(\cdot)$ such that the following holds:

$$\Pr [(x_0, x_1) \in \mathcal{R}_k : (k, t_k) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda), (x_0, x_1) \leftarrow \mathcal{A}(k)] \leq \text{negl}(\lambda)$$

3 Proof of Quantumness Protocol

We will now present our protocol. Throughout the protocol, we will ignore dependence on the security parameter when clear from context. Let \mathcal{F} be a NTCF family with domain \mathcal{X} , range \mathcal{Y} described by the algorithms $\text{GEN}_{\mathcal{F}}, \text{INV}_{\mathcal{F}}, \text{CHK}_{\mathcal{F}}, \text{SAMP}_{\mathcal{F}}$. Let w denote the length of bit decomposition of elements of \mathcal{X} . Finally, let H be a hash function that maps \mathcal{X} to $\{0, 1\}$.

Proof of Quantumness Protocol

The protocol is parameterized by a hash function $H : \{0, 1\}^n \rightarrow \{0, 1\}$ (which will be modeled as a random oracle in the security proof).

1. The verifier generates $(k, \kappa) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda)$ and sends k to the prover.
 2. The prover sends λ tuples $\{(y_i, m_i, d_i)\}_{i \in [\lambda]}$. The verifier initializes **count** = 0 and performs the following checks:
 - a. It checks that all values in $\{y_i\}_i$ are distinct.
 - b. It computes $x_{i,b} = \text{INV}_{\mathcal{F}}(\kappa, b, y_i)$ for each $i \in [\lambda]$, $b \in \{0, 1\}$. Next, it checks if $m_i = d_i^T \cdot (\text{BitDecomp}(x_{i,0}) + \text{BitDecomp}(x_{i,1})) + H(x_{i,0}) + H(x_{i,1})$. If this check passes, it increments the value of **count** by 1.
 3. If **count** > 0.75λ , the verifier outputs 1, else it outputs \perp .
-

■ **Figure 1** Protocol for Proof of Quantumness.

► **Theorem 4.** *Let \mathcal{F} be a family of NTCF functions satisfying Definition 3. Then Protocol 1 satisfies the following properties:*

- *Completeness:* *There exists a quantum polynomial-time prover \mathcal{P} and a negligible function $\text{negl}(\cdot)$ such that for all $\lambda \in \mathbb{N}$ and hash functions H , \mathcal{P} succeeds in the protocol with probability at least $1 - \text{negl}(\lambda)$.*
- *Proof of Quantumness:* *For any PPT (classical) adversary \mathcal{A} , there exists a negligible function $\text{negl}(\cdot)$ such that for all $\lambda \in \mathbb{N}$, \mathcal{A} succeeds in the protocol with probability at most $\text{negl}(\lambda)$ where H is modeled as a random oracle.*

3.1 Completeness

In this section, we show that the honest (quantum) prover is accepted by the verifier.

The honest prover receives NTCF key k . It does the following:

1. It starts with λ copies of the state $|0\rangle|0\rangle|0\rangle|-\rangle$. For each $i \in [\lambda]$, let $|\psi_i\rangle = |0\rangle|0\rangle|0\rangle|-\rangle$. It then applies $\text{SAMP}_{\mathcal{F}}$ to the first three registers of $|\psi_i\rangle$ for each i , resulting in the state $|\psi_i^{(1)}\rangle$, where

$$|\psi_i^{(1)}\rangle = \left(\frac{1}{\sqrt{2|\mathcal{X}|}} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}, b \in \{0,1\}} \sqrt{(f'_{k,b}(x))(y)} |b\rangle |x\rangle |y\rangle \right) |-\rangle. \quad (2)$$

This quantum state is at distance at most 0.2 from the following quantum state:

$$|\psi_i^{(1)}\rangle = \left(\frac{1}{\sqrt{2|\mathcal{X}|}} \sum_{x \in \mathcal{X}, y \in \mathcal{Y}, b \in \{0,1\}} \sqrt{(f_{k,b}(x))(y)} |b\rangle |x\rangle |y\rangle \right) |-\rangle. \quad (3)$$

2. Next, it measures the third register, obtaining measurement $y \in \mathcal{Y}$. Let $x_0, x_1 \in \mathcal{X}$ be the unique elements such that y is in the support of $f_{k,b}(x_b)$. Applying this operation to the state in (3), the resulting state (ignoring the measured register) is

$$|\psi_i^{(2)}\rangle = \left(\frac{1}{\sqrt{2}} (|0\rangle |x_0\rangle + |1\rangle |x_1\rangle) \right) |-\rangle. \quad (4)$$

3. Let U_H be a unitary that maps $|a\rangle|b\rangle$ to $|a\rangle|b + H(a)\rangle$. The prover applies U_H to the second and third register. On applying this operation to the state in (4), the new state is

$$|\psi_i^{(3)}\rangle = \frac{1}{2} \left(\sum_{b,b'} (-1)^{b'} |b\rangle |x_b\rangle |b' + H(x_b)\rangle \right). \quad (5)$$

4. The prover then evaluates the function BitDecomp on the second register. Applying this to (5), the resulting state is

$$|\psi_i^{(4)}\rangle = \frac{1}{2} \left(\sum_{b,b'} (-1)^{b'} |b\rangle |\text{BitDecomp}(x_b)\rangle |b' + H(x_b)\rangle \right). \quad (6)$$

5. Finally, the prover applies the Hadamard operator to all registers. On applying this to (6), this produces the state (where $h_b = H(x_b)$ and $\bar{x}_b = \text{BitDecomp}(x_b)$)

$$\begin{aligned} |\psi_i^{(5)}\rangle &= \frac{1}{\sqrt{2^{w+4}}} \sum_{b,b' \in \{0,1\}} \sum_{\substack{m,m' \in \{0,1\}, \\ d \in \{0,1\}^w}} (-1)^{m \cdot b + d^T \cdot \bar{x}_b + m' \cdot b' + m' \cdot h_b + b'} |m\rangle |d\rangle |m'\rangle \\ &= \frac{1}{\sqrt{2^{w+2}}} \sum_{m \in \{0,1\}, d \in \{0,1\}^w} |m\rangle |d\rangle |1\rangle \left((-1)^{d^T \cdot \bar{x}_0 + h_0} + (-1)^{m + d^T \cdot x_1 + h_1} \right) \end{aligned} \quad (7)$$

Upon measurement of the state in (7), the output tuple $(m, d, 1)$ satisfies $m = d^T \cdot (\bar{x}_0 + \bar{x}_1) + h_0 + h_1$ (with probability 1). As a result, applying the above operations to $|\psi_i'^{(1)}\rangle$ results in a tuple (y, m, d) that is accepted with probability at least 0.8. Using a Chernoff bound it is straightforward to argue that there exists a negligible function $\text{negl}(\cdot)$ such that with probability at least $1 - \text{negl}(\lambda)$, at least $3/4$ fraction of the tuples in $\{(y_i, m_i, d_i)\}$ pass the verification.

3.2 Proof of Quantumness : Classical Prover's Advantage in the Random Oracle Model

Here, we will show that if the function H is replaced with a random oracle, then any classical algorithm that has non-negligible advantage in Protocol 1 can be used to break the claw-free property of \mathcal{F} . Consider the following security experiment which captures the interaction between a (classical) prover and a challenger in the random oracle model; the challenger represents the verifier in the protocol.

Experiment 1

In this experiment, the challenger represents the verifier in Protocol 1 and also responds to the random oracle queries issued by the prover.

1. The challenger (verifier) chooses an NTCF key $(k, \kappa) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda)$ and sends k to the prover. The prover and challenger have access to a random oracle $H : \{0, 1\}^n \rightarrow \{0, 1\}$.
2. The prover sends $\{(y_i, m_i, d_i)\}_{i \in [\lambda]}$. For each $i \in [\lambda]$, the challenger computes $x_{i,b} \leftarrow \text{INV}_{\mathcal{F}}(\kappa, b, y_i)$ for $b \in \{0, 1\}$, queries the random oracle H on $x_{i,0}, x_{i,1}$ and receives $h_{i,0}, h_{i,1}$ respectively. Next, it checks if $m_i = d_i^T \cdot (\text{BitDecomp}(x_{i,0}) + \text{BitDecomp}(x_{i,1})) + h_{i,0} + h_{i,1}$. If at least 0.75λ tuples satisfy the check, it outputs 1, else it outputs \perp .

Experiment 2

This experiment is similar to the previous one, except that the challenger implements the random oracle, and does not use the trapdoor for performing the final λ checks.

1. The challenger (verifier) chooses an NTCF key $(k, \kappa) \leftarrow \text{GEN}_{\mathcal{F}}(1^\lambda)$ and sends k to the prover. The challenger also implements the random oracle as follows. It maintains a database which is initially empty. On receiving a query x , it checks if there exists a tuple (x, h) in the database. If so, it outputs h , else it chooses a random bit $h \leftarrow \{0, 1\}$, adds (x, h) to the database and outputs h .
2. The prover sends $\{(y_i, m_i, d_i)\}_{i \in [\lambda]}$. On receiving this set from the prover, the challenger does not compute the inverses of y_i . Instead, it initializes $\text{count} = 0$, and for each i , it looks for tuples $(x_{i,0}, h_{i,0})$ and $(x_{i,1}, h_{i,1})$ in the table such that $\text{CHK}_{\mathcal{F}}(y_i, 0, x_{i,0}) = \text{CHK}_{\mathcal{F}}(y_i, 1, x_{i,1}) = 1$. If such $(x_{i,0}, x_{i,1})$ do not exist, then the challenger chooses a random bit r_i and sets $\text{count} = \text{count} + r_i$. Else, it checks if $m_i = d_i^T \cdot (\text{BitDecomp}(x_{i,0}) + \text{BitDecomp}(x_{i,1})) + h_{i,0} + h_{i,1}$. If so, it increments count . Finally, it checks if $\text{count} > 0.75\lambda$. If so, it outputs 1, else outputs \perp .

Experiment 3

This experiment is identical to the previous one, except that the challenger, after receiving $\{(y_i, m_i, d_i)\}_i$, outputs \perp if for all $i \in [\lambda]$, there does not exist two entries $(x_{i,0}, h_{i,0}), (x_{i,1}, h_{i,1})$ in the database such that $\text{CHK}_{\mathcal{F}}(y_i, 0, x_{i,0}) = \text{CHK}_{\mathcal{F}}(y_i, 1, x_{i,1}) = 1$.

3.2.1 Analysis

For any classical PPT prover \mathcal{A} , let $p_{\mathcal{A}}$ denote the probability that the verifier outputs 1 in Protocol 1 (when H is replaced with a random oracle), and for $w \in \{1, 2, 3\}$, let $p_{\mathcal{A},w}$ denote the probability that the challenger interacting with \mathcal{A} in Experiment w outputs 1. From the definition of Experiment 1 it follows that $p_{\mathcal{A}} = p_{\mathcal{A},1}$.

▷ **Claim 5.** For any prover \mathcal{A} , $p_{\mathcal{A},1} = p_{\mathcal{A},2}$.

Proof. The main differences between Experiment 1 and Experiment 2 are that the challenger implements the random oracle, and secondly, after receiving $\{(y_i, m_i, d_i)\}_i$, the challenger does not use the trapdoor for checking. Note that in Experiment 1, if either $x_{i,0}$ or $x_{i,1}$ are not queried to the random oracle H , then $H(x_{i,0}) + H(x_{i,1})$ is a uniformly random bit. Moreover, since the y_i values are distinct, if there exist two indices i, j such that both the preimages of y_i and y_j are not queried, then $H(x_{i,0}) + H(x_{i,1})$ is independent of $H(x_{j,0}) + H(x_{j,1})$. As a result, for each index i such that the preimages of y_i are not queried, the value of `count` is incremented with probability $1/2$.

In Experiment 2, the challenger checks for pairs corresponding to $x_{i,0}$ and $x_{i,1}$ in the database, and if either of them is missing, it increments `count` with probability $1/2$. As a result, the probability of `count` $> 0.75\lambda$ is identical in both experiments. ◁

▷ **Claim 6.** There exists a negligible function $\text{negl}(\cdot)$ such that for any prover \mathcal{A} and any security parameter $\lambda \in \mathbb{N}$, $p_{\mathcal{A},2} \leq p_{\mathcal{A},3} + \text{negl}(\lambda)$.

Proof. The only difference between these two experiments is that the challenger, at the end of the experiment, outputs \perp if for all $i \in [\lambda]$, either $x_{i,0}$ or $x_{i,1}$ has not been queried to the random oracle. The only case in which the challenger outputs 1 in Experiment 2 but outputs \perp in Experiment 3 is when for all $i \in [\lambda]$, either $x_{i,0}$ or $x_{i,1}$ has not been queried, but there exist $t \geq 0.75\lambda$ indices $\{i_1, \dots, i_t\}$ such that `count` was incremented. Using Chernoff bounds, we can show that this happens with negligible probability. ◁

▷ **Claim 7.** Assuming \mathcal{F} is a secure claw-free trapdoor family, for any PPT prover \mathcal{A} , there exists a negligible function $\text{negl}(\cdot)$ such that for all $\lambda \in \mathbb{N}$, $p_{\mathcal{A},3}(\lambda) \leq \text{negl}(\lambda)$.

Proof. Suppose there exists a PPT prover \mathcal{A} and a non-negligible function $\epsilon(\cdot)$ such that for all $\lambda \in \mathbb{N}$, the challenger outputs 1 with probability $\epsilon = \epsilon(\lambda)$ in Experiment 3. This means with probability at least ϵ , there exists an index $i^* \in [\lambda]$ such that \mathcal{A} queries the random oracle on $x_{i^*,0}, x_{i^*,1}$ and finally outputs $\{(y_i, m_i, d_i)\}_i$ such that $\text{CHK}_{\mathcal{F}}(y_{i^*}, 0, x_{i^*,0}) = \text{CHK}_{\mathcal{F}}(y_{i^*}, 1, x_{i^*,1}) = 1$.

We will construct a reduction algorithm \mathcal{B} that breaks the claw-free property of \mathcal{F} with probability ϵ . The reduction algorithm receives the key k from the NTCF challenger, which it forwards to \mathcal{A} . Next, \mathcal{A} makes polynomially many random oracle queries, which are answered by the reduction algorithm by maintaining a database. Eventually, \mathcal{A} sends $\{(y_i, m_i, d_i)\}_i$. The reduction algorithm checks if there exist tuples $(x_{i^*,0}, h_{i^*,0})$ and $(x_{i^*,1}, h_{i^*,1})$ in its database such that $\text{CHK}_{\mathcal{F}}(y_{i^*}, 0, x_{i^*,0}) = \text{CHK}_{\mathcal{F}}(y_{i^*}, 1, x_{i^*,1}) = 1$. If so, it sends $(x_{i^*,0}, x_{i^*,1})$ to the NTCF challenger. ◁

Using the above claims, it follows for every classical prover \mathcal{A} , there exists a negligible function $\text{negl}(\cdot)$ such that for all $\lambda \in \mathbb{N}$, $p_{\mathcal{A}} \leq \text{negl}(\lambda)$.

4 Construction of NTCFs based on Ring LWE

Our construction is similar to the one in [3]. Let λ be the security parameter, $n = 2^{\lceil \log \lambda \rceil}$. The following are other parameters chosen by our scheme (we will ignore dependence on security parameter/ n):

- Ring $R = \mathbb{Z}[X]/(X^n + 1)$.
- Modulus $q = \text{poly}(n)$, $R_q = R/qR$
- $m = \Omega(\log q)$: determines the dimension of range space
- χ : the noise distribution. In our case, χ is a Discrete Gaussian over \mathbb{Z}^n with parameter B_V .
- B_P : the noise bound for function evaluation. We require the following constraints on B_P :
 - $B_P \geq \Omega(n \cdot m \cdot B_V)$
 - $2B_P\sqrt{n \cdot m} \leq q/(C_T \cdot \sqrt{n \log q})$ for some constant C_T

The domain is $\mathcal{X} = R_q$, and range is $\mathcal{Y} = R_q^m$.

Each function key $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$, where $s \in R_q$, $a_i, e_i \in R_q$ for all $i \in [m]$, $\mathbf{a} = [a_1 \dots a_m]^T$, $\mathbf{e} = [e_1 \dots e_m]^T$. For $b \in \{0, 1\}$, $x \in \mathcal{X}$, $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$, the density function $f_{k,b}(x)$ is defined as follows:

$$\forall \mathbf{y} \in \mathcal{Y}, (f_{k,b}(x))(\mathbf{y}) = D_{\mathbb{Z}^{n \cdot m}, B_P}(\mathbf{y} - \mathbf{a} \cdot x - b \cdot \mathbf{a} \cdot s), \quad (8)$$

where $\mathbf{y} = [y_1 \dots y_m]^T$, and each \mathbf{y}_i can be represented as an element in \mathbb{Z}_q^n (using the coefficient representation); similarly for $\mathbf{a} \cdot x$ and $\mathbf{a} \cdot s$.

We will now show that each of the properties of NTCFs hold.

1. **Efficient Key Generation:** The key generation algorithm $\text{GEN}_{\mathcal{F}}(1^\lambda)$ first chooses $(\mathbf{a}, \tau) \leftarrow \text{GENTRAP}(1^n, 1^m, q)$, $s \leftarrow R_q$ and $\mathbf{e} \leftarrow \chi^m$. It outputs key $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$, and the trapdoor is $\kappa = (\tau, k, s)$.

2. **Trapdoor Injective Pair:**

- a. *Trapdoor:* For $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$, $x \in \mathcal{X}$ and $b \in \{0, 1\}$, the support of $f_{k,b}(x)$ is

$$\text{SUPP}(f_{k,b}(x)) = \{\mathbf{y} \in \mathcal{Y} : \|\mathbf{y} - \mathbf{a} \cdot x - b \cdot \mathbf{a} \cdot s\| \leq B_P\sqrt{n \cdot m}\}$$

The inversion algorithm $\text{INV}_{\mathcal{F}}$ takes as input the lattice trapdoor τ , $b \in \{0, 1\}$, $\mathbf{y} \in \mathcal{Y}$ and outputs $\text{INVERT}(\tau, \mathbf{a}, \mathbf{y}) - b \cdot s$. From Theorem 2, it follows that with overwhelming probability over the choice of \mathbf{a} , for all $\mathbf{y} \in \text{SUPP}(f_{k,b}(x))$, $\text{INVERT}(\tau, \mathbf{a}, \mathbf{y}) = x + b \cdot s$. Hence, it follows that $\text{INV}_{\mathcal{F}}(\kappa, b, \mathbf{y}) = x$.

- b. *Injective Pair:* Let $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$. From the construction, it follows that $f_{k,0}(x_0) = f_{k,1}(x_1)$ if and only if $x_1 = x_0 + s$. Hence the set $\mathcal{R}_k = \{(x, x + s) : x \in \mathcal{X}\}$.
3. **Efficient Range Superposition:** The function $f'_{k,0}$ is same as $f_{k,0}$, while $f'_{k,1}$ is defined as follows (recall $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$):

$$\forall \mathbf{y} \in \mathcal{Y}, (f'_{k,1}(x))(\mathbf{y}) = D_{\mathbb{Z}^{n \cdot m}, B_P}(\mathbf{y} - \mathbf{a} \cdot x - (\mathbf{a} \cdot s + \mathbf{e})) \quad (9)$$

- a. Since $f'_{k,0} = f_{k,0}$, it follows that for all $(x_0, x_1) \in \mathcal{R}_k$ and $\mathbf{y} \in \text{SUPP}(f'_{k,0}(x_0))$, $\text{INV}_{\mathcal{F}}(\kappa, 0, \mathbf{y}) = x_0$ and $\text{INV}_{\mathcal{F}}(\kappa, 1, \mathbf{y}) = x_1$. We need to show the same for $f'_{k,1}$; that is, for all $(x_0, x_1) \in \mathcal{R}_k$ and $\mathbf{y} \in \text{SUPP}(f'_{k,1}(x_1))$, $\text{INV}_{\mathcal{F}}(\kappa, 1, \mathbf{y}) = x_1$ and $\text{INV}_{\mathcal{F}}(\kappa, 0, \mathbf{y}) = x_0$. For all $x \in \mathcal{X}$,

$$\text{SUPP}(f'_{k,1}(x)) = \{\mathbf{y} \in \mathcal{Y} : \|\mathbf{y} - \mathbf{a} \cdot x - \mathbf{a} \cdot s - \mathbf{e}\| \leq B_P\sqrt{n \cdot m}\}$$

Hence for any $\mathbf{y} \in \text{SUPP}(f'_{k,1}(x))$, $\|\mathbf{y} - \mathbf{a} \cdot x_1 - \mathbf{a} \cdot s\| \leq 2B_P\sqrt{n \cdot m}$; using Theorem 2, we can conclude that $\text{INV}_{\mathcal{F}}(\kappa, 1, \mathbf{y}) = x_1$.

- b. The procedure $\text{CHK}_{\mathcal{F}}$ takes as input $\mathbf{y} \in \mathcal{Y}, k = (\mathbf{a}, \mathbf{v}), b \in \{0, 1\}, x \in \mathcal{X}$ and checks if $\|\mathbf{y} - \mathbf{a} \cdot x - b \cdot \mathbf{v}\| \leq B_P \sqrt{n \cdot m}$.
 - c. The definition of $\text{SAMP}_{\mathcal{F}}$ is identical to the one in [3], and the Hellinger distance can be bounded by $1 - e^{-\frac{2\pi m \cdot n \cdot B_V}{B_P}}$. From our setting of parameters, this quantity is at most $1/50$.
4. **Claw-Free Property** Suppose there exists an adversary \mathcal{A} that, on input $k = (\mathbf{a}, \mathbf{a} \cdot s + \mathbf{e})$ can output $(x_0, x_1) \in \mathcal{R}_k$. Then this adversary can be used to break the Ring LWE assumption, since $x_1 - x_0 = s$.

References

- 1 Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. In *Proceedings of the 43rd ACM Symposium on Theory of Computing, STOC 2011, San Jose, CA, USA, 6-8 June 2011*, pages 333–342, 2011.
- 2 Gorjan Alagic, Andrew M. Childs, Alex B. Grilo, and Shih-Han Hung. Non-interactive classical verification of quantum computation, 2019. [arXiv:1911.08101](#).
- 3 Zvika Brakerski, Paul Christiano, Urmila Mahadev, Umesh V. Vazirani, and Thomas Vidick. A cryptographic test of quantumness and certifiable randomness from a single quantum device. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 320–331, 2018.
- 4 Michael J Bremner, Richard Jozsa, and Dan J Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2126):459–472, 2010.
- 5 Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *J. of the ACM*, 51(4):557–594, 2004.
- 6 Nai-Hui Chia, Kai-Min Chung, and Takashi Yamakawa. Classical verification of quantum computations with efficient verifier. *ArXiv*, abs/1912.00990, 2019.
- 7 Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In *Advances in Cryptology - CRYPTO '86, Santa Barbara, California, USA, 1986, Proceedings*, pages 186–194, 1986.
- 8 Craig Gidney and Martin Ekerå. How to factor 2048 bit rsa integers in 8 hours using 20 million noisy qubits. *arXiv preprint arXiv:1905.09749*, 2019.
- 9 Shafi Goldwasser, Silvio Micali, and Ronald L. Rivest. A “paradoxical” solution to the signature problem. In George Robert Blakley and David Chaum, editors, *Advances in Cryptology*, 1985.
- 10 Vadim Lyubashevsky, Chris Peikert, and Oded Regev. On ideal lattices and learning with errors over rings. In *Advances in Cryptology - EUROCRYPT 2010, 29th Annual International Conference on the Theory and Applications of Cryptographic Techniques, Monaco / French Riviera, May 30 - June 3, 2010. Proceedings*, pages 1–23, 2010.
- 11 Vadim Lyubashevsky, Chris Peikert, and Oded Regev. A toolkit for ring-lwe cryptography. *IACR Cryptology ePrint Archive*, 2013:293, 2013.
- 12 Urmila Mahadev. Classical verification of quantum computations. In *59th IEEE Annual Symposium on Foundations of Computer Science, FOCS 2018, Paris, France, October 7-9, 2018*, pages 259–267, 2018.
- 13 Daniele Micciancio and Chris Peikert. Trapdoors for lattices: Simpler, tighter, faster, smaller. In *Advances in Cryptology - EUROCRYPT 2012 - 31st Annual International Conference on the Theory and Applications of Cryptographic Techniques, Cambridge, UK, April 15-19, 2012. Proceedings*, pages 700–718, 2012.
- 14 NIST. Candidate quantum-resistant cryptographic algorithms publicly available. URL: <https://www.nist.gov/news-events/news/2017/12/candidate-quantum-resistant-cryptographic-algorithms-publicly-available>.

8:14 **Simpler Proofs of Quantumness**

- 15 Oded Regev. On lattices, learning with errors, random linear codes, and cryptography. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing, Baltimore, MD, USA, May 22-24, 2005*, pages 84–93, 2005.
- 16 Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *35th Annual Symposium on Foundations of Computer Science, Santa Fe, New Mexico, USA, 20-22 November 1994*, pages 124–134, 1994.

Quantum Algorithms for Computational Geometry Problems

Andris Ambainis 

Faculty of Computing, University of Latvia, Raina bulvaris 19, Riga, LV-1586, Latvia
ambainis@lu.lv

Nikita Larka

Faculty of Computing, University of Latvia, Raina bulvaris 19, Riga, LV-1586, Latvia
nikitalarka@gmail.com

Abstract

We study quantum algorithms for problems in computational geometry, such as POINT-ON-3-LINES problem. In this problem, we are given a set of lines and we are asked to find a point that lies on at least 3 of these lines. POINT-ON-3-LINES and many other computational geometry problems are known to be 3SUM-HARD. That is, solving them classically requires time $\Omega(n^{2-o(1)})$, unless there is faster algorithm for the well known 3SUM problem (in which we are given a set S of n integers and have to determine if there are $a, b, c \in S$ such that $a + b + c = 0$).

Quantumly, 3SUM can be solved in time $O(n \log n)$ using Grover's quantum search algorithm. This leads to a question: can we solve POINT-ON-3-LINES and other 3SUM-HARD problems in $O(n^c)$ time quantumly, for $c < 2$?

We answer this question affirmatively, by constructing a quantum algorithm that solves POINT-ON-3-LINES in time $O(n^{1+o(1)})$. The algorithm combines recursive use of amplitude amplification with geometrical ideas. We show that the same ideas give $O(n^{1+o(1)})$ time algorithm for many 3SUM-HARD geometrical problems.

2012 ACM Subject Classification Theory of computation \rightarrow Quantum query complexity; Theory of computation \rightarrow Computational geometry

Keywords and phrases Quantum algorithms, quantum search, computational geometry, 3SUM problem, amplitude amplification

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.9

Funding Supported by the ERDF project 1.1.1.5/18/A/020 “Quantum algorithms: from complexity theory to experiment”.

1 Introduction

The 3SUM problem is as follows: given a set of numbers S , do there exist $a, b, c \in S$ such that $a + b + c = 0$? There is a nearly trivial classical algorithm that solves this problem in time $O(n^2)$. More advanced algorithms give only a logarithmic improvement to this quadratic complexity [15]. It is conjectured that no classical algorithm can solve 3SUM problem in $O(n^{2-\epsilon})$ time.

Many problems in computational geometry (for example, determining whether a given set of points contains 3 points that lie on a line) also seem to require $\Omega(n^2)$ time classically. Gajentaan and Overmars [12] showed that the 3SUM problem can be embedded into them. This implies that they cannot be solved in $O(n^{2-\epsilon})$ time, unless the 3SUM problem can also be solved in $O(n^{2-\epsilon})$ time. Such problems are called 3SUM-HARD. Besides 3 points on a line, examples of 3SUM-HARD problems include determining whether a given set of points contain 3 points that lie on a line, determining whether a given set of triangles covers given polygon, and determining whether a given set of axis-parallel segments are separable with a line into two nonempty subsets [12].



© Andris Ambainis and Nikita Larka;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 9; pp. 9:1–9:10

Leibniz International Proceedings in Informatics



LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Quantum computing allows designing quantum algorithms that outperform classical algorithms. One such example is Grover search [14] which achieves quadratic speedup over classical exhaustive search and can be used as a subroutine to speedup more complicated problems [3, 17].

In particular, the 3SUM problem can be solved by a quantum algorithm in $O(n \log n)$ time, by a fairly simple application of Grover search procedure. Indeed, we can do an exhaustive search over pairs $a, b \in S$ and look for $-(a + b) \in S$ using some data structure (for example, we can use a balanced search tree). However, a direct application of Grover search does not give a quadratic speedup for many geometrical 3SUM-HARD class problems. For example, if we need to determine whether a set of points contain three points that lie on the same line, we need to search for all possible triplets of points, which results in $O(n^{\frac{3}{2}})$ time quantum algorithm [11].

In this paper we combine quantum effects with more sophisticated geometric techniques to design a quantum algorithm with complexity $O(n^{1+o(1)})$ for POINT-ON-3-LINES problem. We use ideas from this algorithm to solve many other 3SUM-HARD problems in time $O(n^{1+o(1)})$.

Related work. The 3SUM problem has been studied in the context of query complexity and it can be solved with $O(n^{3/4})$ queries, as it is a special case of the subset finding problem of Childs and Eisenberg [10] in which one has to find constant-size subset S of an n -element set, with S satisfying a certain property. A matching $\Omega(n^{3/4})$ quantum query lower bound is known [6]. However, the subset finding algorithm of [10] does not have a time-efficient implementation in the general case. Some special cases (for example, the element distinctness and k -distinctness algorithms of [2]) can be implemented efficiently but no efficient implementation is known for the 3-SUM case.

We think that it is unlikely that this line of work would lead to an $o(n)$ time quantum algorithm for the 3SUM problem. The element distinctness algorithm [2] and the subset finding algorithm [10] are special cases of a quadratic speedup for hitting times of Markov chains [18, 4, 5]. It is unlikely that these methods will lead to a quantum algorithm that is more than quadratically faster than the best classical algorithm.

More generally, we conjecture that the 3SUM problem cannot be solved in $O(n^{1-\epsilon})$ quantum time in the QRAM model, neither with methods based on subset finding nor any other approach. This could serve as a basis for a quantum version of fine-grained complexity, similarly to recent quantum fine grained lower bounds of [1, 8] based on quantum versions of Strong Exponential Time Hypothesis (SETH).

2 Preliminaries

2.1 Problems

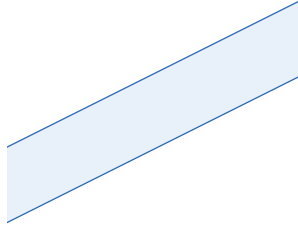
Here we define problems we focus on in this paper. All of them belong to 3SUM-HARD class. [12]

- POINT-ON-3-LINES: Given a set of lines in the plane, is there a point that lies on at least three of them? [12]
- 3-POINTS-ON-LINE: Given a set of points in the plane, is there a line that contains at least three points? [12]
- STRIPS-COVER-BOX: Given a set of strips in the plane (strip is defined as an infinite area between two parallel lines (see Figure 1)), does their union contain a given axis-parallel rectangle? [12]

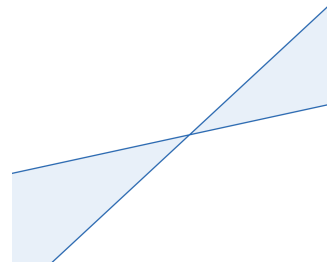
- TRIANGLES-COVER-TRIANGLE: Given a set of triangles in the plane, does their union contain another given triangle? [12]
- POINT-COVERING: Given a set of n half-planes and a number t , determine whether there is a point that is covered by at least t half-planes. [12]
- SEGMENT-SEPARATOR: Given a set of vertical line segments, does there exist a non vertical line that does not intersect any of given segments and contains at least one given segment in each of two half-planes? [12]
- VISIBILITY-BETWEEN-SEGMENTS: Given a set of n vertical line segments S and two particular line segments s_1 and s_2 , determine whether there is a point on s_1 and a point on s_2 , such that segment between these two points doesn't intersect any segment from S . [12]

We also define GENERAL-COVERING problem. We will design quantum algorithm for this problem with $O(n^{1+o(1)})$ complexity and then reduce many 3SUM-HARD problems to this problem.

- GENERAL-COVERING: We are given a set of n strips and angles (angle is defined as an infinite area between two non-parallel lines (see Figure 2)) in the plane. The task is to find a point X that satisfies the following conditions:
 - the point X is an intersection of two angle or strip boundary lines ℓ_1, ℓ_2 (ℓ_1 and ℓ_2 may be boundary lines of two different angles/strips);
 - the point X does not belong to the interior of any angle or strip;
 - the point X satisfies a given predicate $P(X)$ that can be computed in $O(1)$ time.



■ Figure 1 Strip.



■ Figure 2 Angle.

2.2 Model

We assume a query model in which the query returns the description D_i of the i^{th} object (point, line, strip, triangle, etc.), given i . The description consists of several numbers that specify the i^{th} object (e.g. coordinates of a point or values of coefficients in the equation that specifies a line). In the quantum case, we can query superpositions of indices i . The input to quantum query Q consists of two registers, with one register holding i and the other register provides the space for D_i . The query transformation acts as $Q|i, x\rangle = |i, x \oplus D_i\rangle$. In particular, given a superposition $|\psi\rangle = \sum_i \alpha_i |i, 0\rangle$ in which $x = 0$, applying Q gives the state $Q|\psi\rangle = \sum_i \alpha_i |i, D_i\rangle$. Applying Q to $|\phi\rangle = \sum_i \alpha_i |i, D_i\rangle$ gives the state $Q|\phi\rangle = \sum_i \alpha_i |i, 0\rangle$ in which the descriptions D_i are erased from the second register.

Our algorithms work in the commonly used QRAM (quantum random access memory) model of computation [13] which assumes quantum memory can be accessed in a superposition. QRAM has the property that any time- T classical algorithm that uses random access memory

can be invoked as a subroutine for a quantum algorithm in time $O(T)$. We can thus use primitives for quantum search (e.g., Grover's quantum search or Amplitude amplification) with conditions checking which requires data stored in a random access memory.

2.3 Tools

We will use two well known quantum procedures in our algorithm.

► **Theorem 1** (Grover search [16]). *Given a set of n elements $X = \{x_1, x_2, \dots, x_n\}$ and a boolean function $f : X \rightarrow \{0, 1\}$. The task is to find $x \in X$ such that $f(x) = 1$. There is a bounded-error quantum procedure that solves this problem using $O(\sqrt{n})$ quantum queries.*

► **Theorem 2** (Amplitude amplification [7]). *Let A be a quantum procedure with one-sided error and success probability at least ϵ . Then, there is a quantum procedure B that solves the same problem with success probability $\frac{2}{3}$ invoking A $O(\frac{1}{\sqrt{\epsilon}})$ times.*

Note that any constant success probability $1 - \epsilon$ can be achieved with repeating Amplitude Amplification constantly many times.

We will also use the following well known computational geometry algorithm.

► **Theorem 3** (Arrangement of lines [9]). *Given a set of n lines in the plane, we can compute partition of the plane formed by those lines in time $O(n^2)$.*

We will also use point-line dualization for problem reductions. Point-line dualization is a plane transformation that maps points to lines and lines to points in the following way:

- Line $\ell : y = ax + b$ is mapped to point $\ell^* = (a, -b)$
- Point $P = (a, b)$ is mapped to line $P^* : y = ax - b$

One may note, that the following properties are true:

1. $(P^*)^* = P$ and $(\ell^*)^* = \ell$
2. $P \in \ell \iff \ell^* \in P^*$
3. If point A, B, C lie on one non-vertical line, then lines A^*, B^*, C^* meet at one point.
4. If non-vertical lines p, q, r meet at one point, then points p^*, q^*, r^* lie on one line.
5. Points from vertical line segment are mapped to a strip.
6. Points from non-vertical line segment are mapped to an angle.

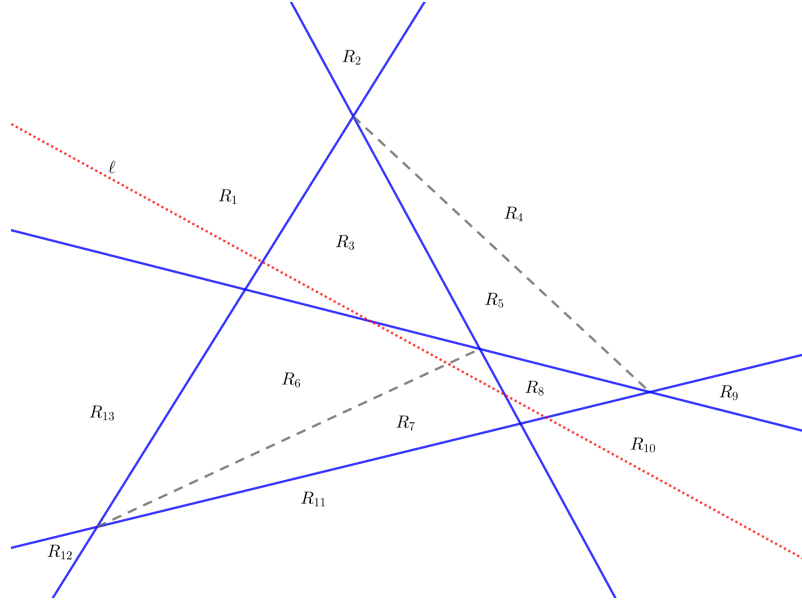
3 Point on three lines

In this part we describe a quantum algorithm which solves POINT-ON-3-LINES problem in $O(n^{1+o(1)})$ time, improving over the $O(n^{3/2})$ time quantum algorithm of Furrow [11]. The idea behind our algorithm is as follows. Suppose we are given a set S of lines in the plane. From this set we randomly pick k lines. Those k lines split plane into no more than $\binom{k+1}{2} + 1$ regions. Each line from set S intersects $k + 1$ regions. So, on average, a region contains $O(\frac{n}{k})$ lines crossing that region.

These facts give an opportunity to design a quantum algorithm that improves over the complexity $O(n^{\frac{3}{2}})$ of simple quantum search (as in Furrow's algorithm [11]). If we have a point A , then we can decide if A belongs to some line from S using $O(\sqrt{\frac{n}{k}} + k)$ quantum time. To do so, we need to find a region which contains point A and check if A belongs to a line which crosses that region. For finding the region, $O(k)$ time suffices. Checking if A belongs to a line from S can be done by Grover's search over $O(\frac{n}{k})$ lines that cross this region. For this, $O(\sqrt{\frac{n}{k}})$ time suffices.

If we need to find three lines that intersect in one point, we run Grover search over all pairs of lines (ℓ_i, ℓ_j) . For each pair, we find the intersection point $P_{i,j} = \ell_i \cap \ell_j$ and check if point $P_{i,j}$ belongs to a third line using algorithm described earlier. If the subdivision of the plane into regions can be done in time $O(nk^2)$, this algorithm runs in time $O(nk^2 + \sqrt{n^2}(k + \sqrt{\frac{n}{k}}))$. Setting $k = n^{\frac{1}{5}}$ gives $O(n^{\frac{7}{5}})$. But we can find even better algorithm. After dividing the plane into regions, instead of searching for an intersection point of three lines, we search for a region which has this point (search is done using Grover search) and we recursively apply $O(n^{\frac{7}{5}})$ algorithm to find the intersection point of three lines inside that region. We can then add more levels of recursion to decrease the complexity further. We now describe the final algorithm (in which we recurse at the optimal choice of k and the number of levels of recursion grows with n).

Let S be the given set of lines in a plane. Let P be a subset of S containing exactly k lines. Lines in P divide plane into convex (possibly infinite) polygons. We arbitrarily triangulate regions, which are bounded by at least 4 lines. This results in a subdivision of the plane into regions R_1, R_2, \dots, R_t where each region is bounded by at most 3 segments (see Figure 3). Let $s(R_i) = \{\ell \mid \ell \in S \text{ and } \ell \text{ intersect } R_i\}$.



■ **Figure 3** Blue lines divide plane into 13 regions: R_1, R_2, \dots, R_{13} . Red line ℓ passes through regions: $\ell \in s(R_1), s(R_3), s(R_6), s(R_7), s(R_8), s(R_{10})$.

We start with the following three observations:

► **Lemma 4.** *If after the triangulation we get t regions: R_1, R_2, \dots, R_t , then $t \leq 2|P|^2 = 2k^2$*

Proof. If f_i is the number of faces bounded by i lines before the triangulation, then

$$t = f_1 + f_2 + f_3 + 2f_4 + 3f_5 + \dots + (k-2)f_k \leq \sum_{i=1}^k i f_i \leq 2k^2 \quad (1)$$

The last inequality holds because $\sum_{i=1}^k i f_i$ is equal to twice the total number of line segments before the triangulation (because each line segment is on the boundary of two faces, one on each side) and the number of line segments is at most k^2 (because each of k lines is split by the other $k-1$ lines into at most k segments). ◀

► **Lemma 5.** Sets $s(R_1), \dots, s(R_t)$ can be built classically in time $O(|S| \times |P|^2)$.

Proof. We can construct regions R_1, R_2, \dots, R_t in time $O(|P|^2)$ using Theorem 3. We build $s(R_i)$ by iterating through each line from S and checking whether the line intersects region R_i . This step takes time $O(|S| \times t) = O(|S| \times |P|^2)$. ◀

► **Lemma 6.** If P is chosen uniformly at random, then

$$\Pr[\max_i |s(R_i)| \geq 3 \frac{|S|}{|P|} (5 \log(|S|) + \log(\epsilon^{-1}))] \leq \epsilon \quad (2)$$

Proof. Let ℓ be an arbitrary line (possibly not from the set S). Lines from set S intersect line ℓ in points X_1, X_2, \dots, X_m in this order (if two lines ℓ_i and ℓ_j intersect ℓ in the same point, then $X_i = X_j$). We note that $m \leq |S|$, since some lines from S might be parallel to ℓ . We color a point X_i with white color if the corresponding line ℓ_i from S is in the set P . Otherwise we color the point X_i with black color. We define $L = \left\lceil \frac{|S|}{|P|} (5 \log(|S|) + \log(\epsilon^{-1})) \right\rceil$ and we assume that $L \leq |S|$, since otherwise lemma is obviously true.

We say that a line ℓ is bad, if there exists index i , such that X_{i+j} is black for all $j \in [0 \dots L-1]$. The probability of ℓ being bad can be upper bounded as follows

$$\Pr \left[\bigvee_{i=1}^{m-L+1} (X_i, X_{i+1}, \dots, X_{i+L-1} \text{ are all black}) \right] = \quad (3)$$

$$\leq \sum_{i=1}^{m-L+1} \Pr[X_i, X_{i+1}, \dots, X_{i+L-1} \text{ are all black}] = (m-L+1) \frac{\binom{|S|-L}{|P|}}{\binom{|S|}{|P|}} \quad (4)$$

$$\leq |S| \left(\frac{|S|-L}{|S|} \right)^{|P|} \leq |S| \left[\left(1 - \frac{L}{|S|} \right)^{\frac{|S|}{L}} \right]^L \frac{|P|}{|S|} \leq \frac{|S|}{e^{5 \log |S| + \log \epsilon^{-1}}} = \frac{\epsilon}{|S|^4} \quad (5)$$

Consider set S' which consists of lines from S and lines that pass through at least two intersection points of lines from S . Then, every edge e of every region R_i lies on a line that belongs to S' (because e is either a segment of one of original lines from S or is created during the triangulation and, in the second case, both endpoints of e are intersection points of two lines from S). Since there are at most $\binom{|S|}{2}$ intersection points of lines from S , we have $|S'| \leq |S|^4$ and

$$\Pr[S' \text{ contains bad line}] \leq \sum_{\ell \in S'} \Pr[\ell \text{ is bad}] \leq |S'| \times \frac{\epsilon}{|S|^4} \leq \epsilon \quad (6)$$

To finish the proof, it is enough to see that the fact that S' doesn't contain bad line implies $|s(R_i)| \leq 3(L-1) < 3 \frac{|S|}{|P|} (5 \log(|S|) + \log(\epsilon^{-1}))$ for all i . Indeed, if no lines in S' is bad, then each side of each region R_i contains less than L black points. Since a black point corresponds to a line which intersects a region R_i and each region is bounded by at most three segments, the region R_i is intersected by no more than $3(L-1)$ lines from S . ◀

► **Theorem 7.** There is a bounded error quantum algorithm for POINT-ON-3-LINES problem, that runs in time $O(|S|^{1+o(1)})$.

Proof. The algorithm has a parameter k and allowable error probability ϵ . The algorithm consists of a recursive procedure that takes a set of lines X as input and returns 3 lines from the set X which intersects at one point or tells that there are no such 3 lines.

Procedure $Algo_k(X)$.

```

if  $|X| < k$  then
  | Check for an intersection of 3 lines classically, by exhaustive search
end
 $R_1, R_2, \dots, R_t = \text{RandomPlaneSeparation}_{\frac{\epsilon}{2}}(X)$ 
Build sets:  $s(R_1), s(R_2), \dots, s(R_t)$ 
if  $\max_i |s(R_i)| > 3 \frac{|X|}{k} (5 \log(|X|) + \log(\frac{2}{\epsilon}))$  then
  | return error
end
Let  $A$  be the algorithm that randomly chooses  $j \in [t]$  and runs  $Algo_k(s(R_j))$ .
With Amplitude amplification, run  $A$  with the success probability amplified to at
least  $1 - \frac{\epsilon}{2}$ .

```

The recursive procedure can be described as follows. If the input set X contains less than k lines, we solve POINT-ON-3-LINES classically in time $O(|X|^2)$. Otherwise, we split the plane into regions R_1, R_2, \dots, R_t with k random lines from X and build sets $s(R_1), s(R_2), \dots, s(R_t)$. If there are three lines that intersect at one point, then this point is located in one of the regions R_i (if this point is on the boundary of a region, then this point can be found during the construction of sets $s(R_i)$). We use amplitude amplification to find the region R_i which contains intersection point of three lines and recursively apply the procedure to lines from $s(R_i)$.

Suppose that S contains three lines that intersect at one point. If $|S| < k$, the algorithm finds those 3 lines with certainty. Let $|S| \geq k$. The probability that $|s(R_i)| > 3 \frac{|S|}{k} (5 \log(|S|) + \log(\frac{2}{\epsilon}))$ for some R_i and the algorithm returns an error is less than $\frac{\epsilon}{2}$. By executing amplitude amplification (running $Algo_k(s(R_i))$ $O(\sqrt{t}) = O(k)$ times), we can reduce probability of not finding the 3 lines to $\frac{\epsilon}{2}$. So, $Algo_k(S)$ finds desired three lines with probability at least $1 - \epsilon$.

If $T(|X|)$ is a runtime of $Algo_k(X)$, then:

$$T(|X|) = O(|X| \times k^2) + O(\sqrt{k^2}) \times T\left(3 \frac{|X|}{k} (5 \log(|X|) + \log(\frac{2}{\epsilon}))\right) \quad (7)$$

If $k = |S|^{\frac{1}{\alpha}} \cdot 3(5 \log(|S|) + \log(\frac{2}{\epsilon}))$, then:

$$T(|X|) \leq O(|X| \times |S|^{\frac{2}{\alpha}} \log^2(|S|)) + O(\sqrt{k^2}) \times T(|X| \times |S|^{-\frac{1}{\alpha}}) \quad (8)$$

There are $(C_1 k)^{2j}$ problems on recursion level j for some constant C_1 and each problem size is at most $|S|^{1-\frac{j}{\alpha}}$.

$$\begin{aligned}
 T(|S|) &\leq \sum_{j=0}^{\alpha} \sqrt{(C_1 k)^{2j}} \times \left[|S|^{1-\frac{j}{\alpha}} \times |S|^{\frac{2}{\alpha}} \log^2(|S|) \right] = \\
 &= \sum_{j=0}^{\alpha} \left(\frac{C_1 k}{|S|^{\frac{1}{\alpha}}} \right)^j (|S|^{1+\frac{2}{\alpha}} \log^2(|S|)) \leq \\
 &\leq \alpha (C_2 \log(|S|))^{\alpha} (|S|^{1+\frac{2}{\alpha}} \log^2(|S|))
 \end{aligned} \quad (9)$$

If $\alpha = \sqrt{\frac{2 \log(|S|)}{\log(C_2) + \log \log |S|}}$, then $T(|S|) = O(\alpha |S|^{1+\frac{4}{\alpha}} \log^2(|S|)) = O(|S|^{1+o(1)})$ ◀

4 Other 3SUM hard problems

In this section, we show how it is possible to apply plane separation ideas, described in the previous section, to speed up other 3-SUM-HARD problems defined in [12].

■ 3-POINTS-ON-LINE

This problem is dual to the POINT-ON-3-LINES problem [12], and so is solvable with the same quantum algorithm in time $O(n^{1+o(1)})$, as described in the previous section.

■ GENERAL-COVERING

The given n strips and angles form $2n$ lines. We divide the plane into regions by randomly choosing k out of those $2n$ lines, similarly to the algorithm described in the previous section. A region and a strip/angle can be in one of the following relations: the strip/angle fully covers the region, the strip/angle partly covers the region or the strip/angle has no common points with the region. We can identify all regions that are fully covered by some strip/angle in time $O(nk^2)$. For regions that are not fully covered by some strip/angle, we identify the set $s(R_i)$ of strips and angles which cross that region. Non covered regions may contain the desired intersection point, but this intersection point is formed by the lines that are boundary lines of the strips/angles in the set $s(R_i)$. Similarly to the algorithm POINT-ON-3-LINES, our task is divided into $O(k^2)$ tasks, each of which involved $O(\frac{n \log n}{k})$ objects. The time complexity is

$$T(n) = O(nk^2) + O(\sqrt{k^2}) \times T\left(3\frac{2n}{k}(5\log(2n) + \log(\frac{2}{\epsilon}))\right). \quad (10)$$

Similarly to the analysis of POINT-ON-3-LINES problem, we get $T(n) = O(n^{1+o(1)})$.

■ STRIPS-COVER-BOX

This problem is just the special case of the GENERAL-COVERING problem with the predicate $P(X)$ being true if the point X is located inside the given box. Then, the point X from GENERAL-COVERING problem corresponds to an uncovered point in STRIPS-COVER-BOX problem. So, STRIPS-COVER-BOX can also be solved in time $O(n^{1+o(1)})$.

■ TRIANGLES-COVER-TRIANGLE

The given n triangles consist of $3n$ segments. We extend each segment to a line and separate the plane into regions, similarly to the POINT-ON-3-LINES problem with randomly chosen k lines. A triangle and a region can be in one of the following relations: the triangle fully covers the region, the triangle partly covers the region or the triangle has no common point with the region. We can identify all regions that are fully covered by some triangle in time $O(nk^2)$. All other regions may contain a point X which is not covered by any triangle. Similarly to the POINT-ON-3-LINES problem, we search for the region R_i which contains that point. Note that, if a triangle partly covers the region R_i , then at least one of the segments that form this triangle is in $s(R_i)$. So, we can finish our algorithm, just as in GENERAL-COVERING problem, with the predicate $P(X)$ being true, if the point X is located inside the triangle that must be covered.

■ POINT-COVERING

The given n half-planes are specified by n lines. We separate the plane into regions, similarly to the POINT-ON-3-LINES problem, by randomly choosing k out of n given lines. For each region R_i we compute the number of half-planes r_i that fully cover this region. This takes $O(nk^2)$ time. To determine if there exists a point that is covered by at least t half-planes, we need to tell, if there exists a point inside a region R_i that is covered by at least $t - r_i$ half-planes from $s(R_i)$. As in the GENERAL-COVERING problem, the algorithm takes $O(n^{1+o(1)})$ time.

■ VISIBILITY-BETWEEN-SEGMENTS

We dualize the given n vertical segments, to get n strips. We need to find a point that does not belong to any of the strips and has the property that the corresponding line in the initial plane intersects two given segments s_1 and s_2 . This problem is just the special case of the GENERAL-COVERING problem, where the predicate $P(X)$ is true if the line corresponding to X intersects two given segments s_1 and s_2 . Just like in the GENERAL-COVERING problem, this problem can also be solved in $O(n^{1+o(1)})$ time.

■ SEGMENT-SEPARATOR

This problem can be solved in exactly the same way as the VISIBILITY-BETWEEN-SEGMENTS problem, with the only difference being the predicate $P(X)$. Now this predicate is true, if the line corresponding to a point separates the given segments in the way required for the SEGMENT-SEPARATOR problem. Since X is the intersection point of two lines in the dual plane, the line, corresponding to the point X , must go through two endpoints of two different given segments. So, $P(X)$ is false, if the corresponding line goes through an edge of the convex hull of the endpoints of given segments. This can be determined in time $O(1)$, after we precompute the convex hull in time $O(n \log(n))$. This results in an $O(n^{1+o(1)})$ time quantum algorithm.

References

- 1 Scott Aaronson, Nai-Hui Chia, Han-Hsuan Lin, Chunhao Wang, and Ruizhe Zhang. On the quantum complexity of closest pair and related problems. *CoRR*, abs/1911.01973, 2019. [arXiv:1911.01973](#).
- 2 Andris Ambainis. Quantum walk algorithm for element distinctness. *SIAM J. Comput.*, 37(1):210–239, 2007. doi:10.1137/S0097539705447311.
- 3 Andris Ambainis, Kaspars Balodis, Jānis Iraids, Martins Kokainis, Krišjānis Prūsis, and Jevgēnijs Vihrovs. Quantum speedups for exponential-time dynamic programming algorithms. In *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA '19*, pages 1783–1793, Philadelphia, PA, USA, 2019. Society for Industrial and Applied Mathematics. URL: <http://dl.acm.org/citation.cfm?id=3310435.3310542>.
- 4 Andris Ambainis, András Gilyén, Stacey Jeffery, and Martins Kokainis. Quadratic speedup for finding marked vertices by quantum walks. *CoRR*, abs/1903.07493, 2019. [arXiv:1903.07493](#).
- 5 Simon Apers, András Gilyén, and Stacey Jeffery. A unified framework of quantum walk search. *CoRR*, abs/1912.04233, 2019. [arXiv:1912.04233](#).
- 6 Aleksandrs Belovs and Robert Spalek. Adversary lower bound for the k-sum problem. In Robert D. Kleinberg, editor, *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 323–328. ACM, 2013. doi:10.1145/2422436.2422474.
- 7 Gilles Brassard, Peter Hoyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. *Contemporary Mathematics*, 305:53–74, 2002.
- 8 Harry Buhrman, Subhasree Patro, and Florian Speelman. The quantum strong exponential-time hypothesis. *CoRR*, abs/1911.05686, 2019. [arXiv:1911.05686](#).
- 9 Bernard Chazelle, Leo J. Guibas, and D. T. Lee. The power of geometric duality. *BIT Numerical Mathematics*, 25, 1985. doi:10.1007/BF01934990.
- 10 Andrew M. Childs and Jason M. Eisenberg. Quantum algorithms for subset finding. *Quantum Information & Computation*, 5(7):593–604, 2005. URL: <http://portal.acm.org/citation.cfm?id=2011663>.
- 11 Bartholomew Furrow. A panoply of quantum algorithms. *Quantum Information & Computation*, 8(8):834–859, 2008. URL: <http://www.rintonpress.com/xxqic8/qic-8-89/0834-0859.pdf>.
- 12 Anka Gajentaan and Mark H. Overmars. On a class of $O(n^2)$ problems in computational geometry. *Comput. Geom.*, 5:165–185, 1995. doi:10.1016/0925-7721(95)00022-2.

- 13 Vittorio Giovannetti, Seth Lloyd, and Lorenzo Maccone. Quantum random access memory. *Physical review letters*, 100:160501, April 2008. doi:10.1103/PhysRevLett.100.160501.
- 14 Lov K. Grover. A fast quantum mechanical algorithm for database search. In *Proceedings of the Twenty-eighth Annual ACM Symposium on Theory of Computing*, STOC '96, pages 212–219, New York, NY, USA, 1996. ACM. doi:10.1145/237814.237866.
- 15 Allan Grønlund Jørgensen and Seth Pettie. Threesomes, degenerates, and love triangles. *CoRR*, abs/1404.0799, 2014. arXiv:1404.0799.
- 16 Lov K. Grover. Framework for fast quantum mechanical algorithms. *Conference Proceedings of the Annual ACM Symposium on Theory of Computing*, 80, December 1997. doi:10.1145/276698.276712.
- 17 Frédéric Magniez, Miklos Santha, and Mario Szegedy. Quantum algorithms for the triangle problem. In *Proceedings of the Sixteenth Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '05, pages 1109–1117, Philadelphia, PA, USA, 2005. Society for Industrial and Applied Mathematics. URL: <http://dl.acm.org/citation.cfm?id=1070432.1070591>.
- 18 Mario Szegedy. Quantum speed-up of markov chain based algorithms. In *45th Symposium on Foundations of Computer Science (FOCS 2004), 17-19 October 2004, Rome, Italy, Proceedings*, pages 32–41. IEEE Computer Society, 2004. doi:10.1109/FOCS.2004.53.

Quantum Coupon Collector

Srinivasan Arunachalam

IBM Research, Yorktown Heights, NY, USA
Srinivasan.Arunachalam@ibm.com

Aleksandrs Belovs

Faculty of Computing, University of Latvia, Riga, Latvia
aleksandrs.belovs@lu.lv

Andrew M. Childs

Department of Computer Science, Institute for Advanced Computer Studies,
University of Maryland, College Park, MD, USA
Joint Center for Quantum Information and Computer Science,
University of Maryland, College Park, MD, USA
amchilds@umd.edu

Robin Kothari

Microsoft Quantum, Redmond, WA, USA
Microsoft Research, Redmond, WA, USA
robin.kothari@microsoft.com

Ansis Rosmanis

Graduate School of Mathematics, Nagoya University, Japan
ansis.rosmanis@math.nagoya-u.ac.jp

Ronald de Wolf

QuSoft, Amsterdam, The Netherlands
CWI, Amsterdam, The Netherlands
University of Amsterdam, The Netherlands
rdewolf@cwi.nl

Abstract

We study how efficiently a k -element set $S \subseteq [n]$ can be learned from a uniform superposition $|S\rangle$ of its elements. One can think of $|S\rangle = \sum_{i \in S} |i\rangle / \sqrt{|S|}$ as the quantum version of a uniformly random sample over S , as in the classical analysis of the “coupon collector problem.” We show that if k is close to n , then we can learn S using asymptotically fewer quantum samples than random samples. In particular, if there are $n - k = O(1)$ missing elements then $O(k)$ copies of $|S\rangle$ suffice, in contrast to the $\Theta(k \log k)$ random samples needed by a classical coupon collector. On the other hand, if $n - k = \Omega(k)$, then $\Omega(k \log k)$ quantum samples are necessary.

More generally, we give tight bounds on the number of quantum samples needed for every k and n , and we give efficient quantum learning algorithms. We also give tight bounds in the model where we can additionally reflect through $|S\rangle$. Finally, we relate coupon collection to a known example separating proper and improper PAC learning that turns out to show no separation in the quantum case.

2012 ACM Subject Classification Theory of computation \rightarrow Quantum computation theory

Keywords and phrases Quantum algorithms, Adversary method, Coupon collector, Quantum learning theory

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.10

Funding *Srinivasan Arunachalam*: Part of this work was done while a PhD student at CWI supported by ERC Consolidator Grant 615307-QPROGRESS, and a postdoc at MIT funded by the MIT-IBM Watson AI Lab under the project *Machine learning in Hilbert space*.

Aleksandrs Belovs: Supported by the ERDF project number 1.1.1.2/I/16/113.

Andrew M. Childs: Supported by the Army Research Office (grant W911NF-20-1-0015); the De-



© Srinivasan Arunachalam, Aleksandrs Belovs, Andrew M. Childs, Robin Kothari, Ansis Rosmanis, and Ronald de Wolf;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).
Editor: Steven T. Flammia; Article No. 10; pp. 10:1–10:17



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

partment of Energy, Office of Science, Office of Advanced Scientific Computing Research, Quantum Algorithms Teams and Accelerated Research in Quantum Computing programs; and the National Science Foundation (grant CCF-1813814).

Ansis Rosmanis: Supported by the JSPS International Research Fellowship program and by the JSPS KAKENHI Grant Number JP19F19079.

Ronald de Wolf: Partially supported by ERC Consolidator Grant 615307-QPROGRESS (which ended Feb 2019), and by the Dutch Research Council (NWO) through Gravitation-grant Quantum Software Consortium 024.003.037, and QuantERA project QuantAlgo 680-91-034.

1 Introduction

Learning from quantum states is a major topic in quantum machine learning. While this task has been studied extensively [15, 27, 7, 8, 6, 17, 4], many fundamental questions about the power of quantum learning remain. Determining properties of quantum states has potential applications not only in the context of machine learning, but also as a basic primitive for other types of quantum algorithms and for quantum information processing more generally.

In this paper we study a very simple and natural quantum learning problem. We are given copies of the uniform superposition

$$|S\rangle := \frac{1}{\sqrt{|S|}} \sum_{i \in S} |i\rangle$$

over the elements of an unknown set $S \subseteq [n] := \{1, \dots, n\}$ (sometimes referred to as a uniform *quantum sample* from S [2]). Assume we know the size $k := |S| < n$. Our goal is to learn S exactly. How many copies of $|S\rangle$ do we need for this? And given the information-theoretically minimal number of copies needed, can we learn S gate-efficiently (i.e., using a quantum circuit with gate count polynomial in k and $\log n$)?

As a warm-up, first consider what happens if we just measure our copies of $|S\rangle$ in the computational basis, giving uniform samples from S . How many such samples do we need before we learn S ? As long as there is some element of S that we have not seen, we cannot even guess S with constant success probability, so we need to sample until we see all k distinct elements. This is known as the “coupon collector problem.” Analyzing the required number of samples is easy to do in expectation, as follows. Suppose we have already seen $i < k$ distinct elements from S . Then the probability that we see a new element in the next sample is $(k - i)/k$, and the expected number of samples to see an $(i + 1)$ st element is the reciprocal of that probability, $k/(k - i)$. By linearity of expectation we can add this up over all i from 0 to $k - 1$, obtaining the expected number of samples to see all k elements:

$$\sum_{i=0}^{k-1} \frac{k}{k-i} = k \sum_{j=1}^k \frac{1}{j} \sim k \ln k.$$

With a bit more work one can show that $\Theta(k \log k)$ samples are necessary and sufficient to identify S with high probability [25, Chapter 3.6]:

► **Proposition 1** (Classical coupon collector). *Given uniformly random samples from a set $S \subseteq [n]$ of size $k < n$, the number of samples needed to identify S with high probability is $\Theta(k \log k)$.*

The relationship between the probability of seeing all elements of S and the number of samples is extremely well understood. In particular, we can achieve probability arbitrarily close to 1 using only $k \ln k + O(k)$ samples [25, Theorem 3.8].

Of course, measuring $|S\rangle$ in the computational basis is not the only approach a quantum computer could take. The goal of this paper is to identify when and how we can do better, reducing the number of copies of $|S\rangle$ that are used to solve this “quantum coupon collector problem.” It turns out that we can asymptotically beat the classical threshold of $\Theta(k \log k)$ if and only if the number $m = n - k$ of “missing elements” is small (whereas classically the parameter m is irrelevant). Specifically, we give a simple, gate-efficient quantum algorithm that learns S from $O(n \log(m + 1))$ copies of $|S\rangle$. For small m this is significantly more efficient than classical coupon collection. In particular, for $m = O(1)$ we only need $O(k)$ quantum samples, saving a factor of $O(\log k)$.

As we explain in Section 5, this result is relevant for the comparison of *proper* and *improper* learning in the PAC model. A “proper” learner is one that only outputs hypotheses from the same concept class that its target function comes from. The coupon collector problem can be viewed as a learning task where the sample complexity of proper learners from classical random examples is asymptotically higher than that of proper learners from quantum examples.

We also prove *lower* bounds on the number T of copies needed, using the general (i.e., negative-weights) adversary bound of quantum query complexity [20]. This approach may be surprising, since no queries are involved when trying to learn S from copies of $|S\rangle$.¹ However, the adversary bound also characterizes the quantum query complexity of “state conversion” [21] and “state discrimination.” Our learning problem may be viewed as the problem of converting the state $|S\rangle^{\otimes T}$ to a basis state that gives a classical description of the k -set S . To employ the general adversary bound, we exploit the underlying symmetries of the problem using the mathematical machinery of association schemes (see also [3, 23] for prior uses of association schemes in proving adversary lower bounds). Using this, we show that, unless the number of missing elements $m = n - k$ is very small, the $O(k \log k)$ classical coupon collector algorithm is essentially optimal even in the quantum case. This means that the quantum coupon collector might as well just measure the copies of the state in the computational basis, unless m is very small.

We also study the situation where, in addition to copies of $|S\rangle$, we can also apply a unitary operation $R_S = 2|S\rangle\langle S| - \text{Id}$ that reflects through the state $|S\rangle$ (i.e., $R_S|S\rangle = |S\rangle$ and $R_S|\phi\rangle = -|\phi\rangle$ for all states $|\phi\rangle$ orthogonal to $|S\rangle$). This model is reasonable to consider because if we had a unitary that prepared $|S\rangle$, or even $|S\rangle|\psi\rangle$ for some garbage state $|\psi\rangle$, starting from some canonical state $|0\rangle$, then we could use this unitary to create the unitary R_S in a black-box manner. For example, if $U|0\rangle = |S\rangle$, then $R_S = U(2|0\rangle\langle 0| - \text{Id})U^\dagger$.

This model gives us extra power and enables more efficient learning of the set S : $\Theta(\sqrt{km})$ states and reflections are necessary and sufficient to learn S for large k (i.e., small m), and $\Theta(k)$ states and reflections are necessary and sufficient for small k .

The following table summarizes our main results. Sections 2 and 3 prove the upper and lower bounds in the first row, respectively, while Section 4 proves the results in the second row.

¹ A natural approach is to analyze the success probability of the “pretty good measurement” (PGM) for discriminating the states $|S\rangle^{\otimes T}$. The PGM is an explicit measurement whose average success probability is no more than quadratically worse than that of the optimal measurement [10] (in fact, one can show that the PGM is optimal in our case because our set of states is “geometrically uniform”). One can write down the average success probability of the PGM explicitly, and upper bounding it would establish a lower bound on the required number of copies of $|S\rangle$. However, we have been unable to suitably bound this expression despite considerable effort.

■ **Table 1** Main results about the complexity of learning the set S with $m = n - k$ missing elements.

	$k \geq n/2$	$k \leq n/2$
Number of copies of $ S\rangle$:	$\Theta(k \log(m+1))$ Theorem 2 and Theorem 6	$\Theta(k \log k)$ Proposition 1 and Theorem 6
Number of copies and reflections:	$\Theta(\sqrt{km})$ Theorem 10 and Theorem 11	$\Theta(k)$ Theorem 12 and Theorem 13

We contrast our work with recent results on the quantum query complexity of approximate counting by Aaronson, Kothari, Kretschmer, and Thaler [1]. They consider a similar model, given copies of the state $|S\rangle$, the ability to reflect through $|S\rangle$, and also the ability to query membership in S . However, in their work the size of S is unknown and the goal is to approximately *count* this set up to small multiplicative error. They obtain tight upper and lower bounds on the complexity of this approximate-counting task using techniques quite different from ours (specifically, Laurent polynomials for the lower bounds). This allows them to give an oracle separation between the complexity classes SBP and QMA. In contrast, in our case the size k of the set S is already known to the learner from the start, and the goal is to *identify* S exactly.

2 Upper bound on quantum samples

In this section we prove upper bounds on the number of copies of $|S\rangle$ that suffice to identify the k -element set $S \subseteq [n]$ with high probability.

The easiest way to recover S is by measuring $O(k \log k)$ copies of $|S\rangle$ in the computational basis. By the classical coupon collector problem (Proposition 1), we will (with high probability) see all elements of S at least once. As we will show later, this number of copies of $|S\rangle$ turns out to be asymptotically optimal if the number of missing elements $m = n - k$ is large (at least polynomial in n). However, here we show that something better is possible for very small m .

► **Theorem 2** (Upper bound for small m). *Let $S \subseteq [n]$ be a set of size $k < n$ and let $m = n - k$. We can identify S with high probability using $O(k \log(m+1))$ copies of $|S\rangle$ by a gate-efficient quantum algorithm.*

Proof. This bound is trivial when m is polynomial in n , since an upper bound of $O(k \log k)$ follows from Proposition 1. So let us now assume that $m \leq n^{1/4}$ and hence $k \geq n - n^{1/4}$.

Consider the uniform superposition over all elements of the universe $[n]$:

$$|[n]\rangle = \frac{1}{\sqrt{n}} \sum_{i \in [n]} |i\rangle.$$

Performing the 2-outcome projective measurement with operators $|[n]\rangle\langle[n]|$ and $\text{Id} - |[n]\rangle\langle[n]|$ is no harder than preparing $|[n]\rangle$, so it can be implemented gate-efficiently. If we apply this measurement to a copy of $|S\rangle$, then we get the first outcome with probability $|\langle S|[n]\rangle|^2 = k/n$ and the second outcome with probability m/n . In the latter case, the post-measurement state is

$$|\psi\rangle = \sqrt{\frac{m}{n}} |S\rangle - \sqrt{\frac{k}{n}} |\bar{S}\rangle$$

which is close to $-\sqrt{\frac{k}{n}} |\bar{S}\rangle$ if $m \ll n$.

We use an expected number of $O(\frac{n}{m} \cdot m \log(m+1)) = O(n \log(m+1))$ copies of $|S\rangle$ to prepare $O(m \log(m+1))$ copies of $|\psi\rangle$. If $|\psi\rangle$ were exactly equal to $|\bar{S}\rangle$, then measuring in the computational basis would sample uniformly over the set \bar{S} of m missing elements, and $O(m \log(m+1))$ such samples suffice to recover \bar{S} by the classical coupon collector problem (Proposition 1). Instead, $|\psi\rangle$ only approximately equals $|\bar{S}\rangle$: if we measure it then each $i \in \bar{S}$ has probability $\frac{k}{nm}$, while each $i \in S$ has (much smaller but nonzero) probability $\frac{m}{nk}$. Suppose we prepare and measure $T = 10m \log(m+1)$ copies of $|\psi\rangle$. Then the expected number of occurrences of each $i \in \bar{S}$ is $T \cdot \frac{k}{nm} \geq 5 \log(m+1)$ since $k \geq n/2$, while the expected number of occurrences of each $i \in S$ is $T \cdot \frac{m}{nk} = O(\log(n)/n^{3/2})$. In both cases the number of occurrences is tightly concentrated.² Hence if we keep only the elements that appear, say, at least $\log(m+1)$ times among the T outcomes, then with high probability we will have found \bar{S} , and hence learned $S = [n] \setminus \bar{S}$. ◀

3 Lower bound on quantum samples

In this section we prove lower bounds on the number of copies of $|S\rangle$ needed to identify S with high probability. Before establishing the lower bounds claimed in Table 1, we introduce some preliminary concepts, namely the γ_2 -norm (Section 3.1), association schemes (Section 3.2), the Johnson scheme (Section 3.3), and the adversary bound for state discrimination (Section 3.4). The lower bound itself is established in Section 3.5.

3.1 γ_2 -norm

The γ_2 -norm of a $D_1 \times D_2$ matrix A with entries $A(x, y)$ for $x \in [D_1]$ and $y \in [D_2]$ can be defined in two equivalent ways [12, Section 3]. The primal definition is

$$\begin{aligned} \text{minimise} \quad & \max \left\{ \max_{x \in [D_1]} \|u_x\|^2, \max_{y \in [D_2]} \|v_y\|^2 \right\} \\ \text{subject to} \quad & A(x, y) = \langle u_x, v_y \rangle \quad \text{for all } x \in [D_1] \text{ and } y \in [D_2], \end{aligned} \tag{1}$$

where $\{u_x : x \in [D_1]\}$ and $\{v_y : y \in [D_2]\}$ are vectors of the same dimension. The dual definition is

$$\begin{aligned} \text{maximise} \quad & \|\Gamma \circ A\| \\ \text{subject to} \quad & \|\Gamma\| \leq 1 \end{aligned} \tag{2}$$

where Γ ranges over $D_1 \times D_2$ -matrices, \circ denotes the Hadamard (entrywise) product of matrices, and $\|\cdot\|$ is the spectral norm of a matrix. Note that $\gamma_2(A \circ B) \leq \gamma_2(A)\gamma_2(B)$: consider the vectors obtained from the optimal feasible solutions of $\gamma_2(A)$ and $\gamma_2(B)$ in Eq. (1) and observe that the tensor product of these vectors forms a feasible solution for the primal problem for $\gamma_2(A \circ B)$ with (not necessarily minimal) value $\gamma_2(A)\gamma_2(B)$.

² Suppose we flip T 0/1-valued coins, each taking value 1 with probability p . Let X be their sum (i.e., the number of 1s), which has expectation $\mu = pT$. The Chernoff bound implies $\Pr[X \leq (1-\delta)\mu] \leq \exp(-\delta^2\mu/2)$. To get concentration for the number of occurrences of a specific $i \in \bar{S}$, apply this tail bound with $p = k/(nm)$, $\mu = Tp \geq 5 \log(m+1)$, $\delta = 4/5$ to obtain $\Pr[X \leq \log(m+1)] \ll 1/m$. Hence, by a union bound, the probability that among the m elements $i \in \bar{S}$ there is one of which we see fewer than $\log(m+1)$ occurrences, is $\ll 1$. For an $i \in S$, by Markov's inequality the probability to see at least $\log(m+1)$ occurrences of this i among the T samples is $\ll 1/n$, and we can use a union bound over all $i \in S$.

3.2 Association schemes

Here we present a quick introduction to association schemes (see, for example, [16, Chapter 1] for a more thorough treatment).

► **Definition 3.** An association scheme on the set U is a finite set of real symmetric $U \times U$ matrices $\{A_0, A_1, \dots, A_s\}$ satisfying all the following properties:

- each A_j only has entries 0 and 1;
 - A_0 is the identity matrix Id ;
 - $\sum_{j=0}^s A_j$ is the all-1 matrix J ; and
 - for every i and j , the product $A_i A_j$ is a linear combination of the matrices $\{A_0, \dots, A_s\}$.
- The space spanned by the set $\{A_0, A_1, \dots, A_s\}$ forms an algebra, which is called the Bose–Mesner algebra corresponding to the scheme. By abuse of terminology, we may also refer to this algebra as the association scheme.

We now state a few properties of $\{A_0, \dots, A_s\}$. First, observe that A_j has zero diagonal for $j > 0$. Additionally, $\{A_0, \dots, A_s\}$ form a basis of the corresponding Bose–Mesner algebra, since for every (x, y) , there is exactly one j for which $A_j(x, y) \neq 0$. Also, the basis $\{A_0, \dots, A_s\}$ satisfies $A_i \circ A_j = \mathbf{1}_{[i=j]} A_i$, where $\mathbf{1}_{[P]}$ is the indicator function of predicate P (i.e., 1 if P is true and 0 if P is false). It is possible to find another basis $\{E_0, \dots, E_s\}$ consisting of idempotent matrices for $\text{span}\{A_0, \dots, A_s\}$ that satisfy $E_i E_j = \mathbf{1}_{[i=j]} E_i$, with respect to the usual product of matrices. The operators E_i are orthogonal projectors onto the *eigenspaces* of the association scheme. We have

$$E_0 = J/N \quad \text{and} \quad \sum_{j=0}^s E_j = \text{Id},$$

where $N = |U|$. Since both $\{A_i\}$ and $\{E_j\}$ are bases for the space of $N \times N$ matrices, it is possible to write

$$A_i = \sum_{j=0}^s p_i(j) E_j \quad \text{and} \quad E_j = \sum_{i=0}^s \frac{q_j(i)}{N} A_i, \quad (3)$$

where $p_i(j)$ and $q_j(i)$ are called the *eigenvalues* and *dual eigenvalues* of the association scheme, respectively.

It is easy to show that the Hadamard product and the usual product of any two elements of the association scheme also belong to the association scheme. Clearly for every i, j we know that $A_i \circ A_j$ and $E_i \cdot E_j$ are elements of the basis of the scheme. Also observe that $A_i \cdot A_j$ and $E_i \circ E_j$ are elements of the scheme by writing out these products using Eq. (3) and observing that $A_i \cdot A_j$ (resp. $E_i \circ E_j$) is a linear combination of elements of $\{A_0, \dots, A_s\}$ (resp. $\{E_0, \dots, E_s\}$). In particular, we can write

$$E_i \circ E_j = \frac{1}{N} \sum_{\ell=0}^s q_{i,j}(\ell) E_\ell. \quad (4)$$

The real numbers $q_{i,j}(\ell)$ are called the *Krein parameters* of the association scheme.

3.3 Johnson scheme

In the Johnson association scheme $\mathcal{J}(n, k)$, the set U is the set of all k -subsets of $[n]$. Therefore, $N = |U| = \binom{n}{k}$. Let $m = \min\{k, n - k\}$. For $j = 0, 1, \dots, m$, define $A_j(x, y) := \mathbf{1}_{[|x \cap y| = k - j]}$.

The idempotent E_j is defined as follows: for $x \in U$, let $e_x \in \mathbb{R}^U$ be the indicator vector defined as $e_x(y) = \mathbf{1}_{[x=y]}$ for $y \in U$, and let

$$\mathcal{V}_j := \begin{cases} \text{span}\left\{\sum_{x \supseteq z} e_x : z \subseteq [n] \text{ with } |z| = j\right\} & \text{if } k \leq n/2, \\ \text{span}\left\{\sum_{x \subseteq z} e_x : z \subseteq [n] \text{ with } |z| = n - j\right\} & \text{if } k > n/2, \end{cases}$$

where the sums are over $x \in U$. These spaces satisfy $\mathcal{V}_0 \subset \mathcal{V}_1 \subset \dots \subset \mathcal{V}_m = \mathbb{R}^U$ and the dimension of \mathcal{V}_j is $\binom{n}{j}$. For $j \in \{1, \dots, m\}$, the idempotent E_j is defined as the orthogonal projector on $\mathcal{V}_j \cap \mathcal{V}_{j-1}^\perp$, and E_0 is the orthogonal projector on \mathcal{V}_0 . Hence, for $j \in \{0, 1, \dots, m\}$, the dimension of the j th eigenspace is

$$d_j := \text{Tr}[E_j] = \binom{n}{j} - \binom{n}{j-1}. \quad (5)$$

We do not require explicit expressions for most eigenvalues and valencies of $\mathcal{J}(n, k)$, the only exceptions being the dual eigenvalues

$$q_0(i) = 1 \quad \text{and} \quad q_1(i) = \frac{n(n-1)}{n-k} \left(\frac{k-i}{k} - \frac{k}{n} \right). \quad (6)$$

See [26, Eq. 1.24] for the latter. We are only interested in the following Krein parameters of this association scheme. When one idempotent is E_0 , we have

$$q_{i,0}(j) = \mathbf{1}_{[i=j]}. \quad (7)$$

When one idempotent is E_1 , we have

$$q_{j-1,1}(j) = \frac{j(n-1)n(k-j+1)(m-j+1)}{mk(n-2j+1)(n-2j+2)}, \quad (8a)$$

$$q_{j,1}(j) = \frac{j(n-1)(n-j+1)(m-k)^2}{mk(n-2j)(n-2j+2)}, \quad (8b)$$

$$q_{j+1,1}(j) = \frac{n(n-1)(n-j+1)(k-j)(m-j)}{mk(n-2j)(n-2j+1)}, \quad (8c)$$

and $q_{i,1}(j) = 0$ whenever $|i-j| > 1$ (see [9, Section 3.2]).

3.4 Adversary lower bound for state discrimination

Consider the following state-discrimination problem.

(*) Let $f: D \rightarrow R$ be a function for some finite sets D and R . Let $\{|\psi_x\rangle : x \in D\}$ be a family of quantum states of the same dimension. Given a copy of $|\psi_x\rangle$ for an arbitrary $x \in D$, the goal is to determine $f(x)$ with high success probability.

Let A be the Gram matrix of the states, namely

$$A(x, y) = \langle \psi_x | \psi_y \rangle,$$

and let F be the $D \times D$ matrix with

$$F(x, y) = \mathbf{1}_{[f(x) \neq f(y)]}.$$

Informally, the main result of this section is that the above state-discrimination problem can be solved with small error probability if and only if

$$\gamma_2(A \circ F)$$

is small. We start with the proof of the lower bound. With constants refined, it reads as follows:

► **Proposition 4.** *If the above state-discrimination problem (*) can be solved with success probability $1 - \varepsilon$, then $\gamma_2(A \circ F) \leq 4\sqrt{\varepsilon}$.*

Proof. This is essentially the result of [12, Claim 3.27], which is also closely related to [20]. For completeness we repeat the proof, with slight modifications.

Without loss of generality we may assume the measurement is projective (this follows from Neumark's theorem). Thus, there exist orthogonal projectors $\{\Pi_a\}_{a \in R}$ such that $\|\Pi_{f(x)} |\psi_x\rangle\|^2 \geq 1 - \varepsilon$ for all $x \in D$. Denote $\Pi_a^\perp = \text{Id} - \Pi_a$, so that $\|\Pi_{f(x)}^\perp |\psi_x\rangle\|^2 \leq \varepsilon$ for all $x \in D$. We first write

$$\begin{aligned} A(x, y) &= \langle \psi_x | \psi_y \rangle = \langle \psi_x | \Pi_{f(y)} |\psi_y\rangle + \langle \psi_x | \Pi_{f(y)}^\perp |\psi_y\rangle \\ &= \langle \psi_x | \Pi_{f(x)} \Pi_{f(y)} |\psi_y\rangle + \langle \psi_x | \Pi_{f(x)}^\perp \Pi_{f(y)} |\psi_y\rangle + \langle \psi_x | \Pi_{f(x)}^\perp \Pi_{f(y)}^\perp |\psi_y\rangle. \end{aligned}$$

Note that if $f(x) \neq f(y)$, then the first term is 0 because $\Pi_{f(x)}$ and $\Pi_{f(y)}$ project onto orthogonal subspaces. This motivates us to define the $D \times D$ matrix

$$B(x, y) = \langle \psi_x | \Pi_{f(x)}^\perp \Pi_{f(y)} |\psi_y\rangle + \langle \psi_x | \Pi_{f(y)}^\perp |\psi_y\rangle.$$

We have $A(x, y) = B(x, y)$ whenever $f(x) \neq f(y)$, and hence $A \circ F = B \circ F$. Note that $\gamma_2(B) \leq 2\sqrt{\varepsilon}$ by taking the vectors $u_x = (\varepsilon^{-1/4} \Pi_{f(x)}^\perp |\psi_x\rangle, \varepsilon^{1/4} |\psi_x\rangle)$ and $v_y = (\varepsilon^{1/4} \Pi_{f(y)} |\psi_y\rangle, \varepsilon^{-1/4} \Pi_{f(y)}^\perp |\psi_y\rangle)$. Now we have

$$\gamma_2(A \circ F) = \gamma_2(B \circ F) \leq \gamma_2(B) \gamma_2(F) \leq 4\sqrt{\varepsilon},$$

where we used the composition property of the γ_2 -norm in the first inequality and in the second inequality we used $\gamma_2(F) \leq 2$, which follows by considering the vectors $u_x, v_y \in \{0, 1\}^{|R|+1}$ whose last coordinate is always 1, and where u_x has a 1 at coordinate $f(x)$ and v_y has a -1 at coordinate $f(y)$ (identifying R with $\{1, \dots, |R|\}$ for the purposes of indexing these vectors), and whose remaining entries are all 0. ◀

► **Proposition 5.** *The above state-discrimination problem (*) can be solved with success probability at least $1 - \gamma_2(A \circ F)$.*

Proof. If B is the Gram matrix of the collection of states $\{|\psi_x\rangle \otimes |f(x)\rangle\}_{x \in D}$, then

$$A - B = A \circ F.$$

Using [22, Claim 3.10], there exists a unitary U such that

$$(\langle \psi_x | \otimes \langle f(x) |) U (|\psi_x\rangle \otimes |0\rangle) \geq 1 - \varepsilon/2$$

where $\varepsilon := \gamma_2(A \circ F)$. Thus, if we measure the second register of $U(|\psi_x\rangle \otimes |0\rangle)$, we get $f(x)$ with probability at least $(1 - \varepsilon/2)^2 \geq 1 - \varepsilon$. ◀

3.5 Lower bound

For $x \subseteq [n]$ of size k , let

$$|\psi_x\rangle = \frac{1}{\sqrt{k}} \sum_{i \in x} |i\rangle.$$

This is what we denoted by $|S\rangle$ earlier ($x = S$); we use $|\psi_x\rangle$ here for consistency with the common notation in lower bounds. The task is to identify the subset x using as few copies of the state $|\psi_x\rangle$ as possible. We prove the following lower bound.

► **Theorem 6.** *To find x with success probability $\Omega(1)$, it is necessary to have $\Omega(k \log(\min\{k, n - k\}))$ copies of the state $|\psi_x\rangle$.*

Let $m = n - k$. Since we could add more elements to the ambient space artificially, the problem becomes no easier as n grows with k fixed. Thus, it suffices to prove the lower bound of $\Omega(k \log(m + 1))$ under the assumption $m \ll k$.

Define the Gram matrix Ψ by $\Psi(x, y) = \langle \psi_x | \psi_y \rangle$. The Gram matrix corresponding to $|\psi_x\rangle^{\otimes \ell}$ is $\Psi^{\circ \ell}$ (where $\Psi^{\circ \ell}$ is the Hadamard product of Ψ with itself ℓ times). The function we want to compute is $f : x \mapsto x$, so we have $F(x, y) = \mathbf{1}_{[f(x) \neq f(y)]} = \mathbf{1}_{[x \neq y]}$, i.e., $F = J - \text{Id}$. By Proposition 4, it thus suffices to prove that for some $\ell = \Omega(k \log(m + 1))$ we have

$$\gamma_2(\Psi^{\circ \ell} \circ (J - \text{Id})) = \Omega(1).$$

To that end, we use the dual formulation of the γ_2 -norm (in Eq. (2)) and construct a matrix Γ such that

$$\|\Gamma\| = 1, \quad \Gamma \circ \text{Id} = 0, \quad \text{and} \quad \|\Gamma \circ \Psi^{\circ \ell}\| = \Omega(1).$$

We now construct a Γ that satisfies the constraints above. To do so, we first write Γ in terms of the idempotents $\{E_j\}_{j=0}^m$ of the Johnson association scheme (as defined above Eq. (5)): for $\{\gamma_j\}_j$ which we define shortly, let

$$\Gamma = \sum_{j=0}^m \gamma_j E_j. \tag{9}$$

To satisfy $\Gamma \circ \text{Id} = 0$, we would like Γ to have zero diagonal. Note that Γ has zero diagonal if and only if $\text{Tr}[\Gamma] = \sum_{j=0}^m \gamma_j \text{Tr}[E_j] = \sum_{j=0}^m \gamma_j d_j = 0$, where d_j was defined in Eq. (5). We now fix $\{\gamma_j\}_j$ as follows: since $d_m = \binom{n}{m} - \binom{n}{m-1}$ is larger than the sum of the remaining d_j s, we let

$$\gamma_0 = \gamma_1 = \dots = \gamma_{m-1} = 1, \quad \gamma_m \in [-1, 0] \tag{10}$$

so that $\text{Tr}[\Gamma] = 0$ and $\|\Gamma\| = 1$. Thus, it remains to show that

$$\|\Gamma \circ \Psi^{\circ \ell}\| = \Omega(1). \tag{11}$$

For that, we use the following technical result.

► **Lemma 7.** *For each $j = 0, 1, \dots, m$, we have*

$$E_j \circ \Psi = p_{j+1,-1} E_{j+1} + p_{j,0} E_j + p_{j-1,+1} E_{j-1},$$

where

$$\begin{aligned} p_{j,-1} &= \frac{j(k-j+1)(m-j+1)}{(n-2j+1)(n-2j+2)k}, \\ p_{j,0} &= \frac{k}{n} + \frac{j(n-j+1)(m-k)^2}{nk(n-2j)(n-2j+2)}, \\ p_{j,+1} &= \frac{(n-j+1)(k-j)(m-j)}{(n-2j)(n-2j+1)k}. \end{aligned}$$

Before we proceed with the proof of this lemma, let us state a simple consequence.

10:10 Quantum Coupon Collector

► **Corollary 8.** For each $j \in \{0, \dots, m\}$, the numbers $p_{j,-1}$, $p_{j,0}$, and $p_{j,+1}$ are non-negative, and satisfy $p_{j,-1} + p_{j,0} + p_{j,+1} = 1$.

Proof. The non-negativity is obvious. For the last property note that

$$\sum_{j=0}^m E_j = \text{Id} = \Psi \circ \text{Id} = \Psi \circ \left(\sum_{j=0}^m E_j \right) = \sum_{j=0}^m (p_{j,-1} + p_{j,0} + p_{j,1}) E_j,$$

where the first equality uses the definition of an association scheme, the second equality follows because $\Psi(x, x) = 1$ by definition, and the last equality is by the assumption of Lemma 7. ◀

Proof of Lemma 7. It suffices to write out Ψ in the basis $\{E_j\}_{j=0}^m$ and use the Krein parameters. By definition of $|\psi_x\rangle = \frac{1}{\sqrt{k}} \sum_{i \in x} |i\rangle$, we have that $\Psi(x, y) = \langle \psi_x | \psi_y \rangle$ equals $\frac{1}{k}$ times the intersection of x and y , and

$$\Psi = \sum_{i=0}^m \left(1 - \frac{i}{k}\right) A_i,$$

where A_i was defined at the beginning of Section 3.3 as $A_i(x, y) := \mathbf{1}_{|x \cap y| = k-i}$. We now rewrite Ψ as follows: using Eq. (6), we have

$$\frac{k}{n} E_0 + \frac{n-k}{n(n-1)} E_1 = \frac{1}{N} \sum_{i=0}^m \left(\frac{k}{n} q_0(i) + \frac{n-k}{n(n-1)} q_1(i) \right) A_i = \frac{1}{N} \sum_{i=0}^m \frac{k-i}{k} A_i = \frac{1}{N} \Psi,$$

where the first equality used Eq. (3). Additionally observe that

$$N E_j \circ E_0 = q_{j,0}(j) E_j \quad \text{and} \quad N E_j \circ E_1 = q_{j,1}(j-1) E_{j-1} + q_{j,1}(j) E_j + q_{j,1}(j+1) E_{j+1}.$$

Plugging in the values of $q_{j,\cdot}$ from Eq. (8), we get the required equality. ◀

We are now ready to prove our main lower bound in Theorem 6.

Proof of Theorem 6. We prove this by induction on the number of copies of the state $|\psi_x\rangle$, which we denote by s . Let us define $\gamma_j^{(s)}$ via

$$\Gamma \circ \Psi^{\circ s} = \sum_{j=0}^m \gamma_j^{(s)} E_j.$$

Since the E_j are pairwise-orthogonal projections, the norm of $\Gamma \circ \Psi^{\circ s}$ equals $\max_j |\gamma_j^{(s)}|$. Hence to lower bound $\|\Gamma \circ \Psi^{\circ s}\|$, it suffices to lower bound $\gamma_0^{(s)}$.

We have

$$\Gamma \circ \Psi^{\circ(s+1)} = \sum_{j=0}^m \gamma_j^{(s)} E_j \circ \Psi$$

and using Lemma 7 we get

$$\gamma_j^{(s+1)} = p_{j,-1} \gamma_{j-1}^{(s)} + p_{j,0} \gamma_j^{(s)} + p_{j,+1} \gamma_{j+1}^{(s)}. \quad (12)$$

For every $j \in \{0, \dots, m\}$, we now consider the following probabilistic sequence $\{B_j^{(s)}\}$. For $s = 0$, we let $B_j^{(0)} = \gamma_j$ and

$$B_j^{(s+1)} = \begin{cases} B_{j-1}^{(s)} & \text{with probability } p_{j,-1}, \\ B_j^{(s)} & \text{with probability } p_{j,0}, \\ B_{j+1}^{(s)} & \text{with probability } p_{j,+1}, \end{cases}$$

using the fact that $p_{j,-1} + p_{j,0} + p_{j,+1} = 1$. Note that $B_j^{(s)}$ only takes values from $\{\gamma_0, \dots, \gamma_m\}$ and there are only two distinct such values, namely 1 and γ_m (since $\gamma_0 = \gamma_1 = \dots = \gamma_{m-1} = 1$ as defined in Eq. (10)). Also note that $p_{0,-1} = p_{m,+1} = 0$, so we do not have to explicitly handle the boundaries. Induction on s using Eq. (12) shows that $\mathbb{E}[B_j^{(s)}] = \gamma_j^{(s)}$, which is the motivation behind defining these variables.

Define similarly $C_j^{(s)}$ as $C_j^{(0)} = \gamma_j$ and

$$C_j^{(s+1)} = \begin{cases} C_j^{(s)} & \text{with probability } p_{j,-1} + p_{j,0}, \\ C_{j+1}^{(s)} & \text{with probability } p_{j,+1}. \end{cases}$$

Let us give an intuitive description of how the random variables $C_j^{(s)}$ behave. For each s , the head of the sequence $C_0^{(s)}, C_1^{(s)}, \dots$, up to some $C_\ell^{(s)}$ consists purely of 1s, and the tail $C_{\ell+1}^{(s)}, \dots, C_m^{(s)}$ consists purely of γ_m . Initially, for $s = 0$, the tail consists of one element $C_m^{(0)}$ only, but the tail gradually extends as s grows (and the head, respectively, shrinks). The probability of growing the length of the tail from $m - j$ to $m - j + 1$ in one step is $p_{j,+1}$.

The random variables $B_j^{(s)}$ behave similarly, but are slightly more complicated, since the tail can also shrink and 1s can get into the tail. This is the reason why we replace $B_j^{(s)}$ with $C_j^{(s)}$ in our analysis: $C_j^{(s)}$ is easier to analyze, and it suffices to lower bound its expectation because $B_j^{(s)}$ dominates $C_j^{(s)}$, i.e., for each s and j and real t we have $\Pr[B_j^{(s)} \geq t] \geq \Pr[C_j^{(s)} \geq t]$. The latter is proven by induction, as follows. The base case $s = 0$ is trivial, and the inductive step is

$$\begin{aligned} \Pr[B_j^{(s+1)} \geq t] &= p_{j,-1} \Pr[B_{j-1}^{(s)} \geq t] + p_{j,0} \Pr[B_j^{(s)} \geq t] + p_{j,+1} \Pr[B_{j+1}^{(s)} \geq t] \\ &\geq p_{j,-1} \Pr[C_{j-1}^{(s)} \geq t] + p_{j,0} \Pr[C_j^{(s)} \geq t] + p_{j,+1} \Pr[C_{j+1}^{(s)} \geq t] \\ &\geq (p_{j,-1} + p_{j,0}) \Pr[C_j^{(s)} \geq t] + p_{j,+1} \Pr[C_{j+1}^{(s)} \geq t] = \Pr[C_j^{(s+1)} \geq t], \end{aligned}$$

since $C_{j-1}^{(s)} \geq C_j^{(s)}$ by our above analysis.

The analysis of $C_j^{(s)}$ is very similar to the classical coupon collector problem if we interpret the length of the tail as the number of acquired coupons. We briefly repeat the argument. For each j , define random variable T_j as the first value of s such that $C_j^{(s)} = \gamma_m$. Obviously, $T_m = 0$. We can interpret T_j as the first value of s such that the length of the tail becomes $m - j + 1$. The random variable $T_j - T_{j+1}$ is the number of steps required to grow the length of the tail from $m - j$ to $m - j + 1$. Clearly, these variables are independent for different j . Also, each of them is distributed according to a geometric distribution and standard probability theory gives us that $\mathbb{E}[T_j - T_{j+1}] = 1/p_{j,+1}$ and $\text{Var}[T_j - T_{j+1}] = (1 - p_{j,+1})/p_{j,+1}^2$. We have $p_{j,+1} = \Theta((m - j)/k)$ from Lemma 7, so

$$\mathbb{E}[T_0] = \sum_{j=0}^{m-1} \frac{1}{p_{j,+1}} = \Theta(k) \left(\sum_{j=0}^{m-1} \frac{1}{m - j} \right) = \Theta(k \log(m + 1)).$$

10:12 Quantum Coupon Collector

Similarly,

$$\text{Var}[T_0] = \sum_{j=0}^{m-1} \frac{1 - p_{j+1}}{p_{j+1}^2} = \Theta(k^2) \left(\sum_{j=0}^{m-1} \frac{1}{(m-j)^2} \right) = \Theta(k^2).$$

Hence, using Chebyshev's inequality, there exists $\ell = \Theta(k \log(m+1))$ such that

$$\Pr[T_0 > \ell] \geq 3/4.$$

Since $C_0^{(\ell)}$ can take only two values (1 and $\gamma_m \in [-1, 0]$), we have that

$$\gamma_0^{(\ell)} = \mathbb{E}[C_0^{(\ell)}] \geq 3/4 \cdot 1 + 1/4 \cdot \gamma_m \geq 1/2.$$

Finally, since $B_0^{(\ell)}$ dominates $C_0^{(\ell)}$, we get

$$\gamma_0^{(\ell)} = \mathbb{E}[B_0^{(\ell)}] \geq \mathbb{E}[C_0^{(\ell)}] \geq 1/2,$$

implying Eq. (11). This shows the existence of $\ell = \Theta(k \log(m+1))$ such that the error probability of any measurement on ℓ copies of $|\psi_x\rangle$ has error probability $\Omega(1)$ in identifying x . ◀

4 Learning from quantum samples and reflections

In the previous sections we assumed we were given a number of copies of the unknown state $|S\rangle$. In this section we assume a stronger model: in addition to a number of copies of the state $|S\rangle$, we are also given the ability to apply the reflection $R_S = 2|S\rangle\langle S| - \text{Id}$ through $|S\rangle$. The key additional tool we will use is (exact) amplitude amplification, encapsulated by the next theorem, which follows from [14]:

► **Theorem 9** (Exact amplitude amplification). *Let $|\phi\rangle$ and $|\psi\rangle$ be states such that $\langle\phi|\psi\rangle = \alpha > 0$. Suppose we know α exactly, and we can implement reflections through $|\phi\rangle$ and $|\psi\rangle$. Then we can convert $|\phi\rangle$ into $|\psi\rangle$ (exactly) using $O(1/\alpha)$ reflections and $\tilde{O}(1/\alpha)$ other gates.*

We distinguish the two regimes of $k \geq n/2$ and $k < n/2$.

4.1 Tight bound if $k \geq n/2$

► **Theorem 10** (Upper bound for small m). *Let $S \subseteq [n]$ be a set of size $k \geq n/2$ and let $m = n - k$. We can identify S with probability 1 using $O(\sqrt{km})$ uses of $R_S = 2|S\rangle\langle S| - \text{Id}$.*

Proof. Our algorithm sequentially finds all m missing elements. We would like to use amplitude amplification to prepare a copy of $|\bar{S}\rangle$, which is the uniform superposition over the m missing elements. Consider the uniform state over the n -element universe:

$$|[n]\rangle = \sqrt{\frac{k}{n}} |S\rangle + \sqrt{\frac{m}{n}} |\bar{S}\rangle.$$

This state is easy to prepare, and hence also easy to reflect through. Note that in the 2-dimensional plane spanned by $|S\rangle$ and $|\bar{S}\rangle$, reflection through $|\bar{S}\rangle$ is the same as a reflection through $|S\rangle$ up to an irrelevant global phase. The inner product between $|[n]\rangle$ and $|\bar{S}\rangle$ equals $\sqrt{m/n}$. Accordingly, using $O(\sqrt{n/m})$ rounds of exact amplitude amplification (which only rotates in the 2-dimensional space spanned by $|S\rangle$ and $|\bar{S}\rangle$; each round “costs” one application of R_S) we can turn $|[n]\rangle$ into $|\bar{S}\rangle$, up to a global phase.

Measuring $|\bar{S}\rangle$ gives us one of the missing elements, uniformly at random. Now we remove this element from the universe. Note that $|S\rangle$ does not change since we removed an element of the universe that was missing from S . We then repeat the above algorithm on a universe of size $n - 1$ with $m - 1$ missing elements in order to find another missing element at the cost of $O(\sqrt{(n-1)/(m-1)})$ rounds of amplitude amplification, and so on. This finds all missing elements (and hence S) with probability 1, using

$$\sum_{i=0}^{m-1} O\left(\sqrt{\frac{n-i}{m-i}}\right) = O(\sqrt{n}) \sum_{j=1}^m \frac{1}{\sqrt{j}} = O(\sqrt{nm}) = O(\sqrt{km})$$

applications of R_S , where we used $k \geq n/2$. Note that in this regime we do not need any copies of $|S\rangle$, just reflections R_S . ◀

► **Theorem 11** (Lower bound for small m). *Let $S \subseteq [n]$ be a set of size $k < n$ and let $m = n - k$. Any quantum algorithm that identifies S with high probability using a total of T copies of $|S\rangle$ and uses of R_S , must satisfy $T = \Omega(\sqrt{km})$. The lower bound holds even if we allow T copies of $|S\rangle$, uses of R_S , and membership queries to S .*

Proof. We prove a matching lower bound in a stronger model, namely in a model where we can make queries to the n -bit characteristic vector x for S . That is, we now assume we have a unitary U_S that maps

$$U_S : |i, b\rangle \mapsto |i, b \oplus x_i\rangle \quad \text{for all } i \in [n], b \in \{0, 1\},$$

where $x_i = 1$ iff $i \in S$.

We first argue that this is indeed a stronger model, by showing how we can unitarily prepare a copy of $|S\rangle$ using $O(1)$ applications of U_S . Note that $\langle [n] | S \rangle = \sqrt{k/n} \geq 1/\sqrt{2}$ under the current assumption that $k \geq n/2$. Also note that, in the 2-dimensional space spanned by $|S\rangle$ and $|\bar{S}\rangle$, a reflection through $|S\rangle$ corresponds to a “phase query” to x , which can be implemented by one query to U_S (setting the target qubit to $(|0\rangle - |1\rangle)/\sqrt{2}$). Hence using $O(1)$ rounds of exact amplitude amplification suffices to prepare a copy of $|S\rangle$ starting from the state $|[n]\rangle$, which is easy to prepare and reflect through. Thus we can implement the state-preparation map $G_S : |0\rangle \mapsto |S\rangle$ using $O(1)$ applications of U_S . Note that one application of G_S^{-1} , followed by a reflection through $|0\rangle$ and an application of G_S , implements a reflection through $|S\rangle$. Thus preparing a copy of $|S\rangle$ and reflecting through $|S\rangle$ each “cost” only $O(1)$ queries to x (i.e., applications of U_S).

Accordingly, an algorithm that learns S using at most T copies of $|S\rangle$ and at most T applications of R_S implies a quantum algorithm that can learn an n -bit string x of weight $k \geq n/2$ using $O(T)$ queries to x . But it is known that this requires $\Omega(\sqrt{nm}) = \Omega(\sqrt{km})$ queries to x , even when allowing bounded error probability. This follows, for instance, from [11, Theorem 4.10]. Hence we obtain the same lower bound on the number of copies of $|S\rangle$ plus the number of reflections through $|S\rangle$. ◀

4.2 Tight bound if $k < n/2$

► **Theorem 12** (Upper bound for small k). *Let $S \subseteq [n]$ be a set of size $k < n$. We can identify S with probability 1 using $O(k)$ copies of $|S\rangle$ and uses of $R_S = 2|S\rangle\langle S| - \text{Id}$.*

Proof. Our algorithm sequentially finds all elements of S . We start with a copy of $|S\rangle$ and measure to find one $i_1 \in S$. Then we use exact amplification to convert a fresh copy of $|S\rangle$ into $|S \setminus \{i_1\}\rangle$. This requires being able to reflect through $|S\rangle$ (i.e., apply R_S), and reflect

10:14 Quantum Coupon Collector

through $|S \setminus \{i_1\}\rangle$. In the 2-dimensional plane spanned by $|S\rangle$ and $|S \setminus \{i_1\}\rangle$, the latter reflection is equivalent to putting a minus in front of $|i_1\rangle$, which is easy to do. We measure $|S \setminus \{i_1\}\rangle$ and learn (with probability 1) another element $i_2 \in S \setminus \{i_1\}$. Then we change a fresh copy of $|S\rangle$ into $|S \setminus \{i_1, i_2\}\rangle$, measure, and learn some $i_3 \in S \setminus \{i_1, i_2\}$. We repeat this until we have seen all k elements.

The amplitude amplifications get more costly as we find more elements of S : If we have already found a set $I \subseteq S$, then changing a fresh copy of $|S\rangle$ to $|S \setminus I\rangle$ uses $O(\frac{1}{\sqrt{|S \setminus I|}}) = O(\sqrt{k/(k - |I|)})$ reflections, and hence $O(\sqrt{k/(k - |I|)})$ applications of R_S . Overall, this procedure finds S using $k = |S|$ copies of $|S\rangle$, and

$$\sum_{i=0}^{k-1} O\left(\sqrt{\frac{k}{k-i}}\right) = O(\sqrt{k}) \sum_{j=1}^k \frac{1}{\sqrt{j}} = O(k)$$

applications of R_S . ◀

► **Theorem 13** (Lower bound for small k). *Let $S \subseteq [n]$ be a set of size $k < n$. Any quantum algorithm that identifies S with high probability using a total of T copies of $|S\rangle$, and uses of R_S must satisfy $T = \Omega(k)$. The lower bound holds even if we allow T copies of $|S\rangle$, uses of R_S , and membership queries to S .*

Proof. To prove a matching lower bound, suppose our algorithm receives advice in the form of $n - 2k$ of the missing elements. This advice reduces the problem to one with universe size $n' = n - (n - 2k) = 2k$ and $m' = m - (n - 2k) = k$ missing elements. Importantly, note that $|S\rangle$, and hence R_S , do not change after learning these missing elements. But in Theorem 11 we already proved an $\Omega(\sqrt{n'm'}) = \Omega(k)$ lower bound on the number of copies of $|S\rangle$, reflections, and queries to S needed to solve this special case. Since the extra advice cannot have made the original problem harder, the same lower bound applies to our original problem. ◀

5 Proper PAC learning

As mentioned briefly in the introduction, one of the motivations for this research is the question whether the sample complexity of *proper* quantum PAC learning is higher than that of improper PAC learning. Let us precisely define Valiant's PAC model [28]. We are trying to learn an unknown element f from a *concept class* \mathcal{C} . For simplicity we only consider f s that are Boolean-valued functions on $[n]$. Our access to f is through random examples, which are pairs of the form $(x, f(x))$, where x is distributed according to a distribution $D: [n] \rightarrow [0, 1]$ that is unknown to the learner. A learning algorithm takes a number T of such i.i.d. examples as input, and produces a hypothesis $h: [n] \rightarrow \{0, 1\}$ that is supposed to be close to the target function f . The error of the hypothesis h (with respect to the target f , under distribution D) is defined as

$$\text{err}_D(f, h) := \Pr_{x \sim D}[f(x) \neq h(x)].$$

We say that a learning algorithm is an (ε, δ) -PAC learner for \mathcal{C} , if it probably (i.e., with probability at least $1 - \delta$) outputs an approximately correct (i.e., with error at most ε) hypothesis h :

$$\forall f \in \mathcal{C}, \forall D : \Pr[\text{err}_D(f, h) > \varepsilon] \leq \delta,$$

where the probability is taken over the sequence of T D -distributed examples that the learner receives, as well as over its internal randomness. The (ε, δ) -PAC *sample complexity* of \mathcal{C} is the minimal T for which such a learning algorithm exists.³

The PAC sample complexity of \mathcal{C} is essentially determined by its VC-dimension d as⁴

$$\Theta\left(\frac{d}{\varepsilon} + \frac{\log(1/\delta)}{\varepsilon}\right). \quad (13)$$

See Blumer et al. [13] for the lower bound and Hanneke [18] for the upper bound.

The above upper bound on sample complexity allows the learner to be improper, i.e., to sometimes output hypotheses $h \notin \mathcal{C}$. The following folklore example, which we learned from Steve Hanneke [19], shows that the sample complexity of *proper* learning can be asymptotically larger.⁵ Consider the concept class $\mathcal{C} = \{f: [n] \rightarrow \{0, 1\} \mid \exists! i \text{ s.t. } f(i) = 0\}$ of functions that are all-1 except on one “missing element” i . The VC-dimension of this class is 1, hence $\Theta(\frac{\log(1/\delta)}{\varepsilon})$ classical examples are necessary and sufficient for PAC learning \mathcal{C} by (13). With $\varepsilon = 1/n$ and $\delta = 1/3$, this bound becomes $\Theta(n)$. Now fix an (ε, δ) -PAC *proper* learner for this class that uses some T examples; we will show that $T = \Omega(n \log n)$, exhibiting an asymptotic separation between the sample complexities of proper and improper PAC learning.

For every $i \in [n]$, consider a distribution D_i that is uniform over $[n] \setminus \{i\}$. If the target concept f has i as its missing element then the learner has to output that f , since any other $g \in \mathcal{C}$ will make an error on its own missing element and hence would have error at least $1/(n-1) > \varepsilon$ under D_i . In other words, when sampling from D_i the learner has to identify the one missing element i with success probability $\geq 2/3$. But we know from the coupon collector argument that this requires $\Omega(n \log n)$ samples. Note that a D_i -distributed $(x, f(x))$ is equivalent to sampling uniformly from $[n] \setminus \{i\}$, since the label $f(x)$ is always 1 under D_i .

What about *quantum* PAC learning? Bshouty and Jackson [15] generalized the PAC model by considering superposition states

$$|\psi_{D,f}\rangle = \sum_x \sqrt{D(x)} |x, f(x)\rangle$$

instead of random samples. The learner now receives T copies of this “quantum example” state, and has to output a probably approximately correct hypothesis. Measuring a quantum example gives a classical example, so quantum examples are at least as useful as classical examples, but one of the questions in quantum learning theory is in what situations they are significantly more useful. Two of us [6] have shown that the bound of (13) also applies to learning from quantum examples, so for improper learning the quantum and classical sample complexities are equal up to constant factors. However, quantum examples *are* beneficial for learning \mathcal{C} under the D_i distributions. Note that $|\psi_{D_i, f_i}\rangle$ is just the uniform superposition over the set $S = [n] \setminus \{i\}$, tensored with an irrelevant extra $|1\rangle$. As we showed in Section 2, given $O(n)$ copies of $|\psi_{D_i, f_i}\rangle$ we can identify the one missing element i with probability $\geq 2/3$. So the example that separates the sample complexities of *classical* proper and improper

³ This definition uses the information-theoretic notion of *sample* complexity. We do not consider the *time* complexity of learning here. For more on sample and time complexity of quantum learning, we refer the reader to [5].

⁴ The VC-dimension of \mathcal{C} is the maximum size among all sets $T \subseteq [n]$ that are “shattered” by \mathcal{C} . A set T is shattered by \mathcal{C} if for all $2^{|T|}$ labelings $\ell: T \rightarrow \{0, 1\}$ of the elements of T , there is an $f \in \mathcal{C}$ that has that labeling (i.e., where $f|_T = \ell$).

⁵ In a recent result, Montasser et al. [24] proved another separation between proper and improper learning.

learning, does not separate *quantum* proper and improper learning. This naturally raises the question of whether the quantum sample complexities of proper and improper PAC learning are asymptotically equal (which, as mentioned, they provably are not in the classical case).

References

- 1 S. Aaronson, R. Kothari, W. Kretschmer, and J. Thaler. Quantum lower bounds for approximate counting via Laurent polynomials. arXiv:1904.08914, supersedes arXiv:1808.02420 and arXiv:1902.02398, 2019.
- 2 D. Aharonov and A. Ta-Shma. Adiabatic quantum state generation and statistical zero knowledge. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing, STOC '03*, page 20–29, 2003. doi:10.1145/780542.780546.
- 3 A. Ambainis, L. Magnin, M. Roetteler, and J. Roland. Symmetry-assisted adversaries for quantum state generation. In *2011 IEEE 26th Annual Conference on Computational Complexity*, pages 167–177, June 2011. doi:10.1109/CCC.2011.24.
- 4 S. Arunachalam, S. Chakraborty, T. Lee, M. Paraashar, and R. de Wolf. Two new results about quantum exact learning. In *46th International Colloquium on Automata, Languages, and Programming, ICALP*, pages 16:1–16:15, 2019. arXiv:1810.00481.
- 5 S. Arunachalam and R. de Wolf. Guest column: A survey of quantum learning theory. *SIGACT News*, 48(2):41–67, 2017. arXiv:1701.06806.
- 6 S. Arunachalam and R. de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19, 2018. Earlier version in CCC'17. arXiv:1607.00932.
- 7 A. Atıcı and R. Servedio. Improved bounds on quantum learning algorithms. *Quantum Information Processing*, 4(5):355–386, 2005. quant-ph/0411140.
- 8 A. Atıcı and R. Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2009. arXiv:0707.3479.
- 9 E. Bannai and T. Itô. *Algebraic Combinatorics I: Association Schemes*. Mathematics lecture note series. Benjamin/Cummings Pub. Co., 1984.
- 10 H. Barnum and E. Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43:2097–2106, 2002. quant-ph/0004088.
- 11 R. Beals, H. Buhrman, R. Cleve, M. Mosca, and R. de Wolf. Quantum lower bounds by polynomials. *jacm*, 48(4):778–797, 2001. Earlier version in FOCS'98. quant-ph/9802049.
- 12 A. Belovs. *Applications of the Adversary Method in Quantum Query Algorithms*. PhD thesis, University of Latvia, 2014. 1402.3858.
- 13 A. Blumer, A. Ehrenfeucht, D. Haussler, and M. K. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *JACM*, 36(4):929–965, 1989.
- 14 G. Brassard, P. Høyer, M. Mosca, and A. Tapp. Quantum amplitude amplification and estimation. In *Quantum Computation and Quantum Information: A Millennium Volume*, volume 305 of *AMS Contemporary Mathematics Series*, pages 53–74. AMS, 2002. quant-ph/0005055.
- 15 N. H. Bshouty and J. C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. *SIAM Journal on Computing*, 28(3):1136–1153, 1999. Earlier version in COLT'95.
- 16 C. Godsil. Association schemes. Available at <https://www.math.uwaterloo.ca/~cgodsil/assocs/pdfs/Assoc.pdf>, 2018.
- 17 A. B. Grilo, I. Kerenidis, and T. Zijlstra. Learning with Errors is easy with quantum samples. *Physical Review Letters A*, 99:032314, 2019. arXiv: 1702.08255.
- 18 S. Hanneke. The optimal sample complexity of PAC learning. *Journal of Machine Learning Research*, 17(38):1–15, 2016. arXiv:1507.00473.
- 19 S. Hanneke. Personal communication with Srinivasan Arunachalam, February 2018. See also <https://cstheory.stackexchange.com/questions/40161/proper-pac-learning-vc-dimension-bounds>.

- 20 P. Høyer, T. Lee, and R. Špalek. Negative weights make adversaries stronger. In *Proceedings of 39th ACM STOC*, pages 526–535, 2007. quant-ph/0611054.
- 21 T. Lee, R. Mittal, B. Reichardt, R. Špalek, and M. Szegedy. Quantum query complexity of state conversion. In *Proceedings of 52nd IEEE FOCS*, pages 344–353, 2011. arXiv:1011.3020.
- 22 T. Lee and J. Roland. A strong direct product theorem for quantum query complexity. *Computational Complexity*, 22(2):429–462, 2013. Earlier version in CCC’12. arXiv:1104.4468.
- 23 N. Lindzey and A. Rosmanis. A Tight Lower Bound For Non-Coherent Index Erasure. In *11th Innovations in Theoretical Computer Science Conference (ITCS 2020)*, volume 151 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 59:1–59:37, 2020. doi:10.4230/LIPIcs.ITCS.2020.59.
- 24 O. Montasser, S. Hanneke, and N. Srebro. VC classes are adversarially robustly learnable, but only improperly. *Journal of Machine Learning Research*, 99:1–19, 2019.
- 25 R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge International Series on Parallel Computation. Cambridge University Press, 1995. doi:10.1017/CB09780511814075.
- 26 A. Rosmanis. *Lower Bounds on Quantum Query and Learning Graph Complexities*. PhD thesis, University of Waterloo, July 2014. Available at <http://hdl.handle.net/10012/8577>.
- 27 R. Servedio and S. Gortler. Equivalences and separations between quantum and classical learnability. *SIAM Journal on Computing*, 33(5):1067–1092, 2004. Combines earlier papers from ICALP’01 and CCC’01. quant-ph/0007036.
- 28 L. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984.

Fast and Effective Techniques for T-Count Reduction via Spider Nest Identities

Niel de Beaudrap 

Department of Computer Science, University of Oxford, United Kingdom
niel.debeaudrap@cs.ox.ac.uk

Xiaoning Bian

Department of Mathematics & Statistics, Dalhousie University, Halifax, Canada
bian@dal.ca

Quanlong Wang

Department of Computer Science, University of Oxford, United Kingdom
Cambridge Quantum Computing Ltd., Cambridge, United Kingdom
quanlong.wang@cs.ox.ac.uk

Abstract

In fault-tolerant quantum computing systems, realising (approximately) universal quantum computation is usually described in terms of realising Clifford+T operations, which is to say a circuit of CNOT, Hadamard, and $\pi/2$ -phase rotations, together with T operations ($\pi/4$ -phase rotations). For many error correcting codes, fault-tolerant realisations of Clifford operations are significantly less resource-intensive than those of T gates, which motivates finding ways to realise the same transformation involving T -count (the number of T gates involved) which is as low as possible. Investigations into this problem [5, 21, 4, 3, 10, 6] has led to observations that this problem is closely related to NP-hard tensor decomposition problems [23] and is tantamount to the difficult problem of decoding exponentially long Reed-Muller codes [6]. This problem then presents itself as one for which must be content in practise with approximate optimisation, in which one develops an array of tactics to be deployed through some pragmatic strategy. In this vein, we describe techniques to reduce the T -count, based on the effective application of “spider nest identities”: easily recognised products of parity-phase operations which are equivalent to the identity operation. We demonstrate the effectiveness of such techniques by obtaining improvements in the T -counts of a number of circuits, in run-times which are typically less than the time required to make a fresh cup of coffee.

2012 ACM Subject Classification Computer systems organization → Quantum computing

Keywords and phrases T-count, Parity-phase operations, Phase gadgets, Clifford hierarchy, ZX calculus

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.11

Supplementary Material The software which produced our results may be found on GitHub, at <https://github.com/njross/optimizer>, commit 46b8ce0873ff09bf54cf704080b7daa252c48eba.

Funding N. de Beaudrap was supported in part by a Fellowship funded by a gift from Tencent Holdings (tencent.com), and by the EPSRC National Hub in Networked Quantum Information Technologies (NQIT.org). X. Bian is supported by NSERC and by AFOSR under Award No. FA9550-15-1-0331. Q. Wang is supported by Cambridge Quantum Computing Ltd. and by the AFOSR grant FA2386-18-1-4028. Our results were made possible in part by the use of the Dalhousie University Mathstat Cluster [11].

Acknowledgements We thank Earl Campbell, Luke Heyfron, Alexander Cowtan, Aleks Kissinger, and John van de Wetering for helpful discussions. We extend a very special thanks to Matthew Amy, who wrote a small extension of `feynver` [2] to allow verification of procedures which post-select the $|+\rangle$ state, for the express purpose of helping us to independently verify the correctness of reductions such as appear in this work and in Ref. [14]. X. Bian would like to thank his Ph.D. supervisor Peter Selinger for his support.



© Niel de Beaudrap, Xiaoning Bian, and Quanlong Wang;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 11; pp. 11:1–11:23



Leibniz International Proceedings in Informatics

LIPICs Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

To achieve practical scalable quantum computation, it is important to find effective (both useful and efficient) techniques to reduce the resources required to perform computations. Error correction, and in particular realising operations in a fault-tolerant way, is expected to be a particularly significant source of resource overheads. In most quantum error-correcting codes, Clifford group operations involve less overhead than non-Clifford gates, such as the T (or $\pi/4$ phase-rotation) gate. As the set of Clifford+ T circuits is approximately universal for quantum computation [32], this motivates the T -count — or the number of T gates — as a quantity of interest in the resources required to realise a quantum computation.

On the other hand, in order to test the effectiveness of quantum technologies, it is helpful to be able to simulate the outcomes of quantum computations inasmuch as this is feasible. As circuits of Clifford operations can be efficiently simulated [22, 1], this motivates the approach of simulating quantum circuits by extending those efficient simulation techniques [9, 8], this again motivates the T -count as a measure of interest in the complexity of quantum circuits.

In this article, we consider the problem of reducing the T -count required to represent a unitary circuit provided as input. Following Heyfron and Campbell [23], we consider transformations of circuits which isolate a subcircuit of diagonal operations which is the only part of the algorithm with non-trivial T -count. The approach of Heyfron and Campbell [23] is to transform Clifford+ T circuits, to circuits with the following structure:

1. An initial stage of CNOT gates; followed by
2. A stage of diagonal non-Clifford operations; followed by
3. A sequence of (possibly classically controlled) Clifford operations.

This allows Ref. [23] to reduce the problem of T -count reduction to an analysis of the diagonal non-Clifford portion of this circuit, in terms of *phase polynomials*. This builds on a sequence of results which revolve around such operations [5, 21, 4, 3, 10, 6] presented in various but similar ways, and in particular establishes a connection between T -count optimisation and difficult coding problems and tensor decomposition problems [6, 23]. Our approach is to elaborate on that of Campbell and Heyfron as follows:

- Reduce the complexity of the diagonal non-Clifford operation by more flexible (but essentially elementary) separation of the circuit into stages by allowing the first stage to contain arbitrary Clifford gates;
- Analyse the diagonal non-Clifford portion of the circuit directly in terms of “ $\pi/4$ -parity-phase operations” — essentially operators of the form $\exp(i\frac{\pi}{8}(Z \otimes \cdots \otimes Z))$ — rather than as phase polynomials, simplifying them through the efficient application of identities of such operations.

We call these “ $\pi/4$ -parity-phase operations” as they induce a $e^{i\pi/4}$ relative phase on standard basis states, depending on some parity computation $f(x) = x_{k_1} \oplus x_{k_2} \oplus \cdots \oplus x_{k_m}$. As each $\pi/4$ -parity-phase gate can be realised in principle using a single T or T^\dagger gate (and some CNOT gates), simplifying $\pi/4$ -parity-phase circuits is directly productive to reducing T -count.

This line of investigation, first identified in the context of T -count by Amy, Maslov, and Mosca [4], was further developed upon by Gosset *et al.* [21], Amy and Mosca [6], Kissinger and van de Wetering [26], and Zhang and Chen [34]. In previous work [14], we described a family of identities of $\pi/4$ -parity-phase operations — “spider nest identities” — which, when used in combination with Heyfron and Campbell’s “TODD” subroutine [23], led to new records in T -count for several benchmark circuits.

In this work, we report new techniques for T -count reduction through the use of spider nest identities, and compare their effectiveness (the reduced T count and run-times) against the best previous result found in the literature. While these techniques could easily be

combined with other high-performance reduction subroutines such as TODD, our results do not involve any other recently developed techniques beyond those of Ref. [14]. We obtain a number of new records for the T -count, obtained almost exclusively¹ in very practical run-times on a consumer-grade laptop. (For example, the second-largest circuit, on 768 qubits, was simplified in less than 3 minutes.) This opens the door to further improvements through the identification of further useful identities of $\pi/4$ -parity-phase operations, and improved techniques for deploying these identities.

2 Preliminaries

We first set out some basic or existing results, using the following notation. Let $[n] := \{1, 2, \dots, n\}$ and $\mathbb{1}$ be the 2×2 identity matrix. For sets $S, T \subseteq V$ we write $S \Delta T$ for the symmetric difference $(S \cup T) \setminus (S \cap T)$, and $\mathbf{x}^{(S)} \in \{0, 1\}^V$ denote the incidence vector of S , where $x_j^{(S)} = 1$ if and only if $j \in S$. We let $\mathcal{P}^n := \{i^k P_1 \otimes \dots \otimes P_n \mid k \in \mathbb{Z} \text{ \& } P_j \in \{\mathbb{1}, X, Y, Z\}\}$ denote the n -qubit Pauli group. We define the Clifford hierarchy (on n qubits) by defining $\mathcal{C}_1^n := \mathcal{P}_n$, and

$$\mathcal{C}_k^n = \{U \in \text{U}_n(\mathbb{C}) \mid \forall P \in \mathcal{P}^n. UPU^\dagger \in \mathcal{C}_{k-1}^n\} \quad (1)$$

for $k > 1$; we call \mathcal{C}_k^n (for arbitrary n) the k^{th} level of the Clifford hierarchy. As an abuse of notation, we identify \mathcal{C}_k^n as a subset of \mathcal{C}_k^N for $n < N$; we may then write $S \in \mathcal{C}_2^n$ and $T \in \mathcal{C}_3^n$ for all $n \geq 1$.

Let $\mathcal{D}_k^n \subseteq \mathcal{C}_k^n$ be the subset of diagonal operations in the k^{th} level of the Clifford hierarchy. (We again identify \mathcal{D}_k^n as a subset of \mathcal{D}_k^N for $n < N$.) It is easy to show that \mathcal{D}_k^n forms an abelian group. In particular: consider any diagonal operation as a product of operators $\exp(i\theta_x |x\rangle\langle x|)$ for various $x \in \{0, 1\}^n$, and expand each $|x\rangle\langle x|$ as a linear combination of Pauli operators. Then one may show (see Ref. e.g. [14, Appendix A]) that \mathcal{D}_k^n is generated by the operators $\omega \cdot \mathbb{1}^{\otimes n}$ for any global phase ω , together with all operations of the form $D_{S,k}$ for sets $S = \{s_1, \dots, s_m\} \subseteq [n]$ for $m \geq 1$, defined by

$$D_{S,k} = \exp\left(-\frac{i\pi}{2^k} (Z_{s_1} \otimes \dots \otimes Z_{s_m})\right) = \exp\left(-\frac{i\pi}{2^k} Z_S\right) = \cos\left(\frac{\pi}{2^k}\right) \mathbb{1} - i \sin\left(\frac{\pi}{2^k}\right) Z_S, \quad (2)$$

where $Z_S = \bigotimes_{j \in S} Z_j$.² Note that $X_a Z_S X_a^\dagger = (-1)^{x_a^{(S)}} Z_S$, and that $\text{CNOT}_{a,b} Z_S \text{CNOT}_{a,b}^\dagger = Z_{S'}$, where here $S' = S \Delta \{a\}$ if $b \in S$ and $S' = S$ otherwise. From this it follows that

$$X_b D_{S,k} X_b^\dagger = D_{S,k}^{-1} \in \mathcal{D}_k^n \quad (3a)$$

if $b \in S$ (and $X_b D_{S,k} X_b^\dagger = D_{S,k}$ otherwise); and

$$\text{CNOT}_{a,b} D_{S,k} \text{CNOT}_{a,b}^\dagger = D_{S',k} \in \mathcal{D}_k^n \quad (3b)$$

so that \mathcal{D}_k^n is preserved under conjugation by CNOT and X operations. Also note that $D_{S,k}^2 = D_{S,k-1}$, from which it follows that $\mathcal{D}_{k-1}^n \subseteq \mathcal{D}_k^n$.

¹ The one circuit which we did not simplify on a laptop was the largest benchmark circuit that we tested, acting on 1536 qubits and involving nearly two million T gates alone. This was instead simplified on Dalhousie University's Mathstat Cluster [11], which took less than 15 minutes to realise a 43% reduction in T -count.

² We define $D_{S,k}$ for all $k \in \mathbb{Z}$; however, as one may easily show $D_{S,0} = -\mathbb{1}^{\otimes n}$ and $D_{S,k} = \mathbb{1}^{\otimes n}$ for all $k < 0$ and $S \subseteq [n]$, these operations are of interest principally for $k > 0$.

We refer to the operators $D_{S,k+1}$, and their inverses, as “ $\pi/2^k$ -parity-phase” operations, as the action of $D_{S,k+1}$ on standard basis states is given by

$$D_{S,k+1} |z\rangle = e^{i\pi/2^{k+1}} \exp\left(i [\mathbf{x}^{(S)} \cdot z] \pi/2^k\right) |z\rangle \quad (4)$$

inducing a relative phase of $\pi/2^k$ depending on the result of a parity computation $\mathbf{x}^{(S)} \cdot z = z_{s_1} \oplus z_{s_2} \oplus \dots \oplus z_{s_m}$. More generally, we may refer to $\exp(\pm \frac{1}{2} i \theta Z_S)$ as a θ -parity-phase operation.

From Eqn. (3b), it follows that any operation $D_{S,k}$ can be reduced to an operation $D_{j,k} \propto \text{diag}(1, e^{2\pi i/2^k})$ acting on a single qubit j , by conjugation with an appropriate CNOT circuit. In particular, it follows that the operation $D_{S,3}$ can be easily realised with a T -count of 1. This allows us to approach the question of reducing T count by considering decompositions of unitaries involving few $\pi/4$ -parity-phase operations, acting on many qubits. Amy and Mosca [6] noted the relevance of the operators $D_{S,k}$ in this context, and both Kissinger and van de Wetering [26] and Zhang and Chen [34] make direct use of them in their analysis of T count to achieve their results. (Litinski [27] similarly considers these operators in the context of compilation of quantum circuits to lattice surgery [24]).

An important role of $D_{S,3}$ gates for $S \subseteq [n]$ is their relationship to diagonal gates in \mathcal{D}_3^n which are controlled-unitaries in a more straightforward sense, such as CS and CCZ :

$$CS = \exp\left(\frac{i\pi}{2} |11\rangle\langle 11|\right), \quad CCZ = \exp\left(i\pi |111\rangle\langle 111|\right); \quad (5)$$

we may describe how to generate these from $D_{k,3}$ operations by decomposing the projectors $|11\rangle\langle 11|$ or $|111\rangle\langle 111|$ into tensor products of $|1\rangle\langle 1| = \frac{1}{2}(\mathbb{1} - Z)$, and expanding to obtain a product of $D_{S,3}$ gates. Disregarding any $D_{\emptyset,3}$ factors, which realise global phases, we obtain

$$CS_{h,j} \propto D_{\{h\},3} D_{\{j\},3} D_{\{h,j\},3}^{-1}; \quad CCZ_{g,h,j} \propto D_{\{g\},3} D_{\{h\},3} D_{\{j\},3} D_{\{g,h\},3}^{-1} D_{\{g,j\},3}^{-1} D_{\{h,j\},3}^{-1} D_{\{g,h,j\},3}. \quad (6)$$

More generally, we may relate $(t-1)$ -controlled $\pi/2^k$ -phase gates to $\pi/2^{k-t+1}$ -phase parity gates:

$$\prod_{\substack{S \in \wp(V) \\ S \neq \emptyset}} D_{S,k}^{(-1)^{|S|}} \propto \exp\left(\frac{i\pi}{2^{k-|V|+1}} |1\rangle\langle 1|^{\otimes V}\right), \quad (7)$$

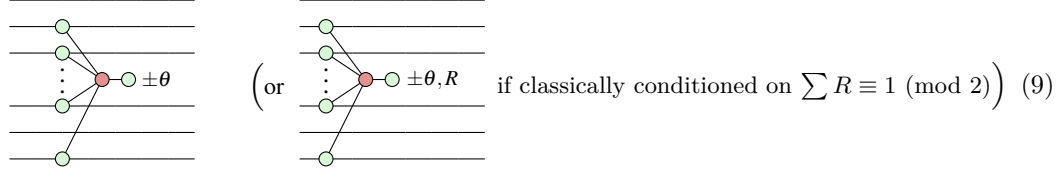
where the right-hand operator applies a phase of $\pi/2^{k-|T|+1}$ to those components of a state in which all of the qubits in T are in the state $|1\rangle$.

Circuits of parity-phase operations on n qubits which realise the identity, correspond in the notation of Amy and Mosca [6] to operators $U_{P_{\mathbf{a}}}$ for $\mathbf{a} \in \mathcal{C}_n \subseteq \mathbb{Z}_8^{2^n-1}$, where

$$P_{\mathbf{a}}(z) = \sum_{\substack{\mathbf{x} \in \{0,1\}^n \\ \mathbf{x} \neq \mathbf{0}}} a_{\mathbf{x}} (x_1 z_1 \oplus x_2 z_2 \oplus \dots \oplus x_n z_n) \quad (8)$$

and where $U_{P_{\mathbf{a}}} |\mathbf{z}\rangle = \exp\left(\frac{i\pi}{4} P_{\mathbf{a}}(\mathbf{z})\right) |\mathbf{z}\rangle$, which is identically $|\mathbf{z}\rangle$ for all $\mathbf{z} \in \{0,1\}^n$ when $\mathbf{a} \in \mathcal{C}_n$. Let $\text{supp}(\mathbf{a}) = \{\mathbf{x} \in \{0,1\}^n : a_{\mathbf{x}} \neq 0\}$. In this notation, each element $\mathbf{y} \in \text{supp}(\mathbf{a})$ corresponds to a single phase-parity operator acting on the qubits j for which $y_j = 1$; the relative phase induced by this operator is $a_{\mathbf{y}}\pi/4$; and the polynomial $P_{\mathbf{a}}$ describes a commuting product of such operations, for which $P_{\mathbf{a}} : \{0,1\}^n \rightarrow \mathbb{Z}_8$ is the all-zero function when $\mathbf{a} \in \mathcal{C}_n$.

We remark that a θ -phase parity operation U (such as an operator $D_{S,k}$) can be easily represented as tensor networks, using ZX diagrams (see Appendix A for an introduction to this notation),³ with structure such as the following:



where horizontal wires represent qubits which are acted on by U , and $S \subseteq [n]$ is the subset of those qubits which have (light, green) degree-3 nodes on them. These are “phase gadgets”, using the terminology of Kissinger and van de Wetering [26]. When the number of qubits acted on is m , we may refer to it as an “ m -gadget”. (If θ is an odd multiple of $\pi/4$, we may refer to it as a “ T -phase m -gadget”; for θ an integer multiple of $\pi/2$, we refer to it as a “Clifford-phase m -gadget”. If $m = 1$, we may also mildly abuse this terminology to refer to a simple green phase node as a “1-gadget”.)

Remark.

The role played by the ZX calculus in our work is not an essential one, nor is expertise in the ZX calculus required to understand our results. However, in practice it did inform our line of investigation, by allowing us to obtain our results more quickly by identifying the objects of interest, and by making it easy to reason directly about the operators $D_{S,k}$. As the ZX calculus also provides a useful notation for visually representing the (non-local) unitary gates $D_{S,k}$ in a readable way, as in Eqn. (9), we use this notation in the article below. Readers should be able to understand our results by reading ZX diagrams simply as a straightforward alternative notation for quantum circuits (see Appendix A), the transformations of which are the subject of our work.

3 Phase gadget elimination tactics & spider nest identities

Reducing the T -count while preserving the meaning of a circuit, implicitly involves applying a mathematical identity. These are often identities of diagonal unitary circuits [4, 6, 34], though not always [21, 26].) In the special case of unitary circuits consisting solely of $\pi/4$ -parity-phase operations, such a mathematical identity may be described in terms of a commuting product of operations which are proportional to the identity operator; and for any such identity, there is the question of how to effectively apply it to realise a significant reduction of T -count, as efficiently as possible.

In this section, we describe a broad framework for the reduction of T -count by means of the application of mathematical identities of commuting \mathcal{D}_3^n operations. We also present some mathematical identities of this form — called “spider nest identities” — first presented in Ref. [14], and describe new techniques to use these identities to reduce T -count.

In the following, we use the terms “identity of $\pi/4$ -parity-phase operations” or “identity of phase gadgets” (or simply “an identity”) to refer to a circuit \mathcal{J} , whose T -count is at least 1 but which nevertheless realises the identity operation.

³ In this article, where they occur, ZX diagrams may be read essentially as circuit diagrams, and in particular are read from left to right as with other circuit diagrams.

3.1 PHAGE tactics

We consider a particular approach to the reduction of \mathcal{D}_3^n circuits by an analysis of families of non-trivial circuits which realise the identity transformation, which may be applied more broadly than we do here (and which in principle can be used to describe some existing techniques [6, 23]). For any family \mathcal{F} of identities of $\pi/4$ -parity-phase operations, there is an associated “phase gadget elimination tactic” (or PHAGE tactic) to reduce the T -count in a circuit \mathbf{C} of such phase gadgets:

► **Phage Tactic** (\mathcal{F}).

1. Determine whether there is an identity $\mathcal{J} \in \mathcal{F}$, such that \mathbf{C} contains at least half of the T -gadgets which occur in \mathcal{J} (or their inverses).
2. For any such identity \mathcal{J} , compute a circuit $\mathbf{C}_{\mathcal{J}}$ as the product of \mathbf{C} and \mathcal{J}^{-1} . This may allow for simplifications (using the fact noted in Section 2 that $D_{S,k}^2 = D_{S,k-1}$), where by T -gadgets accumulate to form Clifford gadgets or to cancel altogether. Determine the resulting T -count.
3. Replace \mathbf{C} with the circuit $\mathbf{C}_{\mathcal{J}}$ with the smallest T -count, if this is less than the T -count of \mathbf{C} itself.

The behaviour of a PHAGE tactic is in a sense “greedy”, in that it selects some circuit $\mathbf{C}_{\mathcal{J}}$ which minimises the T count after a single application, ignoring the possibility of a more complicated sequence of reductions. The main principle of a PHAGE tactic is in that it selects a way to reduce the T -count, based on the comparison of a few different applicable identities of phase gadgets from a specific family \mathcal{F} . Such a tactic can then be applied again, or followed by other such “tactics”.

In principle, the **Tpar** subroutine of Ref. [6], the **TOOL** and **TODD** subroutines of Ref. [23], and the results of Zhang and Chen [34] may be interpreted as algorithms to deploy PHAGE tactics, possibly more than once in sequence, and possibly with a random choice of family \mathcal{F} (and where \mathcal{F} itself may on occasion be a singleton set). This approach to T -count reduction can be distinguished from that of Kissinger and van de Wetering [26], in which phases may be reduced in unitary circuits (or more general tensor networks) which are not diagonal.

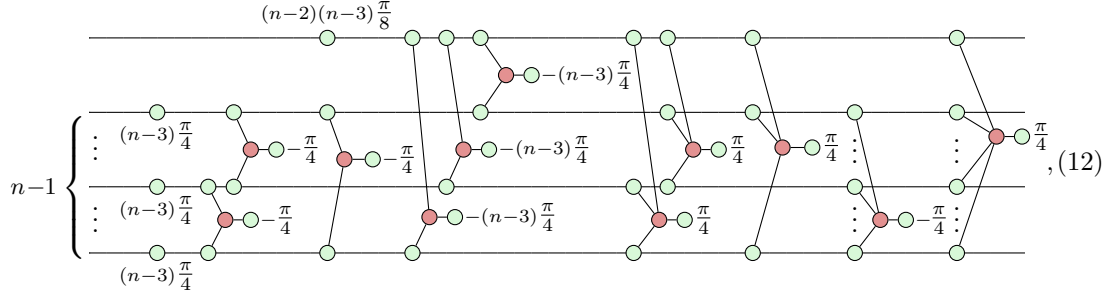
The difficulty in reducing the T -count arises from the fact that there are a very large number of identities of $\pi/4$ -parity-phase operations, and a large number of subsets $S \subseteq [n]$ which one may consider. As Amy and Mosca observe [6], reducing the T -count is formally equivalent to decoding a length $2^n - 1$ punctured Reed-Muller code, in that the smallest T -count of a circuit amounts to the distance of a ciphertext to a valid codeword of such a code. However, no polynomial-time algorithms are known for the decoding problem on such codes. The difficulty is in formulating a successful *strategy* — a means of selecting an appropriately-sized family \mathcal{F} of identities to try on a particular circuit. The question is then one of having a variety of tactics which one may efficiently explore and deploy to reduce the T -count.

3.2 Spider nest identities

We consider PHAGE tactics arising from identities of $\pi/4$ -parity-phase operations (*i.e.*, of T -phase gadgets) which can be composed from some specific circuits — introduced in Ref. [14], and which we call “spider nest identities” — which realise the identity operator.

In qualitative terms, a “spider nest identity” consists of any circuit of phase-parity operations which realises an operation on n qubits which is proportional to the identity, in which only “very few” operations act on “many” qubits, and the vast majority act on “very

\mathcal{N}_S are similarly cancelled.) By collecting together the actions of the phase gadgets on each subset, we may show that $\mathcal{N}_S \mathcal{N}_{S'}^{-1}$ simplifies to a circuit of the following form:



If $r = S \setminus S'$ represents the top qubit in the circuit above, note in particular that the dominant contributions to the size of the circuit are the phase 2-gadgets on all size-2 subsets of S' , and the phase 3-gadgets which involve r and some size-2 subset of S' . If \tilde{T}_n denotes the T -count of the circuit above, we then have

$$\tilde{T}_n = \begin{cases} n^2 - n + 2 + \delta_n & \text{for } n \text{ even;} \\ n^2 - 3n + 4 + \delta_n & \text{for } n \text{ odd,} \end{cases} \quad (13)$$

where again $\delta_n = 1$ if $n \equiv 0$ or $n \equiv 1$ modulo 4, and $\delta_n = 0$ if $n \equiv 2$ or $n \equiv 3$ modulo 4. In any case, we have $\tilde{T}_n = n^2 - O(n)$.

3.3 Simple PHAGE tactics based on spider nest identities

Combining the two ideas above, we describe the PHAGE tactics which are used to achieve the T -count reductions seen in our results.

The first tactic is the reduction of phase-parity circuits by merging together $\pi/4$ -parity-phase operations which act on sets of qubits in common, which may be described as the PHAGE tactic associated to the circuits consisting of mutually inverse pairs of T -phase gadgets on all possible sets of qubits. To do this to greatest effect (and also as simply as possible), we first use a circuit transformation procedure along the lines of Heyfron and Campbell [23], with modifications to improve performance. (In the context of reasoning about T count in terms of $\pi/4$ -parity-phase operations, this technique was introduced in Ref. [14].) We describe this in more detail in the following Section, which describes our T -count reduction procedure.

Our other PHAGE tactic (or tactics, as they are similar but technically numerous) are novel, and are best described in terms of the following two sets of spider-nest identities on N qubit circuits:

- The family $\mathcal{F}_N^{(4)} = \{\mathcal{N}_S \mid S \subseteq [N] \text{ and } |S| = 4\}$, consisting of versions of the identity of Eqn. (10) applied to all subsets of $[N]$ of size 4
- The family

$$\mathcal{F}_N^{(5)} = \left\{ \mathcal{N}_S^{p_0} \mathcal{N}_{S_1}^{p_1} \mathcal{N}_{S_2}^{p_2} \mathcal{N}_{S_3}^{p_3} \mathcal{N}_{S_4}^{p_4} \mathcal{N}_{S_5}^{p_5} \left| \begin{array}{l} S = \{q_1, q_2, q_3, q_4, q_5\} \text{ for distinct } q_j \in [N], \\ S_j = S \setminus \{q_j\} \text{ for } 1 \leq j \leq 5, \text{ and} \\ p_0 p_1 p_2 p_3 p_4 p_5 \in \{0, 1\}^6 \setminus \{000000\} \end{array} \right. \right\}, \quad (14)$$

consisting of the 63 distinct identities for each set $S \subseteq [N]$ with $|S| = 5$, consisting of \mathcal{N}_{S_j} applied to some or all subsets $S_j \subseteq S$ of size 4, and possibly also a copy of \mathcal{N}_S on all the qubits of S , fusing together those phase-parity operations which act on common subsets $S' \subseteq S$.

These are the sets of all possible spider-nest identities on 4 or 5 qubits.⁴

For increasing values of N , the cardinalities of these families grow as $\frac{1}{24}n^4 + O(n^3)$ and $\frac{1}{120}n^5 + O(n^4)$ respectively — polynomial in size, but impractical to exhaustively iterate through for values of N which occur in common benchmark tests. This raises the question of how best to use them to realise T -count reductions. Our approach is to construct a list of 64 identities on four or five qubits, consisting of the elements of the sets $\mathcal{F}_4^{(4)} \cup \mathcal{F}_5^{(5)}$, and performing the following for each element \mathcal{J} of this list:

1. Let s be the number of qubits on which \mathcal{J} acts.
2. Repeat the following R times, for some fixed $R > 0$:
 - a. Select a subset $S \subseteq [N]$ of size s uniformly at random.
 - b. Select (from $\mathcal{F}_N^{(4)}$ if $s = 4$, or $\mathcal{F}_N^{(5)}$ if $s = 5$) the identity \mathcal{K} acting on S , which is equivalent to \mathcal{J} up to relabelling of the qubits.
 - c. Apply the tactic **PHAGE**($\{\mathcal{K}\}$) associated with the singleton set $\{\mathcal{K}\}$.

This technique implicitly provides opportunities for identities to be applied in proportion to the number of isomorphic images of it exist in $\mathcal{F}_4^{(4)} \cup \mathcal{F}_5^{(5)}$. (For instance, isomorphic copies of the simplest identity $\mathcal{N}_{[4]}$ occurs six times in this set, and the identity of Eqn. (12) occurs five times.) As the probability that any one such identity will be useful when applied to a particular set $S \subseteq [N]$ of size 4 or 5 is small, it is important to choose a significantly large value of R : for our results, we took $R = 20\,000$.

We note that this particular strategy for T -count reduction is not particularly strongly suggested by the framework of PHAGE tactics induced by spider nest identities. Both the concept of a PHAGE tactic, and the range of possibilities for assembling spider nest identities, are broad enough that there is potential for much more sophisticated strategies to deploy them. Despite this, as we show in Section 5, in many cases we obtain the best known T -count for a number of circuits. Our result may therefore be considered a further proof of principle of the usefulness of spider nest identities, beyond the results of Ref. [14].

4 Reduction of T -count through simplification of parity-phase circuits

In this section, we describe how we applied the concept of T -count reduction via PHAGE tactics as part of a complete procedure to transform unitary circuits provided as input.

Remark.

Our results do not make heavy (explicit) use of the re-write rules of the ZX calculus: a reader who is content with circuits which involve intermediate measurements, and who is comfortable with reading a parity-phase gadget such as that of Eqn. (9) as a unitary operator, may interpret every diagram below as a circuit diagram. (See Appendix A for a guide to reading ZX diagrams.)

We take unitary circuits with gate-set $\{X, \text{CNOT}, \text{CCNOT}, Z, CZ, CCZ, H, S, T, \text{SWAP}\}$ as input. For the sake of simplicity, we suppose that any multiply-controlled NOT gates with more than two controls are decomposed into CCNOT gates, for instance by computation and uncomputation on auxiliary qubits initialised to $|0\rangle$, or some more advanced technique.⁵

⁴ The set $\mathcal{F}^{(5)}$ in particular is motivated by the reduction in T -count of the spider-nest identity shown in Eqn. (12), which is represented in five different ways in $\mathcal{F}^{(5)}$: once for each subset S_j of size 4.

⁵ In our benchmarks, we consider the simple computation-uncomputation approach; other techniques (see *e.g.* Refs. [25, 20, 29]) are advisable in serious production work for optimising T -count.

Our procedure follows and extends the approach of Heyfron and Campbell [23], of performing a transformation on circuits $\mathbf{C} \rightarrow \mathbf{C}_F \circ \mathbf{C}_\phi \circ \mathbf{C}_I$, where \mathbf{C}_F and \mathbf{C}_I consist entirely of Clifford gates, stabiliser state preparations, and stabiliser state measurements, and where \mathbf{C}_ϕ can be realised using only CNOT and T gates. We express the circuit \mathbf{C}_ϕ entirely in terms of phase gadgets, and so we describe as a “homogeneous” circuit. The objective of isolating such a circuit is that it provides us with the best opportunities to apply PHAGE tactics to reduce the T -count.

4.1 Circuit translation techniques

Our procedure, which we describe more explicitly in the next section, makes use of the following techniques.

H gate gadgetisation.

One of the techniques involved in isolating a D_3^N circuit is to replace Hadamard gates with a measurement-based gadget:

$$[H] \equiv \text{circuit with } |+ \rangle, \text{ SWAP, CZ, X, and measurement} \equiv \text{ZX diagram with phases} \equiv \text{decomposed ZX diagram} \quad (15)$$

In the circuit second from the left, the two qubits are subject to a SWAP operation, followed by a $CZ = \exp(i\pi |11\rangle\langle 11|)$ operation. The bottom qubit is measured finally with an X observable measurement (*i.e.*, in the $|\pm\rangle$ basis), and the top operation is acted on finally by an X operation only if the outcome is $|-\rangle$. The two diagrams on the right are ZX diagrams with additional annotations in the style of Ref. [18] (see also Appendix A). In particular, measurement is represented as a projection with a random outcome s which is heralded and may be used to control phase operations elsewhere. The leftmost ZX diagram describes the decomposition of the controlled- Z operation, using $CZ_{h,j} \propto D_{\{h,j\},2}^{-1} D_{\{h\},2}^{-1} D_{\{j\},2}^{-1}$. The final ZX diagram propagates the single-qubit $D_{\{*\},2}^{-1}$ operations towards the preparation and measurement of the second qubit, so that the second qubit is prepared in the $|-\mathbf{y}\rangle \propto |0\rangle - i|1\rangle$ state.

Extracting H gates from the circuit.

An obvious drawback of gadgetising H gates in this way is that it requires the use of auxiliary qubits. More directly important to our results is that, as the number of wires in a circuit increases, the more difficult it may be to successfully find opportunities to reduce the T count. Therefore, we attempt to transform the circuit in such a way that reduces the number of H gates from the part of the circuit with non-trivial T -count. This motivates us to define a subroutine `moveH` (which we describe at a high level in Appendix B), which transforms a circuit \mathbf{C} over our gate-set, into a pair of circuits $(\mathbf{C}_F, \mathbf{C}')$, obtained by attempting to commute as many Hadamard gates of \mathbf{C} to the end of the circuit as possible.

- We define $(\mathbf{C}_F, \mathbf{C}') = \text{moveH}(\mathbf{C})$ in such a way that $\mathbf{C}_F \circ \mathbf{C}' \cong \mathbf{C}$ realises the same unitary, \mathbf{C}_F contains only Clifford gates, \mathbf{C}' contains no CCNOT gates, and where the total number of Hadamard gates in $(\mathbf{C}_F \circ \mathbf{C}')$ is at most the number of Hadamard gates in \mathbf{C} .

- We may use `moveH` twice, to attempt to extract Hadamard gates either from the end of the circuit \mathbf{C} , and also the beginning of the circuit \mathbf{C} . If we compute

$$(\mathbf{C}_F, \mathbf{C}') = \text{moveH}(\mathbf{C}); \quad (\tilde{\mathbf{C}}_I, \tilde{\mathbf{C}}_M) = \text{moveH}((\mathbf{C}')^{-1}); \quad (\mathbf{C}_I, \mathbf{C}_M) = (\tilde{\mathbf{C}}_I^{-1}, \tilde{\mathbf{C}}_M^{-1}), \quad (16)$$

then $(\mathbf{C}_F \circ \mathbf{C}_M \circ \mathbf{C}_I) \cong \mathbf{C}$, the number of Hadamard gates in $(\mathbf{C}_F \circ \mathbf{C}_M \circ \mathbf{C}_I)$ is at most the number of Hadamard gates in \mathbf{C} , and \mathbf{C}_I and \mathbf{C}_F only contain Clifford gates.

We call \mathbf{C}_I and \mathbf{C}_F the initial and final Clifford stages of the circuit, respectively, and \mathbf{C}_M the main body of the circuit. We use this tripartite decomposition to allow us to condense the part of the circuit with non-trivial T -count in the main body, and to remove Clifford gates (H gates in particular) to the initial and final Clifford phases to the extent that this is possible.

Phase-gadgetisation.

Through appropriate substitution of H gates by gadgets as in Eqn. (15), and substitution of CCZ with $\pi/4$ -parity-phase operations as in Eqn. (6), we may transform the main body of the circuit so that it only contains SWAP gate, X gates, CNOT gates, CZ gates, and various phase gadgets (including powers of the T gate). We wish to transform this into a circuit consisting only of phase gadgets, by commuting everything apart from phase gadgets either to the beginning of the main body (and then removing it to the initial Clifford phase) or to the end of the main body (and then removing it to the final Clifford phase). In particular, we commute all SWAP, measurement, and X operations to the end of the circuit; we commute all preparation operations to the beginning of the circuit; and we commute each CNOT operation either to the beginning or the end according to a simple heuristic (described in Appendix B). This may transform various $D_{S,t}$ gates by Eqns. (3), changing the set S involved and/or negating the phase, according to the following commutation relations:

$$(17)$$

$$(18)$$

$$(19)$$

Phase gadget fusion.

A final simplifying technique is to simply multiply together any phase gadgets acting on the same set S of qubits:



In some cases, this will reduce the T count by turning two gadgets with phases $\alpha = \frac{1}{4}k_1\pi$ and $\beta = \frac{1}{4}k_2\pi$ (for k_1 and k_2 odd) into a single gadget with phase $\alpha + \beta = \frac{1}{4}(k_1 + k_2)\pi$, where $k_1 + k_2$ is even.

4.2 Circuit translation procedure

Given a unitary circuit \mathbf{C} over the gate-set

$\{X, \text{CNOT}, \text{CCNOT}, Z, \text{CZ}, \text{CCZ}, H, S, T, \text{SWAP}\}$, we transform \mathbf{C} as follows:

1. We first replace CCNOT operations in \mathbf{C} with $(1 \otimes 1 \otimes H) \text{CCZ} (1 \otimes 1 \otimes H)$, yielding a circuit \mathbf{C}' .
2. Transform $\mathbf{C}' \rightarrow \mathbf{C}'_F \circ \mathbf{C}'_M \circ \mathbf{C}'_I$, with an initial Clifford stage \mathbf{C}'_I , a final Clifford stage \mathbf{C}'_F , and a main body \mathbf{C}'_M , using the procedure `moveH` to reduce the number of Hadamard gates in \mathbf{C}'_M as much as possible.
3. Substitute the H gates in \mathbf{C}'_M with Hadamard gadgets as in Eqn. (15), using a fresh bit label for each measurement outcome; and decompose CCZ operations in \mathbf{C} using the formula of Eqn. (6), and represent T gates (on some qubit j) by $D_{\{j\},3}$. Call the resulting circuit \mathbf{C}_M .
4. We gadgetize \mathbf{C}_M by commuting all gates which are not single-qubit phase gates or phase gadgets to the beginning or the end, removing these to the initial or final Clifford stages. This will generally add some number of measurements, and classically-conditioned Clifford operations, to the final Clifford stage, and some qubit preparations to the initial Clifford stage. This realises a transformation of circuits $\mathbf{C}'_F \circ \mathbf{C}_M \circ \mathbf{C}'_I \rightarrow \mathbf{C}_F \circ \mathbf{C}'_\phi \circ \mathbf{C}_I$.
5. As \mathbf{C}'_ϕ is now a homogeneous circuit of phase gadgets, we may commute them past one another to fuse gadgets on common subsets, yielding a circuit \mathbf{C}_ϕ .
6. Apply the randomised procedure for applying PHAGE tactics based on spider nest identities described in Section 3.3.

Steps 1–5 realise a transformation $\mathbf{C} \rightarrow \mathbf{C}_F \circ \mathbf{C}_\phi \circ \mathbf{C}_I$. If the original circuit \mathbf{C} acted on n qubits and had m Hadamard gates, then the number of Hadamard gates in \mathbf{C}'_M which are replaced in Step 3 is some $\delta n \leq m$. Then the circuits \mathbf{C}_I , \mathbf{C}_ϕ , and \mathbf{C}_F all act on $N = n + \delta n$ qubits, and \mathbf{C}_F has internal structure

$$\mathbf{C}_F = \tilde{\mathbf{C}}_F \mathbf{D}_{\delta n} \cdots \mathbf{D}_2 \mathbf{D}_1, \quad (21)$$

where $\tilde{\mathbf{C}}_F$ is some general Clifford circuit, and the circuits \mathbf{D}_j (for $1 \leq j \leq \delta n$) consist of the j^{th} measurement in the $|\pm\rangle$ basis with outcome s_j (denoted in ZX notation by a light green “ $\pi, \{s_j\}$ ” node), followed by \mathcal{D}_k^N operations conditioned on the outcome s_j .

In some instances, we find a significant reduction in the T -count simply from the fusion of phase gadgets in Step 5 of this transformation. These improvements are similar to those seen in Refs. [26, 34]. However, the purpose of this circuit transformation (as Ref. [23]) is to isolate a circuit \mathbf{C}_ϕ consisting entirely of \mathcal{D}_3^N operations for some N , on which we can apply the PHAGE tactic of Step 6.

Note that δn , the number of additional “auxiliary” qubits involved in the circuit, is bounded above by how many Hadamard gates are either involved in \mathbf{C} or are introduced from the decomposition of CCNOT gates. More precisely, it depends on how many of these gates can be commuted from the “main body” of \mathbf{C} to the initial or final Clifford stages. For a circuit consisting of M gates, a bound for $N = n + \delta n$ which is substantially better than $N \leq n + M$ will be difficult to obtain, without some knowledge of the structure of \mathbf{C} . In several cases, we find that many or all of these Hadamard gates can be eliminated from the main body of the circuit: so, $N \leq n + M$ is likely a loose upper bound in a large number of practical examples.

The largest contributions to the asymptotic run-time of the procedure above are the complexity of `moveH` in Step 2; the cumulated complexity of computing the heuristic for moving Clifford gates out of the main body of the circuit in Step 4; and the complexity of performing a PHAGE tactics in Steps 5 and 6. For M the number of gates in the input circuit, the procedure `moveH` and the procedure to commute CNOT gates out the main body take time $O(M^2)$, essentially due to repeatedly commuting individual gates past $O(M)$ other gates (or computing the potential cost of doing so, in the case of the heuristic used for determining the direction to move CNOT gates). We use a hash table to store homogeneous circuits, allowing essentially $O(1)$ time to look up the phase associated with a phase gadget acting on a particular subset (which we set to 0 when no such phase gadget is present). In Step 5, fusing together pairs of phase gadgets can be made a part of initialising this hash table, and so takes time $O(M)$. In Step 6, applying a PHAGE tactic associated with some given identity \mathcal{K} (which acts on at most 5 qubits) takes time $O(1)$; performing this for each of the 64 identities in $\mathcal{F}_4^{(4)} \cup \mathcal{F}_5^{(5)}$ on R uniformly random subsets takes time $O(R) = O(1)$, for R independent of M . Thus our procedure runs in time $O(M^2)$.

5 Results

We realised our techniques in Haskell code [7]. All but two of the circuits were obtained from Ref. [30]: the circuits “GF(2²⁵⁶) Mult” and “GF(2⁵¹²) Mult” were obtained instead from Ref. [28]. With one exception, we ran our code for these benchmarks on a 2.5 GHz Intel Core i7 processor and 8 GB of 1867 MHz LPDDR3 memory, running Ubuntu Linux 18.04.4. The largest single benchmark circuit, “GF(2⁵¹²) Mult”, was instead reduced on Dalhousie University’s Mathstat Cluster [11]. The results are presented as Table 1, including comparisons to the best known reductions for which recorded times are available.⁶

Our results do not include an account of the cost of the Clifford group operations. These are also of interest in principle, though these will likely be significantly less expensive than T gates in the error-corrected setting in which the T -count becomes a meaningful quantity to reduce. We also do not describe the T -depth of our circuits, which may also be independently optimised from the T -count itself [4].

The circuits which were obtained using our techniques may be found on GitHub [7]. As our main aim was to consider reductions in T -count, our algorithm ignores the possibility that the measurement outcomes on the auxiliary qubits could be anything but $|+\rangle$: in the event of a $|-\rangle$ outcome, additional Clifford group operations would be required, which however would not affect the T -count. We verified our circuits using `feynver` [2], which was recently extended to accomodate circuits involving post-selection of $|+\rangle$ states on qubits which are maximally entangled with a set of other qubits.

⁶ We do not present the best known T counts which do *not* have recorded times. We do note that for two of our results (for the circuits Mod Red₂₁ and RC Adder₆) which improve on the known timed results, there are recorded untimed results which are still better: these may be found in Ref. [14].

■ **Table 1** Comparison of our techniques to previously reported results. • In each case, “prev. opt.” represents the best T -count with a time record (an asterisk indicates that the recorded time is an upper bound). For some circuits, better reductions without times have been reported: those indicated by ^(a) have a better reduction reported in Ref. [26], and those indicated by ^(b) have a better reduction reported in Ref. [14]. Where it was feasible to verify the correctness of our reduction with `feynver`, this is indicated; in all other cases the verification was too computationally expensive to carry out. • In each case, we also compare the number δn of additional “auxiliary” qubits required by our decomposition, to that of Ref. [23] (where results are available); in the case of ^(c), we may only infer an upper bound on the number of auxiliary qubits used by Ref. [23]. • In our results, those T -counts which are indicated in bold are those which reproduce or surpass the T -count of the best previously known timed result. Those with an asterisk also match or surpass the best previously known untimed result. • All results of Ref. [23] were obtained on the University of Sheffield’s Iceberg HPC cluster. All results of Ref. [31] were obtained on a machine with a 2.9 GHz Intel Core i5 processor and 8 GB of 1867 MHz DDR3 memory, running OS X El Capitan. All of our results were obtained on a machine with a 2.5 GHz Intel Core i7 processor and 8 GB of 1867 MHz LPDDR3 memory, running Ubuntu Linux 18.04.4 — except those indicated by ^(d), which were obtained on Dalhousie University’s Mathstat Cluster [11].

Circuit	# qubits			T count & optimisation						
	n input	δn [23]	δn (ours)	init. # T	final # T (prev. opt.)	Ref.	time (s)	final # T (our results)	time (s)	Verified? (<code>feynver</code>)
Adder ₈	24	71	41	399	212 ^(a)	[23]	227.81	176*	24.62	✓
Barenco Tof ₃	5	3	3	28	14 ^(b)	[23]	0.01*	13*	0.07607	✓
Barenco Tof ₄	7	7	7	56	24	[23]	0.45	25	1.884	✓
Barenco Tof ₅	9	11	11	84	34	[23]	1.94	37	13.76	✓
Barenco Tof ₁₀	19	31	31	224	84	[23]	460.33	97	24.49	✓
CSLA MUX ₃	15	17	6	70	40 ^(b)	[23]	3.73	44	18.01	✓
CSUM MUX ₉	30	12	12	196	74 ^(a)	[23]	36.57	84	23.98	✓
GF(2 ⁴) Mult	12	7	0	112	56 ^(b)	[23]	0.55	53*	0.8180	✓
GF(2 ⁵) Mult	15	9	0	175	90 ^(b)	[23]	6.96	88*	4.279	✓
GF(2 ⁶) Mult	18	11	0	252	132 ^(b)	[23]	121.16	128*	7.894	✓
GF(2 ⁷) Mult	21	13	0	343	185 ^(a)	[23]	153.75	167*	27.21	✓
GF(2 ⁸) Mult	24	15	0	448	216 ^(a)	[23]	517.63	229	95.63	✓
GF(2 ⁹) Mult	27	17	0	567	295	[23]	3213.53	306	24.79	✓
GF(2 ¹⁰) Mult	30	19	0	700	351	[23]	23969.1	357	24.65	✓
GF(2 ¹⁶) Mult	48	31	0	1 792	922	[23]	76312.5	972	25.65	✓ ^(d)
GF(2 ³²) Mult	96	—	0	7 168	4 128	[31]	1.834	3 936*	26.07	✓ ^(d)
GF(2 ⁶⁴) Mult	192	—	0	28 672	16 448	[31]	58.341	15 865*	29.73	—
GF(2 ¹²⁸) Mult	384	—	0	114 688	65 664	[31]	1744.746	64 461*	48.78	—
GF(2 ¹³¹) Mult	393	—	0	120 127	69 037	[31]	1953.353	67 772*	53.39	—
GF(2 ¹⁶³) Mult	489	—	0	185 983	106 765	[31]	4955.927	105 182*	66.27	—
GF(2 ²⁵⁶) Mult	768	—	0	458 752	—	—	—	260 539*	137.1	—
GF(2 ⁵¹²) Mult	1536	—	0	1 835 008	—	—	—	1 046 964*	850.7 ^(d)	—
Mod5 ₄	5	6	0	28	16 ^(b)	[31]	0.001*	7*	0.00899	✓
Mod Adder ₁₀₂₄	28	≤ 270 ^(c)	304	1 995	978	[23]	665.5	1 010	27.56	✓ ^(d)
Mod Mult ₅₅	9	10	3	49	28 ^(a)	[23]	0.02	19*	0.5775	✓
Mod Red ₂₁	11	17	17	119	69 ^(b)	[23]	0.59	65	27.68	✓
QCLA Adder ₁₀	36	28	25	238	157	[23]	366.1	147*	24.96	✓
QCLA Com ₇	24	19	18	203	81	[23]	170.77	84	24.21	✓
QCLA Mod ₇	26	58	58	413	221 ^(a)	[23]	289.77	233	24.26	✓ ^(d)
RC Adder ₆	14	21	10	77	45 ^(b)	[23]	0.97	38	30.70	✓
NC Toff ₃	5	2	2	21	13	[23]	0.01*	13*	0.04005	✓
NC Toff ₄	7	4	4	35	19	[23]	0.06	19*	0.5322	✓
NC Toff ₅	9	11	6	49	25	[23]	0.4	26	2.910	✓
NC Toff ₁₀	19	16	16	119	55	[23]	44.78	60	28.01	✓
VBE Adder ₃	10	4	4	70	20	[23]	0.15	20*	1.896	✓

6 Discussion

Our results show that our techniques, simple as they are, are competitive with the best known techniques for reducing T count. We expect that better results should be achievable by a more refined approach to using these concepts, within the more general framework which we have described of deploying PHAGE tactics. It is not clear which further ideas may prove useful, however. For instance, in experiments for how we might choose subsets to apply PHAGE tactics to, we found that it was not helpful to bias the sets of qubits towards those qubits which were acted on by many T -phase gadgets. More work will be required to find effective ways to bias or to narrow down the ways in which spider nest identities are used to simplify homogeneous circuits.

It is remarkable that the run-times for our results in Table 1 are not more varied. Over half of our results were obtained in an amount of time between 1 and 100 seconds, for circuits over which other leading techniques [23, 31] had times which ranged over more than six orders of magnitude. Indeed, in our tests on larger circuits (and in line with the asymptotic analysis of Section 4.2), we found that the most computationally expensive part of our procedure was the relatively mundane `moveH` and CNOT-commutation subroutines. Refining these elementary steps may provide yet further gains. Expanding the complexity of the subroutines to apply PHAGE tactics may also yield further gains without substantial increases in run-time.

We note an optimisation problem of interest is motivated by gadgetizing Hadamard gates as in Step 3. Simply put: given an n -qubit circuit with M gates over the gate-set $\{X, Z, S, \text{CNOT}, \text{CZ}, T, \text{CCZ}\}$, to obtain an equivalent (unitary) circuit with the minimum number of H gates in between the first and the last non-Clifford gate.⁷ Should this problem be solvable in $O(M^2 \text{poly log}(M))$ time, this may further contribute to the effectiveness of our approach to T -count reduction.

Finally, we remark that while the benchmarks which we have tested have become a commonplace standard for the evaluation of such techniques, they consist entirely of circuits to realise permutation operations which would not in themselves be difficult to realise classically. (Some of these, such as the “GF(2ⁿ) Mult” series, may be motivated on the grounds of cryptography; albeit this motivation may become less important if standard cryptographic practise incorporates post-quantum cryptography.) A larger range of circuits, including ones are motivated by the more likely practical applications of fault-tolerant quantum computation, should be of general interest for future benchmark tests.

References

- 1 S. Aaronson and D. Gottesman. Improved simulation of stabilizer circuits. *Phys. Rev. A*, 70:052328, 2004. [arXiv:quant-ph/0406196](#).
- 2 Matthew Amy. Towards large-scale functional verification of universal quantum circuits. In *Proceedings of QPL 2018*, pages 1–21, 2018. [[arXiv:1901.09476](#)]; see also [<https://github.com/meamy/feynman>]. doi:10.4204/EPTCS.287.1.
- 3 Matthew Amy, Jianxin Chen, and Neil J. Ross. A finite presentation of cnot-dihedral operators. *Electronic Proceedings in Theoretical Computer Science*, 266:84–97, 2018. [[arXiv:1701.00140](#)]. doi:10.1007/978-3-642-12821-9_4.
- 4 Matthew Amy, Dmitri Maslov, and Michele Mosca. Polynomial-time T-depth optimization of Clifford+T circuits via matroid partitioning. *IEEE Transactions on Computer-Aided*

⁷ It seems plausible that this problem would remain equally difficult without CCZ gates.

- Design of Integrated Circuits and Systems*, 33(10):1476–1489, 2014. [arXiv:1303.2042]. doi:10.1109/TCAD.2014.2341953.
- 5 Matthew Amy, Dmitri Maslov, Michele Mosca, and Martin Roetteler. A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits. *IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems*, 32(6):818–830, 2013. [arXiv:1206.0758]. doi:10.1109/TCAD.2013.2244643.
 - 6 Matthew Amy and Michele Mosca. T-count optimization and Reed-Muller codes. *IEEE Transactions on Information Theory*, 65(8):4771–4784, 2019. [arXiv:1601.07363]. doi:10.1109/TIT.2019.2906374.
 - 7 Xiaoning Bian. “stomp-code” github. URL: <https://github.com/onestruggler/stomp-code/tree/8df4f46228c2f413e0cf5f8b6d25c20b6460fc0e>.
 - 8 Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, September 2019. [arXiv:1808.00128]. doi:10.22331/q-2019-09-02-181.
 - 9 Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Physical Review Letters*, 116:250501, 2016. [arXiv:1601.07601]. doi:10.1103/PhysRevLett.116.250501.
 - 10 Earl T. Campbell and Mark Howard. A unified framework for magic state distillation and multi-qubit gate-synthesis with reduced resource cost. *Physical Review A*, 95:022316, 2017. [arXiv:1606.01904]. doi:10.1103/PhysRevA.86.022316.
 - 11 Dalhousie University Mathstat Cluster. URL: <https://www.mathstat.dal.ca/cluster/doku.php>.
 - 12 Bob Coecke and Ross Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13:043016, 2011. [arXiv:0906.4725]. doi:10.1088/1367-2630/13/4/043016.
 - 13 Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes: A First Course in Quantum Theory and Diagrammatic Reasoning*. Cambridge University Press, 2017. doi:10.1017/9781316219317.
 - 14 Niel de Beaudrap, Xiaoning Bian, and Quanlong Wang. Techniques to reduce $\pi/4$ -parity phase circuits, motivated by the zx calculus. In *to appear in Proceedings of QPL 2019*, 2019. [arXiv:1911.09039].
 - 15 Niel de Beaudrap, Ross Duncan, Dominic Horsman, and Simon Perdrix. Pauli fusion: a computational model to realise quantum transformations from zx terms. In *Proceedings of QPL 2019*, page to appear, 2019. [arXiv:1904.12817].
 - 16 Niel de Beaudrap and Dominic Horsman. The zx calculus is a language for surface code lattice surgery. *Quantum*, 4:218, 2020. [arXiv:1704.08670]. doi:10.22331/q-2020-01-09-218.
 - 17 Ross Duncan, Aleks Kissinger, Simon Perdrix, and John van de Wetering. Graph-theoretic simplification of quantum circuits with the zx-calculus. [arXiv:1902.03178], 2019.
 - 18 Ross Duncan and Simon Perdrix. Rewriting measurement-based quantum computations with generalised flow. In Samson Abramsky, Cyril Gavioille, Claude Kirchner, Friedhelm Meyer auf der Heide, and Paul G. Spirakis, editors, *Automata, Languages and Programming*, pages 285–296, Berlin, Heidelberg, 2010. Springer Berlin Heidelberg. doi:10.1007/s10472-009-9141-x.
 - 19 Simon Perdrix Emmanuel Jeandel and Renaud Vilmart. Completeness of the zx-calculus. [arXiv:1903.06035], 2019.
 - 20 Craig Gidney. Halving the cost of quantum addition. *Quantum*, 2:74, 2018. [arXiv:1709.06648]. doi:10.1007/s11128-011-0297-z.
 - 21 David Gosset, Vadym Kliuchnikov, Michele Mosca, and Vincent Russo. An algorithm for the t-count. *Quantum Info. Comput.*, 14(15–16):1261–1276, November 2014. [arXiv:1308.4134].
 - 22 D. Gottesman. Stabilizer codes and quantum error correction. 1997. Ph.D thesis. arXiv:quant-ph/9705052.

- 23 Luke E. Heyfron and Earl T. Campbell. An efficient quantum compiler that reduces t count. *Quantum Science and Technology*, 4(1):015004, 2018. [arXiv:1712.01557]. doi:10.1038/srep01939.
- 24 C. Horsman, A. G Fowler, S. Devitt, and R. Van Meter. Surface code quantum computing by lattice surgery. *New Journal of Physics*, 14(12):123011, 2012. [arXiv:1111.4022].
- 25 Cody Jones. Low-overhead constructions for the fault-tolerant toffoli gate. *Phys. Rev. A*, 87:022328, February 2013. [arXiv:1212.5069]. doi:10.1103/PhysRevA.87.022328.
- 26 Aleks Kissinger and John van de Wetering. Reducing t-count with the zx-calculus. [arXiv:1903.10477], 2019.
- 27 Daniel Litinski. A Game of Surface Codes: Large-scale quantum computing with lattice surgery. *Quantum*, 3:128, 2019. [arXiv:1808.02892]. doi:10.1103/PhysRevB.96.205413.
- 28 Dmitri Maslov. *Reversible Logic Synthesis Benchmarks page*. Accessed February 2020. URL: <http://webhome.cs.uvic.ca/~dmaslov>.
- 29 Giulia Meuli, Mathias Soeken, Earl Campbell, Martin Roetteler, and Giovanni De Micheli. The role of multiplicative complexity in compiling low T-count oracle circuits. [arXiv:1908.01609], 2019.
- 30 Yunseong Nam, Neil J. Ross, Yuan Su, Andrew M. Childs, and Dmitri Maslov. “optimiser” github. <https://github.com/njross/optimizer>.
- 31 Yunseong Nam, Neil J. Ross, Yuan Su, Andrew M. Childs, and Dmitri Maslov. Automated optimization of large quantum circuits with continuous parameters. *npj Quantum Information*, 4(1):23, 2018. [arXiv:1710.07345]. doi:10.1038/s41534-018-0072-4.
- 32 Michael A Nielsen and Isaac Chuang. *Quantum computation and quantum information*. Cambridge University Press, Cambridge UK, 2000.
- 33 Renaud Vilmart. A Near-Optimal Axiomatisation of ZX-Calculus for Pure Qubit Quantum Mechanics. In *34th Annual ACM/IEEE Symposium on Logic in Computer Science*, pages 1–10, 2019. [arXiv:1812.09114]. doi:10.1109/LICS.2019.8785765.
- 34 Fang Zhang and Jianxin Chen. Optimizing t gates in clifford+t circuit as $\pi/4$ rotations around paulis. [arXiv:1903.12456], 2019.

A ZX diagram reference

The ZX calculus — first developed by Coecke and Duncan [12] (see also Refs. [13, 33, 19] for updated treatments, and Refs. [17, 16, 15, 26] for applications to quantum technology) — is a relatively recently developed notation for quantum operations, equipped with rules (the “calculus” part) to compute with this notation. This article does not make explicit use of the “calculus” part of the ZX calculus: while it *does* make statements about equivalences of diagrams which *could* be shown using the calculus, these can and should be understood in the same way that other papers make statements of equivalences of conventional circuit diagrams.

We use ZX notation at various points to describe quantum circuits, including circuits with classically controlled operations and non-local unitaries such as $\pi/4$ -parity-phase operations. The ZX diagrams in this article can be read merely as a slightly unusual (but convenient) circuit notation. In this Appendix, we provide a reference for this notation, serving also as a glossary of sorts for various operations as they are represented in ZX diagrams, to allow readers to understand our results as well as conventional circuit diagrams would.

A.1 General statements

For the purposes of this article (and essentially all other practical purposes), ZX diagrams are representations of tensor networks. To represent quantum circuits, it is common to choose a direction in which to read the diagrams from “input” to “output”. (In our paper, we draw

these diagrams with input on the left and output on the right, as with the usual circuit notation.) The ZX diagrams of our work are composed of three different kinds of tensor nodes:

- **“Green” nodes** (which are lighter coloured in our article), which may have any number of indices, and as a tensor represents a sort of GHZ state over the standard basis. If a phase parameter θ is provided, the tensor also involves a relative phase of $e^{i\theta}$ between the two terms; otherwise $\theta = 0$ is assumed (and there is no relative phase).

$$\theta \left(\begin{array}{c} \text{---} \\ \vdots \\ \text{---} \end{array} \right) \Bigg\}^n = |0\rangle^{\otimes n} + e^{i\theta} |1\rangle^{\otimes n} \quad (22)$$

In principle, we also permit the border case of $n = 0$, in which case this represents the “tensor” $|0\rangle^{\otimes 0} + e^{i\theta} |1\rangle^{\otimes 0} = (1 + \cos(\theta)) + i \sin(\theta)$; though we don’t make use of such nodes in our results.

- **“Red” nodes** (which are darker coloured in our article), which may have any number of indices, and are similar to green nodes except that they are expressed in terms of the $\{|+\rangle, |-\rangle\}$ basis.

$$\theta \left(\begin{array}{c} \text{---} \\ \vdots \\ \text{---} \end{array} \right) \Bigg\}^n = |+\rangle^{\otimes n} + e^{i\theta} |-\rangle^{\otimes n} \quad (23)$$

- **“Hadamard” boxes**, which represent the usual 2×2 unitary Hadamard matrix.

$$\text{---} \boxed{H} \text{---} = |+\rangle\langle 0| + |-\rangle\langle 1| \quad (24)$$

To represent operations taking some qubits as input, we change of some of the “kets” in the tensor nodes to “bras” — but as $|0\rangle, |1\rangle, |+\rangle$ and $|-\rangle$ are real vectors, this change does not affect any of the tensor coefficients. This allows us to be flexible with our diagrams, and avoid committing to the indices of each node as being explicitly an “input” or an “output”, unless it is a free index of the whole diagram. (In particular, this allows us to draw some closed indices by *vertical* wires, without confusion.)

In the rest of this appendix, we describe some simple examples (and simple extensions) of this notation, which the interested reader should find themselves able to verify by routine calculation.

A.2 Single-node diagrams

With (light) green or (dark) red nodes of degree 1, we may easily represent states of the $\{|0\rangle, |1\rangle\}$ basis or $\{|+\rangle, |-\rangle\}$ basis, albeit supernormalised by a factor of $\sqrt{2}$.

$$\text{red node} = |+\rangle^{\otimes 1} + |-\rangle^{\otimes 1} = \sqrt{2} |0\rangle; \quad \pi \text{ red node} = |+\rangle^{\otimes 1} - |-\rangle^{\otimes 1} = \sqrt{2} |1\rangle; \quad (25)$$

$$\text{green node} = |0\rangle^{\otimes 1} + |1\rangle^{\otimes 1} = \sqrt{2} |+\rangle; \quad \pi \text{ green node} = |0\rangle^{\otimes 1} - |1\rangle^{\otimes 1} = \sqrt{2} |-\rangle. \quad (26)$$

More generally, green degree-1 nodes may be used to represent newly prepared qubits in the XY plane of the Bloch sphere, and red degree-1 nodes may be used to represent newly prepared qubits in the YZ plane of the Bloch sphere, up to the same supernormalisation of $\sqrt{2}$. This additional factor of $\sqrt{2}$ does not affect our results: the additional factor may be accounted for any time we represent the preparation of a qubit in one of these states.

We may also represent single-qubit measurements by degree-1 nodes oriented in the opposite direction. As re-orienting edges from the right of a node to the left corresponds to turning $|0\rangle$ to $\langle 0|$, turning $|1\rangle$ to $\langle 1|$, and so forth, we then have

$$\text{---}\bullet = \sqrt{2}\langle 0|; \quad \text{---}\bullet\pi = \sqrt{2}\langle 1|; \quad (27)$$

$$\text{---}\circ = \sqrt{2}\langle +|; \quad \text{---}\circ\pi = \sqrt{2}\langle -|. \quad (28)$$

Again, the additional factor of $\sqrt{2}$ may be accounted for any time we represent a projection of a qubit in one of these states. To represent a measurement which may yield either $|0\rangle$ or $|1\rangle$, or either $|+\rangle$ or $|-\rangle$, we may introduce a variable $s \in \{0, 1\}$ representing whether a relative phase of π is absent in the result ($s = 0$, for the states $|0\rangle$ or $|+\rangle$) or present in the result ($s = 1$, for the states $|1\rangle$ or $|-\rangle$). We then represent measurement in the $\{|0\rangle, |1\rangle\}$ basis and the $\{|+\rangle, |-\rangle\}$ basis respectively as

$$\text{---}\bullet\pi, \{s\} = \langle +| + e^{is\pi}\langle -| \in \{\sqrt{2}\langle 0|, \sqrt{2}\langle 1|\}; \quad (29)$$

$$\text{---}\circ\pi, \{s\} = \langle 0| + e^{is\pi}\langle 1| \in \{\sqrt{2}\langle +|, \sqrt{2}\langle -|\}. \quad (30)$$

The bit s is in effect a random variable representing the measurement outcome.

In other ZX diagrams (including on nodes of degree 2 or higher), we may use a set $S = \{s_1, s_2, \dots\}$ in place of the set $\{s\}$. This indicates a node in which the presence or absense of the phase of π depends on the parity ($s_1 \oplus s_2 \oplus \dots$) of the entire set S , rather than on the single bit s . For example, we may represent Z rotations and X rotations each by a single node of degree 2:

$$\text{---}\circ^\theta = |0\rangle\langle 0| + e^{i\theta}|1\rangle\langle 1| = R_z(\theta), \quad \text{---}\bullet^\theta = |+\rangle\langle +| + e^{i\theta}|-\rangle\langle -| = R_x(\theta); \quad (31)$$

Then, the following diagrams represent the same operations, conditioned on the parity $\mathbf{s} = \bigoplus_j s_j$ of a set of bits $S = \{s_1, s_2, \dots\}$:

$$\text{---}\circ^{\theta, S} = R_z(\mathbf{s}\theta) = R_z(\theta)^{\mathbf{s}}, \quad \text{---}\bullet^{\theta, S} = R_x(\mathbf{s}\theta) = R_x(\theta)^{\mathbf{s}}. \quad (32)$$

This feature of the ZX calculus does not play a prominent role in our work, but is present in our treatment of the Hadamard gadget (Eqn. (15) on page 10) and in principle useful to represent the circuits which we would obtain by representing conditionally-controlled Clifford operations in the ZX calculus.

A.3 Two-node diagrams

Diagrams of more than one node can be easily constructed simply by composing nodes on their edges. In many cases, this has the same meaning as in conventional quantum circuit diagrams (with the same “feature” that the algebra is read right-to-left, even though the diagram is read left-to-right): for example,

$$\text{---}\circ^\theta \text{---} \boxed{H} \text{---} = HR_z(\theta) = R_x(\theta)H = \text{---} \boxed{H} \text{---} \bullet^\theta \text{---} \quad (33)$$

$$\text{---}\circ^\theta \text{---} \bullet^\pi \text{---} = XR_z(\theta) = R_z(-\theta)X = \text{---} \bullet^\pi \text{---} \circ^{-\theta} \text{---} \quad (34)$$

As with circuit diagrams, we may also represent the tensor product of operations by representing operations happening on different wires in parallel — for example:

$$\begin{array}{c} \theta \\ \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \varphi \end{array} = R_z(\theta) \otimes R_x(\varphi). \quad (35)$$

Not all “compositions” of nodes take these forms, however: in general we may compose any two nodes simply by connecting their edges (corresponding to contracting the shared indices of the tensor nodes). An especially important case in point is the way that CNOT operators are represented as ZX terms. As with single-qubit states, the usual representation of CNOT by ZX diagrams is not precisely normalised:

$$\begin{aligned} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} &= |0\rangle\langle 0| \otimes \langle 0|_+ \otimes |+\rangle\langle +| + |0\rangle\langle 0| \otimes \langle 0|_- \otimes |-\rangle\langle -| \\ &\quad + |1\rangle\langle 1| \otimes \langle 1|_+ \otimes |+\rangle\langle +| + |1\rangle\langle 1| \otimes \langle 1|_- \otimes |-\rangle\langle -| \\ &= \frac{1}{\sqrt{2}} |0\rangle\langle 0| \otimes \mathbb{1} + \frac{1}{\sqrt{2}} |1\rangle\langle 1| \otimes X = \frac{1}{\sqrt{2}} \text{CNOT}. \end{aligned} \quad (36)$$

(Again, the subnormalisation of this diagram does not affect our analysis, and can in principle be accounted for in the ZX representation of any circuit involving CNOT gates.) Note that the shared wire between the red and green dot does not have a specific interpretation as an “input” or an “output” of either — nor is this necessary to provide the interpretation of the diagram as an operator.

A.4 Multi-node diagrams

Composing the diagrams above, in series or in parallel (and with appropriate accounting for normalisation), suffices to represent an arbitrary unitary operation by the (slightly redundant) gate set consisting of arbitrary X and Z rotations, Hadamard gates, and CNOT operations. We may also more directly represent somewhat more “exotic” operators using ZX diagrams, and $\pi/4$ -parity-phase operations are in this case the most relevant example: for instance,

$$\begin{aligned} \begin{array}{c} \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \\ \text{---} \text{---} \text{---} \text{---} \text{---} \text{---} \end{array} &= \left(\langle 0|_d + e^{i\theta} \langle 1|_d \right) \left(|+\rangle_d \langle +++|_{abc} + |-\rangle_d \langle --|_{abc} \right) \\ &\quad \times \left(|0\rangle\langle 0|_1 \otimes |0\rangle_a + |1\rangle\langle 1|_1 \otimes |1\rangle_a \right) \left(|0\rangle\langle 0|_3 \otimes |0\rangle_b + |1\rangle\langle 1|_3 \otimes |1\rangle_b \right) \\ &\quad \times \left(|0\rangle\langle 0|_5 \otimes |0\rangle_c + |1\rangle\langle 1|_5 \otimes |1\rangle_c \right) \\ &= \frac{1}{2\sqrt{2}} \sum_{a,b,c \in \{0,1\}} \left(\langle 0|_d + e^{i\theta} \langle 1|_d \right) \left(|+\rangle_d + (-1)^{a+b+c} |-\rangle_d \right) \otimes |a,b,c\rangle \langle a,b,c|_{1,3,5} \\ &= \frac{1}{4} \sum_{a,b,c \in \{0,1\}} \left(1 + e^{i\theta} + (-1)^{a+b+c} - (-1)^{a+b+c} e^{i\theta} \right) |a,b,c\rangle \langle a,b,c|_{1,3,5} \\ &= \frac{1}{2} \sum_{\substack{a,b,c \in \{0,1\} \\ a \oplus b \oplus c = 0}} |a,b,c\rangle \langle a,b,c|_{1,3,5} + \frac{1}{2} \sum_{\substack{a,b,c \in \{0,1\} \\ a \oplus b \oplus c = 1}} e^{i\theta} |a,b,c\rangle \langle a,b,c|_{1,3,5} \\ &= \frac{1}{2} e^{i\theta/2} \exp\left(\frac{1}{2} i\theta (Z \otimes \mathbb{1} \otimes Z \otimes \mathbb{1} \otimes Z)\right). \end{aligned} \quad (37)$$

Again, the subnormalisation by a factor of $\frac{1}{2}$ does not affect our analysis, which is in principle about products of $D_{S,3}$ operators — merely denoted in our work by these phase gadgets, for convenience — which are proportional to the identity by a global phase.

The existence of rules for transforming ZX diagrams allows us to reason (*i.e.*, to compute) effectively about these diagrams without the need to expand their meaning algebraically as we have been doing in this Appendix. This has particularly motivated our use of the ZX calculus in our work, as a convenient notational tool and also as a means by which we performed our analysis.

For more information about the ZX calculus, and in particular for resources to learn about these diagrammatic computational methods, the interested reader is invited to visit [zxcalculus.com].

B Details of the moveH subroutine and CNOT-commutation heuristic

In this Appendix, we describe our procedures for H gate extraction and CNOT gate extraction (used in Steps 2 and 4 of the procedure described in Section 4.2) on a high level. For more details, the interested reader may view our source code on Github [https://github.com/onestruggler/fast-stomp].

B.1 The moveH subroutine

Our procedure for extracting H gates from a circuit are built on a subroutine `moveH`, which attempts to move each H gate as far to the right (the end of the circuit) as possible.

Representing the circuit as a list of gates in a particular order (without parallelisation), this procedure looks for the first H gate, and attempts to move it to the right. In doing so, it makes use of several simple commutation relations or opportunities for cancellation, for example:

$$\begin{array}{lll}
 \boxed{H} \boxed{H} \rightarrow \text{---}; & \boxed{H} \boxed{X} \rightarrow \boxed{Z} \boxed{H}; & \boxed{H} \boxed{Z} \rightarrow \boxed{X} \boxed{H}; \\
 \boxed{H} \oplus \rightarrow \text{---}; & \boxed{H} \bullet \rightarrow \oplus \boxed{H}; & \boxed{H} \text{---} \rightarrow \text{---} \boxed{H}.
 \end{array} \tag{38}$$

If the procedure moves the H gate to a point that it precedes a second H gate, it proceeds recursively to attempt to move the second H gate before continuing with the first. When the procedure is finished attempting (successfully or otherwise) to move the second H gate, it returns to the task of moving the first — moving this gate past the other H gate, if the attempt to move it ended in failure. This process continues until the procedure has stopped trying to move what originally was the first H gate.

In attempting to move H gates, `moveH` may encounter situations in which no progress is possible, without trying to move or cancel other kinds of gates. For instance: in a circuit consisting only of an H gate followed by four T gates on a single wire, it is possible to move the H gate to the end, but only after “pushing” the T gate which follows it to the right, accumulating the other phase gates to form a Z gate. In general, if `moveH` encounters a gate G for which there is no commutation rule provided, it attempts instead to push G forward, to commute with, accumulate with, or cancel against gates further to its right. In doing so, `moveH` may encounter yet another gate F for which G has no provided commutation relation, in which case `moveH` will attempt to move F further to the right, and so on.

In some cases, there are fruitful opportunities for multi-gate substitutions which either reduce the number of H gates, or allows an H gate to be moved further to the right. For instance:

- If in moving an H gate to the right we encounter an S gate followed by an H gate, `moveH` first tries to move the second H gate. If this fails, we may apply the transformation

$$- \boxed{H} - \boxed{S} - \boxed{H} - \rightarrow - \boxed{S} - \boxed{Z} - \boxed{H} - \boxed{S} - \boxed{Z} - . \quad (39)$$

This reduces the number of H gates by 1. We then move the S and Z gates further to the right, then continue by moving the new H gate to the right.

- If in moving an H gate to the right we encounter the *control* qubit of a CNOT gate, followed by an H gate on either the control or target, we again first try to move the H gate. If this fails, we may apply one of the transformations

$$\begin{array}{c} \text{---} \oplus \text{---} \\ | \\ \text{---} \boxed{H} \bullet \boxed{H} \text{---} \end{array} \rightarrow \begin{array}{c} \text{---} \boxed{H} \bullet \text{---} \\ | \\ \text{---} \oplus \text{---} \end{array} ; \quad \begin{array}{c} \text{---} \oplus \boxed{H} \text{---} \\ | \\ \text{---} \boxed{H} \bullet \text{---} \end{array} \rightarrow \begin{array}{c} \text{---} \boxed{H} \bullet \text{---} \\ | \\ \text{---} \oplus \boxed{H} \text{---} \end{array} . \quad (40)$$

This doesn't directly reduce the number of H gates, but may make it possible to move the later H gate and the CNOT gate to the right before continuing further, thereby providing an alternative for at least one of the two H gates to be moved further to the right.

The details of all commutation relations which we define for all of the gates are not important, except that it is important to define these rules in such a way that the procedure terminates (rather than repeatedly commute two gates such as T and CCZ past one another, in an attempt to cancel them so that an H gate can be moved to the right of both). Different techniques will lead to different performances in the ability of `moveH` to reduce the number of H gates which precede any non-Clifford gate.

B.2 CNOT movement heuristic

In Step 4, we move all operations which are not single-qubit phase operations or phase-parity operations out of the main body of the circuit. The way that CNOT gates are treated aims, roughly, to avoid generating phase-parity operations on very large subsystems, but does so in a way that attempts to avoid performing too much computation.

The heuristic used to determine which direction to move a CNOT operation is as follows. For each CNOT gate, from the first in the circuit to the last, we compute the following:

1. Compute the set P_L of all phase-parity gadgets to the left which act on the target but not the control of the CNOT, and the set M_L of phase-parity gadgets to the left which act on its target and control both.
2. Similarly, compute the set P_R of phase-parity gadgets to the right which act on the target but not the control of the CNOT, and the set M_R of phase-parity gadgets to the left which act on its target and control both.
3. If $P_L - M_L < P_R - M_R$, we prefer to move the CNOT to the left; otherwise we prefer to move it to the right.

If no other CNOT gate acted on any qubits in common with this left-most CNOT gate, the quantity P_L (respectively, M_L) would correctly indicate how many phase-parity gadgets would act on one more qubit (respectively, one fewer) if we commuted that CNOT to the left. The difference $P_L - M_L$ then indicates the net change in the cumulative number of qubits acted on by the phase-gadgets to the left of the CNOT. Similar remarks apply for $P_R - M_R$, albeit with the important caveat that this figure may be inaccurate if there are further CNOT gates to the right whose targets coincide with the control of the CNOT under consideration.

The approach taken to produce our results is as follows. For the left-most CNOT in the circuit, compute P_L , M_L , P_R , and M_R . Commute the CNOT gate to the left if $P_L - M_L < P_R - M_R$, and otherwise to commute it to the right. If in commuting it to the right we encounter another CNOT gate with which it does not commute, we also commute that CNOT gate to the right (and any CNOT gates with which *those* do not commute, *etc.*) Having done this, we compute P_L , M_L , P_R , and M_R for the leftmost remaining CNOT gate in the circuit, where these may depend on the commutations which occurred for the previous CNOT gate. We proceed in this way, recursively from left to right, until no more CNOT gates are in the main body of the circuit.


A Device-Independent Protocol for XOR Oblivious Transfer

Srijita Kundu

Centre for Quantum Technologies, National University of Singapore, Singapore
srijita.kundu@u.nus.edu

Jamie Sikora

Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada
jsikora@perimeterinstitute.ca

Ernest Y.-Z. Tan 

Institute for Theoretical Physics, ETH Zürich, Switzerland
ernestt@ethz.ch

Abstract

Oblivious transfer is a cryptographic primitive where Alice has two bits and Bob wishes to learn some function of them. Ideally, Alice should not learn Bob's desired function choice and Bob should not learn any more than logically implied by the function value. While decent quantum protocols for this task are known, many quickly become insecure if an adversary were to control the quantum devices used in the implementation of the protocol. Here we present how some existing protocols fail in this device-independent framework, and give a fully-device independent quantum protocol for XOR oblivious transfer which is provably more secure than any classical protocol.

2012 ACM Subject Classification Security and privacy → Cryptography; Theory of computation → Cryptographic primitives

Keywords and phrases Quantum cryptography, device independence, oblivious transfer, semidefinite programming, security analysis

Digital Object Identifier 10.4230/LIPIcs.TQC.2020.12

Related Version A full version of the paper is available at <https://arxiv.org/abs/2006.06671>.

Supplementary Material The code used for the SDP computations is available at <https://github.com/ernesttyz/dixot>.

Funding *Srijita Kundu*: Supported by the Singapore Ministry of Education and the National Research Foundation, Prime Minister's Office, Singapore. Part of this work was done when S. K. was visiting the Institute for Quantum Computing at the University of Waterloo, Canada.

Jamie Sikora: Supported by Government of Canada through the Department of Innovation, Science and Economic Development Canada, and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade.

Ernest Y.-Z. Tan: Supported by the Swiss National Science Foundation via the National Center for Competence in Research for Quantum Science and Technology (QSIT), the Air Force Office of Scientific Research (AFOSR) via grant FA9550-19-1-0202, and the QuantERA project eDICT.

Acknowledgements We thank Jean-Daniel Bancal, Andrea Coladangelo, Lúcia del Río, Honghao Fu, Anand Natarajan, Christopher Portmann, Xingyao Wu and Vilasini Venkatesh for helpful discussions.



© Srijita Kundu, Jamie Sikora, and Ernest Y.-Z. Tan;
licensed under Creative Commons License CC-BY

15th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2020).

Editor: Steven T. Flammia; Article No. 12; pp. 12:1–12:16



Leibniz International Proceedings in Informatics

Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

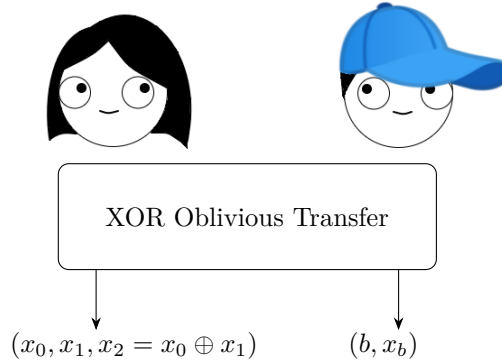
1 Introduction

Oblivious transfer is an important cryptographic primitive in two-party computation as it can be used as a universal building block for constructing more elaborate protocols [10]. Indeed, some quantum protocols for this task are known [22, 18, 6, 5]. It can be shown that there do not exist classical protocols with any level of information-theoretic security, and there do not exist quantum protocols with perfect security [6, 11].

In this paper, we consider a variant of oblivious transfer called *XOR oblivious transfer (XOT)*. This is the two-party cryptographic primitive in which two spatially separated parties, Alice and Bob, wish to do the following task: Alice outputs two bits (x_0, x_1) , which are uniformly random in $\{0, 1\}^2$, and Bob outputs b which is uniformly random in $\{0, 1, 2\}$, as well as x_b where we define $x_2 = x_0 \oplus x_1$. In other words, Alice and Bob communicate and Bob learns one bit of information from Alice's two bits (either the first bit, second bit, or their XOR). When designing XOT protocols, the security goals are:

1. *Completeness*: If both parties are honest, then their outcomes are consistent (i.e., x_b is the correct value), uniformly random, and neither party aborts.
2. *Soundness against cheating Bob*: If Alice is honest, then a dishonest (i.e., cheating) Bob cannot learn both x_0 and x_1 by digressing from protocol.
3. *Soundness against cheating Alice*: If Bob is honest, then a dishonest (i.e., cheating) Alice cannot learn b by digressing from protocol.

► **Remark 1.** One could imagine a situation where Alice already has a fixed choice of (x_0, x_1) that she wishes to input into a XOT protocol (perhaps from the result of an earlier computation). However, we can use the outcomes of an XOT protocol as described above as a one-time pad to convey the information to Bob. For more details, see [6].



■ **Figure 1** Desired outputs for XOR oblivious transfer.

In this paper we are concerned with *information-theoretic security*, meaning that Alice and Bob are only bounded by the laws of quantum mechanics. In other words, Alice and Bob can perform arbitrarily complicated computations, have arbitrarily large quantum memories, and so on. We shall have occasion to change how much control Alice and Bob have over the protocol, but precisely what actions are allowed to be performed by dishonest parties should be clear from context, and will be described shortly.

We focus on studying XOT protocols from the perspective of assuming perfect completeness and trying to make them as sound as possible.¹ To this end, we choose to quantify the soundness via *cheating probabilities*, which we define as follows:

- P_B^{XOT} : The maximum probability with which a dishonest Bob can learn both of honest Alice's outcome (x_0, x_1) and the protocol does not abort.
- P_A^{XOT} : The maximum probability with which a dishonest Alice can learn honest Bob's choice outcome b and the protocol does not abort.

Any XOT protocol with perfect completeness necessarily has $P_A^{\text{XOT}} \geq \frac{1}{3}$ and $P_B^{\text{XOT}} \geq \frac{1}{2}$, since a dishonest Alice could always guess one of three choices for Bob's outcome b uniformly at random; similarly, dishonest Bob can follow the honest protocol to gain perfect knowledge of x_0 , x_1 , or $x_0 \oplus x_1$, and then randomly guess the unknown bit in Alice's outcome (x_0, x_1) .

As mentioned earlier, XOT is a variation on the more well-known cryptographic primitive, oblivious transfer (OT). In an oblivious transfer protocol, Alice and Bob wish to do the following task: Alice outputs two bits (x_0, x_1) , which are uniformly random in $\{0, 1\}^2$, and Bob outputs (b, x_b) where b is uniformly random in $\{0, 1\}$. The completeness and soundness against cheating Alice and Bob for OT are the same as in XOT, the only difference being that here Alice is trying to learn which one of two possible values of b Bob has. Any OT protocol with perfect completeness has $P_A^{\text{OT}} \geq 1/2$, $P_B^{\text{OT}} \geq \frac{1}{2}$.

► **Remark 2.** In this work, we chose to quantify soundness via cheating probabilities, but we note that such a measure of security is not necessarily *composable* [23, 21]. Unfortunately, it can be very challenging to prove that a protocol is composable secure, and in some settings such protocols are in fact impossible [21]. As a first analysis of the protocol proposed in this work, we will restrict ourselves to studying the cheating probabilities only.

Since the lowest possible bounds on P_A^{XOT} and P_B^{XOT} for perfectly complete XOT protocols are asymmetric, we shall consider them in pairs and will not concern ourselves with finding an “optimal protocol”. That is, we consider the security of XOT protocols a partial ordering. Instead, we motivate our work by asking the following question:

“Is it possible to create quantum protocols where both $P_A^{\text{XOT}}, P_B^{\text{XOT}} < 1$ when Alice and Bob do not even trust their own quantum devices?”

Taken literally, this statement cannot be true, since arbitrarily malicious devices could simply broadcast all desired information to a dishonest party. However, it turns out that there exist quantum protocols that can be proven secure using almost no assumptions other than ruling out this extreme scenario (which seems a rather necessary assumption in any case). This is the notion of *device-independent* security, which typically exploits *nonlocal games* played using entangled states. In a fully device-independent model, one only assumes that the parties' devices do not directly broadcast certain information to the dishonest party and/or each other (we shall explain this in more detail in Section 1.4). In particular, one does not assume that the states and/or measurements implemented by the devices are known, and even the dimensions of the quantum systems are not specified. Device-independent security analyses exist for other cryptographic tasks such as quantum key distribution [17, 3], bit commitment [20, 2], and coin-flipping [1].

¹ To contrast, the task of finding protocols with perfect soundness and the best possible completeness was considered in [19].

While we have described the fully device-independent framework above, one can instead choose to trust some subset of the properties described, leading to various levels of semi-device-independent security. For instance, Alice and Bob could trust state preparation devices, measurement devices, quantum operations, or any combination of the above.

In this work, we examine the security of XOT quantum protocols in semi-device-independent and device-independent scenarios. By a slight abuse of notation, we use the same notation P_A^{XOT} and P_B^{XOT} to denote the cheating probabilities of Alice and Bob in all the different scenarios, corresponding to differently defined cheating capabilities of the dishonest party. For example, if we were to allow a dishonest Alice to control Bob's measurements, it may lead to a different value of P_A^{XOT} . The cheating capabilities of cheating parties should be clear when we discuss P_A^{XOT} and P_B^{XOT} .

1.1 Trivial protocols

For readers new to oblivious transfer, we present two bad classical protocols and one bad quantum protocol.

► **Protocol 1** (Bad XOT Protocol 1).

1. Alice chooses (x_0, x_1) uniformly at random and sends (x_0, x_1) to Bob.
2. Bob chooses b uniformly at random.
3. Alice outputs (x_0, x_1) and Bob outputs (b, x_b) .

A moment's thought shows that Bob has full information (he clearly learns (x_0, x_1)) while Alice has no information. Therefore, we have

$$P_A^{\text{XOT}} = 1/3 \quad \text{and} \quad P_B^{\text{XOT}} = 1 \tag{1}$$

which is as insecure concerning cheating Bob as possible.

► **Protocol 2** (Bad XOT Protocol 2).

1. Bob chooses b uniformly at random and tells Alice his choice.
2. Alice chooses and outputs (x_0, x_1) uniformly at random and sends x_b to Bob.
3. Alice outputs (x_0, x_1) and Bob outputs (b, x_b) .

Here Alice has full information while Bob has none. Therefore, here we have

$$P_A^{\text{XOT}} = 1 \quad \text{and} \quad P_B^{\text{XOT}} = 1/2. \tag{2}$$

► **Remark 3.** Surprisingly, these protocols can be useful for protocol design. For instance, suppose Alice wishes to test Bob to see if he has been cheating, and aborts if and only if the test fails. Then if the test passes, the parties need a way to finish executing the protocol, which they could do using Protocol 1 (which is independent of previous steps in the tested protocol).

1.2 A quantum protocol for XOR oblivious transfer with no device-independent security

It is tricky creating a protocol in which Alice and Bob both cannot cheat with high probability. To this end, we often enlist the aid of quantum mechanics. Here we present the oblivious transfer protocol from [6] adapted to the XOT setting.

For $b \in \{0, 1, 2\}$, let $|\psi_b^\pm\rangle \in \mathcal{XY}$ denote the following two-qutrit state:

$$|\psi_b^\pm\rangle = \begin{cases} \frac{1}{\sqrt{2}}(|00\rangle_{\mathcal{XY}} \pm |22\rangle_{\mathcal{XY}}) & \text{if } b = 0, \\ \frac{1}{\sqrt{2}}(|11\rangle_{\mathcal{XY}} \pm |22\rangle_{\mathcal{XY}}) & \text{if } b = 1, \\ \frac{1}{\sqrt{2}}(|00\rangle_{\mathcal{XY}} \pm |11\rangle_{\mathcal{XY}}) & \text{if } b = 2. \end{cases} \quad (3)$$

Note that for every $b \in \{0, 1, 2\}$, we have that $|\psi_b^+\rangle$ and $|\psi_b^-\rangle$ are orthogonal. We are now ready to state the protocol.

► **Protocol 3.**

1. Bob chooses $b \in \{0, 1, 2\}$ uniformly at random, prepares the state $|\psi_b^+\rangle$ in registers \mathcal{XY} , and sends the register \mathcal{X} to Alice.
2. Alice chooses (x_0, x_1) uniformly at random, performs the unitary

$$U_{(x_0, x_1)} = (-1)^{x_0}|0\rangle\langle 0| + (-1)^{x_1}|1\rangle\langle 1| + |2\rangle\langle 2| \quad (4)$$

on \mathcal{X} , and then sends it back to Bob.

3. Bob performs the 2-outcome measurement $\{|\psi_b^+\rangle\langle\psi_b^+|, \mathbb{1} - |\psi_b^+\rangle\langle\psi_b^+|\}$ and records his outcome as $c = 0$ if he gets $|\psi_b^+\rangle\langle\psi_b^+|$ and $c = 1$ otherwise.
4. Alice outputs (x_0, x_1) , Bob outputs (b, c) .

Protocol 3 can be checked to be complete (i.e., Bob gets the correct outcome). The cheating probabilities in this protocol in the cases of trusted and untrusted devices are given by Theorem 4 below. We give a proof for the trusted case Section 3, and the relatively simple proof for the untrusted case is given here.

► **Theorem 4.** *In Protocol 3, the cheating probabilities are as listed in the following table. (In the untrusted setting, Alice controls Bob's state preparation and Bob controls Alice's unitary.)*

	P_A^{XOT}	P_B^{XOT}
Trusted devices	1/2	3/4
Untrusted devices	1	1

We show here that this protocol breaks down when they do not trust their own devices. For example, assume Bob has full control over Alice's unitary. It could be a unitary which implements a superdense coding protocol:

$$U_{(x_0, x_1)}^{\text{cheat}} = \begin{cases} \mathbb{1}_{\mathcal{X}} & \text{if } (x_0, x_1) = (0, 0), \\ \sigma_X & \text{if } (x_0, x_1) = (0, 1), \\ \sigma_Y & \text{if } (x_0, x_1) = (1, 0), \\ \sigma_Z & \text{if } (x_0, x_1) = (1, 1). \end{cases} \quad (5)$$

Note that this unitary acts on a qubit (which we can assume Bob sends if he wishes), or just define it such that it acts trivially on the $|2\rangle$ subspace. Now, if Bob creates $|\psi_2^+\rangle$, he is left with a state from the Bell basis. In other words, he can perfectly learn (x_0, x_1) .

In the case of cheating Alice, she can simply control Bob's state preparation device to prepare the state $|b\rangle$ on input b , and have Bob send that. From this state, Alice simply measures it to learn b . Thus, we have

$$P_A^{\text{XOT}} = 1 \quad \text{and} \quad P_B^{\text{XOT}} = 1 \quad [\text{Devices are NOT trusted}]. \quad (6)$$

Thus, we need a clever way to design protocols where Alice and Bob cannot cheat in the above way. This motivates the need for device-independent XOT protocols and some protocol design ideas that should be avoided.

1.3 XOR oblivious transfer from the magic square game – A case of untrusted sources

Similar to device-independent protocols which exist for other cryptographic tasks, we design our protocols using nonlocal games. In this work, we shall make use of the nonlocal game known as the Mermin-Peres magic square game. In the magic square game,

- Alice and Bob receive respective inputs $a \in \{0, 1, 2\}$ and $b \in \{0, 1, 2\}$ independently and uniformly at random.
- Alice outputs $(x_0, x_1, x_2) \in \{0, 1\}^3$ such that $x_0 \oplus x_1 \oplus x_2 = 0$ and Bob outputs $(y_0, y_1, y_2) \in \{0, 1\}^3$ such that $y_0 \oplus y_1 \oplus y_2 = 1$.
- Alice and Bob win the game if $x_b = y_a$.

If Alice and Bob are allowed only classical strategies (using e.g. shared randomness), the magic square game cannot be won with probability greater than $8/9$. However, there exists a quantum strategy where Alice and Bob share prior entanglement, with which the magic square game can be won with probability 1. We shall refer to this strategy as the *magic square strategy* which is detailed in Section 2.1.

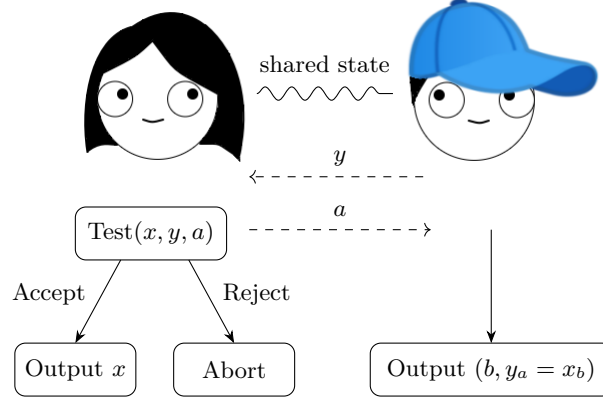
Now, suppose Alice and Bob play the magic square game according to the above description. Notice that x_2 will always be equal to $x_0 \oplus x_1$, similar to the definition of XOT, and that (x_0, x_1) is uniformly distributed (see Section 2.1). Also, for each of Bob's input choices, he learns either x_0 , x_1 , or $x_2 = x_0 \oplus x_1$ depending on the choice of input a for Alice. Since a is chosen uniformly at random, this is almost a proper XOT protocol (putting aside soundness for now). The only missing ingredient is that Bob knows he has x_b , but does not know which of the bits of (y_0, y_1, y_2) it is. To fix this small issue, we simply have Alice tell Bob which bit it is.

We formalize this protocol below and add in a *test* step that helps to prevent cheating. Strictly speaking, Protocol 4 should be thought of as a protocol framework, as we have not specified who creates the entangled state that Alice and Bob share – either party can. We consider different security analyses of Protocol 4, corresponding to each of these different cases. In the trusted state analysis, the honest party (whomever that may be) creates the state, and in the untrusted state analysis it is the cheating party who does so.

► **Protocol 4** (XOR oblivious transfer from the magic square game).

1. Alice and Bob share the bipartite state used in the magic square strategy.
2. Bob chooses $b \in \{0, 1, 2\}$ uniformly at random, performs the measurements corresponding to b in the magic square strategy to his state to get the outcome (y_0, y_1, y_2) , and sends (y_0, y_1, y_2) to Alice.
3. Alice chooses $a \in \{0, 1, 2\}$ uniformly at random and sends a to Bob.
4. Alice performs the measurement corresponding to a in her magic square strategy on her state to get outcome (x_0, x_1, x_2) .
5. **Test:** If $(x_0, x_1) = (0, 0)$ and Bob has sent (y_0, y_1) such that $y_a = 1$, then Alice aborts.
6. Alice outputs (x_0, x_1) and Bob outputs (b, y_a) .

Intuitively, the test step in Protocol 4 serves as a weak test that the magic square winning condition is fulfilled (though the test only occurs with somewhat small probability). This provides a way to partially certify that Bob has measured his share of the state before



■ **Figure 2** Schematic depiction of the messages sent in Protocol 4.

learning Alice’s input. (Note that if Bob has not measured his state yet, then even when the states are trusted, he has the potential to perfectly learn Alice’s output after learning her input, by simply performing the same measurement as Alice on his state. However, forcing him to perform the magic square measurement “deletes” some of this information; a notion referred to as *certified deletion* in [8].)

It can be checked that Protocol 4 accomplishes what XOT wants and that it never aborts in the honest case; in other words, the protocol is complete. To prove its soundness, we bound the cheating probabilities using appropriate SDPs, as described in the full version. Our numerical results are shown as Theorem 5 below.

► **Theorem 5.** *We assume Alice and Bob play the “canonical” strategy for the magic square game (see Section 2.1). Then the cheating probabilities for Alice and Bob in Protocol 4 are bounded by the values in the table below, rounded to 5 decimal places. The bounds for cheating Alice are tight.*

Upper bounds	P_A^{XOT}	P_B^{XOT}
Trusted state	0.83333	0.93628
Untrusted state	0.87268	0.94096
Untrusted measurement	1	1

As the last row of the table in Theorem 5 indicates, Protocol 4 is not fully device-independent. To see this, note that if Bob were to control Alice’s measurements, he can force $(x_0, x_1) = (0, 1)$ to always occur (regardless of the state, by performing a trivial measurement), and then he will never be tested. Moreover, he will learn (x_0, x_1) perfectly. Conversely, if Alice controls Bob’s measurement, she can force the output to be such that $y_0 + y_1 = b$ (note the sum is not modulo 2). Then, Bob’s message will reveal b to Alice.

1.4 A device-independent protocol

We now aim to find an XOT protocol based on the magic square game that is fully device-independent. We shall first clarify the premises and assumptions in such a setting. Specifically, we shall suppose that Alice and Bob each possess one of a pair of black boxes, each of which will accept a classical input in $\{0, 1, 2\}$ and return a classical output in $\{0, 1\}^3$. We shall

only require a single use of these boxes. In the honest scenario, the boxes will simply be implementing the ideal magic square states and measurements. If either party is dishonest, however, we shall suppose only that the boxes' behaviour can be modelled as follows: the boxes share some entangled state between them, and when one of the boxes receives an input, it returns the output of some measurement (conditioned on the input) performed on its share of the quantum state, *without broadcasting either its input or output* to any party other than the one holding the box². While the honest party can only interact with the box as specified, the dishonest party is able to “open” any box they possess and perform arbitrary quantum operations or measurements on the share of the state held by that box. However, the dishonest party cannot interact with or change the behaviour of a box while it is in the honest party's possession.

This describes the general device-independent setting. For the purposes of this work, we shall impose a small additional assumption on the states and measurements the boxes implement, namely that they are described by a tensor product of Hilbert spaces, one for each box. More general scenarios could be considered (for instance, one could require only that the two boxes' measurements commute), but these are outside the scope of this work.

It would seem difficult to design a secure protocol under such weak assumptions. However, we can exploit the fact that many nonlocal games (including the magic square game) exhibit the important property of *self-testing* or *rigidity*: if the boxes win the game with probability equal to the maximum quantum value, then they must be implementing the ideal state and measurements (up to trivial isometries). A robust version of this statement is formally expressed as Lemma 9 in the next section.

This suggests the following idea to make Protocol 4 fully device-independent: we introduce an initial step where with some probability, either party may ask the other to send over their box, so they can perform a single-shot test of whether the boxes win the magic square game. To prevent a dishonest party from always calling for a test, we shall enforce that a party calling for a test must then cede all control if the test is passed, performing a XOT protocol that is perfectly secure against them. If a test is not called, the parties simply perform Protocol 4. We describe this idea more formally as Protocol 5 below.

Qualitatively, Protocol 5 imposes a “tradeoff” for the cheating party between passing the test (if it is called) and the extent to which they deviate from the ideal implementation of Protocol 4. More specifically, if (say) Bob is dishonest, he could cheat perfectly if Alice decides to test, by having both boxes implement the honest behaviour, and he could also cheat perfectly if Alice decides not to test, but to do so he needs to modify Alice's box's behaviour at least (since our device-dependent arguments show that perfect cheating is impossible when Alice's box is honest). Since Alice's box must behave differently in the two scenarios and Bob cannot change how that box behaves once the protocol starts, Alice can constrain his cheating probability by randomly choosing between testing and not testing. A similar argument applies to cheating Alice.

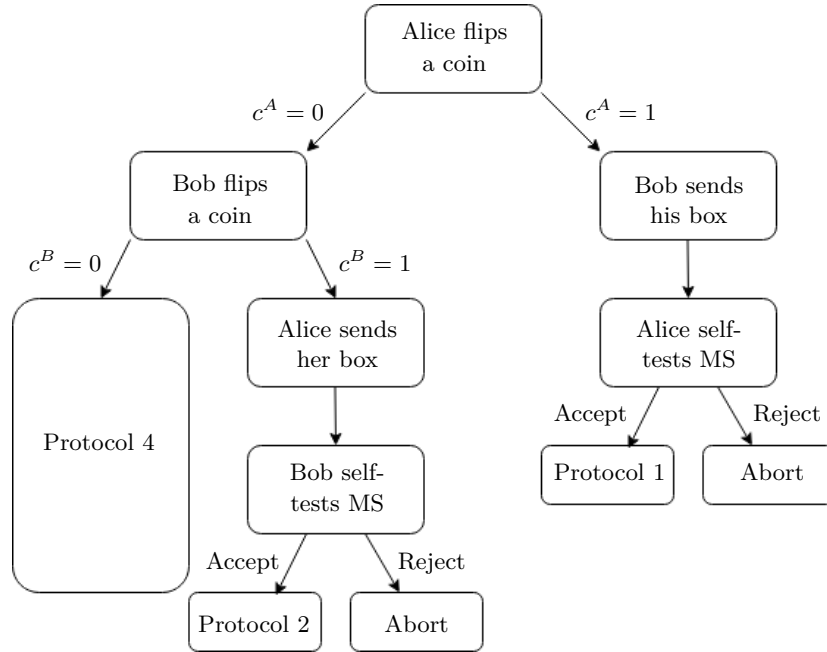
Note that for this reasoning to be valid, it is important that the honest party's box cannot be allowed to detect whether it is being subjected to a magic square test or whether it is being used for Protocol 4 (we are implicitly assuming that the honest party's box behaves the same way in both situations). An assumption of this nature is typically required in device-independent protocols that involve performing a test with some probability, e.g. [1, 3].

² To take a slightly different perspective (used in e.g. [3]), we could suppose that the honest party is able to “shield” their lab in a way such that signals cannot be broadcast out of it once they have supplied the input to their box.

In particular, as observed in [1], we note that if the behaviour of the boxes could be time-dependent, then the honest party must ensure they provide the input to their box at a fixed pre-determined time, regardless of whether the box is being tested or used for Protocol 4.

► **Protocol 5** (XOR oblivious transfer from the magic square game with an extra test step).

1. Alice flips a coin whose outcome is 0 with probability $1 - q^A$, to obtain $c^A \in \{0, 1\}$, which she sends to Bob.
2. a. If $c^A = 0$, Bob flips a coin whose outcome is 0 with probability $1 - q^B$, to obtain $c^B \in \{0, 1\}$, which he sends to Alice.
b. If $c^A = 1$, Bob sends his box to Alice.
3. a. If $c^A = 0, c^B = 0$, Alice and Bob perform Protocol 4 henceforth.
b. If $c^A = 0, c^B = 1$, Alice sends her box to Bob.
c. If $c^A = 1$, Alice receives Bob's box, picks $a^A, b^A \in \{0, 1, 2\}$ uniformly at random to input into her and Bob's boxes, and checks if the outputs x^A, y^A satisfy $x_{b^A}^A = y_{a^A}^A$. If not, she aborts.
4. If $c^A = 0, c^B = 1$, Bob receives Alice's box, picks $a^B, b^B \in \{0, 1, 2\}$ uniformly at random to input into his and Alice's boxes, and checks if the outputs x^B, y^B satisfy $x_{b^B}^B = y_{a^B}^B$. If not, he aborts.
5. a. If $c^A = 1$ and Alice has not aborted, Alice and Bob perform Protocol 1 henceforth.
b. If $c^A = 0, c^B = 1$ and Bob has not aborted, Alice and Bob perform Protocol 2 henceforth.



■ **Figure 3** Flowchart for Protocol 5.

We have required that when either party calls for a test, the tested party must send over their box so that the testing party supplies an input to both boxes themselves, rather than having the tested party self-report an input-output pair for their box. This is to ensure that the inputs to the boxes are indeed uniformly chosen. Also, while it would be convenient if Protocol 4 did not involve Alice sending her input to Bob (thereby more closely resembling a standard nonlocal game), it would seem this step is necessary to allow an honest Bob to know which bit of his output he should use, as previously mentioned regarding Protocol 4.

We give two soundness arguments for Protocol 5. The first consists of explicit numerical bounds on the cheating probabilities, based on the family of SDPs known as the Navascués-Pironio-Acín (NPA) hierarchy [16]. We state the results as Theorem 6 below, and give the proof in the full version. The second is an analytic proof that the cheating probabilities are bounded away from 1, based on the robust self-testing bounds for the magic square game [24, 7]. We state this result formally as Theorem 8 below, and give the proof in the full version.

► **Theorem 6.** *Upper bounds on Alice and Bob’s cheating probabilities for Protocol 5 (the fully device-independent scenario) with $q^A = 0.6$, $q^B = 0.6$ are given below, rounded to 5 decimal places.*

Upper bounds	P_A^{XOT}	P_B^{XOT}
Fully Device-Independent	0.96440	0.99204

► **Remark 7.** The choices for q^A and q^B in Theorem 6 were made by computing the bounds for different choices of q^A and q^B in intervals of 0.1, then simply taking the value that yields the best bounds on the cheating probabilities. We note that the result obtained for P_B^{XOT} is rather close to 1; however, the significant figures shown here are within the tolerance levels of the solver.

► **Theorem 8.** *For any $q^A, q^B > 0$ in Protocol 5, there exists some $\delta > 0$ such that $P_A^{\text{XOT}}, P_B^{\text{XOT}} \leq 1 - \delta$.*

2 Background

In this section, we give the necessary background to prove the results mentioned in the introduction.

2.1 The magic square game

The optimal quantum strategy for magic square can be described as follows. Alice and Bob share the state

$$|\Psi^{\text{MS}}\rangle = \frac{1}{\sqrt{2}}(|00\rangle_{\mathcal{X}_0\mathcal{Y}_0} + |11\rangle_{\mathcal{X}_0\mathcal{Y}_0}) \frac{1}{\sqrt{2}}(|00\rangle_{\mathcal{X}_1\mathcal{Y}_1} + |11\rangle_{\mathcal{X}_1\mathcal{Y}_1}) \quad (7)$$

with Alice holding the registers $\mathcal{X}_0\mathcal{X}_1$, and Bob holding the registers $\mathcal{Y}_0\mathcal{Y}_1$. The measurements of Alice and Bob are given by the following table.

On input a , Alice simultaneously performs the three 2-outcome measurements $\{(\Pi_{ab}^0, \Pi_{ab}^1)\}_b$ in the row indexed by a in Table 1 (it can be checked that the three measurements in every row are compatible, so they can be performed simultaneously) on her registers $\mathcal{X}_0\mathcal{X}_1$. Her output (x_0, x_1, x_2) is the outputs of the three measurements (in order). Similarly, on input b , Bob simultaneously performs the three 2-outcome measurements $\{(\Pi_{ab}^0, \Pi_{ab}^1)\}_a$ in the column indexed by b (the three measurements in every column are also compatible, so they can be performed simultaneously) on his registers $\mathcal{Y}_0\mathcal{Y}_1$, and gives the outcomes of the three measurements as his output (y_0, y_1, y_2) .

Clearly, the measurement Alice performs to output x_b is the same as the measurement Bob performs to output y_a . Since these measurements are performed on maximally entangled states, one can show Alice and Bob always get the same outcome for x_b and y_a . Also, it can

■ **Table 1** Possible measurements for either party in the quantum strategy for magic square.

$a \backslash b$	0	1	2
0	$\Pi_{00}^0 = 0\rangle\langle 0 \otimes \mathbb{1}$ $\Pi_{00}^1 = 1\rangle\langle 1 \otimes \mathbb{1}$	$\Pi_{01}^0 = \mathbb{1} \otimes 0\rangle\langle 0 $ $\Pi_{01}^1 = \mathbb{1} \otimes 1\rangle\langle 1 $	$\Pi_{02}^0 = 0\rangle\langle 0 \otimes 0\rangle\langle 0 $ $\quad + 1\rangle\langle 1 \otimes 1\rangle\langle 1 $ $\Pi_{02}^1 = 0\rangle\langle 0 \otimes 1\rangle\langle 1 $ $\quad + 1\rangle\langle 1 \otimes 0\rangle\langle 0 $
1	$\Pi_{10}^0 = \mathbb{1} \otimes +\rangle\langle + $ $\Pi_{10}^1 = \mathbb{1} \otimes -\rangle\langle - $	$\Pi_{11}^0 = +\rangle\langle + \otimes \mathbb{1}$ $\Pi_{11}^1 = -\rangle\langle - \otimes \mathbb{1}$	$\Pi_{12}^0 = +\rangle\langle + \otimes +\rangle\langle + $ $\quad + -\rangle\langle - \otimes -\rangle\langle - $ $\Pi_{12}^1 = +\rangle\langle + \otimes -\rangle\langle - $ $\quad + -\rangle\langle - \otimes +\rangle\langle + $
2	$\Pi_{20}^0 = 1\rangle\langle 1 \otimes +\rangle\langle + $ $\quad + 0\rangle\langle 0 \otimes -\rangle\langle - $ $\Pi_{20}^1 = 0\rangle\langle 0 \otimes +\rangle\langle + $ $\quad + 1\rangle\langle 1 \otimes -\rangle\langle - $	$\Pi_{21}^0 = +\rangle\langle + \otimes 1\rangle\langle 1 $ $\quad + -\rangle\langle - \otimes 0\rangle\langle 0 $ $\Pi_{21}^1 = +\rangle\langle + \otimes 0\rangle\langle 0 $ $\quad + -\rangle\langle - \otimes 1\rangle\langle 1 $	$\Pi_{22}^0 = + i\rangle\langle + i \otimes + i\rangle\langle + i $ $\quad + - i\rangle\langle - i \otimes - i\rangle\langle - i $ $\Pi_{22}^1 = + i\rangle\langle + i \otimes - i\rangle\langle - i $ $\quad + - i\rangle\langle - i \otimes + i\rangle\langle + i $

be verified that these measurements always produce outcomes satisfying the parity conditions $x_0 \oplus x_1 \oplus x_2 = 0$ and $y_0 \oplus y_1 \oplus y_2 = 1$ (this holds regardless of the state). Some tedious calculation shows that in fact the output distribution is uniform over all combinations that win the magic square game, i.e. $\Pr(xy|ab) = 1/8$ if $x_b = y_a$ (and x, y satisfy the parity conditions), and $\Pr(xy|ab) = 0$ otherwise.

The above description views Alice and Bob as performing 8-outcome measurements (via a sequence of three 2-outcome measurements). However, since for *any* state the measurements always produce outputs satisfying the parity conditions, we can equivalently suppose Alice and Bob measure to determine only (x_0, x_1) and (y_0, y_1) , with the last bit for each determined by the parity conditions. (This is consistent with the way we defined the magic square game earlier.) These are 4-outcome measurements that can be expressed in terms of the Π operators from Table 1 as

$$M_{x_0 x_1 | a}^{\text{MS}} = \Pi_{a0}^{x_0} \Pi_{a1}^{x_1} \quad N_{y_0 y_1 | b}^{\text{MS}} = \Pi_{0b}^{y_0} \Pi_{1b}^{y_1}. \quad (8)$$

It can be checked that each $M_{x_0 x_1 | a}^{\text{MS}}$ and $N_{y_0 y_1 | b}^{\text{MS}}$ is a rank-1 projector. Since the measurements for x_0 and x_1 (resp. y_0 and y_1) commute for every a (resp. b), the product of the Π operators in each case can be taken in either order.

Certain nonlocal games exhibit the property that the quantum strategies achieving their optimal values are essentially unique. That is, if a quantum strategy achieves within ε of the optimal value of the game, that strategy must be $\delta(\varepsilon)$ -close to the ideal strategy for the game, up to certain local operations. This property of rigidity or self-testing was shown first for the CHSH game [13, 14] and has been shown for other nonlocal games since.

[24] originally gave a proof of the rigidity of a version of the magic square game which is slightly different from ours. [7] showed that the rigidity statement also holds for the version of the magic square game we use. However, both of these results show the self-testing of some operators that are related to Alice and Bob's measurement operators in the magic square game, but not the measurement operators themselves. It is not immediately clear

how to self-test the measurement operators themselves from their results. In the full version we derive the following lemma for self-testing of the measurement operators of the magic square strategy.

► **Lemma 9.** *Consider any state $|\rho\rangle$ on registers $\mathcal{X}\mathcal{Y}$ and projective measurements $M_{x|a}, N_{y|b}$ such that $M_{x|a}$ act only on \mathcal{X} and $N_{y|b}$ act only on \mathcal{Y} . If this state and measurements win the magic square game with probability $1 - \varepsilon$, then there exist local isometries $V_A : \mathcal{X} \rightarrow \mathcal{X}_0\mathcal{X}_1\mathcal{J}_A$ and $V_B : \mathcal{Y} \rightarrow \mathcal{Y}_0\mathcal{Y}_1\mathcal{J}_B$ and a state $|\text{junk}\rangle$ on $\mathcal{J}_A\mathcal{J}_B$ such that for all a, b, x, y ,*

$$\begin{aligned} \|(V_A \otimes V_B)|\rho\rangle - |\Psi^{\text{MS}}\rangle \otimes |\text{junk}\rangle\|_2 &\leq O(\varepsilon^{1/4}), \\ \|(V_A \otimes V_B)(M_{x|a} \otimes \mathbb{1})|\rho\rangle - ((M_{x|a}^{\text{MS}} \otimes \mathbb{1})|\Psi^{\text{MS}}\rangle) \otimes |\text{junk}\rangle\|_2 &\leq O(\varepsilon^{1/4}), \\ \|(V_A \otimes V_B)(\mathbb{1} \otimes N_{y|b})|\rho\rangle - ((\mathbb{1} \otimes N_{y|b}^{\text{MS}})|\Psi^{\text{MS}}\rangle) \otimes |\text{junk}\rangle\|_2 &\leq O(\varepsilon^{1/4}), \end{aligned}$$

where $|\Psi^{\text{MS}}\rangle$, $M_{x|a}^{\text{MS}}$, $N_{y|b}^{\text{MS}}$ denote the ideal state and measurements in the magic square game.

2.2 Semidefinite programming

A semidefinite program is an optimization problem of the form

$$p^* = \sup\{\langle C, X \rangle : \Phi(X) = B, X \succeq 0\} \quad (9)$$

where Φ is a linear transformation, C and B are Hermitian. When we write $X \succeq Y$, it means that $X - Y$ is (Hermitian) positive semidefinite, noting the special case that $X \succeq 0$ simply means X is positive semidefinite. We use $X \succ Y$ to mean that $X - Y$ is positive definite.

We can define the *dual* of the above SDP as the optimization problem below

$$d^* = \inf\{\langle B, Y \rangle : \Phi^*(Y) = C + S, S \succeq 0, Y \text{ is Hermitian}\} \quad (10)$$

where we use the notation Φ^* to mean the adjoint of the linear operator Φ .

When we deal with an SDP and its dual, we refer to the original SDP as the *primal* SDP. The primal is called *feasible* if the constraints are satisfiable, that is, if

$$\Phi(X) = B \quad \text{and} \quad X \succeq 0 \quad (11)$$

has a solution. Similarly, if

$$\Phi^*(Y) = C + S, \quad S \succeq 0, \quad \text{and} \quad Y \text{ is Hermitian} \quad (12)$$

has a solution, then the dual is said to be feasible.

We can use primal and dual solutions to show *weak duality*, i.e. if X is primal feasible and (Y, S) is dual feasible, then

$$\langle C, X \rangle \leq \langle B, Y \rangle. \quad (13)$$

In particular, $p^* \leq d^*$. Under mild conditions, we can guarantee equality in Eq. (13). For example, if there exists $X \succeq 0$ which is primal feasible and (Y, S) which is dual feasible, then one can show that $p^* = d^*$. Alternatively, if there exists (Y, S) with $S \succ 0$ which is dual feasible and X which is primal feasible, then we also have $p^* = d^*$. Either of these conditions is known as *strong duality* and the feasible solution with the positive definite solution is known as a *Slater point*. We refer the reader to the book [4] for a proof of weak and strong duality and for other useful information on the subject.

3 A sample proof

In this section, we prove that when the devices are trusted in Protocol 3, then $P_A^{\text{XOT}} = 1/2$ and $P_B^{\text{XOT}} = 3/4$. Before continuing, recall that Protocol 3 is an adaptation of the protocol in [6] where Alice's actions are the exact same and so are the intentions of a dishonest Bob. Therefore, we can also import $P_B^{\text{XOT}} = 3/4$ directly from the security analysis of that protocol.

The rest of this section is devoted to proving $P_A^{\text{XOT}} = 1/2$. Since Bob never aborts, all we need to ascertain is the ability for Alice to learn b from her information contained in the first message. To do this, she must infer b from the ensemble

$$\left\{ \left(\frac{1}{3}, \rho_b = \text{Tr}_{\mathcal{Y}}(|\psi_b^+\rangle\langle\psi_b^+|) \right) : b \in \{0, 1, 2\} \right\}. \quad (14)$$

This is known as the *quantum state discrimination problem*, and the optimal guessing probability can be written as the following SDP:

Primal problem	Dual problem
$\sup \quad \frac{1}{3} \sum_{b=1}^3 \text{Tr}(E_b \rho_b)$	$\inf \quad \text{Tr}(\sigma)$
$\text{subject to:} \quad \sum_{b=0,1,2} E_b = \mathbb{1}$	$\text{subject to:} \quad \forall b \quad \frac{1}{3} \rho_b \preceq \sigma$
$\forall b \quad E_b \in \text{Pos}(\mathcal{X})$	$\sigma \in \text{Herm}(\mathcal{X}).$

Note that the success probability can easily be seen to be equal to the value of the primal problem as it is a maximization over POVMs and the objective function is the success probability of that POVM measurement.

Now, if Alice uses the POVM

$$\{E_0, E_1, E_2\} = \{|0\rangle\langle 0| + |2\rangle\langle 2|, |1\rangle\langle 1|, 0\}, \quad (15)$$

we can see that

$$P_A^{\text{XOT}} \geq \frac{1}{3} \sum_{b=1}^3 \langle E_b, \rho_b \rangle = \frac{1}{3} + \frac{1}{3} \cdot \frac{1}{2} + 0 = \frac{1}{2}. \quad (16)$$

Effectively what this measurement does is measure in the computational basis, then assign the outcomes $|0\rangle$ and $|2\rangle$ to the guess $b = 0$ and the outcome $|1\rangle$ to the guess $b = 1$. Note that $b = 2$ is never guessed in this strategy. All that remains to show is that $P_A^{\text{XOT}} \leq 1/2$. For this, we use the dual problem. Consider the dual feasible solution

$$\sigma = \frac{1}{6} \mathbb{1}_{\mathcal{X}}. \quad (17)$$

It can be checked that σ satisfies the dual constraints, i.e., $\sigma \succeq \frac{1}{3} \rho_b$ for all $b \in \{0, 1, 2\}$. Since $\text{Tr}(\sigma) = 1/2$, we have that $P_A^{\text{XOT}} \leq 1/2$ by weak duality.

Computational platform

Computations were performed using the MATLAB packages QETLAB [9] and YALMIP [12] with solver MOSEK [15]. Some of the calculations reported here were performed using the Euler cluster at ETH Zürich.

References

- 1 Nati Aharon, André Chailloux, Iordanis Kerenidis, Serge Massar, Stefano Pironio, and Jonathan Silman. Weak coin flipping in a device-independent setting. In *Theory of Quantum Computation, Communication, and Cryptography*, pages 1–12, Berlin, Heidelberg, 2014. doi:10.1007/978-3-642-54429-3_1.
- 2 Nati Aharon, Serge Massar, Stefano Pironio, and Jonathan Silman. Device-independent bit commitment based on the CHSH inequality. *New Journal of Physics*, 18(2):025014, 2016. doi:10.1088/1367-2630/18/2/025014.
- 3 Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. Practical device-independent quantum cryptography via entropy accumulation. *Nature Communications*, 9(1):459, 2018. doi:10.1038/s41467-017-02307-4.
- 4 Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, New York, NY, USA, 2004.
- 5 André Chailloux, Gus Gutoski, and Jamie Sikora. Optimal bounds for semi-honest quantum oblivious transfer. *Chicago Journal of Theoretical Computer Science*, 2016(13), 2016. URL: <http://cjtcs.cs.uchicago.edu/articles/2016/13/contents.html>.
- 6 André Chailloux, Iordanis Kerenidis, and Jamie Sikora. Lower bounds for quantum oblivious transfer. *Quantum Information & Computation*, 13(1-2):158–177, 2013. URL: <http://dl.acm.org/citation.cfm?id=2481591.2481600>.
- 7 Matthew Coudron and Anand Natarajan. The parallel-repeated Magic Square game is rigid, 2016. arXiv:1609.06306.
- 8 Honghao Fu and Carl A. Miller. Local randomness: Examples and application. *Physical Review A*, 97:032324, 2018. doi:10.1103/PhysRevA.97.032324.
- 9 Nathaniel Johnston. QETLAB: A MATLAB toolbox for quantum entanglement, version 0.9. <http://qetlab.com>, 2016. doi:10.5281/zenodo.44637.
- 10 Joe Kilian. Founding cryptography on oblivious transfer. In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, STOC ’88, pages 20—31, New York, NY, USA, 1988. doi:10.1145/62212.62215.
- 11 Hoi-Kwong Lo. Insecurity of quantum secure computations. *Physical Review A*, 56:1154–1162, 1997. doi:10.1103/PhysRevA.56.1154.
- 12 Johan Löfberg. YALMIP : A toolbox for modeling and optimization in MATLAB. In *Proceedings of the CACSD Conference*, Taipei, Taiwan, 2004.
- 13 Dominic Mayers and Andrew Yao. Quantum cryptography with imperfect apparatus. In *Proceedings of the 39th Annual Symposium on Foundations of Computer Science*, FOCS ’98, page 503, USA, 1998. URL: <https://dl.acm.org/doi/10.5555/795664.796390>.
- 14 Matthew McKague, Tzyh Haur Yang, and Valerio Scarani. Robust self-testing of the singlet. *Journal of Physics A: Mathematical and Theoretical*, 45(45):455304, 2012. doi:10.1088/1751-8113/45/45/455304.
- 15 MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.1.*, 2019.
- 16 Miguel Navascués, Stefano Pironio, and Antonio Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7):073013, 2008. URL: <http://stacks.iop.org/1367-2630/10/i=7/a=073013>.
- 17 Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. Device-independent quantum key distribution secure against collective attacks. *New Journal of Physics*, 11(4):045021, 2009. doi:10.1088/1367-2630/11/4/045021.
- 18 Christian Schaffner. Cryptography in the bounded-quantum-storage model, 2007. arXiv:0709.0289.
- 19 Jamie Sikora, André Chailloux, and Iordanis Kerenidis. Strong connections between quantum encodings, nonlocality, and quantum cryptography. *Physical Review A*, 89:022334, 2014. doi:10.1103/PhysRevA.89.022334.

- 20 Jonathan Silman, André Chailloux, Nati Aharon, Iordanis Kerenidis, Stefano Pironio, and Serge Massar. Fully distrustful quantum bit commitment and coin flipping. *Physical Review Letters*, 106:220501, 2011. doi:10.1103/PhysRevLett.106.220501.
- 21 Vilasini Venkatesh, Christopher Portmann, and Lídia del Rio. Composable security in relativistic quantum cryptography. *New Journal of Physics*, 21(4):043057, 2019. doi:10.1088/1367-2630/ab0e3b.
- 22 Stephanie Wehner, Christian Schaffner, and Barbara M. Terhal. Cryptography from noisy storage. *Phys. Rev. Lett.*, 100:220502, 2008. doi:10.1103/PhysRevLett.100.220502.
- 23 Stephanie Wehner and Jürg Wullschleger. Composable security in the bounded-quantum-storage model. In *Automata, Languages and Programming*, pages 604–615, 2008. doi:10.1007/978-3-540-70583-3_49.
- 24 Xingyao Wu, Jean-Daniel Bancal, Matthew McKague, and Valerio Scarani. Device-independent parallel self-testing of two singlets. *Physical Review A*, 93:062121, 2016. doi:10.1103/PhysRevA.93.062121.

Note of the Publisher

Unfortunately, this article was accidentally skipped in the first version of the conference proceedings published on June 8, 2020 and was subsequently published on August 19, 2020.

Dagstuhl Publishing – August 19, 2020.