11th Conference on the Theory of Quantum Computation, Communication and Cryptography

TQC 2016, September 27–29, 2016, Berlin, Germany

Edited by

Anne Broadbent

Editor

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

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

ISBN 978-3-95977-019-4

Published online and open access by
Schloss Dagstuhl – Leibniz-Zentrum für Informatik GmbH, Dagstuhl Publishing, Saarbrücken/Wadern, Germany. Online available at http://www.dagstuhl.de/dagpub/978-3-95977-019-4.

Publication date
September, 2016

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

Digital Object Identifier: 10.4230/LIPIcs.TQC.2016.0

LIPIcs – Leibniz International Proceedings in Informatics

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

ISSN 1868-8969

http://www.dagstuhl.de/lipics

Contents

Conference Track Papers

11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016).
Editor: Anne Broadbent

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

Preface

The 11th Conference on the Theory of Quantum Computation, Communication and Cryptography was organized by the Freie Universität Berlin from the 27th to the 29th of September 2016. 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 2015, Université libre de Bruxelles, Belgium
- TQC 2014, National University of Singapore, Singapore
- TQC 2013, University of Guelph, Canada
- TQC 2012, The University of Tokyo, Japan
- TQC 2011, Universidad Complutense de Madrid, Spain
- TQC 2010, University of Leeds, UK
- TQC 2009, Institute for Quantum Computing, University of Waterloo, Canada
- TQC 2008, University of Tokyo, Japan
- TQC 2007, Nara Institute of Science and Technology, Nara, Japan
- TQC 2006, NTT R&D Center, Atsugi, Kanagawa, Japan

The conference consisted of invited talks, contributed talks and a poster session. This year, contributed talks were solicited for two tracks: Conference Track (talk + proceedings) and Workshop Track (talk only). The accepted submissions to the Conference Track appear in these Proceedings, while the accepted Workshop Track submissions are only listed here. Accepted submissions for both tracks are listed in their order of submission.

The invited talks were given by Andris Ambainis (University of Latvia), Ronald Hanson (TU Delft), Lidia del Rio (University of Bristol), Andreas Winter (Universitat Autònoma de Barcelona).

The conference was possible thanks to generous donations from Microsoft, Raytheon BBN Technologies, Institute for Quantum Computing, CryptoWorks21, as well as Journal of Physics A and Quantum Science and Technology. I am deeply indebted to the members of the Program Committee and all subreviewers for their precious contribution in reviewing the submissions. I also wish to thank the members of the Local Organizing Committee for their considerable efforts in organizing the conference. I would like to thank Marc Herbstritt and Michael Wagner (Dagstuhl Publishing) for their technical help, as well as Saeid Molladavoudi for his precious help in putting together the proceedings. Finally, I would like to thank the members of the Steering Committee for offering me this opportunity and for their support. And, of course, a big thank you to all contributors and participants!

August 2016 Anne Broadbent

11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016).
Editor: Anne Broadbent

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

List of Contributed Talks

Michal Oszmaniec, Remigiusz Augusiak, Christian Gogolin, Janek Kolodynski, Antonio Acin and Maciej Lewenstein.
Random bosonic states for robust quantum metrology

Mario Berta, Omar Fawzi and Marco Tomamichel.
On Variational Expressions for Quantum Relative Entropies

Mark Wilde, Marco Tomamichel and Mario Berta.
Strong converse rates for private communication over quantum channels

Giacomo De Palma, Dario Trevisan and Vittorio Giovannetti.
Gaussian States Minimize the Output Entropy of the One-Mode Quantum Attenuator

Stacey Jeffery and Shelby Kimmel.
NAND-Trees and Graph Connectivity in Quantum Algorithms

Christian Majenz, Mario Berta, Frédéric Dupuis, Renato Renner and Matthias Christandl.
Catalytic Decoupling of Quantum Information

Tom Cooney, Christoph Hirche, Ciara Morgan, Jonathan Olson, Kaushik Seshadreesan, John Watrous and Mark Wilde.
Operational meaning of quantum measures of recovery

Stacey Jeffery and Francois Le Gall.
Quantum Communication Complexity of Distributed Set Joins

Marco Piani, Marco Cianciaruso, Thomas Bromley, Carmine Napoli, Nathaniel Johnston and Gerardo Adesso.
Robustness of asymmetry and coherence of quantum states

Shelby Kimmel, Cedric Yen-Yu Lin, Guang Hao Low, Maris Ozols and Theodore Yoder.
Hamiltonian Simulation with Optimal Sample Complexity

Rodrigo Gallego, Jens Eisert and Henrik Wilming.
Defining work from a resource-theoretic perspective

Mohammad Bavarian, Thomas Vidick and Henry Yuen.
Parallel Repetition via Fortification: Analytic View and the Quantum Case

Iagoba Apellaniz, Matthias Kleinmann, Otfried Gühne and Geza Toth.
Witnessing metrologically useful entanglement

Alex Bocharov, Shawn Cui, Martin Roetteler and Krysta Svore.
Computing with Qutrits: Comparative Analysis of Two Ternary Architectures

Patrick Hayden, Sepehr Nezami, Xiao-Liang Qi, Nathaniel Thomas, Michael Walter and Zhao Yang.
Holographic duality from random tensor networks

Cecilia Lancien, Sara Di Martino, Marcus Huber, Marco Piani, Gerardo Adesso and Andreas Winter.
Should Entanglement Measures be Monogamous or Faithful?

Juan Bermejo-Vega, Nicolas Delfosse, Dan E. Browne, Cihan Okay and Robert Raussendorf.
Contextuality as a resource for qubit quantum computation

11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016).
Editor: Anne Broadbent
Leibniz International Proceedings in Informatics
LIPICS Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

Conference Organization

Local Organizing Committee

Jens Eisert – chair
Oliver Buerschaper – co-chair
Juan Bermejo-Vega
Dominik Hangleiter
Albert Werner
Carolin Wille
and the entire QMIO group at the FU Berlin

Program Committee

Gorjan Alagic, University of Copenhagen
Gilles Brassard, Université de Montréal
Anne Broadbent, University of Ottawa – chair
André Chailloux, INRIA Paris Rocquencourt
Giulio Chiribella, University of Hong Kong
Frédéric Dupuis, Masaryk University
Joseph Fitzsimons, Singapore University of Technology and Design
Steve Flammia, University of Sydney
Sevag Gharibian, Virginia Commonwealth University
Stacey Jeffery, California Institute of Technology
Elham Kashefi, University of Edinburgh
Iordanis Kerenidis, LIAFA
Xiongfeng Ma, Tsinghua University
Laura Mančinska, University of Bristol
Carl Miller, University of Michigan, Ann Arbor
Mio Murao, University of Tokyo
Marco Piani, University of Strathclyde
Christopher Portmann, ETH Zurich
Robert Raussendorf, University of British Columbia
Christian Schaffner, CWI Amsterdam
Norbert Schuch, Max-Planck Institute of Quantum Optics
Peter Selinger, Dalhousie University
Jamie Sikora, Centre for Quantum Technologies
Barbara Terhal, RTWH Aachen
Mark Wilde, Louisiana State University

Steering Committee

Wim van Dam, University of California, Santa Barbara, USA
Yasuhito Kawano, NTT, Japan
Michele Mosca, IQC and University of Waterloo, Canada
Martin Roetteler, Microsoft Research, USA
Simone Severini, University College London, UK
Vlatko Vedral, University of Oxford, UK & National University of Singapore, Singapore

11th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2016).
Editor: Anne Broadbent

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

On the Power of Quantum Fourier Sampling

Bill Fefferman[*1] **and Christopher Umans**[†2]

1 **Joint Center for Quantum Information and Computer Science, University of Maryland, College Park, MD, USA**
wjf@umd.edu

2 **California Institute of Technology, Pasadena, CA, USA**
umans@cms.caltech.edu

Abstract

A line of work initiated by Terhal and DiVincenzo [19] and Bremner, Jozsa, and Shepherd [6], shows that restricted classes of quantum computation can efficiently sample from probability distributions that cannot be exactly sampled efficiently on a classical computer, unless the **PH** collapses. Aaronson and Arkhipov [3] take this further by considering a distribution that can be sampled efficiently by linear optical quantum computation, that under two feasible conjectures, cannot even be approximately sampled within bounded total variation distance, unless the **PH** collapses.

In this work we use Quantum Fourier Sampling to construct a class of distributions that can be sampled exactly by a quantum computer. We then argue that these distributions cannot be approximately sampled classically, unless the **PH** collapses, under variants of the Aaronson-Arkhipov conjectures.

In particular, we show a general class of quantumly sampleable distributions each of which is based on an "Efficiently Specifiable" polynomial, for which a classical approximate sampler implies an average-case approximation. This class of polynomials contains the Permanent but also includes, for example, the Hamiltonian Cycle polynomial, as well as many other familiar #**P**-hard polynomials.

Since our distribution likely requires the full power of universal quantum computation, while the Aaronson-Arkhipov distribution uses only linear optical quantum computation with noninteracting bosons, why is our result interesting? We can think of at least three reasons:

1. Since the conjectures required in [3] have not yet been proven, it seems worthwhile to weaken them as much as possible. We do this in two ways, by weakening both conjectures to apply to any "Efficiently Specifiable" polynomial, and by weakening the so-called Anti-Concentration conjecture so that it need only hold for one distribution in a broad class of distributions.
2. Our construction can be understood without any knowledge of linear optics. While this may be a disadvantage for experimentalists, in our opinion it results in a very clean and simple exposition that may be more immediately accessible to computer scientists.
3. It is extremely common for quantum computations to employ "Quantum Fourier Sampling" in the following way: first apply a classically efficient function to a uniform superposition of inputs, then apply a Quantum Fourier Transform followed by a measurement. Our distributions are obtained in exactly this way, where the classically efficient function is related to a (presumed) hard polynomial. Establishing rigorously a robust sense in which the central primitive of Quantum Fourier Sampling is classically hard seems a worthwhile goal in itself.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes

Keywords and phrases Quantum Complexity Theory, Sampling Complexity

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.1

[*] BF was supported by NSF CCF-1423544, BSF grant 2010120 and the Department of Defense.
[†] CU was supported by NSF CCF-1423544 and BSF grant 2010120.

1 Introduction

It is a major goal of computational complexity theory to establish "quantum superiority", obtaining provable settings in which quantum algorithms attain speedups over classical algorithms. Despite the importance of this endeavor, the best evidence that quantum computers can efficiently solve *decision* problems outside **NP** comes from oracle results, see, e.g., [1, 11, 10]. A line of work initiated by DiVincenzo and Terhal [19] and Bremner, Jozsa and Shepherd [6] asks whether we can provide a theoretical basis for quantum superiority by studying *distribution sampling problems*. Since then, there have been many other *exact* sampling results, giving examples of distributions with quantum samplers, which cannot be sampled *exactly* by classical randomized algorithms, see, e.g., [9, 12, 15]. These hardness results are restrictive in that they do not hold in the *approximate* setting, whereby the classical algorithm is allowed to sample from any distribution close in total variation distance to the idealized quantum distribution.

Aaronson and Arkhipov took this a step further, by giving a distribution that can be sampled efficiently by a restrictive form of quantum computation, that assuming the validity of two feasible conjectures, cannot be approximately sampled classically[1], unless the **PH** collapses [3]. The equivalent result for decision problems, establishing $\mathbf{BQP} \not\subset \mathbf{BPP}$ unless the **PH** collapses, would be a crowning achievement in quantum complexity theory. In addition, this research has been very popular with experimentalists who hope to perform this task, "Boson Sampling", in their labs. Experimentally, it seems more relevent to analyze the hardness of approximate quantum sampling, since it is unreasonable to expect that any physical realization of a quantum computer can *itself* exactly sample from its idealized distribution.

In addition to experimental motivation, it is also known that if we can find such a quantumly sampleable distribution for which no classical approximate sampler exists, there exists a "search" problem that can be solved by a quantum computer that cannot be solved classically [2]. In a search problem we are given an input $x \in \{0,1\}^n$, and our goal is to output an element in a nonempty set, $A_x \subseteq \{0,1\}^{\mathrm{poly}(n)}$ with high probability. Establishing this separation, which is not known to follow from exact sampling hardness results, would certainly be one of the strongest pieces of evidence to date that quantum computers can outperform their classical counterparts.

In this work we use the same general algorithmic framework used in many quantum algorithms, which we refer to as "Quantum Fourier Sampling", to demonstrate the existence of a general class of distributions that can be sampled exactly by a quantum computer. We then argue that these distributions cannot be approximately sampled classically, unless the **PH** collapses. Perhaps surprisingly, we obtain and generalize many of the same conclusions as Aaronson and Arkhipov [3] with a completely different class of distributions.

Additionally, concurrently, and independent of us, an exciting result by Bremner, Montanaro and Shepherd [7] obtains similar quantum "approximate sampling" results under related but different conjectures. While our construction has the advantage of a broader class of hardness conjectures, their distribution can be sampled by a class of commuting quantum computations known as Instantaneous Quantum Polynomial time, or **IQP**. This is an advantage of their result, since our quantum sampler likely requires the full power of universal quantum computation.

[1] Indeed, this argument and ours hold even if the classical sampler is a randomized algorithm with access to a **PH** oracle. Therefore it can be interpreted as further evidence that quantum computers can solve problems outside the **PH**.

2 Overview

Our goal is to find a class of distributions that can be sampled efficiently on a quantum computer that cannot be approximately sampled classically. A natural methodology toward showing this is to prove that the existence of a classical approximate sampler implies that a #**P**-hard function can be computed in the **PH**. By Toda's Theorem [20], this would imply a collapse of the **PH**.

In this work, we demonstrate a class of distributions that can, at least in principle, be sampled exactly on a quantum computer. We prove that the existence of an approximate sampler for these distributions implies the existence of a procedure that approximates an "Efficiently Specifiable" polynomial on average. Informally, an Efficiently Specifiable polynomial is a sum of multilinear monomials in which the variables in each monomial can be computed efficiently from the index of the monomial. This includes, among others, the Permanent and Hamiltonian Cycle polynomial.

Computing a multiplicative approximation to the Permanent (or the square of Permanent) with integer entries in the worst-case is #**P**-hard, and computing the Permanent on average is #**P**-hard (see e.g., [3] for more details). The challenge to proving our conjectures is to put these together to prove that an average-case multiplicative approximation to the Permanent (or for that matter, any Efficiently Specifiable polynomial) is still a #**P**-hard problem. Since we can't prove these conjectures, and we don't know the ingredients such a proof will require, it seems worthwhile to attempt to generalize the class of distributions that can be sampled quantumly.

The conjectures we need to prove hardness of approximate sampling are weakened analogues of the conjectures in Aaronson and Arkhipov's results [3]. They conjecture that an *additive approximate average-case solution* to the Permanent with respect to the Gaussian distribution with mean 0 and variance 1 is #**P**-hard. They further propose an "Anti-concentration" conjecture which allows them to reduce the hardness of *multiplicative approximate average-case solutions* to the Permanent over the Gaussian distribution to the hardness of *additive average case solutions* to the Permanent over the Gaussian distribution. The parameters of our conjectures match the parameters of theirs, but our conjectures are broader, so that they need only hold for one such Efficiently Specifiable polynomial, (one of which is the Permanent), and any one of a wider class of distributions.

3 Quantum Preliminaries

In this section we cover a few basic priciples of quantum computing needed to understand the content in the paper. For a complete overview there are many references available, e.g., [13, 16].

We first recall the concept of quantum evaluation of an efficiently classically computable function $f : \{0,1\}^n \to \{0,1\}^m$, which in one quantum query to f maps:

$$\sum_{x\in\{0,1\}^n} |x\rangle|z\rangle \to \sum_{x\in\{0,1\}^n} |x\rangle|z \oplus f(x)\rangle \,.$$

Note that this is a unitary map and can be implemented efficiently as long as f is efficiently computable.

We need the following lemma, which will be useful for our quantum sampler.

▶ **Lemma 1.** *Let $h : [m] \to \{0,1\}^n$ be an efficiently computable one-to-one function, and suppose its inverse can also be efficiently computed. Then the superposition $\frac{1}{\sqrt{m}} \sum_{x\in[m]} |h(x)\rangle$ can be efficiently prepared by a quantum algorithm.*

Proof. Our quantum procedure with two quantum registers proceeds as follows:

1. Prepare $\frac{1}{\sqrt{m}} \sum_{x\in[m]} |x\rangle|00...0\rangle$
2. Query h using the first register as input and the second as output:

$$\frac{1}{\sqrt{m}} \sum_{x\in[m]} |x\rangle|h(x)\rangle$$

3. Query h^{-1} using the second register as input and the first as output:

$$\frac{1}{\sqrt{m}} \sum_{x\in[m]} |x \oplus h^{-1}(h(x))\rangle|h(x)\rangle = \frac{1}{\sqrt{m}} \sum_{x\in[m]} |00...0\rangle|h(x)\rangle$$

4. Discard first register ◀

Finally, we will frequently be dealing with the uniform distribution over $\{\pm 1\}^n$ strings, and a natural generalization:

▶ **Definition 2** ($\mathbb{T}_\ell$)**.** Given $\ell > 0$, we define the set $\mathbb{T}_\ell = \{\omega_\ell^0, \omega_\ell^1 ..., \omega_\ell^{\ell-1}\}$ where ω_ℓ is a primitive ℓ-th root of unity.

We note that $\mathbb{T}_\ell$ is just ℓ evenly spaced points on the unit circle, and $\mathbb{T}_2 = \{\pm 1\}$.

4 Efficiently Specifiable Polynomial Sampling on a Quantum Computer

In this section we describe a general class of distributions that can be sampled efficiently on a Quantum Computer.

▶ **Definition 3** (Efficiently Specifiable Polynomial)**.** We say a multilinear homogenous n-variate polynomial Q with coefficients in $\{0, 1\}$ and m monomials is *Efficiently Specifiable* via an efficiently computable, one-to-one function $h : [m] \to \{0, 1\}^n$, with an efficiently computable inverse, if:

$$Q(X_1, X_2..., X_n) = \sum_{z\in[m]} {X_1}^{h(z)_1} {X_2}^{h(z)_2} ... {X_n}^{h(z)_n}.$$

▶ **Definition 4** ($\mathcal{D}_{Q,\ell}$)**.** Suppose Q is an Efficiently Specifiable polynomial with n variables and m monomials. For fixed Q and ℓ, we define the class of distributions $\mathcal{D}_{Q,\ell}$ over ℓ-ary strings $y \in [0, \ell - 1]^n$ given by:

$$\Pr_{\mathcal{D}_{Q,\ell}}[y] = \frac{|Q(Z_y)|^2}{\ell^n m}$$

where $Z_y \in \mathbb{T}_\ell^n$ is a vector of complex values encoded by the string y.

The encoding works by assigning each value $j \in [0, \ell - 1]$ to ω_ℓ^j. For example, notice that when $\ell = 2$ then $y \in \{0, 1\}^n$ and Z_y is simply the corresponding $\{\pm 1\}^n$ assignment with each entry set to 1 if the corresponding entry in y is 0 and -1 if the corresponding entry in y is 1.

▶ **Theorem 5** (Quantum Sampling Theorem)**.** *Given an Efficiently Specifiable polynomial, Q with n variables, m monomials, relative to a function h, and $\ell \leqslant exp(n)$, the resulting $\mathcal{D}_{Q,\ell}$ can be sampled in $poly(n)$ time on a Quantum Computer.*

Proof.

1. We start in a uniform superposition $\frac{1}{\sqrt{m}} \sum_{z \in [m]} |z\rangle$.
2. We then apply Lemma 1 to prepare $\frac{1}{\sqrt{m}} \sum_{z \in [m]} |h(z)\rangle$.
3. Apply Quantum Fourier Transform over $\mathbb{Z}_\ell^n$ to attain $\frac{1}{\sqrt{\ell^n m}} \sum_{y \in [0,\ell-1]^n} \sum_{z \in [m]} \omega_\ell^{<y,h(z)>} |y\rangle$.

Notice that the amplitude of each y basis state in the final state after Step 3 is proportional to the value of $Q(Z_y)$. A measurement in the computational basis will amount to sampling from the distribution $D_{Q,\ell}$ as desired.

Why is each evaluation appearing in the amplitudes of this quantum state? To see this, let's analyze the simple case of $D_{Q,2}$, in which we claim each amplitude of the state after Step 3 is proportional to Q evaluated at a particular $\{\pm 1\}^n$ assignment. Note that in this case the Quantum Fourier Transform we apply in Step 2 is simply $H^{\otimes n}$, where H is the 2×2 Hadamard matrix.

We can think of the Hadamard transform as having columns indexed by all 2^n multilinear monomials $M_1, M_2, ..., M_{2^n}$ on n variables $x_1, x_2, ..., x_n$, and the 2^n rows of the transform as indexed by all possible $\{\pm 1\}^n$ assignments to the n variables. Then the unnormalized (i, j)-th element of the matrix is $M_j(y_i)$, the evaluation of the j-th monomial on the i-th assignment. To prove this, we first observe that the one qubit Hadamard matrix can be seen in this way, where $M_1 = 1$, the "empty monomial" that always evaluates to 1 irrespective of the assignment, and $M_2 = x_1$. The rows of the transform can be indexed by assignments -1 and $+1$, and the unnormalized matrix entries simply correspond to the evaluations of each monomial on the respective assignment, as mentioned earlier. Further, it is easy to see that the tensor product respects this structure, giving rise to our claimed interpretation.

The state we prepare in Step 2, $\frac{1}{\sqrt{m}} \sum_{z \in [m]} |h(z)\rangle$, is simply the quantum state that is uniformly supported over each of the m monomials in Q, and so after applying the Hadamard transform in Step 3, we obtain a state with amplitudes equal to suitably normalized evaluations of Q at each $\{\pm 1\}^n$ assignment. It is not hard to further generalize this argument to the case of $\mathcal{D}_{Q,\ell}$, in which case we apply a similar interpretation to the Quantum Fourier Transform over $\mathbb{Z}_\ell^n$. ◀

5 Classical Hardness of Efficiently Specifiable Polynomial Sampling

We are interested in demonstrating the existence of some distribution that can be sampled exactly by a uniform family of quantum circuits, that cannot be sampled approximately classically. Approximate here means close in Total Variation distance, where we denote the Total Variation distance between two distributions X and Y by $\|X - Y\|$. Thus we define the notion of a Sampler to be a classical randomized algorithm that approximately samples from a given class of distributions:

▶ **Definition 6** (Sampler). Let $\{D_n\}_{n>0}$ be a class of distributions where each D_n is distributed over $\mathbb{C}^n$. Let $r(n) \in \text{poly}(n), \epsilon(n) \in 1/poly(n)$. We say S is a *Sampler* with respect to $\{D_n\}$ if $\|S(0^n, x \sim U_{\{0,1\}^{r(n)}}, 0^{1/\epsilon(n)}) - D_n\| \leqslant \epsilon(n)$ in (classical) polynomial time.

We first recall a theorem due to Stockmeyer [17] on the ability to "approximate count" in the **PH**.

▶ **Theorem 7** (Stockmeyer). *Given as input an efficiently computable function $f : \{0,1\}^n \to \{0,1\}^m$ and $y \in \{0,1\}^m$, there is a procedure that outputs α such that:*

$$(1-\epsilon) \Pr_{x \sim U_{\{0,1\}^n}}[f(x) = y] \leqslant \alpha \leqslant (1+\epsilon) \Pr_{x \sim U_{\{0,1\}^n}}[f(x) = y] .$$

In randomized time $\mathrm{poly}(n, 1/\epsilon)$ *with access to an* **NP** *oracle.*

In this section we use Theorem 7, together with the assumed existence of a Sampler for $\mathcal{D}_{Q,\ell}$ to obtain hardness consequences.

In particular, we show that a Sampler would imply the existence of an efficient approximation to an Efficiently Specifiable polynomial in the following two contexts:

▶ **Definition 8** ($\epsilon-$additive δ-approximate solution). Given a distribution D over $\mathbb{C}^n$ and $P : \mathbb{C}^n \to \mathbb{C}$ we say $T : \mathbb{C}^n \to \mathbb{C}$ is an $\epsilon-$additive approximate $\delta-$average case solution with respect to D, to P, if $\Pr_{x \sim D}[|T(x) - P(x)| \leqslant \epsilon] \geqslant 1 - \delta$.

▶ **Definition 9** ($\epsilon-$multiplicative δ-approximate solution). Given a distribution D over $\mathbb{C}^n$ and a function $P : \mathbb{C}^n \to \mathbb{C}$ we say $T : \mathbb{C}^n \to \mathbb{C}$ is an $\epsilon-$multiplicative approximate $\delta-$average case solution with respect to D, to P, if $\Pr_{x \sim D}[|T(x) - P(x)| \leqslant \epsilon |P(x)|] \geqslant 1 - \delta$.

These definitions formalize a notion that we will need, in which an efficient algorithm computes a particular hard function approximately only on most inputs, and can act arbitrarily on a small fraction of remaining inputs.

Now we prove our main theorem, which informally states that the existence of a Sampler for $\mathcal{D}_{Q,\ell}$ would imply a solution to Q^2 in the following sense: the solution gives a good additive error approximation to $Q^2(X)$ with probability $1 - \delta$ over the choice of assignments X. That is, on a δ-fraction of assignments the output of the solution may not even be additively-close to the desired value of Q^2.

The proof of this theorem is somewhat technical, but the intuition is very clear. If we have access to a classical randomized algorithm that samples from a distribution close in Total Variation distance to $\mathcal{D}_{Q,\ell}$, we would like to use Stockmeyer's Algorithm (Theorem 7) to get a multiplicative estimate to the probability of a particular outcome of the Sampler. After accounting for normalization, this would amount to a multiplicative estimate to the desired evaluation of the Efficiently Specifiable polynomial. Of course, if the Sampler sampled from exactly the distribution $\mathcal{D}_{Q,\ell}$ we'd be able to do this. Unfortunately though, we only know that the distribution sampled by our Sampler is *close* to the ideal distribution $\mathcal{D}_{Q,\ell}$. Therefore, we can't trust that the probability of any particular outcome of the Sampler is exactly the same as the probability of this outcome according to $\mathcal{D}_{Q,\ell}$. One thing we do know, however, is that *most* of the probabilities of the distribution sampled by the Sampler must be additively close to the probabilities of $\mathcal{D}_{Q,\ell}$, since the two distributions are close in Total Variation distance. This will be enough to guarantee that if we use Stockmeyer's algorithm to estimate the probability of a uniformly chosen outcome, with high probability over choice of assignment, we get a decent additive estimate to the evaluation of the Efficiently Specifiable polynomial. Note that our analysis can be thought of as a simplified version of the analysis in [3].

▶ **Theorem 10** (Complexity consequences of Sampler). *Given an Efficiently Specifiable polynomial Q with n variables and m monomials, and a Sampler S with respect to $\mathcal{D}_{Q,\ell}$, there is a randomized procedure computing an $(\epsilon \cdot m)-$additive approximate $\delta-$average case solution with respect to the uniform distribution over $\mathbb{T}_\ell^n$, to the Q^2 function, in randomized time* $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ *with access to an* **NP** *oracle.*

Proof. We need to give a procedure that outputs an ϵm-additive estimate to the Q^2 function evaluated at a uniform setting of the variables, with probability $1-\delta$ over choice of setting. Setting $\nu = \frac{\epsilon\delta}{16}$, suppose S samples from a distribution $\mathcal{D}'$ such that $\|\mathcal{D}_{Q,\ell} - \mathcal{D}'\| \leqslant \nu$. We let p_y be $\Pr_{\mathcal{D}_{Q,\ell}}[y]$ and q_y be $\Pr_{\mathcal{D}'}[y]$.

Our procedure picks a uniformly chosen encoding of an assignment $y \in [0, \ell-1]^n$, and outputs an estimate $\tilde{q}_y$. Note that $p_y = \frac{|Q(Z_y)|^2}{\ell^n m}$. Thus our goal will be to output a $\tilde{q}_y$ that approximates p_y within additive error $\epsilon\frac{m}{\ell^n m} = \frac{\epsilon}{\ell^n}$, in time polynomial in n, $\frac{1}{\epsilon}$, and $\frac{1}{\delta}$.

We need:

$$\Pr_y[|\tilde{q}_y - p_y| > \frac{\epsilon}{\ell^n}] \leqslant \delta\,.$$

First, define for each y, $\Delta_y = |p_y - q_y|$, which by definition gives us $\|\mathcal{D}_{Q,\ell} - \mathcal{D}'\| = \frac{1}{2}\sum_y[\Delta_y]$. Now:

$$E_y[\Delta_y] = \frac{\sum_y[\Delta_y]}{\ell^n} = \frac{2\nu}{\ell^n}\,.$$

And applying Markov's inequality, $\forall k > 1$,

$$\Pr_y[\Delta_y > \frac{k2\nu}{\ell^n}] < \frac{1}{k}\,.$$

Setting $k = \frac{4}{\delta}$ and recalling that $\nu = \frac{\epsilon\delta}{16}$, we have:

$$\Pr_y[\Delta_y > \frac{\epsilon}{2}\cdot\frac{1}{\ell^n}] < \frac{\delta}{4}\,.$$

Then use approximate counting (with an **NP** oracle), using Theorem 7 on the randomness of S to obtain an output $\tilde{q}_y$ so that, for all $\gamma > 0$, in time polynomial in n and $\frac{1}{\gamma}$:

$$\Pr[|\tilde{q}_y - q_y| > \gamma\cdot q_y] < \frac{1}{2^n}\,.$$

Because we can amplify the failure probability of Stockmeyer's algorithm to be inverse exponential. Now because the q_y's are probabilities that sum to 1:

$$E_y[q_y] = \frac{\sum_y q_y}{\ell^n} = \frac{1}{\ell^n} \Rightarrow \Pr_y[q_y > \frac{k}{\ell^n}] < \frac{1}{k}\,.$$

Now, applying the union bound with γ set to $\frac{\epsilon\delta}{8}$:

$$\begin{aligned}\Pr_y[|\tilde{q}_y - p_y| > \frac{\epsilon}{\ell^n}] &\leqslant \Pr_y[|\tilde{q}_y - q_y| > \frac{\epsilon}{2}\cdot\frac{1}{\ell^n}] + \Pr_y[|q_y - p_y| > \frac{\epsilon}{2}\cdot\frac{1}{\ell^n}]\\ &\leqslant \Pr_y[q_y > \frac{k}{\ell^n}] + \Pr[|\tilde{q}_y - q_y| > \gamma\cdot q_y] + \Pr_y[\Delta_y > \frac{\epsilon}{2}\cdot\frac{1}{\ell^n}]\\ &\leqslant \frac{1}{k} + \frac{1}{2^n} + \frac{\delta}{4} = \frac{\delta}{2} + \frac{1}{2^n} \leqslant \delta.\end{aligned}$$

◀

Now, as will be proven in Appendix A, the variance, $\mathrm{Var}\,[Q(X)]$, of the distribution over $\mathbb{C}$ induced by an Efficiently Specifiable Q with m monomials, evaluated at uniformly distributed entries over $\mathbb{T}_\ell^n$ is m, and so the preceeding Theorem 10 promised us we can achieve an $\epsilon\mathrm{Var}\,[Q(X)]$-additive approximation to Q^2, given a Sampler. We now show that, under a conjecture, this approximation can be used to obtain a good multiplicative estimate to Q^2. This conjecture effectively states that the Chebyshev inequality for this random variable is tight.

▶ **Conjecture 11** (Anti-Concentration Conjecture relative to an n-variate polynomial Q and distribution $\mathcal{D}$ over $\mathbb{C}^n$). *There exists a polynomial p such that for all n and $\delta > 0$,*

$$\Pr_{X\sim\mathcal{D}}\left[|Q(X)|^2 < \frac{\mathrm{Var}\left[Q(X)\right]}{p(n,1/\delta)}\right] < \delta\,.$$

▶ **Theorem 12.** *Assuming Conjecture 11, relative to an Efficiently Specifiable polynomial Q and a distribution $\mathcal{D}$, an $\epsilon\mathrm{Var}\left[Q(X)\right]$-additive approximate δ-average case solution with respect to D, to the Q^2 function can be used to obtain an $\epsilon' \leqslant poly(n)\epsilon$-multiplicative approximate $\delta' = 2\delta$-average case solution with respect to $\mathcal{D}$ to Q^2.*

Proof. Suppose λ is, with high probability, an $\epsilon\mathrm{Var}\left[Q(X)\right]$-additive approximation to $|Q(X)|^2$, as guaranteed in the statement of the Theorem. This means:

$$\Pr_{X\sim\mathcal{D}}\left[\left|\lambda - |Q(X)|^2\right| > \epsilon\mathrm{Var}\left[Q(X)\right]\right] < \delta\,.$$

Now assuming Conjecture 11 with polynomial p, we will show that λ is also a multiplicative estimate to $|Q(X)|^2$ with high probability. By the union bound,

$$\begin{aligned}\Pr_{X\sim\mathcal{D}}\left[\frac{\left|\lambda - |Q(X)|^2\right|}{\epsilon p(n,1/\delta)} > |Q(X)|^2\right] &\leqslant \Pr_{X\sim D}\left[\left|\lambda - |Q(X)|^2\right| > \epsilon\mathrm{Var}\left[Q(X)\right]\right] + \\ &\quad \Pr_{X\sim\mathcal{D}}\left[\frac{\epsilon\mathrm{Var}\left[Q(X)\right]}{\epsilon p(n,1/\delta)} > |Q(X)|^2\right] \\ &\leqslant 2\delta\end{aligned}$$

where the second line comes from Conjecture 11. To attain multiplicative error bounds ϵ' and δ' we can set $\delta = \delta'/2$ and $\epsilon = \epsilon'/p(n,1/\delta)$. ◀

For the results in this section to be meaningful, we simply need the Anti-Concentration conjecture to hold for some Efficiently Specifiable polynomial that is #**P**-hard to compute, relative to any distribution we can sample from (either $U_{\{\pm1\}^n}$, or $\mathcal{B}(0,k)^n$). We note that Aaronson and Arkhipov [3] conjectures the same statement as Conjecture 11 for the special case of the **Permanent** function relative to matrices with entries distributed independently from the complex Gaussian distribution of mean 0 and variance 1.

Additionally, we acknowledge a result of Tao and Vu [18] who show:

▶ **Theorem 13** (Tao & Vu). *For all $\epsilon > 0$ and sufficiently large n,*

$$\Pr_{X\sim U_{\{\pm1\}^{n\times n}}}\left[|\mathbf{Permanent}[X]| < \frac{\sqrt{n!}}{n^{\epsilon n}}\right] < \frac{1}{n^{0.1}}\,.$$

Which comes quite close to our conjecture for the case of the **Permanent** function and uniformly distributed $\{\pm1\}^{n\times n} = \mathbb{T}_2^{n\times n}$ matrix. More specifically, for the above purpose of relating the hardness of additive solutions to the hardness of multiplicative solutions, we would need an upper bound of any inverse polynomial δ, instead of a fixed $n^{-0.1}$.

6 Sampling from Distributions with Probabilities Proportional to $[-k,k]$ Evaluations of Efficiently Specifiable Polynomials

In the prior sections we discussed quantum sampling from distributions in which the probabilities are proportional to evaluations of Efficiently Specifiable polynomials evaluated

at points in $\mathbb{T}_\ell^n$. In this section we show how to generalize this to quantumly sampling from distributions in which the probabilities are proportional to evaluations of Efficiently Specifiable polynomials evaluated at polynomially bounded integer values. In particular, we show a simple way to take an Efficiently Specifiable polynomial with n variables and create another Efficiently Specifiable polynomial with kn variables, in which evaluating this new polynomial at $\{-1,+1\}^{kn}$ is equivalent to evaluation of the old polynomial at $[-k,k]^n$.

▶ **Definition 14** (k-valued equivalent polynomial)**.** For every Efficiently Specifiable polynomial Q with m monomials and every fixed $k > 0$ consider the polynomial $Q'_k : \mathbb{T}_2^{kn} \to \mathbb{R}$ defined by replacing each variable x_i in Q with the sum of k new variables $x_i^{(1)} + x_i^{(2)} + ... + x_i^{(k)}$. We will call Q'_k the k-valued equivalent polynomial with respect to Q.

Note that a uniformly chosen $\{\pm 1\}$ assignment to the variables in Q'_k induces an assignment to the variables in Q, distributed from a distribution we call $\mathcal{B}(0,k)$:

▶ **Definition 15** ($\mathcal{B}(0,k)$)**.** For k a positive integer, we define the distribution $\mathcal{B}(0,k)$ supported over the odd integers in the range $[-k,k]$ (if k is odd), or even integers in the range $[-k,k]$ (if k is even), so that:

$$\Pr_{\mathcal{B}(0,k)}[y] = \begin{cases} \frac{\binom{k}{\frac{k+y}{2}}}{2^k} & \text{if y and k are both odd or both even} \\ 0 & \text{otherwise} \end{cases}$$

▶ **Theorem 16.** *Given an Efficiently Specifiable polynomial Q with n variables and m monomials, let Q'_k be its k-valued equivalent polynomial. For all $\ell < exp(n)$, we can quantumly sample from the distribution $\mathcal{D}_{Q'_k,\ell}$ in time* $\mathrm{poly}(n,k)$.

Proof. Our proof follows from the following lemma, which proves that Q'_k is Efficiently Specifiable.

▶ **Lemma 17.** *Suppose Q is an n-variate, homogeneous degree d Efficiently Specifiable polynomial with m monomials relative to a function $h : [m] \to \{0,1\}^n$. Let $k \leqslant poly(n)$ and let Q'_k be the k-valued equivalent polynomial with respect to Q. Then Q'_k is Efficiently Specifiable with respect to an efficiently computable function $h' : [m] \times [k]^d \to \{0,1\}^{kn}$.*

Proof. We first define and prove that h' is efficiently computable. We note that if there are m monomials in Q, there are mk^d monomials in Q'_k. As above, we'll think of the new variables in Q'_k as indexed by a pair of indices, a "top index" in $[k]$ and a "bottom index" in $[n]$. Equivalently we are labeling each variable in Q'_k as $x_i^{(j)}$, the j-th copy of the i-th variable in Q.

We can think of each monomial in Q'_k (and hence the input to h') as being indexed by a value $r \in [m]$ and $y_1, y_2, ..., y_d \in [k]^d$. We can obtain the variables in any particular monomial of Q'_k by simply using the output of $h(r)$ to obtain the "bottom" indices of the variables, and use the values of $y_1, y_2, ..., y_d$ to obtain the "top" indices for each of the d variables.

We will now show that h'^{-1} is efficiently computable. As before we will think of $z \in \{0,1\}^{kn}$ as being indexed by a pair, a "top index" in $[k]$ and a "bottom index" in $[n]$. Then we compute $h'^{-1}(z)$ by first obtaining from z the bottom indices $j_1, j_2, ..., j_d$ and the corresponding top indices, $i_1, i_2, ..., i_d$. Then obtain from the bottom indices the string $x \in \{0,1\}^n$ corresponding to the variables used in Q and output the concatenation of $h^{-1}(x)$ and $j_1, j_2, ..., j_d$. ◀

Theorem 16 now follows from Lemma 17, where we established that Q'_k is Efficiently Specifiable, and Theorem 5, where we established that we can sample from $\mathcal{D}_{Q'_k,\ell}$ quantumly. ◀

▶ **Theorem 18.** *Let* $\mathrm{Var}\left[Q(X)\right] = \mathrm{Var}\left[Q(X_1, X_2, ..., X_n)\right]$ *denote the variance of the distribution over* $\mathbb{R}$ *induced by* Q *with assignments distributed from* $\mathcal{B}(0,k)^n$*. Given a Sampler* S *with respect to* $\mathcal{D}_{Q'_k,2}$*, we can find a randomized procedure computing an* $\epsilon\mathrm{Var}\left[Q(X)\right]$*-additive approximate* δ*-average case solution to* Q^2 *with respect to* $\mathcal{B}(0,k)^n$ *in time* $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ *with access to an* **NP** *oracle.*

Proof. We begin by noting that Q'_k is a polynomial of degree d that has kn variables and $m' = mk^d$ monomials. By Theorem 10 we get that a Sampler with respect to $\mathcal{D}_{Q'_{k,2}}$ implies there exists A, an $\epsilon m'$-additive approximate δ-average case solution to ${Q'_k}^2$ with respect to $U_{\{\pm 1\}^{kn}}$ that runs in time $poly(n, 1/\epsilon, 1/\delta)$ with access to an **NP** oracle. We need to show the existence of an A', an $\epsilon m'$-additive approximate δ-average case solution to ${Q'_k}^2$ with respect to the $\mathcal{B}(0,k)^n$ distribution.

We think of A' as receiving an input, $z \in [-k,k]^n$ drawn from $\mathcal{B}(0,k)^n$. A' picks y uniformly from the orbit of z over $\{\pm 1\}^{kn}$ and outputs $A(y)$. Now:

$$\Pr_{z\sim\mathcal{B}(0,k)^n}\left[\left|A'(z) - Q^2(z)\right| \leqslant \epsilon m'\right] = \Pr_{z\sim\mathcal{B}(0,k)^n, y\sim_R \mathrm{orbit}(z)}\left[\left|A(y) - Q^2(z)\right| \leqslant \epsilon m'\right] \quad (1)$$

$$= \Pr_{y\sim U_{\{\pm 1\}^{kn}}}\left[|A(y) - Q'_k(y)| \leqslant \epsilon m'\right] \geqslant 1 - \delta \quad (2)$$

$$(3)$$

Thus, because a uniformly chosen $\{\pm 1\}^{kn}$ assignment to the variables in Q'_k induces a $\mathcal{B}(0,k)^n$ distributed assignment to the variables in Q, this amounts to an $\epsilon m'$-additive approximate δ-average case solution to Q^2 with respect to $\mathcal{B}(0,k)^n$. In Appendix A we prove that $\mathrm{Var}\left[Q(X)\right]$ is m' as desired. ◀

7 The "Compressed" QFT

In this section we begin to prove that quantum algorithms can sample efficiently from distributions with probabilities proportional to evaluations of Efficiently Specifiable polynomials at points in $[-k,k]^n$ for $k \in \exp(n)$. Note that in the prior quantum algorithm of Section 4 we would need to invoke the QFT over $\mathbb{Z}_2^{kn}$, of dimension doubly-exponential in n. Thus we need to define a new Polynomial Transform that can be obtained from the standard Quantum Fourier Transform over $\mathbb{Z}_2^n$, which we refer to as the "Compressed QFT". Now we describe the unitary matrix which implements the Compressed QFT.

Consider the $2^k \times 2^k$ matrix D_k, whose columns are indexed by all possible 2^k multilinear monomials of the variables $x_1, x_2, ..., x_k$ and the rows are indexed by the 2^k different $\{-1,+1\}$ assignments to the variables. The (i,j)-th entry is then defined to be the evaluation of the j-th monomial on the i-th assignment. As we noted earlier, defining $\bar{D}_k$ to be the matrix whose entries are the entries in D_k normalized by $1/\sqrt{2^k}$ gives us the Quantum Fourier Transform matrix over $\mathbb{Z}_2^k$. It is clear, by the unitarity of the Quantum Fourier Transform, that the columns (and rows) in D_k are pairwise orthogonal.

Now we define the "Elementary Symmetric Polynomials":

▶ **Definition 19** (Elementary Symmetric Polynomials)**.** We define the j-th Elementary Symmetric Polynomial on k variables for $j \in [0,k]$ to be:

$$p_j(X_1, X_2, ..., X_k) = \sum_{1 \leqslant \ell_1 < \ell_2 < ... < \ell_j \leqslant k} X_{\ell_1} X_{\ell_2} ... X_{\ell_j} .$$

In this work we will care particularly about the first two elementary symmetric polynomials, p_0 and p_1 which are defined as $p_0(X_1, X_2, ..., X_k) = 1$ and $p_1(X_1, X_2, ..., X_k) = \sum_{1 \leqslant \ell \leqslant k} X_\ell$. Consider the $(k+1) \times (k+1)$ matrix, $\tilde{D}_k$, whose columns are indexed by elementary symmetric polynomials on k variables and whose rows are indexed by equivalence classes of assignments in $\mathbb{Z}_2^k$ under S_k symmetry. We obtain $\tilde{D}_k$ from D_k using two steps.

First obtain a $2^k \times (k+1)$ rectangular matrix $\tilde{D}_k^{(1)}$ whose rows are indexed by assignments to the variables $x_1, x_2, ..., x_k \in \{\pm 1\}^k$ and columns are the entry-wise sum of the entries in each column of D_k whose monomial is in each respective elementary symmetric polynomial. Then obtain the final $(k+1) \times (k+1)$ matrix $\tilde{D}_k$ by taking $\tilde{D}_k^{(1)}$ and keeping only one representative row in each equivalence class of assignments under S_k symmetry. We label the equivalence classes of assignments under S_k symmetry $o_0, o_1, o_2, ..., o_k$ and note that for each $i \in [k]$, $|o_i| = \binom{k}{i}$. Observe that $\tilde{D}_k$ is precisely the matrix whose (i,j)-th entry is the evaluation of the j-th symmetric polynomial evaluated on an assignment in the i-th symmetry class.

▶ **Theorem 20.** *The columns in the matrix $\tilde{D}_k^{(1)}$ are pairwise orthogonal.*

Proof. Note that each column in the matrix $\tilde{D}_k^{(1)}$ is the sum of columns in D_k each of which are orthogonal. We can prove this theorem by observing that if we take any two columns in $D_k^{(1)}$, called c_1, c_2, where c_1 is the sum of columns $\{u_i\}$ of D_k and c_2 is the sum of columns $\{v_i\}$ of D_k. The inner product, $\langle c_1, c_2 \rangle$ can be written:

$$\langle \sum_i u_i, \sum_j v_j \rangle = \sum_{i,j} \langle u_i, v_j \rangle = 0 .$$ ◀

▶ **Theorem 21.** *Let L be the $(k+1) \times (k+1)$ diagonal matrix with i-th entry equal to $\sqrt{o_i}$. Then the columns of $L \cdot \tilde{D}_k$ are orthogonal.*

Proof. Note that the value of the symmetric polynomial at each assignment in an equivalence class is the same. We have already concluded the orthogonality of columns in $\tilde{D}_k^{(1)}$. Therefore if we let a and b be any two columns in the matrix $\tilde{D}_k$, and their respective columns be $\bar{a}, \bar{b}$ in $\tilde{D}_k^{(1)}$, we can see:

$$\sum_{i=0}^{k} (a_i b_i |o_i|) = \sum_{i=0}^{2^k} \bar{a}_i \bar{b}_i = 0 .$$

From this we conclude that the columns of the matrix $L \cdot \tilde{D}_k$, in which the i-th row of $\tilde{D}_k$ is multiplied by $\sqrt{o_i}$, are orthogonal. ◀

▶ **Theorem 22.** *We have just established that the columns in the matrix $L \cdot \tilde{D}_k$ are orthogonal. Let the $k+1 \times k+1$ diagonal matrix R be such that so that the columns in $L \cdot \tilde{D}_k \cdot R$ are orthonormal, and thus $L \cdot \tilde{D}_k \cdot R$ is unitary. Then the first two nonzero entries in R, which we call r_0, r_1, corresponding to the normalization of the column pertaining to the zero-th and first elementary symmetric polynomial, are $1/\sqrt{2^k}$ and $\frac{1}{\sqrt{\sum_{i=0}^{k} [\binom{k}{i}(k-2i)^2]}}$.*

Proof. First we calculate r_0. Since we wish for a unitary matrix, we want the ℓ_2 norm of the first column of $\tilde{D}_k$ to be 1, and so need:

$$r_0^2 \sum_{i=0}^{k} \left(\sqrt{o_i}\right)^2 = r_0^2 \sum_{i=0}^{k} \binom{k}{i} = 1\,.$$

And so r_0 is $1/\sqrt{2^k}$ as desired.

Now we calculate r_1, the normalization in the column of $\tilde{D}_k$ corresponding to the first elementary symmetric polynomial. Note that in i-th equivalence class of assignments we have exactly i negative ones and $k-i$ positive ones. Thus the value of the first symmetric polynomial is the sum of these values, which for the $i-th$ equivalence class is precisely $k-2i$. Then we note the normalization in each row is $\sqrt{\binom{k}{i}}$. Thus we have

$$r_1^2 \sum_{i=0}^{k} \left[\sqrt{\binom{k}{i}}\,(k-2i)\right]^2 = 1\,.$$

Thus $r_1 = \dfrac{1}{\sqrt{\sum_{i=0}^{k}\left[\binom{k}{i}(k-2i)^2\right]}}$ as desired. ◀

8 Using our "Compressed QFT" to Quantumly Sample from Distributions of Efficiently Specifiable Polynomial Evaluations and Hardness Consequences

In this section we use the unitary matrix developed in Section 7 to quantumly sample distributions with probabilities proportional to evaluations of Efficiently Specifiable polynomials at points in $[-k,k]^n$ for $k \in \exp(n)$. Here we assume that we have an efficient quantum circuit decomposition for this unitary. The prospects for this efficient decomposition are discussed in Section 9.

For convenience, we'll define a map $\psi : [-k,k] \to [0,k]$, for k even, with

$$\psi(y) = \begin{cases} \frac{k+y}{2} & \text{if } y \text{ is even} \\ 0 & \text{otherwise} \end{cases}$$

▶ **Definition 23.** Suppose Q is an Efficiently Specifiable polynomial Q with n variables and m monomials, and, for $k \leqslant exp(n)$, let Q'_k be its k-valued equivalent polynomial. Let $\mathrm{Var}\left[Q(X)\right]$ be the variance of the distribution over $\mathbb{R}$ induced by Q with assignments to the variables distributed over $\mathcal{B}(0,k)^n$ (or equivalently, this is $\mathrm{Var}\left[Q'_k\right]$ where each variable in Q'_k is independently uniformly chosen from $\{\pm 1\}$), as calculated in Appendix A. Then we define the of distribution $\mathcal{D}_{Q'_k}$ over n tuples of integers in $[-k,k]$ by:

$$\Pr_{\mathcal{D}_{Q'_k}}[y] = \frac{Q(y)^2 \binom{k}{\psi(y_1)}\binom{k}{\psi(y_2)}\cdots\binom{k}{\psi(y_n)}}{2^{kn}\mathrm{Var}\left[Q(X)\right]}\,.$$

▶ **Theorem 24.** *By applying $(L \cdot \tilde{D}_k \cdot R)^{\otimes n}$ in place of the Quantum Fourier Transform over $\mathbb{Z}_2^n$ in Section 4 we can quantumly sample from $\mathcal{D}_{Q'_k}$.*

Proof. Since we are assuming Q is Efficiently Specifiable, let $h : [m] \to \{0,1\}^n$ be the invertible function describing the variables in each monomial. We start by producing the

state over $k+1$ dimensional qudits:

$$\frac{1}{\sqrt{m}} \sum_{z \in [m]} |h(z)\rangle$$

which we prepare via the procedure described in Lemma 1.

Instead of thinking of h as mapping an index of a monomial from $[m]$ to the variables in that monomial, we now think of h as taking an index of a monomial in Q to a polynomial expressed in the $\{1, x^{(1)} + x^{(2)} + ... + x^{(k)}\}^n$ basis.

Now take this state and apply the unitary (which we assume can be realized by an efficient quantum circuit) $(L \cdot \tilde{D}_k \cdot R)^{\otimes n}$. Notice each $y \in [-k,k]^n$ has an associated amplitude:

$$\alpha_y = \frac{r_0^{n-d} r_1^d Q(y) \sqrt{\binom{k}{\psi(y_1)}\binom{k}{\psi(y_2)}...\binom{k}{\psi(y_n)}}}{\sqrt{m}}.$$

Letting $p_y = \Pr_{\mathcal{D}_{Q'_k}}[y]$, note that, by plugging in r_0, r_1 from Section 7:

$$\begin{aligned}
\alpha_y^2 &= \frac{Q(y)^2 \binom{k}{\psi(y_1)}\binom{k}{\psi(y_2)}...\binom{k}{\psi(y_n)} r_0^{2(n-d)} r_1^{2d}}{m} \\
&= \frac{Q(y)^2 \binom{k}{\psi(y_1)}\binom{k}{\psi(y_2)}...\binom{k}{\psi(y_n)}}{m 2^{k(n-d)} \left(\sum_{i=0}^{k} \left[\binom{k}{i} (k-2i)^2 \right] \right)^d} \\
&= \frac{Q(y)^2 \binom{k}{\psi(y_1)}\binom{k}{\psi(y_2)}...\binom{k}{\psi(y_n)}}{2^{kn-kd} \mathrm{Var}\left[Q(X)\right] 2^{kd}} = \frac{Q(y)^2 \binom{k}{\psi(y_1)}\binom{k}{\psi(y_2)}...\binom{k}{\psi(y_n)}}{2^{kn} \mathrm{Var}\left[Q(X)\right]} = p_y
\end{aligned}$$

◀

Furthermore, using a similar argument to Theorem 10 we can obtain the following theorem, which now gives our hardness result for the existence of Sampler for this class of distributions, whose proof we give in Appendix C:

▶ **Lemma 25.** *Given an Efficiently Specifiable polynomial Q with n variables and m monomials, let Q'_k be its k-valued equivalent polynomial, for some fixed $k \leqslant exp(n)$. Suppose we have a Sampler S with respect to our quantumly sampled distribution class, $\mathcal{D}_{Q'_k}$, and let* $\mathrm{Var}\left[Q(X)\right]$ *denote the variance of the distribution over $\mathbb{R}$ induced by Q with assignments distributed from $\mathcal{B}(0,k)^n$. Then we can find a randomized procedure computing an $\epsilon \mathrm{Var}\left[Q(X)\right]$-additive approximate δ-average case solution to Q^2 with respect to $\mathcal{B}(0,k)^n$ in time* $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ *with access to an* **NP** *oracle.*

9 Putting it All Together

In this section we put our results in perspective and conclude. As mentioned before, our goal is to find a class of distributions $\{\mathcal{D}_n\}_{n>0}$ that can be sampled exactly in $\mathrm{poly}(n)$ time on a Quantum Computer, with the property that there does not exist a (classical) Sampler relative to that class of distributions, $\{\mathcal{D}_n\}_{n>0}$. Using the results in Sections 5 and 6 we can quantumly sample from a class of distributions $\{\mathcal{D}_{Q'_k}\}_{n>0}$, where $k \in \mathrm{poly}(n)$ with the property that, if there exists a classical Sampler relative to this class of distributions, there exists an $\epsilon \mathrm{Var}\left[Q(X)\right]$-additive δ-average case solution to the Q^2 function with respect to the $\mathcal{B}(0,k)^n$ distribution. If we had an efficient decomposition for the "Compressed QFT" unitary matrix, we could use the results from Sections 8 and Appendix C to make k as large as $\exp(n)$. We would like this to be an infeasible proposition, and so we conjecture:

▶ **Conjecture 26.** *There exists some Efficiently Specifiable polynomial Q on n variables, so that $\epsilon \mathrm{Var}\left[Q(X)\right]$-additive δ-average case solutions with respect to $\mathcal{B}(0,k)^n$, for any fixed $k \leqslant \exp(n)$, to Q^2, cannot be computed in (classical) randomized $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ time with a* **PH** *oracle.*

At the moment we don't know of such a decomposition for the "Compressed QFT". However, we do know that we can classically evaluate a related fast (time $n \log^2 n$) polynomial transform by a theorem of Driscoll, Healy, and Rockmore [8]. We wonder if there is some way to use intuition gained by the existence of this fast polynomial transform to show the existence of an efficient decomposition for our "Compressed QFT".

Additionally, if we can prove the Anti-Concentration Conjecture (Conjecture 11) relative to some Efficiently Specifiable polynomial Q and the $\mathcal{B}(0,k)^n$ distribution, we appeal to Theorem 12 to show that it suffices to prove:

▶ **Conjecture 27.** *There exists some Efficiently Specifiable polynomial Q with n variables, so that Q satisfies Conjecture 11 relative to $\mathcal{B}(0,k)^n$, for some fixed $k \leqslant \exp(n)$, and ϵ-multiplicative δ-average case solutions, with respect to $\mathcal{B}(0,k)^n$, to Q^2 cannot be computed in (classical) randomized $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ time with a* **PH** *oracle.*

We would be happy to prove that either of these two solutions (additive or multiplicative) are #**P**-hard. In this case we can simply invoke Toda's Theorem [20] to show that such a randomized classical solution would collapse **PH** to some finite level. We note that at present, both of these conjectures seem out of reach, because we do not have an example of a polynomial that is #**P**-hard to approximate (either multiplicatively or additively) on average, in the sense that we need.

References

1. Scott Aaronson. BQP and the polynomial hierarchy. In Leonard J. Schulman, editor, *STOC*, pages 141–150. ACM, 2010.
2. Scott Aaronson. The equivalence of sampling and searching. *Electronic Colloquium on Computational Complexity (ECCC)*, 17:128, 2010.
3. Scott Aaronson and Alex Arkhipov. The computational complexity of linear optics. *Theory of Computing*, 9:143–252, 2013.
4. Andrew C Berry. The accuracy of the gaussian approximation to the sum of independent variates. *Transactions of the American Mathematical Society*, 49(1):122–136, 1941.
5. G. E. P. Box and M. E. Muller. A note on the generation of random normal deviates. *Annals of Mathematical Statistics*, 29:610–611, 1958.
6. 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 of London A: Mathematical, Physical and Engineering Sciences*, 2010. doi:10.1098/rspa.2010.0301.
7. Michael J. Bremner, Ashley Montanaro, and Dan J. Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *CoRR*, abs/1504.07999, 2015. URL: http://arxiv.org/abs/1504.07999.
8. James R. Driscoll, Dennis M. Healy Jr., and Daniel N. Rockmore. Fast discrete polynomial transforms with applications to data analysis for distance transitive graphs. *SIAM J. Comput.*, 26(4):1066–1099, 1997.
9. Edward Farhi and Aram W Harrow. Quantum supremacy through the quantum approximate optimization algorithm, 2016.

10 Bill Fefferman, Ronen Shaltiel, Christopher Umans, and Emanuele Viola. On beating the hybrid argument. *Theory of Computing*, 9:809–843, 2013. doi:10.4086/toc.2013.v009a026.

11 Bill Fefferman and Chris Umans. On pseudorandom generators and the BQP vs PH problem. *QIP*, 2011.

12 Richard Jozsa and Marrten Van Den Nest. Classical simulation complexity of extended clifford circuits. *Quantum Info. Comput.*, 14(7&8):633–648, May 2014. URL: http://dl.acm.org/citation.cfm?id=2638682.2638689.

13 A.Y Kitaev, A.H Shen, and M.N Vyalyi. *Quantum and Classical Computation*. AMS, 2002.

14 Donald E. Knuth. *The Art of Computer Programming, Volume III: Sorting and Searching*. Addison-Wesley, 1973.

15 Tomoyuki Morimae, Keisuke Fujii, and Joseph F. Fitzsimons. Hardness of classically simulating the one-clean-qubit model. *Phys. Rev. Lett.*, 112:130502, Apr 2014. doi:10.1103/PhysRevLett.112.130502.

16 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quanum Information*. Cambridge U.P., 2000.

17 Larry J. Stockmeyer. On approximation algorithms for #P. *SIAM J. Comput.*, 14(4):849–861, 1985.

18 Terence Tao and Van Vu. On the permanent of random Bernoulli matrices. In *Advances in Mathematics*, page 75, 2008.

19 Barbara M. Terhal and David P. DiVincenzo. Adaptive quantum computation, constant depth quantum circuits and Arthur-Merlin games. *CoRR*, quant-ph/0205133, 2002. URL: http://arxiv.org/abs/quant-ph/0205133.

20 Seinosuke Toda. PP is as hard as the Polynomial-Time Hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991.

A Computation of the Variance of Efficiently Specifiable Polynomial

In this section we compute the variance of the distribution induced by an Efficiently Specifiable polynomial Q with assignments to the variables chosen independently from the $\mathcal{B}(0,k)$ distribution. We will denote this throughout the section by $\mathrm{Var}\left[Q(X)\right]$. Recall, by the definition of Efficiently Specifiable, we have that Q is an n variate homogenous multilinear polynomial with $\{0,1\}$ coefficients. Assume Q is of degree d and has m monomials. Let each $[-k,k]$ valued variable X_i be independently distributed from $\mathcal{B}(0,k)$.

We adopt the notation whereby, for $j \in [m], l \in [d]$, x_{j_l} is the l-th variable in the j-th monomial of Q.

Using the notation we can express $Q(X_1,...,X_n) = \sum_{j=1}^{m} \prod_{l=1}^{d} X_{j_l}$. By independence of these random variables and since they are mean 0, it suffices to compute the variance of each monomial and multiply by m:

$$\mathrm{Var}\left[Q(X)\right] = \mathrm{Var}\left[Q(X_1,...,X_n)\right] = \mathrm{E}\left[\sum_{j=1}^{m}\prod_{l=1}^{d} X_{j_l}^2\right] = \sum_{j=1}^{m}\mathrm{E}\left[\prod_{l=1}^{d} X_{j_l}^2\right] \tag{4}$$

$$= m\mathrm{E}\left[\prod_{l=1}^{d} X_{1_l}^2\right] = m\prod_{l=1}^{d}\mathrm{E}\left[X_{1_l}^2\right] \tag{5}$$

$$= m\left(\mathrm{E}\left[X_{1_1}^2\right]\right)^d \tag{6}$$

Now since these random variables are independent and identically distributed, we can calculate the variance of an arbitrary X_{j_l} for any $j \in [m]$ and $l \in [d]$:

$$\mathrm{E}\left[X_{jl}^2\right] = \frac{1}{2^k}\sum_{i=0}^{k}\left[(k-2i)^2\binom{k}{i}\right] \tag{7}$$

$$\tag{8}$$

Thus, the variance of Q is:

$$m\frac{1}{2^{kd}}\left(\sum_{i=0}^{k}\left[(k-2i)^2\binom{k}{i}\right]\right)^d.$$

It will be useful to calculate this variance in a different way, and obtain a simple closed form. In this way we will consider the k-valued equivalent polynomial $Q'_k : \mathbb{T}_2^{nk} \to \mathbb{R}$ which is a sum of $m' = mk^d$ multilinear monomials, each of degree d. As before we can write $Q'_k(X_1, ..., X_{nk}) = \sum_{j=1}^{m'}\prod_{l=1}^{d} X_{j_l}$. Note that the uniform distribution over assignments in $\mathbb{T}_2^{kn}$ to Q'_k induces $\mathcal{B}(0,k)^n$ over $[-k,k]^n$ assignments to Q. By the same argument as above, using symmetry and independence of random variables, we have:

$$\begin{aligned}\mathrm{Var}\left[Q(X)\right] = \mathrm{Var}\left[Q(X_1, X_2, ..., X_n)\right] =& \mathrm{Var}\left[Q'_k(X_1, X_2, ..., X_{nk})\right] &(9)\\ =& m'\prod_{l=1}^{d}\mathrm{E}\left[X_{1_l}^2\right] &(10)\\ =& m'\mathrm{E}\left[X_{1_1}^2\right]^d = 1^d m' = m' = k^d m &(11)\end{aligned}$$

B Examples of Efficiently Specifiable Polynomials

In this section we give two examples of Efficiently Specifiable polynomials.

▶ **Theorem 28. Permanent** $(x_1, ..., x_{n^2}) = \sum_{\sigma\in S_n}\prod_{i=1}^{n} x_{i,\sigma(i)}$ *is Efficiently Specifiable.*

Proof. We note that it will be convenient in this section to index starting from 0. The theorem follows from the existence of an $h_{\mathbf{Permanent}} : [0, n! - 1] \to \{0,1\}^{n^2}$ that efficiently maps the i-th permutation over n elements to a string representing its obvious encoding as an $n \times n$ permutation matrix. We will prove that such an efficiently computable $h_{\mathbf{Permanent}}$ exists and prove that its inverse, $h^{-1}_{\mathbf{Permanent}}$ is also efficiently computable.

The existence of $h_{\mathbf{Permanent}}$ follows from the so-called "factorial number system" [14], which gives an efficient bijection that associates each number in $[0, n! - 1]$ with a permutation in S_n. It is customary to think of the permutation encoded by the factorial number system as a permuted sequence of n numbers, so that each permutation is encoded in $n \log n$ bits. However, it is clear that we can efficiently transform this notation into the $n \times n$ permutation matrix.

To go from an integer $j \in [0, n! - 1]$ to its permutation we:

1. Take j to its "factorial representation", an n number sequence, where the i-th place value is associated with $(i-1)!$, and the sum of the digits multiplied by the respective place value is the value of the number itself. We achieve this representation by starting from $(n-1)!$, setting the leftmost value of the representation to $j' = \lfloor \frac{j}{(n-1)!} \rfloor$, letting the

next value be $\lfloor \frac{j-j' \cdot (n-1)!}{(n-2)!} \rfloor$ and continuing until 0. Clearly this process can be efficiently achieved and efficiently inverted, and observe that the largest each value in the i-th place value can be is i.

2. In each step we maintain a list ℓ which we think of as originally containing n numbers in ascending order from 0 to $n-1$.
3. Repeat this step n times, once for each number in the factorial representation. Going from left to right, start with the left-most number in the representation and output the value in that position in the list, ℓ. Remove that position from ℓ.
4. The resulting n number sequence is the encoding of the permutation, in the standard $n \log n$ bit encoding. ◀

Now we show that the Hamiltonian Cycle Polynomial is Efficiently Specifiable.

Given a graph G on n vertices, we say a Hamiltonian Cycle is a path in G that starts at a given vertex, visits each vertex in the graph exactly once and returns to the start vertex.

We define an n-cycle to be a Hamiltonian cycle in the complete graph on n vertices. Note that there are exactly $(n-1)!$ n-cycles.

▶ **Theorem 29.** **HamiltonianCycle** $(x_1, ..., x_{n^2}) = \sum_{\sigma:\, n-cycle} \prod_{i=1}^{n} x_{i,\sigma(i)}$ *is Efficiently Specifiable.*

Proof. We can modify the algorithm for the Permanent above to give us an efficiently computable $h_{HC} : [0, (n-1)! - 1] \to \{0,1\}^{n^2}$ with an efficiently computable h_{HC}^{-1}.

To go from a number $j \in [0, (n-1)! - 1]$ to its n-cycle we:

1. Take j to its factorial representation as above. Now this is an $n-1$ number sequence where the i-th place value is associated with $(i-1)!$, and the sum of the digits multiplied by the respective place value is the value of the number itself.
2. In each step we maintain a list ℓ which we think of as originally containing n numbers in ascending order from 0 to $n-1$.
3. Repeat this step $n-1$ times, once for each number in the factorial representation. First remove the smallest element of the list. Then going from left to right, start with the left-most number in the representation and output the value in that position in the list, ℓ. Remove that position from ℓ.
4. We output 0 as the n-th value of our n-cycle.

To take an n-cycle to a factorial representation, we can easily invert the process:

1. In each step we maintain a list ℓ which we think of as originally containing n numbers in order from 0 to $n-1$.
2. Repeat this step $n-1$ times. Remove the smallest element of the list. Going from left to right, start with the left-most number in the n-cycle and output the position of that number in the list ℓ (where we index the list starting with the 0 position). Remove the number at this position from ℓ. ◀

C The Hardness of Classical Sampling from the Compressed Distribution

In this section, we use the same ideas used in the analysis of Section 5, to invoke Stockmeyer's Theorem (Theorem 7), together with the assumed existence of a Sampler for $\mathcal{D}_{Q'_k}$ to obtain hardness consequences for classical sampling with $k \leqslant \exp(n)$.

▶ **Lemma 30.** *Given an Efficiently Specifiable polynomial Q with n variables and m monomials, let Q'_k be its k-valued equivalent polynomial, for some fixed $k \leqslant exp(n)$. Suppose we have a*

Sampler S with respect to our quantumly sampled distribution class, $\mathcal{D}_{Q'_k}$, and let $\mathrm{Var}[Q(X)]$ *denote the variance of the distribution over $\mathbb{R}$ induced by Q with assignments distributed from $\mathcal{B}(0,k)^n$. Then we can find a randomized procedure computing an $\epsilon\mathrm{Var}[Q(X)]$-additive approximate δ-average case solution to Q^2 with respect to $\mathcal{B}(0,k)^n$ in time* $\mathrm{poly}(n, 1/\epsilon, 1/\delta)$ *with access to an* **NP** *oracle.*

Proof. Setting $\nu = \epsilon\delta/16$, suppose S samples from a distribution $\mathcal{D}'$ so that $\|\mathcal{D}_{Q'_k} - \mathcal{D}'\| \leqslant \nu$. Let $p_y = \Pr_{\mathcal{D}_{Q'_k}}[y]$ and $q_y = \Pr_{\mathcal{D}'}[y]$.

We define $\phi : \{\pm 1\}^{kn} \to [-k,k]^n$ to be the map from each $\{\pm 1\}^{kn}$ assignment to its equivalence class of assignments, which is n blocks of even integral values in the interval $[-k,k]$. Note that, given a uniformly random $\{\pm 1\}^{kn}$ assignment, ϕ induces the $\mathcal{B}(0,k)$ distribution over $[-k,k]^n$.

Our procedure picks a $y \in [-k,k]^n$ distributed[2] via $\mathcal{B}(0,k)^n$, and outputs an estimate $\tilde{q}_y$. Equivalently, we analyze this procedure by considering a uniformly distributed $x \in \{\pm 1\}^{kn}$ and then returning an approximate count, $\tilde{q}_{\phi(x)}$ to $q_{\phi(x)}$. We prove that our procedure runs in time $poly(n, 1/\epsilon, 1/\delta)$ with the guarantee that:

$$\Pr_x\left[\frac{|\tilde{q}_{\phi(x)} - p_{\phi(x)}|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \frac{\epsilon}{2^{kn}}\right] \leqslant \delta\,.$$

And by our above analysis of the quantum sampler:

$$p_{\phi(x)} = \frac{Q(\phi(x))^2\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}}{2^{kn}\mathrm{Var}[Q(X)]}\,.$$

Note that: $\frac{1}{2}\sum_{y\in[-k,+k]^n}|p_y - q_y| \leqslant \nu$, which, in terms of x, because we are summing over all strings in the orbit under $(S_k)^n$ symmetry, can be written:

$$\frac{1}{2}\sum_{x\in\{\pm 1\}^{kn}}\frac{\left|p_{\phi(x)} - q_{\phi(x)}\right|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} \leqslant \nu\,.$$

First we define for each x, $\Delta_x = \frac{|p_{\phi(x)} - q_{\phi(x)}|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}}$ and so $\|\mathcal{D}_{Q'_k} - \mathcal{D}'\| = \frac{1}{2}\sum_x \Delta_x$.

Note that:

$$\mathrm{E}_x[\Delta_x] = \frac{\sum_x \Delta_x}{2^{kn}} = \frac{2\nu}{2^{kn}}\,.$$

And applying Markov, $\forall j > 1$,

$$\Pr_x[\Delta_x > \frac{j2\nu}{2^{kn}}] < \frac{1}{j}\,.$$

Setting $j = \frac{4}{\delta}$ and recalling that $\nu = \frac{\epsilon\delta}{16}$, we have,

$$\Pr_x[\Delta_x > \frac{\epsilon}{2}\cdot\frac{1}{2^{kn}}] < \frac{\delta}{4}\,.$$

[2] We can do this when $k = exp(n)$ by approximately sampling from the Normal distribution, with only $poly(n)$ bits of randomness, and using this to approximate $\mathcal{B}(0,k)$ to within additive error $1/poly(n)$ e.g., [5, 4].

Then use approximate counting (with an **NP** oracle), using Theorem 7 on the randomness of S to obtain an output $\tilde{q}_y$ so that, for all $\gamma > 0$, in time polynomial in n and $\frac{1}{\gamma}$:

$$\Pr[|\tilde{q}_y - q_y| > \gamma \cdot q_y] < \frac{1}{2^n}.$$

Because we can amplify the failure probability of Stockmeyer's algorithm to be inverse exponential.

Equivalently in terms of x:

$$\Pr_x\left[\frac{|\tilde{q}_{\phi(x)} - q_{\phi(x)}|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \gamma \cdot \frac{q_{\phi(x)}}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}}\right] < \frac{1}{2^n}.$$

And we have:

$$\mathrm{E}_x\left[\frac{q_{\phi(x)}}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}}\right] \leqslant \frac{\sum_x \frac{q_{\phi(x)}}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}}}{2^{kn}} = \frac{1}{2^{kn}}.$$

Thus, by Markov,

$$\Pr_x[\frac{q_{\phi(x)}}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \frac{j}{2^{kn}}] < \frac{1}{j}.$$

Now, setting $\gamma = \frac{\epsilon\delta}{8}$ and applying the union bound:

$$\begin{aligned}
&\Pr_x\left[\frac{\left|\tilde{q}_{\phi(x)} - p_{\phi(x)}\right|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \frac{\epsilon}{2^{kn}}\right] \\
&\quad\leqslant \Pr_x\left[\frac{\left|\tilde{q}_{\phi(x)} - q_{\phi(x)}\right|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \frac{\epsilon}{2}\cdot\frac{1}{2^{kn}}\right] \\
&\qquad\quad + \Pr_x\left[\frac{\left|q_{\phi(x)} - p_{\phi(x)}\right|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \frac{\epsilon}{2}\cdot\frac{1}{2^{kn}}\right] \\
&\quad\leqslant \Pr_x\left[\frac{q_{\phi(x)}}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \frac{j}{2^{kn}}\right] \\
&\qquad\quad + \Pr\left[\frac{|\tilde{q}_{\phi(x)} - q_{\phi(x)}|}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}} > \gamma\cdot\frac{q_{\phi(x)}}{\binom{k}{\psi(\phi(x)_1)}\binom{k}{\psi(\phi(x)_2)}\cdots\binom{k}{\psi(\phi(x)_n)}}\right] \\
&\qquad\quad + \Pr_x\left[\Delta_x > \frac{\epsilon}{2}\cdot\frac{1}{2^{kn}}\right] \\
&\quad\leqslant \frac{1}{j} + \frac{1}{2^n} + \frac{\delta}{4} \\
&\quad\leqslant \frac{\delta}{2} + \frac{1}{2^n} \leqslant \delta.
\end{aligned}$$

◀

Quantum-Proof Multi-Source Randomness Extractors in the Markov Model*

Rotem Arnon-Friedman[1], Christopher Portmann[2], and Volkher B. Scholz†[3]

1 **Institute for Theoretical Physics, ETH Zürich, Zürich, Switzerland**
`rotema@itp.phys.ethz.ch`
2 **Institute for Theoretical Physics, ETH Zürich, Zürich, Switzerland**
`chportma@phys.ethz.ch`
3 **Institute for Theoretical Physics, ETH Zürich, Zürich, Switzerland; and Department of Physics and Astronomy, Ghent University, Ghent, Belgium**
`volkher.scholz@ugent.be`

Abstract

Randomness extractors, widely used in classical and quantum cryptography and other fields of computer science, e.g., derandomization, are functions which generate almost uniform randomness from weak sources of randomness. In the quantum setting one must take into account the quantum side information held by an adversary which might be used to break the security of the extractor. In the case of seeded extractors the presence of quantum side information has been extensively studied. For multi-source extractors one can easily see that high conditional min-entropy is not sufficient to guarantee security against arbitrary side information, even in the classical case. Hence, the interesting question is under which models of (both quantum and classical) side information multi-source extractors remain secure. In this work we suggest a natural model of side information, which we call the Markov model, and prove that any multi-source extractor remains secure in the presence of quantum side information of this type (albeit with weaker parameters). This improves on previous results in which more restricted models were considered or the security of only some types of extractors was shown.

1998 ACM Subject Classification E.4 Coding and Information Theory

Keywords and phrases Quantum proof randomness extractors, multisource extractors, device independent quantum cryptography

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.2

1 Introduction

Randomness extractors are of great importance in many applications in computer science, such as derandomization and cryptography. The goal of a randomness extractor is to generate (almost) uniform randomness from weak sources of randomness. A weak source is usually modelled as a distribution X over $\{0,1\}^n$ such that the min-entropy of X is lower bounded by k: $H_{\min}(X) \geq k$. That is, the source is defined via a probability distribution for which the

* This project was supported by the European Research Council (ERC) via grant No. 258932, by the Swiss National Science Foundation (SNSF) via the National Centre of Competence in Research "QSIT", by the European Commission via the project "RAQUEL", and by the US Air Force Office of Scientific Research (AFOSR) via grant FA9550-16-1-0245.
† VBS was also supported by the EC through grants ERC QUTE (NR 197868). The majority of this work was carried out while VBS was at ETH.

probability of any string $x \in \{0,1\}^n$ is at most 2^{-k}. The idea is then to apply a randomness extractor to the weak source, such that the output source Y is indistinguishable from a uniformly random source.

Unfortunately, no deterministic function can extract the randomness from all sources with a given min-entropy, even for sources with high min-entropy [31, 33]. The most common ways to avoid this problem are to consider seeded extractors and multi-source extractors. In the case of seeded extractors one uses an additional truly uniform (but short) and independent seed, together with the weak source, as the input to the extractor (see, e.g., [15, 37, 33]).

Alternatively, and of special importance in applications where a uniform seed is not available (e.g., in quantum randomness amplification protocols, see Appendix C), multi-source randomness extractors can be used. In the multi-source case, instead of starting with one weak source X, one considers several *independent* weak sources $X_1, X_2, \ldots, X_l$ for some $l \geq 2$, with $H_{\min}(X_i) \geq k_i$ for $i \in [l]$, as the input to the extractor (see, for example, [38, 9, 27, 26, 5]).

In all types of extractors the randomness present in the weak sources must be lower bounded for the extractor to work (i.e., a bound on the min-entropy is given as a promise). However, this randomness inherently depends on the information one has about the weak sources, or to put differently, on the *side information* about the sources. For example, extractors are widely used for privacy amplification in cryptographic tasks. There, the starting point is that an adversary holds some side information C about the source such that the *conditional* min-entropy is bounded: $H_{\min}(X|C) \geq k$. The extractor is then used to transform X to a key Y, which should be close to uniform even conditioned on the side information C. If the extractor fulfils this requirement it is said to be secure.

Depending on the application one can consider adversaries with classical or quantum side information and ask whether an extractor remains secure even in the presence of such side information (with slightly weaker parameters). For seeded extractors this question has been extensively studied. In the presence of classical adversaries the side information about X can be translated to a decrease in the min-entropy and the extractor remains secure [17]. In the quantum case, it was further shown in [17] that all one-bit output extractors remain secure. It is still unknown whether all multi-bit output extractors remain secure (although the results of [2, 3] goes in this direction[1]), but several constructions of seeded extractors with good parameters were shown to work also in the presence of quantum side information [30, 8, 36, 13].

When considering multi-source extractors things get more complicated, even in the classical case. To see this, consider any one-bit output two-source extractor and let the adversary hold as side information the output of the extractor $Y = \mathrm{Ext}(X_1, X_2)$. As this is just one bit, $H_{\min}(X_1|Y) \geq k_1 - 1$ and $H_{\min}(X_2|Y) \geq k_2 - 1$. Furthermore, as the sources are independent even $H_{\min}(X_1|YX_2)$ and $H_{\min}(X_2|YX_1)$ remain high. Nevertheless, the extractor obviously fails to produce an output which looks uniform given the side information. In [16] several more examples are given in which a small amount of classical side information breaks the extractor completely.

This implies that one cannot expect to have multi-source extractors which are secure against any classical or quantum side information and thus raises the question: *under which assumptions on the structure of the sources and the side information $X_1 \cdots X_l C$ do multi-*

[1] Note that there is no contradiction between the results of [2, 3] and the famous counter example of a seeded extractor which breaks in the presence of quantum side information given in [11]; for details see [2].

source extractors remain secure even in the presence of C? The main objective of this work is to answer this question. In particular, we define a natural condition on the sources and the side information for which *all* multi-source extractors remain secure in the presence of both classical and quantum side information, but with an increase in the error of the extractor – the distance from uniform of the output.

1.1 Results and contributions

Our first contribution is a new definition of a quantum-proof multi-source extractor, which is simpler than previous proposals [16, 6] and yet sufficient to extract from these models. The original classical extractor definition requires the sources to be independent, i.e., in the two-source case one must have $I(X_1 : X_2) = 0$, where $I(\cdot : \cdot)$ denotes the mutual information. If an adversary is present and holds some side information C, the definition we introduce requires that the two sources be independent from the point of view of this adversary, i.e., $I(X_1 : X_2|C) = 0$. This definition is valid for both classical and quantum side information C. This means that the sources and the side information should form a *Markov chain* $X_1 \leftrightarrow C \leftrightarrow X_2$. For the case of more than two sources a similar Markov-type condition can be defined and we say that the sources and the side information are in the *Markov model*. The formal definitions are given in Section 3.

Compared to previous definitions of quantum-proof multi-source extractors, this has several advantages. Firstly, it is a natural generalization of the original classical extractor definition and the extension to quantum side information from [16], and it connects to the model of [6] in the following sense: any function satisfying our definition of a strong[2] extractor is also an extractor in the model of [6] – a more precise comparison to previous work is given in Section 1.2. Secondly, we consider it much more natural to put a requirement on the structure of the global state $\rho_{X_1X_2C}$, instead of describing permissible adversarial strategies that generate the side information C, as in [16, 6]. Thirdly, Markov chains arise naturally in certain applications. For example, in realisations of quantum randomness amplification protocols one can sometimes assume that the devices on which the experiment is being preformed have a Markov chain structure (for further details see Appendix C).

We also show that extractors in the Markov model can be used to extract randomness from a larger set of states. We prove that a bound on the *smooth* min-entropy [29] suffices for randomness extraction. This can be seen as a robustness property of the model, since in many applications one can only bound the smooth min-entropy rather than the min-entropy itself. In addition, we prove that any CPTP map performed on the side information – which might delete information and thus destroy the Markov property – cannot decrease the security of an extractor, hence extractors in the Markov model are also extractors for such non-Markov states[3].

Our second contribution is to prove that *all* extractors (weak and strong) remain secure in this model, both in the classical and quantum case, albeit with weaker parameters. In the classical case the proof is pretty trivial and standard (and the result is indeed not surprising). Nevertheless, as we could not find it anywhere else in the literature, we give it in this work for completeness and as comparison to the quantum case. More specifically, for classical side information we prove the following theorem:

[2] An extractor is said to be strong in a set of its sources if even conditioned on all the sources in this set the output cannot be distinguished from uniform (see formal definition in Section 3).
[3] This includes, in particular, states constructed according to the model of [6].

▶ **Theorem 1.** *Any* $(k_1, \ldots, k_l, \varepsilon)$*-[strong] l-source extractor is a* $\left(k_1 + \log \frac{1}{\varepsilon}, \ldots, k_l + \log \frac{1}{\varepsilon}, (l+1)\varepsilon\right)$*-[strong] classical-proof l-source extractor in the Markov model.*

The formal definitions of a (strong) l-source extractor and a (strong) classical-proof l-source extractor are given in Section 3.1. The important thing to note is that for the extractor to remain secure, the min-entropy of the sources needs to be just $\log \frac{1}{\varepsilon}$ higher, where ε is the security parameter (or the error) of the extractor. This is exactly the same as in the case of seeded extractors [17] with classical side information.

The main contribution of the current work is the quantum version of the theorem above:

▶ **Theorem 2.** *Any* $(k_1, \ldots, k_l, \varepsilon)$*-[strong] l-source extractor is a* $(k_1 + \log \frac{1}{\varepsilon}, \ldots, k_l + \log \frac{1}{\varepsilon}, \varepsilon')$*-[strong] quantum-proof l-source extractor in the Markov model with* $\varepsilon' = \sqrt{(l+1)\varepsilon 2^{(m-2)}}$*, where m is the output length of the extractor.*

The formal definitions of the extractors are given in Section 3.2. As in the classical case, the min-entropy of the sources needs to be just $\log \frac{1}{\varepsilon}$ higher. The error itself is $\sqrt{(l+1)\varepsilon 2^{(m-2)}}$ where l is the number of sources and m is the number of output bits of the considered extractor[4].

Although the blow-up in the error of the extractor in Theorem 2 might seem relatively high, one must note that many classical multi-source extractors have an error $\varepsilon = 2^{-mc}$ for some constant $c > 1$, hence in the quantum case the new error is $\varepsilon' = \frac{\sqrt{l+1}}{2} 2^{-m\frac{c-1}{2}}$, i.e., both the classical and quantum errors are of the order $2^{-\Omega(m)}$. We provide several explicit constructions in Appendix B, where we show how to achieve similar parameters to the classical case, even if $\varepsilon \gg 2^{-m}$, by composing the multi-source extractor with a quantum-proof seeded extractor.

Apart from presenting the Markov model for extractors and proving the theorems above, we also contribute on the technical level. While previous works use the techniques of [17] for the one-bit output case and then extend it using a quantum XOR lemma [16], we use a completely different proof technique which is based on the recent work of [2]. The advantage of our technique is that it also applies to weak extractors, whereas the techniques of [6] require the extractors to be strong in order to prove that they are secure. We extend on our proof technique in Section 1.3.

1.2 Related work

As far as we are aware, the question of the security of multi-source extractors in the presence of side information was considered only in two works: [16] and [6]. Both works deal with quantum side information, and classical side information can of course be taken as a special case. We are not aware of any works dealing with the case of classical side information directly.

[16] initiated the study of multi-source extractors in the presence of side information. They considered the case of two sources and quantum side information in product form. More specifically, given the two independent sources X_1 and X_2, the side information is given by a state $\rho_{C_1} \otimes \rho_{C_2}$ such that $H_{\min}(X_i|C_1C_2) = H_{\min}(X_i|C_i) \geq k_i$. In this way, the

[4] This matches exactly the bound proven in [16, Corollary 27] for the restricted case of product side information, $l = 2$, and $m = 1$. We note that this is also an improvement over the constructions in the model of [6], for which the error in [6, Theorem 5.3] for $l = 2$ is of the form $2^m\sqrt{\varepsilon}$, i.e., an order of $\sqrt{2^m}$ worse than ours.

side information does not break the independence of the sources[5]. It was proven in [16] that any *one-bit output* two-source extractor remains secure in the presence of product side information. They further show that a specific construction of a multi-bit output two-source extractor, that of [9], is also secure in the considered model, by reducing it to the one-bit case.

Recently, another, more general model for an adversary was considered in [6]. For simplicity, we explain here the model for the case of two sources only; see [6] for the general definition. In [6] the side information of the adversary must be created in the following way: in the beginning the adversary can have any bipartite quantum state $\rho_{E_1E_2}$, independent of the sources. Then, to create her final side information $\rho_{C_1C_2}$, she can correlate her state with the sources by performing an independent "leaking operation" from each source to one of the subsystems. More specifically, they model the leaking operation as a map for $i \in \{1,2\}$, $\Phi_i : \mathrm{L}(X_i \otimes E_i) \to \mathrm{L}(X_i \otimes C_i)$. The resulting classical-quantum state $\rho_{X_1X_2C_1C_2}$ can be written as $\rho_{X_1X_2C_1C_2} = \Phi_1 \otimes \Phi_2(\rho_{X_1X_2E_1E_2})$. For the relevant conditions on the min-entropy see [6].

It was then proven in [6] that for multi-source extractors which are strong in all but one source, this complex adversarial leaking operation is in fact equivalent to providing the adversary with side information about only one source. That is, when using an extractor which is strong in all but one of its sources, any adversary who is restricted to the model of [6] is in fact no stronger than an adversary who has side information about just one source. It is further shown that several known extractor constructions are still secure when the adversary holds quantum side information about one of the sources – with an increase in the error of the extractor. The leaking model of [6] can also be defined for weak extractors. However, the proof techniques of [6] only work for strong extractors, since they rely on the equivalence to side information about one source. Thus, there are currently no known extractor constructions that directly satisfy the weak extractor model from [6], without relying on an underlying strong extractor.

Our work is a natural generalization of [16], since independent sources are a subset of Markov sources. The model from [6] is different from ours in the sense that there exist states $\rho_{X_1X_2C}$ which are Markov chains but cannot be constructed by the leaking model from [6] and vice versa. However, as already proven in [6], for a function to satisfy their strong extractor definition, it is sufficient for it to be secure in the presence of side information about one of the sources. Since side information about one source is a Markov chain, it follows that any strong extractor in the Markov model is also a strong extractor in the leaking model of [6] – for completeness, we provide a proof of this in Appendix A.2. It is currently unknown whether the same statement holds for weak extractors. Interestingly, the converse statement also holds: we (implicitly) prove in this work that any function that is an extractor for side information in product form is an extractor in the Markov model with slightly weaker parameters. Since the leaking model from [6] includes states in product from, an extractor from [6] is also an extractor in the Markov model with slightly weaker parameters.

[5] [16] also considered another model for the adversary, called the bounded storage model, in which an assumption is made on the size of the adversary's storage capacity. In this work we consider only the more general case, in which we make no assumption about the adversary's power. For more details see [16].

1.3 Proof outline and techniques

The proof of the classical result, i.e., Theorem 1, is quite standard. The main part of this work is therefore devoted to the quantum case – the proof of Theorem 2. The main idea is to not consider the most general measurement that could be performed to distinguish the output of the extractor from uniform, but instead consider a specific strategy, which consists in first measuring the quantum side information, then trying to distinguish the output from uniform given the resulting classical side information. We first prove that this specific strategy is not much worse than the optimal strategy. Then we show that this classical side information satisfies the requirements of a classical two-source extractor in the Markov model. Thus, security in the quantum case follows from security in the classical case.

More specifically, the proof can be decomposed in the following steps.

1. We start by considering only *product* side information in Section 4.1. We employ ideas from [2], where the security definition of the extractor is rewritten using operators inequalities, to give a bound in Lemma 12 on the distance from uniform of the extractor output.
2. Next (in Lemmas 13 and 14) we simplify the bound by noting that it can be seen as a *specific* simple distinguishing strategy when trying to distinguish the output of the extractor from uniform using the side information. This specific strategy is one in which the product side information is measured independently of the output of the extractor (while a general distinguisher could use more complicated distinguishing strategies). Hence we obtain a reduction from quantum to classical side information.
3. We put this together in Lemma 15, to show that *any* multi-source extractor is secure in the presence of product side information[6].
4. Finally, in Section 4.2, we extend the result from the product model to the quantum Markov model by exploiting the structure of quantum Markov-chain states, and by this prove that Theorem 2 holds.

1.4 Organisation of the paper

The rest of paper is organised as follows. In Section 2 we give some necessary preliminaries. Section 3 is devoted to the definitions of classical and quantum-proof multi-source extractors in the Markov model. The proof of our main theorem, Theorem 2, is then given in Section 4. We conclude in Section 5 with some open questions.

Do to space restrictions, some additional results have been moved to the appendices. In Appendix A we show that multi-source extractors in the Markov model can be used to extract from some sources that do not directly satisfy the definition, e.g., when only a bound on the smooth min-entropy is given. In Appendix B we give the parameters of explicit constructions of quantum multi-source extractors, i.e., we apply our results to some specific constructions of multi-source extractors. In Appendix C we further motivate the Markov model in the context of quantum randomness amplification protocols. The remaining appendices contain technical proofs.

2 Preliminaries

We assume familiarity with standard notation in probability theory as well as with basic concepts in quantum information theory including density matrices, positive-operator valued

[6] This can be seen as an extension of the result of [16] but the proof is different.

measures (POVMs), and distance measures such as the trace distance. We refer to, e.g., [23] for an introduction to quantum information.

Throughout the paper X, Y and Z denote classical random variables while A, B and C denote quantum systems. All logarithms are in base 2. $[l]$ denotes the set $\{1, 2, \ldots, l\}$ and for $i \in [l]$ we denote $\bar{i} = 1, \ldots, i-1, i+1, \ldots, l$.

If a classical random variable X takes the value x with probability p_x it can be written as the quantum state $\rho_X = \sum_x p_x |x\rangle\langle x|$, where $\{|x\rangle\}_x$ is an orthonormal basis. If the classical system X is part of a composite system XC, any state of that composite system can be written as $\rho_{XC} = \sum_x p_x |x\rangle\langle x| \otimes \rho_C^x$. If C is quantum we say that the state ρ_{XC} is a classical-quantum state, or a cq-state. Similarly, a state $\rho_{X_1X_2C}$ classical on X_1, X_2 and quantum on C is called a ccq-state. For two independent random variables X and Y we often write $X \circ Y$ to denote the joint random variable with product distribution.

For a quantum state ρ_A we denote by $H(A)$ the Von Neumann entropy of ρ_A, i.e., $H(A) = -\mathrm{Tr}(\rho_A \log \rho_A)$. The conditional mutual information is defined as

$$I(A : B|C) = H(AC) + H(BC) - H(C) - H(ABC)\,.$$

In the case of classical systems, the Von Neumann entropy is reduced to the Shannon entropy. That is, for a random variable X, $H(X) = -\sum_x p_x \log p_x$, where p_x is the probability of $X = x$.

Given a cq-state $\rho_{XC} = \sum_x p_x |x\rangle\langle x| \otimes \rho_C^x$ the conditional min-entropy is $H_{\min}(X|C) = -\log p_{\mathrm{guess}}(X|C)$, where $p_{\mathrm{guess}}(X|C)$ is the maximum probability of guessing X given the quantum system C. That is,

$$p_{\mathrm{guess}}(X|C) = \max_{\{E_C^x\}_x} \left(\sum_x p_x \mathrm{Tr}(E_C^x \rho_C^x) \right)\,,$$

where the maximum is taken over all POVMs $\{E_C^x\}_x$ on C. For an empty system C, the conditional min-entropy of X given C reduces to the usual $H_{\min}(X) = -\log \max_x p_x$. Furthermore, if a quantum system C is measured and the measurement outcome is registered in the classical system Z then the min-entropy can only increase, namely, $H_{\min}(X|Z) \geq H_{\min}(X|C)$.

3 Multi-source extractors

3.1 Multi-source extractors in the presence of classical side information

Two-source extractors are defined as follows. The extension of the definition to the case of more than two sources is straightforward.

▶ **Definition 3** (Two-source extractor, [27]). A function $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is called a (k_1, k_2, ε) two-source extractor if for any two independent sources X_1, X_2 with $H_{\min}(X_1) \geq k_1$ and $H_{\min}(X_2) \geq k_2$, we have

$$\frac{1}{2} \| \mathrm{Ext}(X_1, X_2) - U_m \| \leq \varepsilon\,,$$

where U_m is a perfectly uniform random variable on m-bit strings. Ext is said to be *strong in the i'th input* if

$$\frac{1}{2} \| \mathrm{Ext}(X_1, X_2) X_i - U_m \circ X_i \| \leq \varepsilon\,.$$

If Ext is not strong in any of its inputs it is said to be weak.

As explained in Section 1, in the classical case one can also consider the security of the extractor in the presence of classical side information, denoted by Z, held by an adversary. That is, we would like the output of the extractor to be indistinguishable from uniform also *given* some additional classical information.

Since multi-source extractors cannot remain secure in the presence of an arbitrary classical side information (recall the examples presented in Section 1), we require the sources to be independent conditioned on the side information. Formally:

▶ **Definition 4** (Classical Markov model)**.** The random variables X_1, X_2 and Z are said to form a Markov chain, denoted by $X_1 \leftrightarrow Z \leftrightarrow X_2$, if

$$I(X_1 : X_2|Z) = 0\,.$$

For more than two sources $X_1, \ldots, X_l$ and side information Z, we say that they are in the Markov model if

$$\forall i \in [l], \quad I(X_i : X_{\bar{i}}|Z) = 0\,.$$

To see that $I(X_1 : X_2|Z) = 0$ indeed captures the idea that conditioned on Z the sources are independent, note that $I(X_1 : X_2|Z) = 0$ if and only if $\mathrm{p}(x_1, x_2|z) = \mathrm{p}(x_1|z) \cdot \mathrm{p}(x_2|z)$ for all x_1, x_2 and z.

We can now define classical-proof multi-source extractors in the following way. For simplicity, we give the definition in the case of two sources; the extension to more than two sources in the Markov model is straightforward.

▶ **Definition 5** (Classical-proof two-source extractor)**.** A function $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is a (k_1, k_2, ε) classical-proof two-source extractor secure in the Markov model, if for all sources X_1, X_2, and classical side information Z, where $X_1 \leftrightarrow Z \leftrightarrow X_2$ form a Markov chain, and with min-entropy $H_{\min}(X_1|Z) \geq k_1$ and $H_{\min}(X_2|Z) \geq k_2$, we have

$$\frac{1}{2}\|\mathrm{Ext}(X_1, X_2)\, Z - U_m \circ Z\| \leq \varepsilon\,, \tag{1}$$

where U_m is a perfectly uniform random variable on m-bit strings. Ext is said to be *strong in the i'th input* if

$$\frac{1}{2}\|\mathrm{Ext}(X_1, X_2) X_i Z - U_m \circ X_i Z\| \leq \varepsilon\,. \tag{2}$$

Indeed, if one requires that the sources and the side information Z fulfil Definition 4 then all multi-source extractors remain secure also in the presence of the side information Z. This is proven in the following lemma for two sources.

▶ **Lemma 6.** *Any (k_1, k_2, ε)-[strong] two-source extractor is a $(k_1 + \log\frac{1}{\varepsilon}, k_2 + \log\frac{1}{\varepsilon}, 3\varepsilon)$-[strong] classical-proof two-source extractor in the Markov model.*

Proof. Let $X_1 \leftrightarrow Z \leftrightarrow X_2$ be such that $H_{\min}(X_1|Z) \geq k_1 + \log\frac{1}{\varepsilon}$ and $H_{\min}(X_2|Z) \geq k_2 + \log\frac{1}{\varepsilon}$. For any two classical systems X and Z, we have

$$2^{-H_{\min}(X|Z)} = \mathop{\mathbb{E}}_{z \leftarrow Z}\left[2^{-H_{\min}(X|Z=z)}\right],$$

so by Markov's inequality,

$$\Pr_{z \leftarrow Z}\left[H_{\min}(X|Z=z) \leq H_{\min}(X|Z) - \log 1/\varepsilon\right] \leq \varepsilon\,.$$

Applying this to both X_1 and X_2, we have that with probability at least $1-2\varepsilon$ (over Z), $H_{\min}(X_1|Z=z) \geq k_1$ and $H_{\min}(X_2|Z=z) \geq k_2$. Due to the Markov-chain condition, the distributions $X_1|_{Z=z}$ and $X_2|_{Z=z}$ are independent. Hence for any (k_1, k_2, ε) two-source extractor Ext,

$$\frac{1}{2}\|\mathrm{Ext}(X_1,X_2)Z - U_m \circ Z\| = \frac{1}{2}\sum_z P_Z(z)\|\mathrm{Ext}(X_1|_{Z=z}, X_2|_{Z=z}) - U_m\| \leq 3\varepsilon\,.$$

For a strong extractor the proof is identical. ◀

By following the same steps as the proof of Lemma 6 for the case of l sources we get Theorem 1.

3.2 Multi-source extractors in the presence of quantum side information

We now consider multi-source extractors in the presence of quantum side information, i.e., in the following C denotes a quantum system. Similarly to Section 3.1 we restrict the sources and the quantum side information to the quantum Markov model. Formally,

▶ **Definition 7** (Quantum Markov model)**.** A ccq-state $\rho_{X_1X_2C}$ is said to form a Markov chain[7], denoted by $X_1 \leftrightarrow C \leftrightarrow X_2$, if

$$I(X_1 : X_2|C) = 0\,.$$

For more than two sources $X_1, \ldots, X_l$ and C we say that they are in the Markov model if

$$\forall i \in [l], \quad I(X_i : X_{\bar{i}}|C) = 0\,.$$

The following is then the natural analog of Definition 5 to the quantum setting. The extension to the case of more than two sources is straightforward.

▶ **Definition 8** (Quantum-proof two-source extractor)**.** A function $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is a (k_1, k_2, ε) quantum-proof two-source extractor in the Markov model, if for all sources X_1, X_2, and quantum side information C, where $X_1 \leftrightarrow C \leftrightarrow X_2$ form a Markov chain, and with min-entropy $H_{\min}(X_1|C) \geq k_1$ and $H_{\min}(X_2|C) \geq k_2$, we have

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\| \leq \varepsilon\,, \tag{3}$$

where $\rho_{\mathrm{Ext}(X_1,X_2)C} = \mathrm{Ext} \otimes \mathbb{1}_C \rho_{X_1X_2C}$ and ρ_{U_m} is the fully mixed state on a system of dimension 2^m. Ext is said to be *strong in the i'th input* if

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)X_iC} - \rho_{U_m} \otimes \rho_{X_iC}\| \leq \varepsilon\,. \tag{4}$$

If C above is classical then Definition 8 is reduced to Definition 5.

The interesting question is therefore whether there exist quantum-proof multi-source extractors. The main contribution of this work is to show that *any* multi-source extractor is also a quantum-proof multi-source extractor in the Markov model with a bit weaker parameters. The formal statement is given in Theorem 2 above and proven in the following section.

[7] The same definition is also used in the more general case where also the X_i's are quantum. For our purpose the case of classical sources and quantum side information is sufficient.

4 Security of multi-source extractors in the quantum Markov model

For simplicity, in this section we prove that two-source extractors are secure even when considering quantum side information in the form of a Markov chain. The extension to any number of sources, i.e., the proof of Theorem 2, follows by trivially repeating the same steps for more than two sources and using our definition of the Markov model (Definition 7). More specifically, the goal of this section is to prove the following:

▶ **Lemma 9.** *Any (k_1, k_2, ε)-[strong] two-source extractor is a $(k_1 + \log \frac{1}{\varepsilon}, k_2 + \log \frac{1}{\varepsilon}, \sqrt{3\varepsilon} \cdot 2^{(m/2-1)})$-[strong] quantum-proof two-source extractor in the Markov model, where m is the output length of the extractor.*

To prove this, we first show in Section 4.1 that all extractors are still secure in the case of side information in product from. Then in Section 4.2 we generalise this result to any side information in the Markov model.

4.1 Product quantum side information

We start by showing that *any* two-source extractor, as in Definition 3, is secure against product quantum side information. The product extractor as defined below is a special case of the extractor in Definition 8:

▶ **Definition 10** (Quantum-proof product two-source extractor, [16])**.** A function $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ is a (k_1, k_2, ε) quantum-proof product two-source extractor, if for all sources X_1, X_2, and quantum side information C, where $\rho_{X_1X_1C} = \rho_{X_1C_1} \otimes \rho_{X_2C_2}$, and with min-entropy $H_{\min}(X_1|C_1) \geq k_1$ and $H_{\min}(X_2|C_2) \geq k_2$, we have

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\| \leq \varepsilon\,, \tag{5}$$

where $\rho_{\mathrm{Ext}(X_1,X_2)C} = \mathrm{Ext} \otimes \mathbb{1}_C \rho_{X_1X_2C}$ and ρ_{U_m} is the fully mixed state on a system of dimension 2^m. Ext is said to be *strong in the i'th input* if

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)X_iC} - \rho_{U_m} \otimes \rho_{X_iC_i} \otimes \rho_{C_{\bar{i}}}\| \leq \varepsilon\,. \tag{6}$$

In the following we show that any two-source extractor remains secure in the product model, i.e., if the quantum state of the sources and the side information is of the form $\rho_{X_1X_2C} = \rho_{X_1C_1} \otimes \rho_{X_2C_2}$ (see Corollary 16 below for the formal statement). This can be seen as an extension of the results of [16], where only two-source extractors with one-bit output (i.e., $m = 1$ in our notation) and the extractor of [9] were shown to be secure against product quantum side information.

The first step of the proof uses the fact that any ccq-state $\rho_{X_1X_2C}$ can be obtained by performing local measurements on a pure state ρ_{ABC}. We formalise this in the following lemma. The proof of the lemma is trivial and given in Appendix D.

▶ **Lemma 11.** *Let $\rho_{X_1X_2C} = \sum_{x_1,x_2} |x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \rho_C(x_1, x_2)$. Then there exists a pure state ρ_{ABC} and POVMs $\{F_{x_1}\}, \{G_{x_2}\}$ such that*

$$\rho_C(x_1, x_2) = \mathrm{Tr}_{AB}\left[F_{x_1}^{\frac{1}{2}} \otimes G_{x_2}^{\frac{1}{2}} \otimes \mathbb{1}_C \rho_{ABC} F_{x_1}^{\frac{1}{2}} \otimes G_{x_2}^{\frac{1}{2}} \otimes \mathbb{1}_C\right]\,. \tag{7}$$

The following three lemmas are proven for the case of weak extractors. The lemmas and proofs for the strong case are very similar and therefore given in Appendix E. We start with the next lemma where the Cauchy-Schwarz inequality is used, as in [2].

▶ **Lemma 12.** *Let $\rho_{X_1X_2C}$ be any ccq-state, and let ρ_{ABC} and $\{F_{x_1}\},\{G_{x_2}\}$ satisfy Equation (7). Then there exists an alternative purification of ρ_{AB}, namely $\Psi_{ABC_1C_2}$, and two POVMs $\{H_{z_1}\}$, $\{K_{z_2}\}$ acting on C_1 and C_2, such that*

$$\frac{1}{M}\|\rho_{\mathrm{Ext}(X_1,X_2)C}-\rho_{U_m}\otimes\rho_C\|^2 \leq \sum_{\substack{x_1,x_2,\\ z_1,z_2,y}}\left[\delta_{\mathrm{Ext}(x_1,x_2)=y}-\frac{1}{M}\right]\left[\delta_{\mathrm{Ext}(z_1,z_2)=y}-\frac{1}{M}\right]\mathrm{Tr}\left[\Psi_{ABC_1C_2}F_{x_1}\otimes G_{x_2}\otimes H_{z_1}\otimes K_{z_2}\right],$$

where $M=2^m$ and m is the output length of the extractor. Moreover, if the state ρ_{AB} is of tensor product form, the purification $\Psi_{ABC_1C_2}$ also factorises into a tensor product between AC_1 and BC_2.

Proof. First, recall that for a hermitian matrix R we have $\|R\|=\max\{\mathrm{Tr}[RS]:-\mathbb{1}\leq S\leq\mathbb{1}\}$. Applying this to the matrix whose norm specifies the error of the extractor, we find

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C}-\rho_{U_m}\otimes\rho_C\|=\max_{-\mathbb{1}\leq S\leq\mathbb{1}}\mathrm{Tr}\left[\left(\rho_{\mathrm{Ext}(X_1,X_2)C}-\rho_{U_m}\otimes\rho_C\right)S\right].$$

Since $\rho_{\mathrm{Ext}(X_1,X_2)C}$ and $\rho_{U_m}\otimes\rho_C$ are block diagonal with respect to the outcome variable of the extractor y, S can be assumed to be block diagonal as well. Using this and inserting the expression for $\rho_{X_1X_2C}$ in Equation (7) we arrive at

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C}-\rho_{U_m}\otimes\rho_C\|=\max_{-\mathbb{1}\leq S_y\leq\mathbb{1}}\sum_y\mathrm{Tr}\left[\rho_{ABC}\Delta_y\otimes S_y\right],$$

where we used the abbreviation

$$\Delta_y=\sum_{x_1,x_2}\left[\delta_{\mathrm{Ext}(x_1,x_2)=y}-\frac{1}{M}\right]F_{x_1}\otimes G_{x_2}.$$

We now choose a special purification of ρ_{AB}, namely we consider the *pretty good purification* [39]

$$|\psi\rangle_{ABA'B'}=\rho_{AB}^{\frac{1}{2}}\otimes\mathbb{1}_{A'B'}|\Phi_{AA'}\rangle|\Phi_{BB'}\rangle,$$

where $|\Phi\rangle_{AA'}=\sum_a|aa\rangle$ denotes the unnormalised maximally entangled state. Since both $|\psi\rangle_{ABA'B'}$ and ρ_{ABC} are purifications of ρ_{AB} there exists an isometry $V:A'B'\to C$ such that $V|\psi\rangle\langle\psi|V^*=\rho_{ABC}$ and hence

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C}-\rho_{U_m}\otimes\rho_C\|\leq\max_{-\mathbb{1}\leq S_y\leq\mathbb{1}}\sum_y\mathrm{Tr}\left[|\psi\rangle\langle\psi|\Delta_y\otimes S_y\right],$$

since V^*S_yV is bounded in norm by one and hermitian. Inserting the identity $\mathbb{1}\otimes X_{A'}|\Phi_{AA'}\rangle=X_{A'}^T\otimes\mathbb{1}|\Phi_{AA'}\rangle$ for any matrix X (where $(.)^T$ denotes the transpose in the basis of the maximally entangled state), we find

$$\mathrm{Tr}\left[|\psi\rangle\langle\psi|\Delta_y\otimes S_y\right]=\mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}}\Delta_y\rho_{AB}^{\frac{1}{2}}(S_y)^T\right]. \tag{8}$$

The crucial observation is now that the sesquilinear form $(R_y)\times(S_y)\mapsto\sum_y\mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}}R_y^*\rho_{AB}^{\frac{1}{2}}S_y\right]$ on block-diagonal matrices is positive semi-definite and hence fulfils the Cauchy-Schwarz inequality. Applying this gives

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C}-\rho_{U_m}\otimes\rho_C\|^2\leq\left(\sum_y\mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}}\Delta_y\rho_{AB}^{\frac{1}{2}}\Delta_y\right]\right)\left(\sum_y\mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}}S_y\rho_{AB}^{\frac{1}{2}}S_y\right]\right).$$

Since we have that the norm of S_y is bounded by one, the terms in the second sum satisfies

$$\mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}} S_y \rho_{AB}^{\frac{1}{2}} S_y\right] \leq \mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}} S_y \rho_{AB}^{\frac{1}{2}}\right] \leq \mathrm{Tr}\left[\rho_{AB}\right] = 1\,,$$

and we arrive at

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\|^2 \leq M \sum_y \mathrm{Tr}\left[\rho_{AB}^{\frac{1}{2}} \Delta_y \rho_{AB}^{\frac{1}{2}} \Delta_y\right]\,.$$

Inserting the definition of Δ_y and reversing the identity leading to Equation (8) proves the assertion with $C_1 = A'$, $C_2 = B'$, $\Psi_{ABC_1C_2} = |\psi\rangle\langle\psi|$ and $H_{z_1} = F_{z_1}^T$, $K_{z_2} = G_{z_2}^T$. ◀

The upper bound of the preceding lemma can be further simplified (the proof is given in Appendix D):

▶ **Lemma 13.** *For any $\Psi_{ABC_1C_2}$ and positive operators $\{F_{x_1}\}, \{G_{x_2}\}, \{H_{z_1}\}, \{K_{z_2}\}$ which sum up to the identity,*

$$\begin{aligned}\sum_{\substack{x_1,x_2,\\ z_1,z_2,y}} \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right]\left[\delta_{\mathrm{Ext}(z_1,z_2)=y} - \frac{1}{M}\right] \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right] \\ = \sum_{\substack{x_1,x_2,z_1,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(z_1,z_2)}} \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right] - \frac{1}{M}\end{aligned} \tag{9}$$

The quantity in Equation (9) can be seen as a simple distinguishing strategy of a distinguisher trying to distinguish the output of the extractor from uniform given classical side information. We can therefore relate it to the error of the extractor in the case of classical side information, i.e., to Equation (1). This is shown in the following lemma.

▶ **Lemma 14.** *For $i \in \{1,2\}$ let Z_i denote the classical side information about the source X_i such that $\mathrm{p}(x_1,x_2,z_1,z_2) = \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right]$. Then*

$$\sum_{\substack{x_1,x_2,z_1,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(z_1,z_2)}} \mathrm{p}(x_1,x_2,z_1,z_2) - \frac{1}{M} \leq \frac{1}{2}\|\mathrm{Ext}\left(X_1,X_2\right) Z_1Z_2 - U_m \circ Z_1Z_2\|\,.$$

Proof. Define the following random variables over $\{0,1\}^m \times \{0,1\}^{n_1} \times \{0,1\}^{n_2}$:

$$\mathrm{R} = \mathrm{Ext}(X_1,X_2)Z_1Z_2 \quad ; \quad \mathrm{Q} = U_m \circ Z_1Z_2\,.$$

Let $\mathcal{A}^\star = \left\{(a_1,a_2,a_3)\middle| a_1 = \mathrm{Ext}\left(a_2,a_3\right)\right\} \subseteq \{0,1\}^m \times \{0,1\}^{n_1} \times \{0,1\}^{n_2}$. Then, the probabilities that R and Q assign to the event $\mathcal{A}^\star$ are

$$\mathrm{R}(\mathcal{A}^\star) = \sum_{\substack{x_1,x_2,z_1,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(z_1,z_2)}} \mathrm{p}(x_1,x_2,z_1,z_2) \quad ; \quad \mathrm{Q}(\mathcal{A}^\star) = \frac{1}{M}$$

Using the definition of the variational distance we therefore have

$$\begin{aligned}\frac{1}{2}\|\mathrm{Ext}(X_1,X_2)Z_1Z_2 - U_m \circ Z_1Z_2\| &= \sup_{\mathcal{A}} \|\mathrm{R}(\mathcal{A}) - \mathrm{Q}(\mathcal{A})\| \\ &\geq \mathrm{R}(\mathcal{A}^\star) - \mathrm{Q}(\mathcal{A}^\star) \\ &= \sum_{\substack{x_1,x_2,z_1,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(z_1,z_2)}} \mathrm{p}(x_1,x_2,z_1,z_2) - \frac{1}{M}\,.\end{aligned}$$ ◀

Finally, we combine the lemmas together to show that any weak classical-proof two-source extractor in the Markov model is secure against product quantum side information as well.

▶ **Lemma 15.** *Any (k_1, k_2, ε) classical-proof two-source extractor in the Markov model is a $\left(k_1, k_2, \sqrt{\varepsilon \cdot 2^{(m-2)}}\right)$ quantum-proof product two-source extractor, where m is the output length of the extractor.*

Proof. For any state of two classical sources and product side information $\rho_{X_1X_2C} = \rho_{X_1C_1} \otimes \rho_{X_2C_2}$ with $H_{\min}(X_1|C) \geq k_1$ and $H_{\min}(X_2|C) \geq k_2$, let ρ_{ABC} and $\{F_{x_1}\}, \{G_{x_2}\}$ be the state and measurements satisfying Equation (7).

We can now apply Lemmas 12, 13, and 14 to get the bound

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\| \leq \sqrt{\frac{M}{2}\|\mathrm{Ext}\,(X_1, X_2)\, Z_1Z_2 - U_m \circ Z_1Z_2\|}\,, \tag{10}$$

where Z_1, Z_2 are defined via $\mathrm{p}(x_1, x_2, z_1, z_2) = \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right]$, for $\Psi_{ABC_1C_2}$ which is constructed in the proof of Lemma 12.

As $\Psi_{ABC_1C_2} = \Psi_{AC_1} \otimes \Psi_{BC_2}$ and the measurements are all in tensor product we have $\mathrm{p}(x_1, x_2, z_1, z_2) = \mathrm{p}(x_1, z_1) \cdot \mathrm{p}(x_2, z_2)$, which implies:

1. The sources and the classical side information form a Markov chain $X_1 \leftrightarrow Z_1Z_2 \leftrightarrow X_2$.
2. $H_{\min}\left(X_i|Z_1Z_2\right) = H_{\min}\left(X_i|Z_i\right) \geq H_{\min}\left(X_i|C_i\right)$ for $i \in \{1, 2\}$.

Hence, if $H_{\min}\left(X_i|C_i\right) \geq k_i$ then by the definition of a classical-proof two-source extractor,

$$\frac{1}{2}\|\mathrm{Ext}\,(X_1, X_2)\, Z_1Z_2 - U_m \circ Z_1Z_2\| \leq \varepsilon\,. \tag{11}$$

Combining Equations (10) and (11) we get

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\| \leq \frac{1}{2}\sqrt{M\varepsilon} = \sqrt{\varepsilon 2^{(m-2)}}\,. \qquad \blacktriangleleft$$

By combining Lemma 6 together with Lemma 15 (Lemma 36 in Appendix E) for the weak (strong) case we get that any weak (strong) two-source extractor is also secure against product quantum side information. The bound given in Corollary 16 matches exactly the bound given in [16] for the special case of $m = 1$ (see [16, Corollary 27]).

▶ **Corollary 16.** *Any (k_1, k_2, ε)-[strong] two-source extractor is a $(k_1 + \log\frac{1}{\varepsilon}, k_2 + \log\frac{1}{\varepsilon}, \varepsilon')$-[strong] quantum-proof product two-source extractor with $\varepsilon' = \sqrt{3\varepsilon 2^{(m-2)}}$, where m is the output length of the extractor.*

4.2 Extending to the Markov model

We now extend the result of Section 4.1 to the case of the more general Markov model. To do so, we first recall that by the result of [14], Markov states (according to Definition 7) can also be written in the form

$$\rho_{A_1A_2C} = \bigoplus_t \mathrm{p}(t)\rho^t_{A_1C_1^t} \otimes \rho^t_{A_2C_2^t}\,, \tag{12}$$

where the index t runs over a finite alphabet T, $\mathrm{p}(t)$ is a probability distribution on that alphabet, $\mathcal{H}_C = \bigoplus_t \mathcal{H}_{C_1^t} \otimes \mathcal{H}_{C_2^t}$ is the Hilbert space of C, and $\rho^t_{A_iC_i^t}$ denote states on $A_iC_i^t$, $i \in \{1, 2\}$.

Proof of Lemma 9. Let $\rho_{X_1X_2C}$ be a Markov state such that $H_{\min}(X_i|C) \geq k_i + \log\frac{1}{\varepsilon}$. We first deal with the case of weak extractors. Using the decomposition from Equation (12) we can reduce the problem to the product case by writing

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\rho_C\| = \sum_t \mathrm{p}(t)\|\mathrm{Ext}\otimes\mathbb{1}_C\left(\rho^t_{X_1C_1^t}\otimes\rho^t_{X_2C_2^t}\right) - \rho_{U_m}\otimes\rho^t_{C_1^t}\otimes\rho^t_{C_2^t}\|\,.$$

From Equation (10) we thus have

$$\begin{aligned}&\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\rho_C\|\\ &\qquad\leq \sum_t \mathrm{p}(t)\sqrt{\frac{M}{2}\|\mathrm{Ext}(X_1,X_2)Z_1Z_2|T=t - U_m\circ Z_1Z_2|T=t\|}\\ &\qquad\leq \sqrt{\frac{M}{2}\|\mathrm{Ext}(X_1,X_2)Z_1Z_2T - U_m\circ Z_1Z_2T\|}\,,\end{aligned}$$

where in the last line we used Jensen's inequality and Z_1, Z_2 are defined via

$$\mathrm{p}(x_1,x_2,z_1,z_2|t) = \mathrm{Tr}\left[\rho^t_{ABC}F^t_{x_1}\otimes G^t_{x_2}\otimes H^t_{z_1}\otimes K^t_{z_2}\right]\,.$$

That is, Z_1 and Z_2 are derived from C in the following way: from Equation (12) the states $\{\rho^t_{C_1^t}\otimes\rho^t_{C_2^t}\}_t$ are orthogonal, hence there exists an isometry $C\to CT$ which maps $\sum_t \mathrm{p}(t)\rho^t_{C_1^t}\otimes\rho^t_{C_2^t}$ to $\sum_t \mathrm{p}(t)\rho^t_{C_1^t}\otimes\rho^t_{C_2^t}\otimes|t\rangle\langle t|$. The state $\rho^t_{C_1^t}\otimes\rho^t_{C_2^t}$ is then measured in the same way as in Lemma 15 for the product case to get the side information $Z_1Z_2|T$. Hence, the structure $X_1\leftrightarrow Z_1Z_2T\leftrightarrow X_2$ is conserved. Furthermore, we also have $H_{\min}(X_i|Z_1Z_2T)\geq k_i+\log\frac{1}{\varepsilon}$. Using these two conditions, the problem has been reduced to one with classical side information in the Markov model. Using the fact that Ext is a (k_1,k_2,ε) two-source extractor and applying Lemma 6 we conclude the proof for weak extractors.

Similarly, for strong extractors, from Equation (16) we have

$$\begin{aligned}&\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m}\otimes\rho_{X_1C}\|\\ &\qquad\leq \sum_t \mathrm{p}(t)\sqrt{\frac{M}{2}\|\mathrm{Ext}(X_1,X_2)X_1Z_2|T=t - U_m\circ X_1Z_2|T=t\|}\\ &\qquad\leq \sqrt{\frac{M}{2}\|\mathrm{Ext}(X_1,X_2)X_1Z_2T - U_m\circ X_1Z_2T\|}\,.\end{aligned}$$

Again, we can see this as a measurement made on C such that the value of T is measured and then a further measurements of C_2^t is done in the same way as for the product case to get the side information about X_2 (while there is no additional side information about X_1). Hence, as in the weak case, $X_1\leftrightarrow Z_2T\leftrightarrow X_2$ and $H_{\min}(X_i|Z_2T)\geq k_i+\log\frac{1}{\varepsilon}$, so the problem has been reduced to the classical case. ◀

In the case of l sources, a state $\rho_{X_{[l]}C}$ that satisfies the Markov model (Definition 7) can be written as

$$\rho_{X_{[l]}C} = \bigoplus_t \mathrm{p}(t)\rho^t_{X_1C_1^t}\otimes\cdots\otimes\rho^t_{X_lC_l^t}\,. \tag{13}$$

We provide a proof of this in Appendix D as Lemma 32. It follows from Equation (13) that Lemma 9 can be easily generalised to l sources.

5 Conclusions and open questions

In this work a new and natural model for classical and quantum-proof multi-source extractors was defined – the Markov model. We then showed that *all* multi-source extractors, weak and strong, are also secure in the presence of side information that falls into the Markov model, both in the classical and quantum case. As explained in the previous sections, our main result, Theorem 2, can be seen as a continuation, extension and improvement of previously known results [16, 6].

Apart from the result itself, on the technical level, a new proof technique was used, which, in contrast to the previous works is indifferent to whether the extractors are strong or not. In particular this implies that no adaptations are required for any new multi-source extractor that might be proposed in the future.

We finish this work with several open questions:

1. Are there more general models that extend the Markov model in which all extractors remain secure? Some natural extensions are discussed in Appendix A.
2. Are there different families of states $\rho_{X_1X_2C}$ from which it is possible to extract randomness that are relevant for practical applications? The difficulty in extracting randomness comes from the fact that we are not given one (known) state $\rho_{X_1X_2C}$, but that the extractor is expected to work for any state in a given family, e.g., a Markov state with lower bounds on the conditional min-entropy. The standard criterion of independence between the sources X_1 and X_2 has been relaxed in this work to allow for sources that are independent conditioned on C. Other structures might also allow randomness to be extracted.
3. What happens if the sources and the side information are not exactly in the Markov model but only close to it? Even in the case of only two sources, there are different ways to quantify the closeness of a state to a Markov-chain state (see, e.g., [10]). It is interesting to ask which notion of approximation is relevant in applications of multi-source extractors (such as quantum randomness amplification) and under which such notions the quantum-proof extractors remain secure. Note that the recovery map notion of approximation of Markov chains [10] does not guarantee approximate conditional independence of the sources, and seems to provide quite a different structure.
4. It is unclear whether Theorem 2 is tight, i.e., whether the loss in the error of the extractor is inevitable when considering arbitrary extractors. In other words, it is not known if there are multi-source extractors for which the $\sqrt{2^m}$ loss in the error term is necessary[8]. In the other direction, as noted in Appendix B.1, the work of [13] can be used, in combination with our proof technique, to show that for two-universal hashing (when the seed is taken to be the second source) the blow-up in the error term is not necessary.
5. Do multi-source extractors remain secure also in the presence of non-signalling side information? Non-signalling adversaries are in general more powerful than quantum ones. For seeded extractors this does not seem to be the case [12, 1] but for multi-source extractors nothing is known. Note however that our proof technique is not applicable to non-signalling side information.

Acknowledgments. We thank Mario Berta, Omar Fawzi and Thomas Vidick for helpful comments and discussions.

[8] The same question arises in the case of seeded extractors as well [2].

References

1 Rotem Arnon-Friedman and Amnon Ta-Shma. Limits of privacy amplification against nonsignaling memory attacks. *Physical Review A*, 86(6):062333, 2012.

2 Mario Berta, Omar Fawzi, and Volkher B Scholz. Quantum-proof randomness extractors via operator space theory. *arXiv preprint arXiv:1409.3563*, 2014.

3 Mario Berta, Omar Fawzi, and Volkher B. Scholz. Quantum bilinear optimization. *SIAM Journal for Optimization*, 2016. To appear. `arXiv:1506.08810`.

4 Fernando GSL Brandão, Ravishankar Ramanathan, Andrzej Grudka, Karol Horodecki, Michał Horodecki, and Paweł Horodecki. Robust device-independent randomness amplification with few devices. *Nature Communications*, 7:11345, 2016.

5 Eshan Chattopadhyay and David Zuckerman. Explicit two-source extractors and resilient functions. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 22, page 119, 2015.

6 Kai-Min Chung, Xin Li, and Xiaodi Wu. Multi-source randomness extractors against quantum side information, and their applications. *arXiv preprint arXiv:1411.2315*, 2014.

7 Kai-Min Chung, Yaoyun Shi, and Xiaodi Wu. Physical randomness extractors: Generating random numbers with minimal assumptions. *arXiv preprint arXiv:1402.4797*, 2014.

8 Anindya De, Christopher Portmann, Thomas Vidick, and Renato Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41(4):915–940, 2012.

9 Yevgeniy Dodis, Ariel Elbaz, Roberto Oliveira, and Ran Raz. Improved randomness extraction from two independent sources. In *Approximation, randomization, and combinatorial optimization. Algorithms and techniques*, pages 334–344. Springer, 2004.

10 Omar Fawzi and Renato Renner. Quantum conditional mutual information and approximate markov chains. *Communications in Mathematical Physics*, 340(2):575–611, 2015.

11 Dmitry Gavinsky, Julia Kempe, Iordanis Kerenidis, Ran Raz, and Ronald De Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 516–525. ACM, 2007.

12 Esther Hänggi, Renato Renner, and Stefan Wolf. The impossibility of non-signaling privacy amplification. *arXiv preprint arXiv:0906.4760*, 2009.

13 Masahito Hayashi and Toyohiro Tsurumaru. More efficient privacy amplification with less random seeds. In *Information Theory (ISIT), 2015 IEEE International Symposium on*, pages 1786–1790. IEEE, 2015.

14 Patrick Hayden, Richard Jozsa, Denes Petz, and Andreas Winter. Structure of states which satisfy strong subadditivity of quantum entropy with equality. *Communications in mathematical physics*, 246(2):359–374, 2004.

15 Russell Impagliazzo, Leonid A Levin, and Michael Luby. Pseudo-random generation from one-way functions. In *Proceedings of the twenty-first annual ACM symposium on Theory of computing*, pages 12–24. ACM, 1989.

16 Roy Kasher and Julia Kempe. *Two-source extractors secure against quantum adversaries*, volume 6302 of *Lecture Notes in Computer Science*, pages 656–669. Springer, 2010.

17 Robert T Konig and Barbara M Terhal. The bounded-storage model in the presence of a quantum adversary. *Information Theory, IEEE Transactions on*, 54(2):749–762, 2008.

18 Xin Li. Improved constructions of two-source extractors. In *Proceedings of the 2011 IEEE 26th Annual Conference on Computational Complexity (CCC)*, 2015. arXiv preprint arXiv:1508.01115.

19 Xin Li. Personal communication, 2015.

20 Xin Li. Three-source extractors for polylogarithmic min-entropy. In *Proceedings of the 2015 IEEE 56th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 863–882, 2015. arXiv preprint arXiv:1503.02286.

21 Wolfgang Mauerer, Christopher Portmann, and Volkher B. Scholz. A modular framework for randomness extraction based on trevisan's construction. *arXiv preprint arXiv:1212.0520*, 2012.

22 Piotr Mironowicz, Rodrigo Gallego, and Marcin Pawlowski. Amplification of arbitrarily weak randomness. *Physical Review A*, 91(032317), 2015. arXiv preprint arXiv:1301.7722.

23 Michael A Nielsen and Isaac L Chuang. *Quantum computation and quantum information*. Cambridge university press, 2010.

24 Martin Plesch and Matej Pivoluska. Device-independent randomness amplification with a single device. *Physics Letters A*, 378(40):2938–2944, 2014.

25 Jaikumar Radhakrishnan and Amnon Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13(1):2–24, 2000. doi:10.1137/S0895480197329508.

26 Anup Rao. Extractors for a constant number of polynomially small min-entropy independent sources. *SIAM Journal on Computing*, 39(1):168–194, 2009.

27 Ran Raz. Extractors with weak random seeds. In *Proceedings of the 37th Symposium on Theory of Computing, STOC '05*, pages 11–20. ACM, 2005. doi:10.1145/1060590.1060593.

28 Ben W Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, 2013.

29 Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.

30 Renato Renner and Robert König. Universally composable privacy amplification against quantum adversaries. In *Theory of Cryptography*, pages 407–425. Springer, 2005.

31 Miklos Santha and Umesh V Vazirani. Generating quasi-random sequences from semi-random sources. *Journal of Computer and System Sciences*, 33(1):75–87, 1986.

32 Valerio Scarani. The device-independent outlook on quantum physics (lecture notes on the power of Bell's theorem). *arXiv preprint arXiv:1303.3081*, 2013.

33 Ronen Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77(67-95):10, 2002.

34 Marco Tomamichel, Roger Colbeck, and Renato Renner. Duality between smooth min- and max-entropies. *IEEE Transactions on Information Theory*, 56(9):4674–4681, 2010. arXiv:arXiv:0907.5238, doi:10.1109/TIT.2010.2054130.

35 Marco Tomamichel, Charles Ci Wen Lim, Nicolas Gisin, and Renato Renner. Tight finite-key analysis for quantum cryptography. *Nature Communications*, 3:634, 2012. arXiv:1103.4130, doi:10.1038/ncomms1631.

36 Marco Tomamichel, Christian Schaffner, Adam Smith, and Renato Renner. Leftover hashing against quantum side information. *IEEE Transactions on Information Theory*, 57(8):5524–5535, August 2011. A preliminary version appeared at ISIT 2010. arXiv:arXiv:1002.2436, doi:10.1109/TIT.2011.2158473.

37 Luca Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001.

38 Umesh V Vazirani. Strong communication complexity or generating quasi-random sequences from two communicating semi-random sources. *Combinatorica*, 7(4):375–392, 1987.

39 Andreas Winter. "Extrinsic" and "intrinsic" data in quantum measurements: Asymptotic convex decomposition of positive operator valued measures. *Communications in mathematical physics*, 244(1):157–185, 2004.

40 Tzyh Haur Yang, Tamás Vértesi, Jean-Daniel Bancal, Valerio Scarani, and Miguel Navascués. Robust and versatile black-box certification of quantum devices. *Physical review letters*, 113(4):040401, 2014.

A Extending the set of extractable sources

Although the definition of a quantum-proof two-source extractors (Definition 8) requires the source $\rho_{X_1EX_2}$ to be a Markov chain with a bound on the min-entropy, a function proven be such an extractor can also be used to extract randomness from a larger set of sources, e.g., if the adversary were to destroy her side information E, this would not hinder extraction, yet it could destroy the Markov chain property of the source. In this section we consider two extensions of the multi-source extractor definition for which all multi-source extractors in the Markov model can be used. In Appendix A.1 we show that it is not necessary to have a bound on the min-entropy, it is sufficient to bound the smooth min-entropy of the sources X_1 and X_2. Then in Appendix A.2 we show that one can also extract from any source obtained by deleting information from a Markov source, even though the resulting state might not be a Markov chain any longer. The multi-source extractor model for strong extractors from [6] falls in this category.

A.1 Smooth min-entropy

It is standard for the extractor definitions to require a bound on the min-entropy of the source conditioned on the side information, i.e., $H_{\min}(X_i|C) \geq k_i$. In practical situations, however, one often only has a bound on the *smooth* min-entropy – this is defined by maximising the min-entropy over all states δ-close, see Equation (14) below. For example, in quantum key distribution a bound on the smooth min-entropy is obtained by sampling the noise on the quantum channel [35]. In this section we prove that any quantum-proof two-source extractor can be used in a context where only a bound on the smooth min-entropy is known.

The smooth conditional min-entropy with smoothness parameter δ of a state ρ_{XC} is defined as follows.

$$H^{\delta}_{\min}(X|C)_{\rho} = \max_{\sigma \in \mathcal{B}^{\delta}(\rho)} H_{\min}(X|C)_{\sigma}, \tag{14}$$

where $\mathcal{B}^{\delta}(\rho)$ is a ball of radius δ around ρ_{XC}. This ball is defined as the set of *subnormalized* states σ with $P(\rho, \sigma) \leq \delta$, where $P(\cdot, \cdot)$ is the *purified distance* [34]. The exact definition of the purified distance is not needed in this paper, so we omit it for simplicity and refer the interested reader to [34]. The only property of the purified distance that we need in this work is that for any (subnormalized) ρ and σ,

$$P(\rho, \sigma) \geq \frac{1}{2}\|\rho - \sigma\|.$$

This means that if $H^{\delta}_{\min}(X|C)_{\rho} \geq k$, then there exists a subnormalized σ_{XC} such that $\frac{1}{2}\|\rho - \sigma\| \leq \delta$ and $H_{\min}(X|C)_{\sigma} \geq k$.

We can now state our main lemma. This can be generalised to the multi-source case in a straightforward manner.

▶ **Lemma 17.** *Let* $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *be a* $(k_1 - \log 1/\varepsilon_1 - 1, k_2 - \log 1/\varepsilon_2 - 1, \varepsilon)$ *quantum-proof two-source extractor in the Markov model. Then for any Markov state* $\rho_{X_1X_2C}$ *with* $H^{\delta_1}_{min}(X_1|C)_{\rho} \geq k_1$ *and* $H^{\delta_2}_{min}(X_2|C)_{\rho} \geq k_2$,

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\| \leq 6\delta_1 + 6\delta_2 + 2\varepsilon_1 + 2\varepsilon_2 + 2\varepsilon$$

if the extractor is weak, and

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)X_iC} - \rho_{U_m}\otimes\rho_{X_iC}\| \leq 6\delta_1 + 6\delta_2 + 2\varepsilon_1 + 2\varepsilon_2 + 2\varepsilon$$

if the extractor is strong in the source X_i.

To prove that Lemma 17 holds, we first need to prove that if a state $\rho_{X_1X_2C}$ is guaranteed to be a Markov state with bounded smooth min-entropy, then there is a (subnormalized) state $\sigma_{X_1X_2C}$ close by which is also a Markov state with a bound on the min-entropy. This can be seen as a robustness property of the Markov model for extractors.

▶ **Lemma 18.** *Let $\rho_{X_1X_2C}$ be a Markov state $X_1 \leftrightarrow C \leftrightarrow X_2$ such that $H_{min}^{\delta_1}(X_1|C)_\rho \geq k_1$ and $H_{min}^{\delta_2}(X_2|C)_\rho \geq k_2$. Then there exists a subnormalized state $\sigma_{X_1X_2C}$ such that X_1, X_2 and C still form a Markov chain $X_1 \leftrightarrow C \leftrightarrow X_2$, and $H_{min}(X_1|C)_\sigma \geq k_1 - \log\frac{1}{\varepsilon_1}$, $H_{min}(X_2|C)_\sigma \geq k_2 - \log\frac{1}{\varepsilon_2}$ and $\frac{1}{2}\|\rho-\sigma\| \leq \varepsilon_1+\varepsilon_2+3\delta_1+3\delta_2$.*

Proof. By the Markov chain condition, the state $\rho_{X_1X_2C}$ can equivalently be written

$$\rho_{X_1C_1ZE_2X_2} = \sum_{x_1,x_2,z} p(z)p(x_1|z)p(x_2|z)|x_1\rangle\langle x_1|\otimes\rho_{C_1}^{x_1,z}\otimes|z\rangle\langle z|\otimes\rho_{C_2}^{x_2,z}\otimes|x_2\rangle\langle x_2|.$$

Thus $H_{\min}^{\delta_1}(X_1|C)_\rho = H_{\min}^{\delta_1}(X_1|C_1Z)_\rho$ and $H_{\min}^{\delta_2}(X_2|C)_\rho = H_{\min}^{\delta_2}(X_2|C_2Z)_\rho$. In the following we use only this form with the explicit classical register Z.

By the definition of smooth min-entropy, we know that there exist (subnormalized) states

$$\tilde{\sigma}_{X_1C_1Z} = \sum_{x_1,z} q_1(z)q(x_1|z)|x_1\rangle\langle x_1|\otimes\sigma_{C_1}^{x_1,z}\otimes|z\rangle\langle z|$$

and

$$\hat{\sigma}_{X_2C_2Z} = \sum_{x_2,z} q_2(z)q(x_2|z)|x_2\rangle\langle x_2|\otimes\sigma_{C_2}^{x_2,z}\otimes|z\rangle\langle z|$$

such that $\frac{1}{2}\|\rho_{X_1C_1Z}-\tilde{\sigma}_{X_1C_1Z}\| \leq \delta_1$, $\frac{1}{2}\|\rho_{X_2C_2Z}-\hat{\sigma}_{X_2C_2Z}\| \leq \delta_2$, $H_{\min}(X_1|C_1Z)_{\tilde{\sigma}} \geq k_1$ and $H_{\min}(X_2|C_2Z)_{\hat{\sigma}} \geq k_2$.

Since $2^{-H_{\min}(X_1|CZ)_\sigma} = \sum_z q(z)2^{-H_{\min}(X_1|CZ=z)_\sigma}$ also for subnormalized distributions $q(\cdot)$, we can define $2^{-H_{\min}(X_1|CZ=z)_\sigma} := 0$ when $q(z)=0$, then pad $q(\cdot)$ to get a normalized distribution for which $2^{-H_{\min}(X_1|CZ)_\sigma} = \mathbb{E}_z\left[2^{-H_{\min}(X_1|CZ=z)_\sigma}\right]$. We can thus use Markov's inequality and get

$$\Pr_{z\leftarrow Z}\left[H_{\min}(X_1|C_1Z=z)_{\tilde{\sigma}} \leq k_1 - \log\frac{1}{\varepsilon_1}\right] \leq \varepsilon_1$$

and

$$\Pr_{z\leftarrow Z}\left[H_{\min}(X_2|C_2Z=z)_{\hat{\sigma}} \leq k_2 - \log\frac{1}{\varepsilon_2}\right] \leq \varepsilon_2\ .$$

Let $\mathcal{Z}_1$ and $\mathcal{Z}_2$ be the sets of values for which $q_1(z_1)\neq 0$, $q_2(z_2)\neq 0$, and

$$\forall z_1\in\mathcal{Z}_1,\quad H_{\min}(X_1|C_1Z=z_1)_{\tilde{\sigma}} \geq k_1 - \log\frac{1}{\varepsilon_1}$$

and

$$\forall z_2\in\mathcal{Z}_2,\quad H_{\min}(X_2|C_2Z=z_2)_{\hat{\sigma}} \geq k_2 - \log\frac{1}{\varepsilon_2}\ .$$

Let $\bar{\mathcal{Z}} := \mathcal{Z}_1\cap\mathcal{Z}_2$ be their intersection, and let $\bar{p}(z)$ be a subnormalized distribution given by

$$\bar{p}(z) := \begin{cases} p(z) & \text{if } z\in\bar{\mathcal{Z}},\\ 0 & \text{otherwise.}\end{cases}$$

We define the (subnormalized) state

$$\sigma_{X_1C_1ZC_2X_2} := \sum_{x,y,z} \bar{p}(z)q(x_1|z)q(x_2|z)|x_1\rangle\langle x_1| \otimes \sigma_{C_1}^{x_1,z} \otimes |z\rangle\langle z| \otimes \sigma_{C_2}^{x_2,z} \otimes |x_2\rangle\langle x_2|,$$

and prove in the following that it satisfies the conditions of the lemma.

By construction of σ we have $2^{-H_{\min}(X_1|C_1Z)_\sigma} = \sum_z \bar{p}(z)2^{-H_{\min}(X_1|C_1Z=z)_\sigma}$ for values z such that $H_{\min}(X_1|C_1Z=z)_\sigma \geq k_1 - \log\frac{1}{\varepsilon_1}$. Hence $H_{\min}(X_1|C_1Z)_\sigma \geq k_1 - \log\frac{1}{\varepsilon_1}$ and similarly $H_{\min}(X_2|C_2Z)_\sigma \geq k_2 - \log\frac{1}{\varepsilon_2}$.

To bound the distance from $\rho_{X_1C_1ZC_2X_2}$, first note that

$$\frac{1}{2}\sum_z |\bar{p}(z) - p(z)| \leq \varepsilon_1 + \varepsilon_2 + \delta_1 + \delta_2.$$

We also have

$$\begin{aligned}
&\sum_{x_1,z} \frac{p(z)}{2}\|p(x_1|z)\rho_{C_1}^{x_1,z} - q(x_1|z)\sigma_{C_1}^{x_1,z}\| \leq \\
&\sum_{x_1,z} \frac{1}{2}\|p(z)p(x_1|z)\rho_{C_1}^{x_1,z} - q_1(z)q(x_1|z)\sigma_{C_1}^{x_1,z}\| + \frac{1}{2}\|q_1(z)q(x_1|z)\sigma_{C_1}^{x_1,z} - p(z)q(x_1|z)\sigma_{C_1}^{x_1,z}\| \\
&\qquad\qquad \leq 2\delta_1.
\end{aligned}$$

The same holds for X_2C_2Z, namely

$$\sum_{x_2,z} \frac{p(z)}{2}\|p(x_2|z)\rho_{C_2}^{x_2,z} - q(x_2|z)\sigma_{C_2}^{x_2,z}\| \leq 2\delta_2.$$

Putting this together we get

$$\begin{aligned}
&\frac{1}{2}\|\rho_{X_1C_1ZC_2X_2} - \sigma_{X_1C_1ZC_2X_2}\| \\
&\quad= \sum_{x,y,z} \frac{1}{2}\|p(z)p(x_1|z)p(x_2|z)\rho_{C_1}^{x_1,z} \otimes \rho_{C_2}^{x_2,z} - \bar{p}(z)q(x_1|z)q(x_2|z)\sigma_{C_1}^{x_1,z} \otimes \sigma_{C_2}^{x_2,z}\| \\
&\quad= \sum_{x,y,z} \frac{p(z)}{2}\|p(x_1|z)p(x_2|z)\rho_{C_1}^{x_1,z} \otimes \rho_{C_2}^{x_2,z} - q(x_1|z)q(x_2|z)\sigma_{C_1}^{x_1,z} \otimes \sigma_{C_2}^{x_2,z}\| \\
&\qquad\qquad + \varepsilon_1 + \varepsilon_2 + \delta_1 + \delta_2 \\
&\quad\leq \sum_{x,y,z} \frac{p(z)}{2}\|p(x_1|z)p(x_2|z)\rho_{C_1}^{x_1,z} \otimes \rho_{C_2}^{x_2,z} - q(x_1|z)p(x_2|z)\sigma_{C_1}^{x_1,z} \otimes \rho_{C_2}^{x_2,z}\| \\
&\qquad\qquad + \frac{p(z)}{2}\|q(x_1|z)p(x_2|z)\sigma_{C_1}^{x_1,z} \otimes \rho_{C_2}^{x_2,z} - q(x_1|z)q(x_2|z)\sigma_{C_1}^{x_1,z} \otimes \sigma_{C_2}^{x_2,z}\| \\
&\qquad\qquad + \varepsilon_1 + \varepsilon_2 + \delta_1 + \delta_2 \\
&\quad= \sum_{x_1,z} \frac{p(z)}{2}\|p(x_1|z)\rho_{C_1}^{x_1,z} - q(x_1|z)\sigma_{C_1}^{x_1,z}\| \\
&\qquad\qquad + \sum_{x_2,z} \frac{p(z)}{2}\|p(x_2|z)\rho_{C_2}^{x_2,z} - q(x_2|z)\sigma_{C_2}^{x_2,z}\| + \varepsilon_1 + \varepsilon_2 + \delta_1 + \delta_2 \\
&\quad\leq \varepsilon_1 + \varepsilon_2 + 3\delta_1 + 3\delta_2.
\end{aligned}$$

◀

Since Lemma 18 finds a subnormalized state that is close, the next step is to prove that one can extract from subnormalized states. This is done in Appendix F in Lemma 37. Combining this with a simple use of the triangle inequality allows us to prove Lemma 17.

Proof of Lemma 17. We prove the case of a weak extractor Ext. The proof for a strong extractor is identical.

By Lemma 18 there exists a subnormalized Markov state $\sigma_{X_1X_2C}$ such that $\frac{1}{2}\|\rho_{X_1X_2C} - \sigma_{X_1X_2C}\| \leq \varepsilon_1 + \varepsilon_2 + 3\delta_1 + 3\delta_2$ and $H_{\min}(X_i|C)_\sigma \geq k_i - \log 1/\varepsilon_i$. Hence

$$\begin{aligned}
&\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\rho_C\| \\
&\qquad\leq \frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \sigma_{\mathrm{Ext}(X_1,X_2)C}\| + \frac{1}{2}\|\sigma_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\sigma_C\| \\
&\qquad\qquad + \frac{1}{2}\|\rho_{U_m}\otimes\sigma_C - \rho_{U_m}\otimes\rho_C\| \\
&\qquad\leq 2\varepsilon_1 + 2\varepsilon_2 + 6\delta_1 + 6\delta_2 + \frac{1}{2}\|\sigma_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\sigma_C\| \\
&\qquad\leq 2\varepsilon_1 + 2\varepsilon_2 + 6\delta_1 + 6\delta_2 + 2\varepsilon\,,
\end{aligned}$$

where in the last line we used Lemma 37. ◀

A.2 Non-Markov sources

It is trivial to show that if part of the side information E is deleted, this cannot decrease the security of an extractor. As already observed in [6], in the case of an extractor that is strong in the source X_i, any operation on E conditioned on X_i cannot help an adversary either. Intuitively, this holds because the adversary is given the entire source X_i, thus copying information about it to E is pointless. We formalize this in the following lemma.

▶ **Lemma 19.** *Let $\rho_{X_1EX_2}$ be a Markov source with $H_{min}(X_i|E)_\rho \geq k_i$. Let $\mathcal{E}:\mathrm{L}(E)\to\mathrm{L}(E)$ be any CPTP map on E. If* Ext *is a (k_1,k_2,ε) quantum-proof two-source extractor, then it can be used to extract from $\sigma_{X_1EX_2} = \mathcal{E}(\rho_{X_1EX_2})$ with error ε. Let $\mathcal{E}:\mathrm{L}(X_iE)\to\mathrm{L}(X_iE)$ be a CPTP map that leaves X_i unmodified, i.e., $\mathcal{E}(\sum_x p_x|x\rangle\langle x|\otimes\rho_E^x) = \sum_x p_x|x\rangle\langle x|\otimes\mathcal{E}_x(\rho_E^x)$ for some set of CPTP maps $\mathcal{E}_x:\mathcal{L}(\mathcal{H}_E)\to\mathcal{L}(\mathcal{H}_E)$. If* Ext *is a (k_1,k_2,ε) quantum-proof two-source extractor strong in X_i, then it can be used to extract from $\sigma_{X_1EX_2} = \mathcal{E}(\rho_{X_1EX}2)$ with error ε.*

Proof. We prove the case of the strong extractor. The proof for the weak extractor follows the same steps. We need to show that

$$\frac{1}{2}\|\sigma_{\mathrm{Ext(X_1,X_2)X_iE}} - \rho_{U_m}\otimes\sigma_{X_iE}\| \leq \varepsilon\,.$$

This follows from the contractivity of the trace distance and because the maps Ext and $\mathcal{E}$ commute:

$$\begin{aligned}
\frac{1}{2}\|\sigma_{\mathrm{Ext(X_1,X_2)X_iE}} - \rho_{U_m}\otimes\sigma_{X_iE}\| &= \frac{1}{2}\|\mathcal{E}(\rho_{\mathrm{Ext(X_1,X_2)X_iE}}) - \rho_{U_m}\otimes\mathcal{E}(\sigma_{X_iE})\| \\
&\leq \frac{1}{2}\|\rho_{\mathrm{Ext(X_1,X_2)X_iE}} - \rho_{U_m}\otimes\rho_{X_iE}\| \leq \varepsilon\,.
\end{aligned}$$

◀

An equivalent result in [6, Theorem 4.1] allows the authors to prove that their complex information leaking model can be reduced to side information about one of the sources, which implies that a strong extractor in the Markov model is also an extractor in the model of [6]. Note that, as already observed in [6], the entropy of the state $\sigma_{X_1EX_2}$ is not meaningful, since the operation $\mathcal{E}$ might delete information without reducing the capacity to distinguish the output of the extractor from uniform. One has to measure the entropy on the Markov state before $\mathcal{E}$ is applied [6].

B Explicit constructions

In this section we give some examples for explicit constructions of quantum-proof multi-source extractors in the Markov model, as follows from our main theorem, Theorem 2.

In Appendix B.1 we consider a two-source extractor by Dodis et al. [9]. This extractor requires the sum of the entropies in both sources to be larger than n, and we get a construction with nearly identical parameters in the quantum case. In Appendix B.2 we consider a two-source extractor construction by Raz [27], which requires one source to have entropy at least $n/2$, whereas the other can be logarithmic. Here too, the resulting quantum-proof extractor has nearly identical parameters to the classical case. In Appendix B.3 we look at a three source extractor by Li [20], which only requires the sources to have entropy poly-logarithmic in n. Plugging this in our main theorem allows a sublinear amount of entropy to be extracted in the quantum case, and by combining it with Trevisan's extractor [8], we can extract the remaining entropy and thus obtain the same output length as in the classical case. The final construction we analyse in Appendix B.4 is based on a recent two-source extractor by Li [18], which only needs two sources of poly-logarithmic min-entropy. Unfortunately, the error is $n^{-\Omega(1)}$, which means that Theorem 2 only allows $\Omega(\log n)$ bits to be extracted. Composing this with another variant of Trevisan's extractor [8] allows a sublinear amount of randomness to be extracted at the cost of requiring one of the sources to have $k = n^\alpha$ bits of entropy for any constant $\alpha < 1$.

Since the works of Dodis et al. [9] and Raz [27] provide the exact parameters for their extractors, we do the same here below in Appendices B.1 and B.2. In contrast, for the two extractors from [20, 18] the exact parameters are unknown, as only the simplified O-notation form is given in the corresponding papers. For this reason the constructions in Appendices B.3 and B.4 are also given in O-notation.

B.1 High entropy sources

The first extractor we consider is a strong two-source extractor from Dodis et al. [9], which requires both sources together to have at least n bits of entropy.

▶ **Lemma 20** ([9]). *For any $n_1 = n_2 = n$, k_1, k_2 and m there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *that is a (k_1, k_2, ε) two-source extractor, strong in both X_1 and in X_2 (separately), with $\varepsilon = 2^{-(k_1+k_2+1-n-m)/2}$.*

To have an error $\varepsilon < 1$, the total entropy must be $k_1 + k_2 > n - 1$. The difference between $k_1 + k_2$ and $n - 1$ can either be extracted or used to decrease the error. Let $\ell + m = k_1 + k_2 + 1 - n$, then the error is $\varepsilon = 2^{-\ell/2}$.

Plugging Lemma 20 into Theorem 2 we get the following.

▶ **Corollary 21.** *For any $n_1 = n_2 = n$, k_1', k_2' and m there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *that is a $(k_1', k_2', \varepsilon)$ two-source extractor, strong in both sources (separately), with $\varepsilon' = \frac{\sqrt{3}}{2} 2^{-(k_1'+k_2'+1-n-5m)/8}$.*

Proof. From Theorem 2 we have $k_1' = k_1 + \log \frac{1}{\varepsilon}$ and $k_2' = k_2 + \log \frac{1}{\varepsilon}$. Rewriting the error from Lemma 20 in terms of m we get $m = k_1 + k_2 + 1 - n - 2\log \frac{1}{\varepsilon}$. Hence $m = k_1' + k_2' + 1 - n - 4\log \frac{1}{\varepsilon}$, so $\varepsilon = 2^{-(k_1'+k_2'+1-n-m)/4}$. Plugging this in the error from Theorem 2, namely $\varepsilon' = \sqrt{3\varepsilon 2^m}/2$ finishes the proof. ◀

The parameters in the quantum case are very similar to the classical one. We still need $k_1' + k_2' > n - 1$ and the difference can either be extracted or used to decrease the error. But this time for $\ell + \tilde{m} = k_1' + k_2' + 1 - n$ the extractor outputs $m = \tilde{m}/5$ bits with error $2^{-\ell/8}$.

Since the extractor is strong we can compose it with a quantum-proof seeded extractor, e.g., Trevisan's extractor [8], to extractor more randomness from the sources – this procedure is explained in Appendix G. Here we use a variant of Trevisan's extractor with parameters given in Lemma 39 in Appendix G.

▶ **Corollary 22.** *For any $n_1 = n_2 = n$, k_1', k_2', ε', m'', ε'', such that*

$$m = \frac{k_1' + k_2' + 1 - n - 8\log(\sqrt{3}/2\varepsilon')}{5} \geq d\,,$$
$$\max[k_1', k_2'] \geq m'' + 4\log\frac{m''}{\varepsilon''} + 6\,,$$

where d is the seed length needed by the extractor from Lemma 39, there exists an explicit function $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{m+m''}$ *that is a quantum-proof $(k_1', k_2', \varepsilon' + \varepsilon'')$ two-source extractor.*

We remark that the construction of Dodis et al. [9] is based on universal hash functions. These are already known to be good quantum-proof seeded extractors [30, 36, 13]. Recently, Hayashi and Tsurumaru [13] proved that they are also good quantum-proof extractors if the seed is not uniform. Using some of our proof techniques, the result of Hayashi and Tsurumaru can be generalised to obtain a different proof that the construction of Dodis et al. is a two-source extractor in the Markov model. The resulting parameters are better than what we obtain here with the generic reduction from quantum-proof to classical extractors, since the Hayashi-Tsurumaru proof [13] does not have the $\sqrt{2^m}$ factor.

B.2 One high and one logarithmic entropy source

The following construction by Raz [27] improves on Dodis et al. [9]. One of the sources still requires at least $n/2$ bits of entropy, but the other can be logarithmic.

▶ **Lemma 23** ([27, Theorem 1])**.** *For any n_1, n_2, k_1, k_2, m, and any $0 < \delta < 1/2$, such that,*

$$n_1 \geq 6\log n_1 + 2\log n_2,$$
$$k_1 \geq \left(\frac{1}{2} + \delta\right) n_1 + 3\log n_1 + \log n_2,$$
$$k_2 \geq 5\log(n_1 - k_1),$$
$$m \leq \delta \min\left[\frac{n_1}{8}, \frac{k_2}{40}\right] - 1,$$

there exists an explicit function $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *that is a (k_1, k_2, ε)-two-source extractor strong in both inputs (separately) with $\varepsilon = 2^{-3m/2}$.*

Plugging this into Theorem 2 we get the following.

▶ **Corollary 24.** *For any n_1, n_2, k_1', k_2', m, and $0 < \delta' < 19/32$, such that,*

$$n_1 \geq 6\log n_1 + 2\log n_2,$$
$$k_1' \geq \left(\frac{1}{2} + \delta'\right) n_1 + 3\log n_1 + \log n_2,$$
$$k_2' \geq \frac{163}{32}\log\left(\left(1 + \frac{3\delta'}{19}\right) n_1 - k_1'\right),$$
$$m \leq \frac{16\delta'}{19}\min\left[\frac{n_1}{8}, \frac{4k_2'}{163}\right] - 1,$$

there exists an explicit function $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *that is a quantum-proof* $(k_1', k_2', \varepsilon)$*-two-source extractor strong in both inputs (separately) with* $\varepsilon' = \frac{\sqrt{3}}{2} 2^{-m/4}$.

Proof. We need $k_1' \geq k_1 + \log 1/\varepsilon$, so we set

$$k_1' = k_1 + \frac{3}{2}\delta\frac{n_1}{8} \geq \left(\frac{1}{2} + \frac{19\delta}{16}\right) n_1 + 3\log n_1 + \log n_2.$$

We obtain the bound on k_1' given above by setting $\delta' = 19\delta/16$. Similarly, we need $k_2' \geq k_2 + \log 1/\varepsilon$, so we set

$$k_2' = k_2 + \frac{3}{2}\frac{1}{2}\frac{k_2}{40} = \frac{163}{160}k_2 \geq \frac{163}{32}\log(n_1 - k_1).$$

Writing this in terms of k_1' instead of k_1 gives the bound on k_2'. The bound on m is also updated in terms of δ' and k_2'. Finally the new error is given by $\varepsilon' = \sqrt{3\varepsilon 2^m}/2$. ◀

Here too the parameters are very similar to the classical case, only the coefficients change somewhat. As in Appendix B.1, this extractor is strong, hence we can compose it with Lemma 39 as explained in Appendix G.

▶ **Corollary 25.** *For any* n_1, n_2, k_1', k_2', m, *and* $0 < \delta' < 19/32$, *satisfying the constraints from Corollary 24 and any* m'', ε'' *such that*

$$m \geq d(m'', \varepsilon'') \,,$$
$$\max[k_1', k_2'] \geq m'' + 4\log\frac{m''}{\varepsilon''} + 6 \,,$$

where d *– the seed length needed by the extractor from Lemma 39 – is a function of* m'' *and* ε'', *there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{m+m''}$ *that is a quantum-proof* $(k_1', k_2', \frac{\sqrt{3}}{2}2^{-m/4} + \varepsilon'')$ *two-source extractor.*

B.3 Three poly-logarithmic sources

The third extractor we consider can break the barrier of $n/2$ min-entropy – it is sufficient for the sources to have $k = \log^{12} n$ bits of entropy – but requires three sources instead of two.

▶ **Lemma 26** ([20, Theorem 1.5])**.** *For any* n *and* $k \geq \log^{12} n$, *there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *that is a* (k,k,k,ε) *three-source extractor, strong in* X_1 *and in* X_2X_3 *with* $m = 0.9k$ *and* $\varepsilon = 2^{-k^{\Omega(1)}}$.

Since the error of this extractor is not exponential in k, but only in k^c for some $c < 1$, when applying it to a source with quantum side information we cannot extract all of the entropy, but only $k^{c'}$ bits, for any $c' < c$.

▶ **Corollary 27.** *For any* n *and* $k' \geq 2\log^{12} n$, *there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{m'}$ *that is a quantum-proof* $(k', k', k', \varepsilon')$ *three-source extractor, strong in* X_1 *and in* X_2X_3 *with* $m' = k'^{\Omega(1)}$ *and* $\varepsilon' = 2^{-k'^{\Omega(1)}}$.

Proof. Let c be the leading term in $\Omega(1)$ for $\varepsilon = 2^{-k^{\Omega(1)}}$ from Lemma 26. Note that we necessarily have $c < 1$, because otherwise for $k = n$ the error would be $2^{-n+o(n)}$ which is impossible [25]. We thus get $k' = k + \log 1/\varepsilon = k + k^c + o(k^c)$. Requiring $k' \geq 2\log^{12} n$ is sufficient to have $k \geq \log^{12} n$ for large enough k. Picking $m' = k^{c'} = k'^{\Omega(1)}$ for some $c' < c$ implies that $\varepsilon' = \sqrt{4\varepsilon 2^m}/2 = 2^{-k^{\Omega(1)}} = 2^{-k'^{\Omega(1)}}$. ◀

Corollary 27 does not extract as much entropy as Lemma 26, but it extracts enough to use as a seed in Trevisan's construction and extract the entropy of the sources X_2X_3. The parameters below are obtained by composing Corollary 27 with Lemma 40.

▶ **Corollary 28.** *There exists a constant c' such that for any n and $k_3 \geq k_2 \geq k_1 \geq \max\left[2\log^{12} n, \log^{3/c'} n\right]$, there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *that is a quantum-proof $(k_1, k_2, k_3, \varepsilon)$ three-source extractor with $m = k_1^{\Omega(1)} + k_2 + k_3 - o(k_2 + k_3)$ and $\varepsilon = n^{-\Omega(1)}$.*

Proof. The quantum-proof extractor from Lemma 40 requires a seed of length $d = O(\log^3 n)$ for an error $\varepsilon = n^{-\Omega(1)}$. The output length of Corollary 27 is $m' = k_1^{c'} - o(k_1^{c'})$ for some constant c'. Thus, if $k_1^{c'} > \log^3 n$, the output is long enough. ◀

B.4 Two poly-logarithmic sources

In a recent breakthrough Chattopadhyay and Zuckerman constructed a two-source extractor that outputs 1 bit and only requires two sources of poly-logarithmic entropy [5]. This was then generalised to multiple output bits by Li [18].

▶ **Lemma 29** ([18, Theorem 1.3])**.** *There exists a constant c_1 such that for any n and $k \geq \log^{c_1} n$, there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ *that is a (k, k, ε) two-source extractor strong in X_2 with $m = k^{\Omega(1)}$ and $\varepsilon = n^{-\Omega(1)}$.*

Since the error of this extractor is only polynomial in $1/n$, the quantum-proof version can only produce an output of length $m' = \Omega(\log n)$. The constant hidden in $m = \Omega(\log n)$ depends on the constant in $\varepsilon = n^{-\Omega(1)}$. However, Lemma 29 allows the error to be n^{-c_2} for any constant c_2 [19], which means that $m' = c_3 \log n$ for any c_3.

▶ **Corollary 30.** *There exists a constant c_1' such that for any n and $k' \geq \log^{c_1'} n$, there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{m'}$ *that is a quantum-proof (k', k', ε') two-source extractor strong in X_2 with $\varepsilon' = n^{-\Omega(1)}$ and $m' = c_3 \log n$ for any constant $c_3 > 0$ and sufficiently large n.*

Proof. Since for Lemma 29 we have $\varepsilon = n^{-c_2}$ for any c_2, we set $m' = \frac{c_2}{2} \log n$, hence $\varepsilon' = \sqrt{3\varepsilon 2^{m'}}/2 = n^{-\Omega(1)}$. The difference between k' and k is absorbed in the constant c_1'. ◀

This extractor is strong in the second input, hence as previously we can extract the entropy of this source using Trevisan's extractor. However, since the output is only $m' = c_3 \log n$, we compose it with a variant of Trevisan's extractor that only needs a seed of length $O(\log n)$, but requires the source to have entropy $k = n^\alpha$ for some constant $0 < \alpha \leq 1$ and extracts k^β bits for any $0 < \beta < 1$. This extractor is given in Lemma 41. The result given here below allows one of the sources to still have poly-logarithmic entropy, but requires the other to have $k = n^\alpha$ bits of min-entropy.

▶ **Corollary 31.** *There exists a constant c_1' such that for any $0 < \alpha \leq 1$, $0 < \beta < 1$, n, $k_1 \geq \log^{c_1'} n$ and $k_2 \geq n^\alpha$ there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^{m''}$ *that is a quantum-proof $(k_1, k_2, \varepsilon'')$ two-source extractor with $\varepsilon'' = n^{-\Omega(1)}$ and $m'' = k_2^\beta$.*

C The quantum Markov model in quantum randomness amplification protocols

Recently, the interest in quantum-proof two-source extractors (and multi-source in general) was renewed as people wished to use them as part of quantum randomness amplification

(QRA) protocols. As for randomness extractors, the goal of a QRA protocol is to extract an almost uniformly random string from a weak source (which is usually known in public, e.g., NIST's Randomness Beacon). However, in contrast to randomness extractors, the idea is to do it with only *one* weak source (and no seed) by exploiting the power of quantum physics (as mentioned in Section 1 this is impossible in the case of randomness extractors). Of course, once a quantum protocol is considered, it only makes sense to consider quantum side information.

With particular importance are QRA protocols which are device independent. That is, protocols in which one treats the devices as black boxes and does not assume much regarding the underlying quantum states and measurements inside the boxes[9]. One should then prove the security of the protocol only based on the statistics which are observed by the honest user when running the protocol. This seemly impossible task is made possible by the use of Bell inequalities, which allow one to "certify the quantumness" of the considered protocol (for a review on the topic see, e.g., [32]).

In the past couple of years several protocols were suggested for this task. The result presented in [7] was a big breakthrough: they considered a QRA protocol which uses a polynomial (at best) number of devices and a security proof against a general quantum adversary was proven. The main disadvantage of the protocol given in [7] for actual implementations is the number of devices; each device can be thought of as a separate computer (or actually a complete laboratory where a Bell violation experiment can be done) and for the protocol to work one must make sure that the different computers cannot send signals to one another. Hence, a large number of devices amounts to a huge impractical apparatus. It is therefore interesting to ask whether a QRA protocol with a constant number of devices exists, or under which assumptions on the devices it is possible to devise such a protocol which can also be implemented in practice.

Several other works considered the question of QRA with a constant number of devices, e.g., [4, 22, 24]. The general idea in those works was to create two independent weak random sources from devices (under different additional setup assumptions not made in [7]) that violate some Bell inequality, and then to apply a two-source extractor to get a final uniform key. However, as two-source extractors are not secure against general quantum adversaries the security was compromised. Indeed, [22, 24] for example did not give a complete security proof against quantum adversaries. In [4] security was proven[10] by a reduction to the case of a simple classical adversary (and hence the extractor could be used), at the cost of an additional setup assumption, namely that the adversary never has access to the initial weak source, and some loss in parameters.

Following the current work about quantum-proof multi-source extractors it is therefore interesting to consider the Markov model in the context of QRA protocols. More specifically, one can assume that two (or more) separated devices are a priori in product and become correlated only via the adversary or the environment, i.e., the state of the devices and the adversary ρ_{ABC} is a Markov chain $A \leftrightarrow C \leftrightarrow B$. The (unknown but local) measurements in the two devices then create a ccq-state $\rho_{X_1X_2C}$ in the Markov model, to which the quantum-proof two-source extractor is applied.

---

[9] The advantage of this approach is that this stronger notion of security allows for some inevitable unknown imperfections in actual implementations of the protocol.

[10] The security proof of [4] holds against non-signalling adversaries, which are more powerful than quantum ones. Note that two-source extractors are not known to be secure against those more powerful non-signalling adversaries (in any model of the sources and the side information). Furthermore, our proof technique that shows security in specific quantum models cannot be used in the non-signalling case.

Such assumptions about the structure of the devices could be justified in an intermediate device independent manner, e.g., if the devices are produced by two different experimental groups, or if the experimentalists know that a priori the devices are in a product state but might get correlated since they are placed in near by locations and therefore effected from the same temperature fluctuations. In any case, we still consider one quantum adversary and do not restrict her side information to some leakage operation as in [6]. Furthermore, it is well known that for many Bell inequalities, if the observed Bell violation in the QRA protocol is maximal then the devices must be in product with one another (i.e., one does not need to *assume* that this is the case). Taking into account self-testing results like [28, 40], although out of reach of current techniques, it is possible that in the future one could justify an almost tensor product structure (in some appropriate notion of closeness under which the extractors still perform well) from a non-maximal observed Bell violation.

D Proofs of Section 4

▶ **Lemma 11.** *Let* $\rho_{X_1X_2C} = \sum_{x_1,x_2} |x_1\rangle\langle x_1| \otimes |x_2\rangle\langle x_2| \rho_C(x_1,x_2)$. *Then there exists a pure state* ρ_{ABC} *and POVMs* $\{F_{x_1}\},\{G_{x_2}\}$ *such that*

$$\rho_C(x_1,x_2) = \mathrm{Tr}_{AB}\left[F_{x_1}^{\frac{1}{2}} \otimes G_{x_2}^{\frac{1}{2}} \otimes \mathbb{1}_C \rho_{ABC} F_{x_1}^{\frac{1}{2}} \otimes G_{x_2}^{\frac{1}{2}} \otimes \mathbb{1}_C\right] .$$

Proof. Let $\rho_{X_1X_2C} = \sum_{x_1,x_2} p_{x_1,x_2} |x_1\rangle\langle x_1|_{X_1} \otimes |x_2\rangle\langle x_2|_{X_2} \otimes \tilde{\rho}_C^{x_1,x_2}$, where $\tilde{\rho}_C^{x_1,x_2} = \frac{\rho_C(x_1,x_2)}{\mathrm{Tr}\rho_C(x_1,x_2)}$. And let $|\psi^{x_1,x_2}\rangle_{RC}$ be a purification of $\tilde{\rho}_C^{x_1,x_2}$. We define

$$|\rho\rangle_{ABC} = \sum_{x_1,x_2} \sqrt{p_{x_1,x_2}} |x_1\rangle_{X_1} \otimes |x_2\rangle_{X_2} \otimes |\psi^{x_1,x_2}\rangle_{RC}$$

with $A = X_1$ and $B = X_2R$. One can easily verify that this lemma holds for $F_{x_1} = |x_1\rangle\langle x_1|$ and $G_{x_2} = |x_2\rangle\langle x_2| \otimes \mathbb{1}_R$. ◀

▶ **Lemma 13.** *For any* $\Psi_{ABC_1C_2}$ *and positive operators* $\{F_{x_1}\},\{G_{x_2}\},\{H_{z_1}\},\{K_{z_2}\}$ *which sum up to the identity,*

$$\sum_{\substack{x_1,x_2,\\ z_1,z_2,y}} \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right]\left[\delta_{\mathrm{Ext}(z_1,z_2)=y} - \frac{1}{M}\right] \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right]$$
$$= \sum_{\substack{x_1,x_2,z_1,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(z_1,z_2)}} \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right] - \frac{1}{M}$$

Proof. Let $\mathrm{p}(x_1,x_2,z_1,z_2) = \mathrm{Tr}\left[\Psi_{ABC_1C_2} F_{x_1} \otimes G_{x_2} \otimes H_{z_1} \otimes K_{z_2}\right]$. We consider each of the terms of the LHS of the equation separately:

$$\sum_{\substack{x_1,x_2,\\ z_1,z_2,y}} \delta_{\mathrm{Ext}(x_1,x_2)=y}\delta_{\mathrm{Ext}(z_1,z_2)=y}\, \mathrm{p}(x_1,x_2,z_1,z_2) = \sum_{\substack{x_1,x_2,z_1,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(z_1,z_2)}} \mathrm{p}(x_1,x_2,z_1,z_2) ;$$

$$\frac{1}{M}\sum_{\substack{x_1,x_2,\\ z_1,z_2,y}} \delta_{\mathrm{Ext}(x_1,x_2)=y}\, \mathrm{p}(x_1,x_2,z_1,z_2) = \frac{1}{M} ;$$

$$\frac{1}{M}\sum_{\substack{x_1,x_2,\\ z_1,z_2,y}} \delta_{\mathrm{Ext}(z_1,z_2)=y}\, \mathrm{p}(x_1,x_2,z_1,z_2) = \frac{1}{M} ;$$

$$\frac{1}{M^2}\sum_{\substack{x_1,x_2,\\ z_1,z_2,y}} \mathrm{p}(x_1,x_2,z_1,z_2) = \frac{1}{M} .$$

The lemma follows by combining all the terms. ◀

In the following we denote $[l] \setminus \{i\}$ by $\bar{i}$ and $[l] \setminus \{i,j\}$ by $\overline{\{i,j\}}$.

▶ **Lemma 32.** *Let $\rho_{A_{[l]}C}$ be such that for all $i \in [l]$,*

$$I(A_i : A_{\bar{i}}|C) = 0 \ .$$

Then $\rho_{A_{[l]}C}$ can be written as a direct sum of product states,

$$\rho_{A_{[l]}C} = \bigoplus_t \mathrm{p}(t) \rho^t_{A_1 C_1^t} \otimes \cdots \otimes \rho^t_{A_l C_l^t} \ ,$$

where $\mathcal{H}_C = \bigoplus_t \mathcal{H}_{C_1^t} \otimes \cdots \otimes \mathcal{H}_{C_l^t}$.

Proof. We prove this lemmas by recursively applying the result from [14] given in Equation (12) on the structure of quantum Markov chains. We will also use the following facts about conditional mutual information:

1. For any ρ_{ABC}, $I(A:B|C) \geq 0$.
2. For any ρ_{ABCD}, $I(A:BC|D) \geq I(A:B|D)$.
3. For any $\rho_{ABCX} = \sum_x p_x \rho^x_{ABC} \otimes |x\rangle\langle x|$ classical on X, $I(A:B|CX) = \sum_x p_x I(A:B|CX=x)$.

Because $I(A_1 : A_{\bar{1}}|C) = 0$, we know that

$$\rho_{A_{[l]}C} = \bigoplus_{t_1} p_{t_1} \rho^{t_1}_{A_1 C_1^{t_1}} \otimes \rho^{t_1}_{A_{\bar{1}} C_{\bar{1}}^{t_1}} \ .$$

Let T_1 denote a classical system defined by

$$\rho_{A_{[l]}CT_1} = \bigoplus_{t_1} p_{t_1} \rho^{t_1}_{A_1 C_1^{t_1}} \otimes \rho^{t_1}_{A_{\bar{1}} C_{\bar{1}}^{t_1}} \otimes |t_1\rangle\langle t_1| \ .$$

Note that $\rho_{A_{[l]}CT_1}$ is related to $\rho_{A_{[l]}C}$ by an isometry from C to CT_1, hence

$$I(A_2 : A_{\bar{2}}|CT_1) = I(A_2 : A_{\bar{2}}|C) = 0 \ .$$

It follows that for all t_1,

$$I(A_2 : A_{\bar{2}}|CT_1 = t_1) = 0 \ ,$$

and hence

$$I(A_2 : A_{\overline{\{1,2\}}}|CT_1 = t_1) = 0 \ ,$$

which means that the state $\rho^{t_1}_{A_2 A_{\overline{\{1,2\}}} C^{t_1}}$ is a Markov chain $A_2 \leftrightarrow C^{t_1} \leftrightarrow A_{\overline{\{1,2\}}}$. Applying Equation (12) again, we get

$$\rho^{t_1}_{A_2 A_{\overline{\{1,2\}}} C^{t_1}} = \bigoplus_{t_2} p_{t_2} \rho^{t_1,t_2}_{A_2 C_2^{t_1,t_2}} \otimes \rho^{t_1,t_2}_{A_{\overline{\{1,2\}}} C^{t_1,t_2}_{\overline{\{1,2\}}}} \ .$$

Repeating this for all $i \in [l]$ proves the lemma. ◀

E Strong extractors

In this section we give the proofs necessary for the security of quantum-proof two-source extractors, *strong* in the source X_1, against product side information. The same steps can be repeated to prove the same result for multi-source extractors which are strong with respect to other sources.

The following lemma is the analogues of Lemma 12 for the strong case.

▶ **Lemma 33.** *Let $\rho_{X_1X_2C} = \rho_{X_1C_1} \otimes \rho_{X_2C_2}$ be a product ccq-state. Then there exists a POVM $\{G_{z_2}\}$ acting on C_2 such that*

$$\frac{1}{M}\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\|^2 \leq \sum_{\substack{x_1,x_2,\\ z_2,y}} \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right]\left[\delta_{\mathrm{Ext}(x_1,z_2)=y} - \frac{1}{M}\right] \mathbb{P}[X_1 = x_1] \operatorname{Tr}_{C_2}\left[\rho_{C_2}(x_2)G_{z_2}\right],$$

where $M = 2^m$.

Proof. First, recall that for a hermitian matrix R we have $\|R\| = \max\{\mathrm{Tr}[RS] : -\mathbb{1} \leq S \leq \mathbb{1}\}$. Applying this to the matrix which norm specifies the error of the extractor, we find

$$\|\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\| = \max_{-\mathbb{1}\leq S\leq \mathbb{1}} \mathrm{Tr}\left[\left(\rho_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \rho_C\right) S\right].$$

Since $\rho_{\mathrm{Ext}(X_1,X_2)X_1C}$ and $\rho_{U_m} \otimes \rho_{X_1C}$ are block diagonal with respect to the outcome variable of the extractor y, as well as to the classical variable x_1, S can be assumed to be block diagonal as well. Using this and inserting the expression for $\rho_{X_1X_2C}$ in Equation (7) we arrive at

$$\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| = \max_{-\mathbb{1}\leq S_{y,x_1}\leq \mathbb{1}} \sum_{y,x_1,x_2} \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right] \mathrm{Tr}\left[\rho_{C_1}(x_1) \otimes \rho_{C_2}(x_2) S_{x_1,y}\right].$$

Let us denote $G_{x_2} = \bar{\rho}_{C_2}^{-\frac{1}{2}} \rho_{C_2}(x_2) \bar{\rho}_{C_2}^{-\frac{1}{2}}$ with $\bar{\rho}_{C_2} = \sum_{x_2} \rho_{C_2}(x_2)$. Then we find

$$\begin{aligned}
&\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| \\
&\quad= \max_{-\mathbb{1}\leq S_{y,x_1}\leq \mathbb{1}} \sum_{y,x_1,x_2} \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right] \\
&\qquad\qquad \cdot \mathrm{Tr}\left[\left(\rho_{C_1}(x_1) \otimes \bar{\rho}_{C_2}\right)^{\frac{1}{2}} \mathbb{1}_{C_1} \otimes G_{x_2} \left(\rho_{C_1}(x_1) \otimes \bar{\rho}_{C_2}\right)^{\frac{1}{2}} S_{x_1,y}\right] \\
&\quad= \max_{-\mathbb{1}\leq S_{y,x_1}\leq \mathbb{1}} \sum_{y,x_1} \mathrm{Tr}\left[\left(\rho_{C_1}(x_1) \otimes \bar{\rho}_{C_2}\right)^{\frac{1}{2}} \mathbb{1}_{C_1} \otimes \Delta_{x_1,y} \left(\rho_{C_1}(x_1) \otimes \bar{\rho}_{C_2}\right)^{\frac{1}{2}} S_{x_1,y}\right]
\end{aligned}$$

where we used the abbreviation

$$\Delta_{y,x_1} = \sum_{x_2} \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right] G_{x_2}.$$

We now denote

$$\rho_{X_1C_1} = \sum_{x_1} |x_1\rangle\langle x_1| \otimes \rho_{C_1}(x_1)$$

and find $\rho_{X_1C_1}^{\frac{1}{2}} = \sum_{x_1} |x_1\rangle\langle x_1| \otimes \rho_{C_1}(x_1)^{\frac{1}{2}}$. Setting

$$\Delta_y = \sum_{x_1} |x_1\rangle\langle x_1| \otimes \mathbb{1}_{C_1} \otimes \Delta_{y,x_1}\,, \qquad S_y = \sum_{x_1} |x_1\rangle\langle x_1| \otimes S_{y,x_1}$$

we find

$$\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| = \max_{-\mathbb{1}\leq S_y \leq \mathbb{1}} \sum_y \mathrm{Tr}\left[\rho_{X_1C_1}^{\frac{1}{2}} \otimes \bar{\rho}_{C_2}^{\frac{1}{2}} \Delta_y \rho_{X_1C_1}^{\frac{1}{2}} \otimes \bar{\rho}_{C_2}^{\frac{1}{2}} S_y\right].$$

The crucial observation is now that the sesquilinear form

$$(R_y) \times (T_y) \mapsto \sum_y \mathrm{Tr}\left[(\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} R_y^* (\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} T_y\right]$$

on block-diagonal matrices is positive semi-definite and hence fulfils the Cauchy-Schwarz inequality. Applying this gives

$$\begin{aligned}\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\|^2 \leq &\left(\sum_y \mathrm{Tr}\left[(\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} \Delta_y (\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} \Delta_y\right]\right) \\ &\cdot \left(\sum_y \mathrm{Tr}\left[(\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} S_y (\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} S_y\right]\right).\end{aligned}$$

Since we have that the norm of S_y is bounded by one, the terms in the second sum satisfy

$$\mathrm{Tr}\left[(\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} S_y (\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} S_y\right] \leq \mathrm{Tr}\left[\rho_{X_1C_1} \otimes \bar{\rho}_{C_2} S_y\right] \leq 1\,.$$

Hence we arrive at

$$\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| \leq \sqrt{M} \sum_y \mathrm{Tr}\left[(\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} \Delta_y (\rho_{X_1C_1} \otimes \bar{\rho}_{C_2})^{\frac{1}{2}} \Delta_y\right]$$

and expanding the definition of Δ_y yields

$$\begin{aligned}&\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| \leq \\ &\sqrt{M} \sum_{y,x_1,x_2,z_2} \mathrm{Tr}_{C_1}\left[\rho_{C_1}(x_1)\right] \left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right] \left[\delta_{\mathrm{Ext}(x_1,z_2)=y} - \frac{1}{M}\right] \mathrm{Tr}\left[\rho_{C_2}(x_2) G_{z_2}\right],\end{aligned}$$

and since G_{z_2} are positive operators summing up to the identity, the assertion is proven. ◀

Next, let $\mathrm{p}(x_1,x_2,z_2) = \mathbb{P}[X_1 = x_1]\ \mathrm{Tr}_{C_2}\left[\rho_{C_2}(x_2) G_{z_2}\right]$ and note that $\mathrm{p}(x_1,x_2,z_2)$ is indeed a probability distribution. Then, the following lemma is analogues to Lemma 13.

▶ **Lemma 34.** *For* $\mathrm{p}(x_1,x_2,z_2) = \mathbb{P}[X_1 = x_1]\ \mathrm{Tr}_{C_2}\left[\rho_{C_2}(x_2) G_{z_2}\right]$,

$$\begin{aligned}\sum_{\substack{x_1,x_2,\\ z_2,y}} &\left[\delta_{\mathrm{Ext}(x_1,x_2)=y} - \frac{1}{M}\right] \left[\delta_{\mathrm{Ext}(x_1,z_2)=y} - \frac{1}{M}\right]\ \mathrm{p}(x_1,x_2,z_2) \\ &= \sum_{\substack{x_1,x_2,z_2| \\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(x_1,z_2)}} \mathrm{p}(x_1,x_2,z_2) - \frac{1}{M} \qquad (15)\end{aligned}$$

Proof. We follow a similar line as in the proof of Lemma 13.

$$\sum_{\substack{x_1,x_2,\\ z_2,y}} \delta_{\mathrm{Ext}(x_1,x_2)=y}\delta_{\mathrm{Ext}(x_1,z_2)=y}\ \mathrm{p}(x_1,x_2,z_2) = \sum_{\substack{x_1,x_2,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(x_1,z_2)}} \mathrm{p}(x_1,x_2,z_2)\ ;$$

$$\frac{1}{M}\sum_{\substack{x_1,x_2,\\ z_2,y}} \delta_{\mathrm{Ext}(x_1,x_2)=y}\ \mathrm{p}(x_1,x_2,z_2) = \frac{1}{M}\ ;$$

$$\frac{1}{M}\sum_{\substack{x_1,x_2,\\ z_2,y}} \delta_{\mathrm{Ext}(x_1,z_2)=y}\ \mathrm{p}(x_1,x_2,z_2) = \frac{1}{M}\ ;$$

$$\frac{1}{M^2}\sum_{\substack{x_1,x_2,\\ z_2,y}} \mathrm{p}(x_1,x_2,z_2) = \frac{1}{M}\ .$$ ◀

The quantity in Equation (15) can be seen as a simple distinguishing strategy of a distinguisher trying to distinguish the output of the extractor from uniform given classical side information Z_2 about the second source X_2 and the source X_1. We can therefore relate it to the error of the *strong* extractor in the case of classical side information, i.e., to Equation (2). This is shown in the following lemma, which is analogues to Lemma 14.

▶ **Lemma 35.** *Let Z_2 denote the classical side information about the source X_2.*[11] *Then*

$$\sum_{\substack{x_1,x_2,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(x_1,z_2)}} \mathrm{p}(x_1,x_2,z_2) - \frac{1}{M} \leq \frac{1}{2}\|\mathrm{Ext}\,(X_1,X_2)\,X_1Z_2 - U_m \circ X_1Z_2\|\ .$$

Proof. Define the following random variables over $\{0,1\}^m \times \{0,1\}^{n_1} \times \{0,1\}^{n_2}$:

$$\mathrm{R} = \mathrm{Ext}(X_1,X_2)X_1Z_2 \quad ; \quad \mathrm{Q} = U_m \circ X_1Z_2\ .$$

Let $\mathcal{A}^\star = \left\{(a_1,a_2,a_3)\middle| a_1 = \mathrm{Ext}\,(a_2,a_3)\right\} \subseteq \{0,1\}^m \times \{0,1\}^{n_1} \times \{0,1\}^{n_2}$. Then, the probabilities that R and Q assign to the event $\mathcal{A}^\star$ are

$$\mathrm{R}(\mathcal{A}^\star) = \sum_{\substack{x_1,x_2,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(x_1,z_2)}} \mathrm{p}(x_1,x_2,z_2) \quad ; \quad \mathrm{Q}(\mathcal{A}^\star) = \frac{1}{M}$$

Using the definition of the variational distance we therefore have

$$\begin{aligned}\frac{1}{2}\|\mathrm{Ext}(X_1,X_2)X_1Z_2 - U_m \circ X_1Z_2\| &= \sup_{\mathcal{A}} \|\mathrm{R}(\mathcal{A}) - \mathrm{Q}(\mathcal{A})\| \\ &\geq \mathrm{R}(\mathcal{A}^\star) - \mathrm{Q}(\mathcal{A}^\star) \\ &= \sum_{\substack{x_1,x_2,z_2|\\ \mathrm{Ext}(x_1,x_2)=\mathrm{Ext}(x_1,z_2)}} \mathrm{p}(x_1,x_2,z_2) - \frac{1}{M}\ .\end{aligned}$$ ◀

Finally, we combine the lemmas together to show that any strong classical-proof two-source extractor in the Markov model is secure against product quantum side information as well. We follow similar steps to those in the proof of Lemma 15.

[11] There is no side information about the source X_1, since it is made available in full.

▶ **Lemma 36.** *Any (k_1, k_2, ε)-strong classical-proof two-source extractor in the Markov model is a $\left(k_1, k_2, \sqrt{\varepsilon \cdot 2^{(m-2)}}\right)$-strong quantum-proof product two-source extractor, where m is the output length of the extractor.*

Proof. Let $\rho_{X_1X_2C} = \rho_{X_1C_1} \otimes \rho_{X_2C_2}$ be any state of two classical sources and product side information with $H_{\min}(X_1|C_1) \geq k_1$ and $H_{\min}(X_2|C_2) \geq k_2$.

We can apply Lemmas 33, 34, and 35 to get the bound

$$\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| \leq \sqrt{\frac{M}{2}\|\mathrm{Ext}\,(X_1, X_2)\, X_1Z_2 - U_m \circ X_1Z_2\|}\,. \tag{16}$$

As it follows from the proofs of the previous lemmas that Z_2 includes side information about X_2 alone (and there is no additional side information about X_2, i.e., the quantum system C_1 is just thrown away) $\mathrm{p}(x_1, x_2, z_2) = \mathrm{p}(x_1) \cdot \mathrm{p}(x_2, z_2)$, which implies:

1. The sources and the classical side information form a Markov chain $X_1 \leftrightarrow Z_2 \leftrightarrow X_2$.
2. $H_{\min}(X_1|Z_2) = H_{\min}(X_1) \geq H_{\min}(X_1|C_1)$.
3. $H_{\min}(X_2|Z_2) \geq H_{\min}(X_2|C_2)$.

Hence, if $H_{\min}(X_i|C_i) \geq k_i$ then by the definition of a strong classical-proof two-source extractor,

$$\frac{1}{2}\|\mathrm{Ext}\,(X_1, X_2)\, X_1Z_2 - U_m \circ X_1Z_2\| \leq \varepsilon\,. \tag{17}$$

Combining Equations (16) and (17) we get

$$\frac{1}{2}\|\rho_{\mathrm{Ext}(X_1,X_2)X_1C} - \rho_{U_m} \otimes \rho_{X_1C}\| \leq \frac{1}{2}\sqrt{M\varepsilon} = \sqrt{\varepsilon 2^{(m-2)}}\,. \qquad \blacktriangleleft$$

F Extracting from subnormalized states

Extractors are usually defined for normalized states $\rho_{X_1X_2C}$. In applications one might wish to extract from subnormalized states – for example, the smooth min-entropy of a state is a bound on the entropy of a subnormalized state that is close by. Here we prove that if a function is an extractor (for normalized states), then one can use it to extract from subnormalized states as well. We write up the lemma and proof in the case of two-source extractors in the Markov model. Similar statements hold for multiple sources as well as seeded extractors.

▶ **Lemma 37.** *Let $\sigma_{X_1X_2C}$ be a subnormalized Markov state satisfying $H_{min}(X_1|C)_\sigma \geq k_1$ as well as $H_{min}(X_2|C)_\sigma \geq k_2$, and let* $\mathrm{Ext} : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ *be a $(k_1 - 1, k_2 - 1, \varepsilon)$ quantum-proof two-source extractor in the Markov model. If* Ext *is weak, then we have that*

$$\frac{1}{2}\|\sigma_{\mathrm{Ext}(X_1,X_2)C} - \rho_{U_m} \otimes \sigma_C\| \leq 2\varepsilon\,.$$

If Ext *is strong in X_i, then we have that*

$$\frac{1}{2}\|\sigma_{\mathrm{Ext}(X_1,X_2)X_iC} - \rho_{U_m} \otimes \sigma_{X_iC}\| \leq 2\varepsilon\,.$$

Proof. We prove the weak case. The proof for strong extractors is identical.

Define $p = \mathrm{Tr}[\sigma_C]$ and with that the normalized state $\hat{\sigma}_{X_1X_2C} = \frac{1}{p}\sigma_{X_1X_2C}$ as well as the auxiliary normalized state

$$\tilde{\sigma}_{X_1X_2CP} = \sigma_{X_1X_2C} \otimes |0\rangle\langle 0|_P + (1-p)\tau_{X_1X_2} \otimes \hat{\sigma}_C \otimes |1\rangle\langle 1|_P,$$

where $\tau_{X_1X_2}$ is the fully mixed state. Note that $X_1 \leftrightarrow CP \leftrightarrow X_2$ is a Markov chain for the state $\tilde{\sigma}_{X_1X_2CP}$. This state satisfies slightly modified min-entropy conditions:

$$\begin{aligned} p_{\text{guess}}(X_1|CP)_{\tilde{\sigma}_{X_1CP}} &= p_{\text{guess}}(X_1|C)_{\sigma_{X_1C}} + (1-p)p_{\text{guess}}(X_1|C)_{\tau_{X_1}\otimes\hat{\sigma}_C} \\ &= 2^{-k_1} + (1-p)2^{-n_1} \leq 2\cdot 2^{-k_1}\,. \end{aligned}$$

Hence $H_{\min}(X_1|CP)_{\tilde{\sigma}} \geq k_1 - 1$, and the same argument can also be carried out for X_2 showing that $H_{\min}(X_2|CP)_{\tilde{\sigma}} \geq k_2 - 1$. The state $\tilde{\sigma}_{X_1X_2CP}$ is thus a valid Markov state satisfying the min-entropy conditions and hence we have

$$\frac{1}{2}\|\tilde{\sigma}_{\text{Ext}(X_1,X_2)CP} - \rho_{U_m}\otimes\tilde{\sigma}_{CP}\| \leq \varepsilon\,.$$

But since the partial trace over the P system only decreases the trace distance, we infer that

$$\frac{1}{2}\|\sigma_{\text{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\sigma_C + (1-p)\tau_{\text{Ext}(X_1,X_2)}\otimes\hat{\sigma}_C - (1-p)\rho_{U_m}\otimes\hat{\sigma}_C\| \leq \varepsilon\,.$$

Thus starting from the expression $\|\sigma_{\text{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\sigma_C\|$ and then adding and subtracting the term $(1-p)[\tau_{\text{Ext}(X_1,X_2)}\otimes\hat{\sigma}_C - \rho_{U_m}\otimes\hat{\sigma}_C]$ as well as applying the triangle inequality leaves us with

$$\frac{1}{2}\|\sigma_{\text{Ext}(X_1,X_2)C} - \rho_{U_m}\otimes\sigma_C\| \leq \varepsilon + \frac{1-p}{2}\|\tau_{\text{Ext}(X_1,X_2)}\otimes\hat{\sigma}_C - \rho_{U_m}\otimes\hat{\sigma}_C\| \leq 2\varepsilon\,,$$

since $\tau_{X_1X_2}\otimes\hat{\sigma}_C$ is a Markov source satisfying the entropic constraints. ◀

G Composing two-source and seeded extractors

If a multi-source extractor is strong in an input X_1, then the output Y is independent from X_1. This can be interpreted as Y containing the entropy from X_2; the randomness of X_1 served only as a catalyst, but is still contained in that random variable. A very common technique used to extract that randomness is to use another extractor. Since Y is uniform and independent from X_1, it fulfils the conditions needed to use it as a seed in seeded extractor. This is formalised in the following lemma.

▶ **Lemma 38.** *Let* $\text{Ext} : \{0,1\}^{n_1}\times\{0,1\}^{n_2} \to \{0,1\}^d$ *be a quantum-proof* (k_1,k_2,ε)*-two-source extractor strong in the first input. And let* $\text{Ext}' : \{0,1\}^{n_1}\times\{0,1\}^d \to \{0,1\}^m$ *be a quantum-proof* (k_1,ε')*-seeded extractor. Then the function*

$$\begin{aligned} \text{Ext}'' : \{0,1\}^{n_1}\times\{0,1\}^{n_2} &\to \{0,1\}^m \\ (x_1,x_2) &\mapsto \text{Ext}'(x_1,\text{Ext}(x_1,x_2)), \end{aligned}$$

is a quantum-proof $(k_1,k_2,\varepsilon+\varepsilon')$*-two-source extractor.*

Proof. Let $\rho_{U_dX_1C} = \rho_{U_d}\otimes\rho_{X_1C}$, where ρ_{U_d} is a fully mixed state of dimension 2^d. And let $\rho_{\text{Ext}'(X_1,U_d)C}$ denote the state resulting from applying Ext' to X_1 with U_d as seed. From the triangle inequality and contractivity of the trace distance we have

$$\begin{aligned} &\frac{1}{2}\|\rho_{\text{Ext}'(X_1,\text{Ext}(X_1,X_2))C} - \rho_{U_m}\otimes\rho_C\| \\ &\qquad\leq \frac{1}{2}\|\rho_{\text{Ext}'(X_1,\text{Ext}(X_1,X_2))C} - \rho_{\text{Ext}'(X_1,U_d)C}\| + \frac{1}{2}\|\rho_{\text{Ext}'(X_1,U_d)C} - \rho_{U_m}\otimes\rho_C\| \\ &\qquad\leq \frac{1}{2}\|\rho_{\text{Ext}(X_1,X_2)X_1C} - \rho_{U_d}\otimes\rho_{X_1C}\| + \frac{1}{2}\|\rho_{\text{Ext}'(X_1,U_d)C} - \rho_{U_m}\otimes\rho_C\|\,. \end{aligned}$$

The first term above is the error of Ext and the second is the error of Ext'. ◀

Note that Lemma 38 only requires a weak seeded extractor. Hence if a strong extractor is used, the seed can additionally be appended to the output – this is the case for all the following extractors.

Here below we give several seeded quantum-proof extractor constructions – all variants of Trevisan's extractor – that we use in the explicit constructions from Appendix B.

The first construction [8, Corollary 5.3] is one for which the exact parameters have been calculated [21].

▶ **Lemma 39** ([8, Corollary 5.3],[21]). *There exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, which is a quantum-proof* (k,ε)*-strong extractor with*

$$
\begin{aligned}
t &= 2\log\frac{2nm^2}{\varepsilon^2}\ ,\\
a &= 1+\max\left\{0,\frac{\log(m-e)-\log(t-e)}{\log e-\log(e-1)}\right\}\ ,\\
k &= m+4\log\frac{m}{\varepsilon}+6\ ,\\
d &= at^2\ ,
\end{aligned}
$$

where e *is the mathematical constant.*

The entropy loss of this extractor, $k-m=4\log\frac{m}{\varepsilon}+6$, can be reduced by composing it with an almost universal hash function [36].

▶ **Lemma 40** ([8, Corollary 5.4]). *There exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, which is a quantum-proof* (k,ε)*-strong extractor with* $k=m+4\log\frac{1}{\varepsilon}+O(1)$ *and* $d=O(\log^2\frac{n}{\varepsilon}\log m)$*.*

For $\varepsilon=n^{-\Omega(1)}$ in Lemma 40 we get $d=O(\log^3 n)$.

The final construction we consider only requires a seed of length $O(\log n)$, but can only extract a sublinear amount of entropy.

▶ **Lemma 41** ([8, Corollary 5.6]). *For any constant* $0<\gamma<1$ *there exists an explicit function* $\mathrm{Ext} : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$*, which is a quantum-proof* (k,ε)*-strong extractor with* $k=n^\gamma m+8\log\frac{m}{\varepsilon}+O(1)$*,* $d=O(\log n)$ *and* $\varepsilon=n^{-\Omega(1)}$*.*

For example, if $k=n^\alpha$ in Lemma 41 for $\gamma<\alpha\leq 1$, then $m=n^{\alpha-\gamma}-o(1)=k^{1-\frac{\gamma}{\alpha}}-o(1)$.

Lower Bound on Expected Communication Cost of Quantum Huffman Coding

Anurag Anshu[*1], **Ankit Garg**[†2], **Aram W. Harrow**[‡3], **and Penghui Yao**[§4]

1 **Centre for Quantum Technologies, National University of Singapore, Singapore**
a0109169@u.nus.edu

2 **Microsoft Research New England, USA**
garga@microsoft.com

3 **Center for Theoretical Physics, Massachusetts Institute of Technology, USA**
aram@mit.edu

4 **Institute for Quantum Computing, University of Waterloo, Canada; and Department of Combinatorics and Optimization, University of Waterloo, Canada**
phyao1985@gmail.com

Abstract

Data compression is a fundamental problem in quantum and classical information theory. A typical version of the problem is that the sender Alice receives a (classical or quantum) state from some known ensemble and needs to transmit them to the receiver Bob with average error below some specified bound. We consider the case in which the message can have a variable length and the goal is to minimize its expected length.

For classical messages this problem has a well-known solution given by Huffman coding. In this scheme, the expected length of the message is equal to the Shannon entropy of the source (with a constant additive factor) and the scheme succeeds with zero error. This is a single-shot result which implies the asymptotic result, viz. Shannon's source coding theorem, by encoding each state sequentially.

For the quantum case, the asymptotic compression rate is given by the von-Neumann entropy. However, we show that there is no one-shot scheme which is able to match this rate, even if interactive communication is allowed. This is a relatively rare case in quantum information theory when the cost of a quantum task is significantly different than the classical analogue. Our result has implications for direct sum theorems in quantum communication complexity and one-shot formulations of Quantum Reverse Shannon theorem.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Quantum information, quantum communication, expected communication cost, huffman coding

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.3

* A.A. is supported by Core Grants of Centre for Quantum Technologies.
† This work was done when A.G. was a student at Princeton University and his research was partially supported by NSF grants CCF-1149888 and CCF-1525342, a Simons fellowship for graduate students in theoretical computer science and a Siebel scholarship.
‡ A.W.H. was funded by NSF grants CCF-1111382 and CCF-1452616.
§ P.Y. is supported by NSERC and CIFAR.

1 Introduction

The central theme of information theory is compression of messages up to their *information content*. The celebrated work of Shannon [19] initiated this idea by showing that in the asymptotic setting, compression could be achieved up to the *Shannon entropy* of the message source. Subsequently, it was shown by Huffman [14] that by encoding each message into a codeword of different length based on the probability of the occurrence $p(x)$ of message x, one can construct a code whose expected length is at most $H(p)+1$, where $H(\cdot)$ is the Shannon entropy. This led to an operational interpretation of the Shannon entropy of a source in the *one-shot* setting.

The Huffman coding scheme can easily be illustrated in the following way (adapted from the reference [11]). Alice and Bob share infinitely many copies of the joint random variable XY (X with Alice and Y with Bob), such that $p(x,y)=\delta_{x,y}p(x)$. These copies are arranged in a sequence known to both parties. If Alice gets an input x, she measures her half of the copies in this sequence, and sends to Bob the address of the first location where she finds her input x. The average length of the message is can easily be computed to be at most $\log(\frac{1}{p(x)})+1$. Thus, average length of the message in overall protocol is at most $\sum_x p(x)(\log(\frac{1}{p(x)})+1)=H(p)+1$.

The study of compression of messages in terms of *expected communication cost*, rather than *worst case communication cost* has been very fruitful in information theory, both in operational interpretation of fundamental quantities and in applications in communication complexity. In the work [11], the following task was considered (inspired by a result of Wyner [25]): Alice is given an input x with probability $p(x)$ and she needs to send a message to Bob so that Bob can output a y distributed according to $p(y|x)$. This is a joint sampling task of the probability distribution $p(x,y)\stackrel{\text{def}}{=}p(x)p(y|x)$. The authors showed that in the presence of shared randomness, the expected communication cost of jointly sampling $p(x,y)$ is upper and lower bounded by $\mathrm{I}(X:Y)+2\log(\mathrm{I}(X:Y))+\mathcal{O}(1)$ and $\mathrm{I}(X:Y)$, respectively. This served as a natural characterization of mutual information in one-shot setting (different from the one already given by Shannon [19] in terms of channel capacity). Huffman coding can be seen as a special case of the above task by setting $p(y|x)=\delta_{y,x}$. This result also has applications in proving direct sum theorems for communication complexity. The direct sum problem asks whether computing N copies of a function (or a task in general) requires N times as much communication as computing a single copy. [11] used their compression result to prove the following theorem:

▶ **Theorem** (Informal, [11]). *The minimum expected communication cost of an r-round protocol, w.r.t. N iid copies of a product distribution μ, required to compute N copies of a function $f(x,y)$ is at least $N\cdot(CC_r(f)-O(r))$, where $CC_r(f)$ is the minimum expected communication cost (w.r.t μ) of an r-round protocol required to compute a single copy of f.*

The message compression in the presence of side information was first studied in the asymptotic setting by Slepian and Wolf [20]. The work by Braverman and Rao [8] gave its one-shot analogue in the following manner. Given a probability distribution P with Alice and Q with Bob, they constructed an interactive protocol (assisted by shared randomness) that allowed both Alice and Bob to output a distribution P' satisfying $\|P'-P\|_1\leq\varepsilon$, with expected communication cost $\mathrm{D}(P\|Q)+\mathcal{O}(\sqrt{\mathrm{D}(P\|Q)})+2\log(\frac{1}{\varepsilon})$. Here $\mathrm{D}(P\|Q)$ is relative entropy between P and Q. This work thus provided an operational interpretation to *relative entropy* [1] and extended the above theorem to general distributions. The holy grail for such

[1] The work [11] given an operational interpretation of relative entropy as well, but for the task where Alice knows the distribution P and both Alice and Bob know the distribution Q.

direct sum theorems is to remove the dependence on the number of rounds, and the above mentioned results ([11],[8]) along with [4] are important steps in this direction.

The aforementioned discussion points to a generic principle: it is possible to compress communication protocols up to their *Information Cost* (formally introduced in [8, 4], see also references therein) with the aid of shared randomness and consideration of expected communication cost as communication measure.

On the other hand, while many of the above results have their quantum counterpart, a similar principle for entanglement assisted quantum communication protocols has not yet been well established, as we discuss now. Quantum communication protocols typically fall into two classes: non-coherent protocols and coherent protocols.

In the case of coherent quantum protocols, Alice and Bob share a tripartite quantum state with the Referee and their objective is to perform a task while maintaining quantum coherence with the Referee. An example of coherent quantum protocols is the quantum state merging, introduced in [13] as the quantum analogue of Slepian-Wolf protocol [20] (in the asymptotic setting). The most general form of coherent quantum protocols, involving two parties and one Referee, is known as the quantum state redistribution. It is defined as follows: Alice (AC), Bob(B) and Referee (R) share a pure quantum state Ψ_{RABC} and Alice needs to transfer the register C to Bob. This task was originally introduced in [9, 26] to give an operational meaning of the quantum conditional mutual information in the asymptotic setting. Furthermore, as shown by Touchette [22], it nicely captures interactive quantum communication protocols within the framework of quantum communication complexity and leads to a formulation of *quantum information complexity*.

Using the one-shot quantum protocols for quantum state redistribution developed in [6], and the notion of quantum information complexity, Touchette [22] obtains the following direct sum result for entanglement assisted quantum communication complexity.

▶ **Theorem** (Informal, [22]). *The minimum worst case quantum communication cost of an r-round quantum protocol required to compute N copies of a (classical) function $f(x, y)$ is at least $N \cdot (\frac{QCC_r(f)}{r^2} - O(r))$, where $QCC_r(f)$ is the worst case communication cost of an r-round quantum protocol required to compute a single copy of f.*

The above result has a strong dependence on number of rounds (as opposed to a weaker dependence in the the direct sum result by [11]), that comes from the consideration of the worst case quantum communication cost for the quantum state redistribution in the work [6]. Furthermore, it has been shown recently in [1] that the expected quantum communication cost of a protocol achieving quantum state redistribution cannot be substantially better than its worst case quantum communication cost. This leads to a bottleneck in the improvement of the direct sum results for the quantum case within the framework of coherent quantum protocols.

In non-coherent protocols, Alice and Bob perform a task on their inputs without maintaining the coherence with the Referee. The works which exhibit one-shot quantum compression protocols in the non-coherent setting, include [15, 16] (which also show direct sum theorems for entanglement assisted one-way quantum communication complexity) and [3] (which is an extension of Braverman-Rao protocol [8] to the quantum domain). All of these results take into consideration only the worst case quantum communication cost, and it is not clear if the expected communication cost of these message compression task can be substantially improved (to the information cost) over the worst case cost.

In this work, we explore the possibility of having quantum protocols with better expected communication cost in the non-coherent framework. Towards this, we define the following *quantum Huffman task*.

▶ **Definition 1** (Quantum Huffman task)**.** Alice (A) receives an input x and an associated quantum pure state $|\Psi_x\rangle$ with probability $p(x)$. For a given $\eta > 0$, which we shall henceforth identify as 'error parameter', Alice needs to transfer the state $|\Psi_x\rangle$ to Bob, such that the final state Φ_x with Bob satisfies $\sum_x p(x)\mathrm{F}^2(\Psi_x, \Phi_x) \geq 1 - \eta^2$. Here, $\mathrm{F}(\cdot,\cdot)$ is fidelity and η^2 is average error of the protocol.

The above task is a quantum version of the classical one-shot source coding. The expected communication cost in the asymptotic setting is lower bounded by $S(\sum_x p(x)\Psi_x)$ due to [12], which is also the quantum information cost of this task. The main question that we address is whether there exists a communication protocol that achieves the above task with expected communication cost close to $S(\sum_x p(x)\Psi_x)$.

A prior work by Braunstein *et. al.*[7] had considered our question and had noted several issues in generalizing directly the techniques of 'classical' Huffman coding to quantum case. In present work, we show that no such compression scheme is possible.

Our results

We refer to the collection of pairs $\{(p(x), \Psi_x)\}_x$ as an *ensemble* of states and associated probabilities. Following the discussion in introduction, we would like to compare the expected communication cost of any protocol achieving quantum Huffman task with the von-Neumann entropy of average state with Alice : $S(\sum_x p(x)\,|\Psi_x\rangle\langle\Psi_x|)$. Our main result is a large gap between the two quantities, that we state below.

▶ **Theorem 2.** *Fix a positive integer $d > 10^{12}$ and real δ that satisfy $\frac{16}{\sqrt{d}} < \delta < \frac{1}{100}$. There exist a collection of $N \stackrel{\text{def}}{=} (\frac{3}{\delta^2})^d$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to a d dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$ such that following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\,|\Psi_x\rangle\langle\Psi_x|) \leq 4\delta\log(d) + H(\delta) + 1$.*
- *For any one-way protocol achieving quantum Huffman coding of above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by $(1-\eta)\cdot\log(d\delta) - 6$.*
- *For any r-round protocol achieving quantum Huffman coding of above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by $\Omega(\frac{\log(d\delta)}{\log r})$.*

The one-way part of this theorem is proved in Section 5, as a special case of Theorem 19. The r-round part follows argument similar to that of one-way part, and its technical details can be found in the arXiv eprint of this work [2].

For interactive case, we also give a round independent statement for small enough η.

▶ **Theorem 3.** *Fix a positive integer $d > 10^{12}$, real δ that satisfies $\frac{\sqrt{768}}{\log(d)} < \delta < 1$ and a monotonically increasing function $f : \mathbb{R} \to \mathbb{R}$ such that $f(x) \geq x^2$. There exist a collection of $N \stackrel{\text{def}}{=} 2^{f(d)}$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to a d dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$, such that the following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\,|\Psi_x\rangle\langle\Psi_x|) \leq 4\delta\log(d) + H(\delta) + 1$.*
- *For any interactive protocol with error parameter $\eta \stackrel{\text{def}}{=} \frac{1}{\log^2(d)} = \frac{4}{\log^2(f^{-1}(\log(N)))}$, the expected communication cost is lower bounded by $\Omega(\frac{\log(d\delta)}{\log\log(d)})$.*

The proof of this theorem can again be found in the arXiv eprint of this work [2]. It may be noted that the dependence of error parameter η on input size $\log N$ can be made as weak as desired, by choosing an appropriate function f which increases sufficiently fast.

Our techniques

Our proof follows in two main steps, which we illustrate here for the case of one-way protocols for simplicity. All the quantum states appearing below are assumed to belong to a Hilbert space of dimension d. We first show that for every message i sent from Alice to Bob, there exists a quantum state σ_i, such that the probability p_i of this message is upper bounded by $p_i \leq \sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma_i)}$, where η is the error parameter and $\mathrm{D}^{\eta}_{\max}(\cdot\|\cdot)$ is smooth relative max-entropy. This upper bound crucially uses the fact that the quantum states Ψ_x are pure. Section 3 for one-way protocols is built upon this idea. Our aim now is to find an ensemble $\{p(x), \Psi_x\}$ for which the quantity $\sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma_i)}$ is small, as a result of which the expected communication cost must be large.

Our second step is based upon the observation that given the quantum state σ_i (as mentioned above), and a pure state Ψ chosen according to Haar measure, the smooth relative max-entropy ($= \mathrm{D}^{\eta}_{\max}(\Psi \| \sigma_i)$) must attain large value ($\approx \log(d)$) with high probability. This suggests that the ensemble $\{(p(x), \Psi_x)\}_x$ should be constructed by choosing vectors from Haar measure, making the quantity $\sum_x p(x) 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma_i)}$ close to $\mathcal{O}(1) \cdot 2^{-\log(d)}$. This gives the upper bound $p_i \leq \frac{\mathcal{O}(1)}{d}$ and hence expected communication cost is at least $\log(d) - \mathcal{O}(1)$. Unfortunately, this choice of ensemble makes the von-Neumann entropy of the average state $\sum_x p(x) \Psi_x$ equal to $\log(d)$, which is not much smaller than expected communication cost.

We remedy this problem by introducing a free variable δ and letting $|\Psi_x\rangle = \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|x\rangle$, where $|0\rangle$ is some fixed vector and $|x\rangle$ belongs to $d-1$ dimensional subspace orthogonal to $|0\rangle$. We choose $|x\rangle$ according to Haar measure in the $d-1$ dimensional subspace and show that the smooth relative max entropy $\mathrm{D}^{\eta}_{\max}(\Psi_x \| \sigma)$ is still large ($\approx \log(d\delta)$ with high probability) as long as $\eta < \delta/16$. Interestingly, now the von-Neumann entropy of the average state $\sum_x p(x) \Psi_x$ is $\approx \delta \log(d)$, which is much smaller than expected communication cost. Details have been discussed in Section 4, where epsilon nets have been used to make the input size finite.

2 Preliminaries

In this section we present some notations, definitions, facts and lemmas that we will use in our proofs.

Information theory

For a natural number n, let $[n]$ represent the set $\{1, 2, \ldots, n\}$. For a set S, let $|S|$ be the size of S. A *tuple* is a finite collection of positive integers, such as $(i_1, i_2 \ldots i_r)$ for some finite r. We let log represent logarithm to the base 2 and ln represent logarithm to the base e. The ℓ_1 norm of an operator X is $\|X\|_1 \stackrel{\text{def}}{=} \mathrm{Tr}\sqrt{X^\dagger X}$ and ℓ_2 norm is $\|X\|_2 \stackrel{\text{def}}{=} \sqrt{\mathrm{Tr} X X^\dagger}$. A quantum state (or just a state) is a positive semi-definite matrix with trace equal to 1. It is called *pure* if and only if the rank is 1. Let $|\psi\rangle$ be a unit vector. We use ψ to represent the state and also the density matrix $|\psi\rangle\langle\psi|$, associated with $|\psi\rangle$.

A sub-normalized state is a positive semi-definite matrix with trace less than or equal to 1. A *quantum register* A is associated with some Hilbert space $\mathcal{H}_A$. Define $|A| \stackrel{\text{def}}{=} \dim(\mathcal{H}_A)$. We denote by $\mathcal{D}(A)$, the set of quantum states in the Hilbert space $\mathcal{H}_A$ and by $\mathcal{D}_{\leq}(A)$, the set of all sub-normalized states on register A. State ρ with subscript A indicates $\rho_A \in \mathcal{D}(A)$.

For two quantum states ρ and σ, $\rho \otimes \sigma$ represents the tensor product (Kronecker product) of ρ and σ. Composition of two registers A and B, denoted AB, is associated with Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$. If two registers A, B are associated with the same Hilbert space, we shall denote it by $A \equiv B$. Let ρ_{AB} be a bipartite quantum state in registers AB. We define

$$\rho_B \stackrel{\text{def}}{=} \mathrm{Tr}_A(\rho_{AB}) \stackrel{\text{def}}{=} \sum_i (\langle i| \otimes \mathbb{1}_B)\rho_{AB}(|i\rangle \otimes \mathbb{1}_B),$$

where $\{|i\rangle\}_i$ is an orthonormal basis for the Hilbert space A and $\mathbb{1}_B$ is the identity matrix in space B. The state ρ_B is referred to as the marginal state of ρ_{AB} in register B. Unless otherwise stated, a missing register from subscript in a state will represent partial trace over that register. A quantum map $\mathcal{E} : A \to B$ is a completely positive and trace preserving (CPTP) linear map (mapping states from $\mathcal{D}(A)$ to states in $\mathcal{D}(B)$). A completely positive and trace non-increasing linear map $\tilde{\mathcal{E}} : A \to B$ maps quantum states to sub-normalized states. The identity operator in Hilbert space $\mathcal{H}_A$ (and associated register A) is denoted I_A. A *unitary* operator $U_A : \mathcal{H}_A \to \mathcal{H}_A$ is such that $U_A^\dagger U_A = U_A U_A^\dagger = I_A$. An *isometry* $V : \mathcal{H}_A \to \mathcal{H}_B$ is such that $V^\dagger V = I_A$. The set of all unitary operations on register A is denoted by $\mathcal{U}(A)$.

We denote a unit ball in space $\mathbb{R}^d$ as S^d. An element of S^d is a unit vector in $\mathbb{R}^d$. We shall represent an element $x \in S^d$ using the bra-ket notation as $|x\rangle$. Euclidean norm of $|x\rangle$ is $\| |x\rangle\langle x| \|_1$. Given two vectors $|x\rangle, |y\rangle \in S^d$, the *Euclidean distance* between them is $\|(|x\rangle - |y\rangle)(\langle x| - \langle y|)\|_1$.

▶ **Definition 4.** We shall consider the following information theoretic quantities. Let $\varepsilon \geq 0$.

1. **Generalized fidelity.** For $\rho, \sigma \in \mathcal{D}_{\leq}(A)$,

$$\mathrm{F}(\rho, \sigma) \stackrel{\text{def}}{=} \left\| \sqrt{\rho}\sqrt{\sigma} \right\|_1 + \sqrt{(1 - \mathrm{Tr}(\rho))(1 - \mathrm{Tr}(\sigma))}.$$

2. **Purified distance.** For $\rho, \sigma \in \mathcal{D}_{\leq}(A)$,

$$\mathrm{P}(\rho, \sigma) = \sqrt{1 - \mathrm{F}^2(\rho, \sigma)}.$$

3. **ε-ball.** For $\rho_A \in \mathcal{D}(A)$,

$$\mathcal{B}^\varepsilon(\rho_A) \stackrel{\text{def}}{=} \{\rho'_A \in \mathcal{D}(A) |\ \mathrm{F}(\rho_A, \rho'_A) \geq 1 - \varepsilon\}.$$

4. **Entropy.** For $\rho_A \in \mathcal{D}(A)$,

$$\mathrm{H}(A)_\rho \stackrel{\text{def}}{=} -\mathrm{Tr}(\rho_A \log \rho_A).$$

5. **Relative entropy.** For $\rho_A, \sigma_A \in \mathcal{D}(A)$,

$$\mathrm{D}(\rho_A \| \sigma_A) \stackrel{\text{def}}{=} \mathrm{Tr}(\rho_A \log \rho_A) - \mathrm{Tr}(\rho_A \log \sigma_A).$$

6. **Max-relative entropy.** For $\rho_A, \sigma_A \in \mathcal{D}(A)$,

$$\mathrm{D}_{\max}(\rho_A \| \sigma_A) \stackrel{\text{def}}{=} \inf\{\lambda \in \mathbb{R} : 2^\lambda \sigma_A \geq \rho_A\}.$$

7. **Smooth max-relative entropy.** For $\rho_A, \sigma_A \in \mathcal{D}(A)$,

$$\mathrm{D}^\eta_{\max}(\rho_A \| \sigma_A) \stackrel{\text{def}}{=} \inf_{\rho'_A \in \mathcal{B}^\eta(\rho_A)} \mathrm{D}_{\max}(\rho'_A \| \sigma_A)\,.$$

8. Mutual information. For $\rho_{AB} \in \mathcal{D}(AB)$,

$$\mathrm{I}(A:B)_\rho \stackrel{\text{def}}{=} \mathrm{D}(\rho_{AB} \| \rho_A \otimes \rho_B) = \mathrm{H}(A)_\rho + \mathrm{H}(B)_\rho - \mathrm{H}(AB)_\rho .$$

We will use the following facts.

▶ **Fact 5** (Monotonicity of quantum operations). [[18, 5], [21], Theorem 3.4] For states ρ, $\sigma \in \mathcal{D}(A)$, and quantum map $\mathcal{E}(\cdot)$,

$$\|\mathcal{E}(\rho) - \mathcal{E}(\sigma)\|_1 \leq \|\rho - \sigma\|_1 , \mathrm{F}(\rho, \sigma) \leq \mathrm{F}(\mathcal{E}(\rho), \mathcal{E}(\sigma)) \text{ and } \mathrm{D}_{\max}(\rho\|\sigma) \geq \mathrm{D}_{\max}(\mathcal{E}(\rho)\|\mathcal{E}(\sigma)) .$$

▶ **Fact 6** (Joint concavity of fidelity). [[24], Proposition 4.7] Given quantum states $\rho_1, \rho_2 \ldots \rho_k$, $\sigma_1, \sigma_2 \ldots \sigma_k \in \mathcal{D}(A)$ and positive numbers $p_1, p_2 \ldots p_k$ such that $\sum_i p_i = 1$. Then

$$\mathrm{F}(\sum_i p_i \rho_i, \sum_i p_i \sigma_i) \geq \sum_i p_i \mathrm{F}(\rho_i, \sigma_i).$$

▶ **Fact 7** (Fannes inequality). [[10]] Given quantum states $\rho_1, \rho_2 \in \mathcal{D}(A)$, such that $|A| = d$ and $\mathrm{P}(\rho_1, \rho_2) = \varepsilon \leq \frac{1}{2\mathrm{e}}$,

$$|S(\rho_1) - S(\rho_2)| \leq \varepsilon \log(d) + 1.$$

▶ **Fact 8** (Levy's concentration lemma). [[17]] Let $f : S^d \to \mathbb{R}$ be Lipschitz continuous function with Lipschitz constant ℓ, defined as

$$\ell \stackrel{\text{def}}{=} \max_{x,y} \frac{|f(x) - f(y)|}{\|x - y\|_2}.$$

Let $\mathbb{E}(f)$ be expectation value of f with respect to uniform measure over S^d. Then

$$\mathrm{Prob}(|f - \mathbb{E}(f)| \geq \alpha) \leq 2e^{-\frac{d\alpha^2}{18\pi^3\ell^2}}.$$

3 One way communication

A one-way quantum communication protocol P for quantum Huffman coding with error η^2 is described as follows.

Input: Alice gets an input x with probability $p(x)$ and she needs to send the state $|\Psi_x\rangle$ to Bob.

Pre-shared entanglement: They have a pre-shared entanglement $|\theta\rangle_{AB}$.

- Conditioned on the input x, Alice applies a measurement $\{M_1^x, M_2^x \ldots\}$ on her side and sends the outcome i to Bob. Let

$$p_i^x \stackrel{\text{def}}{=} \mathrm{Tr}(M_i^x \theta_A), \quad \rho_i^x \stackrel{\text{def}}{=} \frac{\mathrm{Tr}_A(M_i^x \theta_{AB})}{p_i^x}.$$

- Receiving message i from Alice, Bob applies a quantum channel $\mathcal{E}_i$ based on the message i, to obtain a state σ_i^x in his output register.
- The final state in the output register is $\sum_i p_i^x \sigma_i^x$ and it follows that

$$\sum_x p(x) \sum_i p_i^x \langle\Psi_x| \sigma_i^x |\Psi_x\rangle > 1 - \eta^2$$

due to correctness of protocol.

The expected communication cost of P is $\sum_x p(x) \sum_i p_i^x \lceil \log(i) \rceil$ which can be lower bounded by $\sum_x p(x) \sum_i p_i^x \log(i)$. Since we are interested in lower bounding the expected communication cost, we shall consider the latter quantity.

Define the quantity $t_i \stackrel{\text{def}}{=} \sum_x p(x) 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))}$.

We have the following lemma, proof of which has been given in Appendix A.

▶ **Lemma 9.** *Let a be the largest integer such that $t_i \leq 2^{-a}$ for all i. Then expected communication cost of* P *is lower bounded by $a(1-\sqrt{\eta})^2 - 1$.*

4 Example separating expected communication and information

4.1 An epsilon net over S^d

We will use the epsilon net over S^d, as defined below.

▶ **Definition 10** (Epsilon-nets, [23])**.** Fix an $\varepsilon > 0$. There exists an integer N and a set of vectors $\{|x_1\rangle, |x_2\rangle \dots |x_N\rangle\}$ on S^d such that the following properties hold:

- $N \leq (\frac{2}{\varepsilon})^d$.
- For any two vectors $|x_i\rangle, |x_j\rangle$ it holds that $\|(|x_i\rangle - |x_j\rangle)(\langle x_i| - \langle x_j|)\|_2 \leq \varepsilon$.
- For any vector $|y\rangle \in S^d$, there exists j such that $\|(|y\rangle - |x_j\rangle)(\langle y| - \langle x_j|)\|_2 \leq \varepsilon$

Let the set be denoted as $\mathcal{N}_\varepsilon$.

We recall that μ is a uniform measure over S^d. For every vector $|x_i\rangle \in \mathcal{N}_\varepsilon$, we let $S_i \subset S^d$ be the set of all vectors $|y\rangle \in S^d$ such that $|x_i\rangle$ is one of the closest (in euclidean distance) to $|y\rangle$ among all vectors in $\mathcal{N}_\varepsilon$. Let $\mu(S_i)$ be the measure associated to S_i. $\mu(S_i)$ can also be interpreted as the volume of S_i. Due to the fact that set of vectors in S^d which are equidistant to two or more vectors in $\mathcal{N}_\varepsilon$ have measure zero, we obtain the relation:

$$\sum_i \mu(S_i) = 1, \quad \mu(S_i \cap S_j) = 0 \tag{1}$$

Let λ be a distribution over $\mathcal{N}_\varepsilon$, such that $\lambda(i) \stackrel{\text{def}}{=} \mu(S_i)$. Let $\mathbb{E}_i$ denote the expectation over the set $\mathcal{N}_\varepsilon$ with vectors chosen according to λ. That is, for any function $f(.)$ on $\mathcal{N}_\varepsilon$, we define

$$\mathbb{E}_i f(|x_i\rangle) \stackrel{\text{def}}{=} \sum_i \lambda(i) f(|x_i\rangle).$$

The following lemma follows from the the above definition.

▶ **Lemma 11.** *It holds that*

$$\|(\mathbb{E}_i |x_i\rangle)(\mathbb{E}_i \langle x_i|)\|_1 \leq \varepsilon, \quad \|\mathbb{E}_i |x_i\rangle\langle x_i| - \frac{I}{d}\|_1 \leq 2\sqrt{\varepsilon}.$$

Proof. For the first part, we use the identities

$$\int_y \mu(y) dy\, |y\rangle = 0, \quad \int_y \mu(y) dy\, |y\rangle = \sum_i \mu(S_i) \frac{\int_{y \in S_i} \mu(y) dy\, |y\rangle}{\mu(S_i)},$$

where the second identity follows from Equation 1. Now we notice from the definition of set S_i that

$$\|(|x_i\rangle - \frac{\int_{y \in S_i} \mu(y) dy\, |y\rangle}{\mu(S_i)})(\langle x_i| - \frac{\int_{y \in S_i} \mu(y) dy\, \langle y|}{\mu(S_i)})\|_2 \leq \varepsilon.$$

Applying expectation $\mathbb{E}_i$ to both sides and then using the triangle inequality, we immediately obtain

$$\|(\mathbb{E}_i |x_i\rangle)(\mathbb{E}_i \langle x_i|)\|_2 = \|(\sum_i \mu(S_i) |x_i\rangle)(\sum_i \mu(S_i) \langle x_i|)\| \leq \varepsilon.$$

For the second part, we again notice the identities

$$\int_y \mu(y)dy\, |y\rangle\langle y| = \frac{I}{d}, \quad \int_y \mu(y)dy\, |y\rangle\langle y| = \sum_i \mu(S_i) \frac{\int_{y\in S_i} \mu(y)dy\, |y\rangle\langle y|}{\mu(S_i)}.$$

From the definition of the set S_i, we have that for every $|y\rangle \in S_i$, $|\langle x_i|y\rangle|^2 \geq 1 - 2\varepsilon$. Thus, $\mathrm{F}(|x\rangle\langle x|, \frac{\int_{y\in S_i} \mu(y)dy|y\rangle\langle y|}{\mu(S_i)}) \geq 1 - 2\varepsilon$, which translates to $\| |x\rangle\langle x| - \frac{\int_{y\in S_i} \mu(y)dy|y\rangle\langle y|}{\mu(S_i)}\|_1 \leq 2\sqrt{\varepsilon}$. Now the proof follows along the same lines as first part. ◀

4.2 Our construction

Our construction now proceeds as follows, recalling the quantum Huffman task in Definition 1. Fix a $\delta > 0$. Alice is given the input i with probability $\lambda(i)$, which corresponds to the vector $|x_i\rangle \in \mathcal{N}_\varepsilon$. We embed C^d in a $d+1$ dimensional space C^{d+1} and let P be a projector onto the original space C^d. We define $|\Psi_i\rangle \stackrel{\text{def}}{=} \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|x_i\rangle$, where $|0\rangle$ is a vector satisfying $P|0\rangle = 0$.

We have the following lemma.

▶ **Lemma 12.** *The von-Neumann entropy of the average state $\mathbb{E}_i\Psi_i = \sum_i \lambda(i)\Psi_i$ satisfies $S(\mathbb{E}_i\Psi_i) \leq (\delta + 3\sqrt{\varepsilon})\log(d) + H(\delta) + 1$.*

Proof. Consider,

$$\mathbb{E}_i\, |\Psi_i\rangle\langle\Psi_i| = (1-\delta)\,|0\rangle\langle 0| + \sqrt{\delta(1-\delta)}\mathbb{E}_i(|0\rangle\langle x_i| + |x_i\rangle\langle 0|) + \delta\mathbb{E}_i\, |x_i\rangle\langle x_i|.$$

From Lemma 11, it follows that

$$\|\mathbb{E}_i\Psi_i - (1-\delta)\,|0\rangle\langle 0| + \delta\frac{P}{d}\|_1 \leq 3\sqrt{\varepsilon}.$$

Now we use Fannes inequality (Fact 7) to conclude that $S(\mathbb{E}_i\Psi_i)$ is at most $(\delta + 3\sqrt{\varepsilon})\log(d) + H(\delta) + 1$. ◀

4.3 A property of smooth relative max entropy

Following lower bound on smooth relative entropy shall be crucial for our argument.

▶ **Lemma 13.** *Let σ be any quantum state belonging to C^{d+1}. Let $k < d$ be an integer and Q^- (Q^+) be projector onto subspace where σ has eigenvalues less than (greater than) $\frac{1}{k}$. For any i and $\eta > 0$ such that $\langle\Psi_i| Q^- |\Psi_i\rangle > 2\eta$, it holds that*

$$2^{-\mathrm{D}^\eta_{\max}(\Psi_i\|\sigma)} \leq \frac{1}{k(1-\eta)(\sqrt{(1-\eta)\langle\Psi_i| Q^- |\Psi_i\rangle} - \sqrt{\langle\Psi_i| Q^+ |\Psi_i\rangle\,\eta})^2}.$$

Proof. Since $\dim(Q^+) \leq k$, it holds that $\dim(Q^-) \geq d + 1 - k$. Define the quantity

$$S^\eta(\Psi_i||Q^-) \stackrel{\text{def}}{=} \inf_{|\lambda\rangle:|\langle\lambda|\Psi_i\rangle|^2>1-\eta} \langle\lambda| Q^- |\lambda\rangle.$$

The lemma follows from the following two claims, which have been proved in Appendix B.

▶ **Claim 14.** *For any i, it holds that*

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} < \frac{1}{k(1-\eta)S^{2\eta}(\Psi_i||Q^-)}.$$

We now calculate an explicit expression for $S^{\eta}(\Psi_i||Q^-)$ in the following claim.

▶ **Claim 15.** *If $\langle\Psi_i| Q^- |\Psi_i\rangle > \eta$, then we have*

$$S^{\eta}(\Psi_i||Q^-) = (\sqrt{(1-\eta)\langle\Psi_i| Q^- |\Psi_i\rangle} - \sqrt{\langle\Psi_i| Q^+ |\Psi_i\rangle\,\eta})^2.$$

Else $S^{\eta}(\Psi_i||Q^-) = 0$.

Combining the two claims, our lemma follows. ◀

4.4 Final lower bound

Let μ be uniform measure over S^d. For any vector $|y\rangle$ belonging to subspace of P, let $|\Psi_y\rangle = \sqrt{1-\delta}\,|0\rangle + \sqrt{\delta}\,|y\rangle$ be a vector in C^{d+1}. We have the following claims, the first of which computes the expectation value and the second computes the Lipschitz constant.

▶ **Claim 16.** *It holds that*

$$\int_y \mu(y)dy\,\langle\Psi_y| Q^- |\Psi_y\rangle = (1-\delta-\frac{\delta}{d})\langle 0| Q |0\rangle + \delta(\frac{d+1-k}{d}).$$

Proof. Consider the following analysis, from which the statement follows.

$$\begin{aligned}
\int_y \mu(y)dy\,\langle\Psi_y| Q^- |\Psi_y\rangle &= (1-\delta)\langle 0| Q^- |0\rangle + \delta\int_y \mu(y)dy\,\langle y| Q^- |y\rangle \\
&= (1-\delta)\langle 0| Q^- |0\rangle + \frac{\delta}{d}\mathrm{Tr}(PQ) \\
&= (1-\delta)\langle 0| Q^- |0\rangle + \frac{\delta}{d}(\mathrm{Tr}(Q) - \langle 0| Q |0\rangle) \\
&= (1-\delta-\frac{\delta}{d})\langle 0| Q |0\rangle + \delta(\frac{d+1-k}{d})
\end{aligned}$$

This proves the claim. ◀

▶ **Claim 17.** *Let Q be a projector and $|y\rangle, |y'\rangle$ be any two vectors in S^d . Then it holds that*

$$|\langle\Psi_y| Q^- |\Psi_y\rangle - \langle\Psi_{y'}| Q^- |\Psi_{y'}\rangle| \le (2\sqrt{2\delta(1-\delta)}+2\delta)\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_2 \le 4\sqrt{\delta}\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_2.$$

Proof. Consider the analysis

$$\begin{aligned}
& |\langle\Psi_y| Q^- |\Psi_y\rangle - \langle\Psi_{y'}| Q^- |\Psi_{y'}\rangle| \le \|\Psi_y - \Psi_{y'}\|_1 \\
\le\ & 2\sqrt{\delta(1-\delta)}\|\,|0\rangle(\langle y| - \langle y'|)\|_1 + \delta\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_1 \\
=\ & 4\sqrt{\delta(1-\delta)}(1 - F(|y\rangle\langle y|, |y'\rangle\langle y'|)) + \delta\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_1 \\
\le\ & (2\sqrt{\delta(1-\delta)} + \delta)\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_1 \\
\le\ & (2\sqrt{\delta(1-\delta)} + \delta)\sqrt{2}\|\,|y\rangle\langle y| - |y'\rangle\langle y'|\,\|_2
\end{aligned}$$

This proves the claim. ◀

We now proceed to the main lemma of this section, proof of which is deferred to Appendix C.

▶ **Lemma 18.** *Assume the conditions $\delta > \frac{16}{\sqrt{d}}$ and $d > 10^{12}$. Let η, ε be such that $\eta < \frac{\delta}{16}$ and $\varepsilon < \frac{\delta}{100}$. Let a be the largest real that satisfies $\mathbb{E}_i 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} \le 2^{-a}$. Then it holds that $a \ge \log(\frac{d\delta(1-2\eta)^2}{50})$*

5 Proof of main result

We provide the proof of Theorem 2 in this section. It can easily be obtained from the following more general result by setting $\varepsilon = \delta^2$ and letting $\delta < \frac{1}{100}$.

▶ **Theorem 19.** *Fix a positive integer $d > 10^{12}$ and reals δ, ε that satisfy $\frac{16}{\sqrt{d}} < \delta < 1$ and $\varepsilon < \frac{\delta}{100}$. There exist a collection of $N \stackrel{\text{def}}{=} (\frac{3}{\varepsilon})^d$ states $\{|\Psi_x\rangle\}_{x=1}^N$ that depend on δ and belong to d dimensional Hilbert space, and a probability distribution $\{p(x)\}_{x=1}^N$, such that following holds for the ensemble $\{(p(x), \Psi_x)\}_{x=1}^N$.*

- *The von-Neumann entropy of the average state satisfies $S(\sum_x p(x)\Psi_x) \leq (\delta + 3\sqrt{\varepsilon}) \log(d) + H(\delta) + 1$*
- *For any one-way protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by $(1 - \sqrt{\eta})^2 \log(\frac{d\delta}{300})$.*
- *For any r-round protocol achieving the quantum Huffman coding of the above ensemble with error parameter $\eta < \frac{\delta}{16}$, the expected communication cost is lower bounded by*

$$\frac{1}{20} \cdot \frac{\log(\frac{d\delta}{400})}{(\log r)}.$$

Proof. We use the construction as given in Subsection 4.2.

For the first part of the theorem, we combine Lemma 9 and Lemma 18 to obtain a lower bound on expected communication cost as

$$(1 - \sqrt{\eta})^2 \log(\frac{d\delta(1 - 2\sqrt{\eta})^2}{50}) - 1 > (1 - \sqrt{\eta})^2 \log(\frac{d\delta}{300}).$$

The proof of second part of the theorem follows from the Reference [2] (Theorem 6.1). ◀

The proof of Theorem 3 is given in Reference [2] (Lemma 6.2).

6 Conclusion

In this work, we have shown a large gap between the quantum information complexity and the average/expected communication complexity of the quantum Huffman task (Definition 1). As an application of our main results, we show that in one-shot setting, quantum channels cannot be simulated with a cost as good as their entanglement assisted classical capacity.

We have following questions that we leave open.

- The interactive part of our main theorem, Theorem 2 has a dependence on the number of rounds. We get rid of this dependence in Theorem 3, but at the expense of weaker lower bound on expected communication cost. Can we get rid of dependence on number of rounds in Theorem 2 itself. For comparison, it may be noted that the results in [1] have no dependence on the number of rounds.
- Our lower bounds on expected communication cost and the quantum information complexity of the quantum Huffman tasks that we construct are doubly-logarithmically small in input size N, that is $\mathcal{O}(\log\log(N))$ (see Theorem 2). Can we have examples where the dependence on input size is better?
- What is the correct way to operationally understand fundamental quantum information theoretic quantities in one-shot setting? Our result says that expected communication cost is not the right notion, but naturally we cannot rule out other notions.
- Is there a way to improve the direct sum result for bounded-round entanglement assisted quantum information complexity of [22]?

Acknowledgements. A.A., A.G. and P.Y. would like to thank the Institute for Mathematical Science, Singapore, for their hospitality and their organized workshop "Semidefinite and Matrix Methods for Optimization and Communication". A.A. would like to thank the Institute for Quantum Computing, University of Waterloo, for their hospitality, where part of this work was done. We thank Dave Touchette for helpful comments on the manuscript. A.A. thanks Rahul Jain and Guo Yalei for helpful discussions. A.G. thanks Mohammed Bavarian and Henry Yuen for helpful discussions.

References

1 Anurag Anshu. A lower bound on expected communication cost of quantum state redistribution, 2015. URL: http://arxiv.org/abs/1506.06380.

2 Anurag Anshu, Ankit Garg, Aram Harrow, and Penghui Yao. Lower bound on expected communication cost of quantum huffman coding, 2016. URL: http://arxiv.org/abs/1605.04601.

3 Anurag Anshu, Rahul Jain, Priyanka Mukhopadhyay, Ala Shayeghi, and Penghui Yao. A new operational interpretation of relative entropy and trace distance between quantum states, 2014. URL: http://arxiv.org/abs/1404.1366.

4 Boaz Barak, Mark Braverman, Xi Chen, and Anup Rao. How to compress interactive communication. In *Proceedings of the forty-second ACM symposium on Theory of computing*, STOC'10, pages 67–76, New York, NY, USA, 2010. ACM.

5 Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Phys. Rev. Lett.*, 76:2818–2821, Apr 1996. doi:10.1103/PhysRevLett.76.2818.

6 M. Berta, M. Christandl, and D. Touchette. Smooth entropy bounds on one-shot quantum state redistribution. *IEEE Transactions on Information Theory*, 62(3):1425–1439, March 2016. doi:10.1109/TIT.2016.2516006.

7 S. L. Braunstein, C. A. Fuchs, D. Gottesman, and Hoi-Kwong Lo. A quantum analog of huffman coding. In *Information Theory, 1998. Proceedings. 1998 IEEE International Symposium on*, pages 353–, Aug 1998. doi:10.1109/ISIT.1998.708958.

8 Mark Braverman and Anup Rao. Information equals amortized communication. In *Proceedings of the 52nd Symposium on Foundations of Computer Science*, FOCS'11, pages 748–757, Washington, DC, USA, 2011. IEEE Computer Society.

9 Igor Devetak and Jon Yard. Exact cost of redistributing multipartite quantum states. *Phys. Rev. Lett.*, 100:230501, Jun 2008. doi:10.1103/PhysRevLett.100.230501.

10 M. Fannes. A continuity property of the entropy density for spin lattice systems. *Communications in Mathematical Physics*, 31:291–294, 1973.

11 P. Harsha, R. Jain, D. McAllester, and J. Radhakrishnan. The communication complexity of correlation. *IEEE Transactions on Information Theory*, 56(1):438–449, Jan 2010. doi:10.1109/TIT.2009.2034824.

12 Alexander S. Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problems of Information Transmission*, (9):177–183, 1973.

13 Michał Horodecki, Jonathan Oppenheim, and Andreas Winter. Quantum state merging and negative information. *Communications in Mathematical Physics*, 269:107–136, 2007. doi:10.1007/s00220-006-0118-x.

14 David Huffman. A method for the construction of minimum-redundancy codes. *Proceedings of IRE*, 40(9):1098–1101, 1952.

15 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Prior entanglement, message compression and privacy in quantum communication. In *Proceedings of the 20th Annual IEEE*

Conference on Computational Complexity, pages 285–296, Washington, DC, USA, 2005. IEEE Computer Society. doi:10.1109/CCC.2005.24.

16 Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity, 2008. URL: http://arxiv.org/abs/0807.1267.

17 Michel Ledoux. The concentration of measure phenomenon. Mathematical Surveys and Monographs. American Mathematical Society, 2005.

18 G. Lindblad. Completely positive maps and entropy inequalities. *Commun. Math. Phys.*, 40:147–151, 1975.

19 Claude Elwood Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423, 1948. doi:10.1002/j.1538-7305.1948.tb01338.x.

20 D. Slepian and J. Wolf. Noiseless coding of correlated information sources. *IEEE Transactions on Information Theory*, 19(4):471–480, Jul 1973. doi:10.1109/TIT.1973.1055037.

21 Marco Tomamichel. A framework for non-asymptotic quantum information theory, 2012. PhD Thesis, ETH Zurich. URL: arXiv:1203.2142.

22 Dave Touchette. Quantum information complexity. In *Proceedings of the Forty-Seventh Annual ACM on Symposium on Theory of Computing*, STOC'15, pages 317–326, New York, NY, USA, 2015. ACM. doi:10.1145/2746539.2746613.

23 Roman Vershynin. Introduction to the non-asymptotic analysis of random matrices. In *Compressed Sensing, Theory and Applications*. Cambridge University Press, 2012.

24 John Watrous. Theory of Quantum Information, lecture notes, 2011. URL: https://cs.uwaterloo.ca/~watrous/LectureNotes.html.

25 A. Wyner. The common information of two dependent random variables. *IEEE Transactions on Information Theory*, 21(2):163–179, Mar 1975. doi:10.1109/TIT.1975.1055346.

26 J. T. Yard and I. Devetak. Optimal quantum source coding with quantum side information at the encoder and decoder. *IEEE Transactions on Information Theory*, 55(11):5339–5351, Nov 2009.

A Proof of Lemma 9

Proof. Our proof shall proceed in the steps outlined below.

1. **Pruning away x with low fidelity:**
   Let $\mathcal{G}$ be the set of all x such that $\sum_i p_i^x \langle\Psi_x| \sigma_i^x |\Psi_x\rangle \geq 1-\eta^{3/2}$. Let $\mathcal{B}$ be the set of rest of x. Then we have that $\sum_{x\in\mathcal{G}} p(x) \geq 1-\sqrt{\eta}$ and equivalently $\sum_{x\in\mathcal{B}} p(x) \leq \sqrt{\eta}$.
   Define a new probability distribution $p'(x)$ which is 0 whenever $x \in \mathcal{B}$ and equal to $\frac{p(x)}{\sum_{x\in\mathcal{G}} p(x)}$ for $x \in \mathcal{G}$. Since $\sum_{x\in\mathcal{G}} p(x) \geq 1-\sqrt{\eta}$, it holds that $p'(x) \leq \frac{p(x)}{1-\sqrt{\eta}}$ for all x.
2. **Upper bound on probabilities p_i^x:**
   We upper bound the probabilities p_i^x in the following way. Consider,

$$\theta_B = \mathrm{Tr}_A(M_i^x \theta_{AB}) + \mathrm{Tr}_A((\mathrm{I}-M_i^x)\theta_{AB}) > \mathrm{Tr}_A(M_i^x\theta_{AB}).$$

   Thus,

$$p_i^x \rho_i^x < \theta_B \implies \rho_i^x < \frac{1}{p_i^x}\theta_B.$$

   By definition of max-entropy, this means $2^{\mathrm{D}_{\max}(\rho_i^x\|\theta_B)} < \frac{1}{p_i^x}$. Now we use monotonicity of max-entropy under quantum operations (Fact 5), to obtain

$$p_i^x < 2^{-\mathrm{D}_{\max}(\rho_i^x\|\theta_B)} < 2^{-\mathrm{D}_{\max}(\sigma_i^x\|\mathcal{E}_i(\theta_B))}. \tag{2}$$

3. **Upper bound on probability of each message:**
For every $x \in \mathcal{G}$, let $\mathcal{B}_x$ be set of i such that $\langle\Psi_x| \sigma_i^x |\Psi_x\rangle < 1-\eta$. Let $\mathcal{G}_x$ be rest of the indices. Using the relation

$$\sum_i p_i^x (1 - \langle\Psi_x| \sigma_i^x |\Psi_x\rangle) < \eta^{3/2},$$

we obtain that $\sum_{i\in\mathcal{B}_x} p_i^x < \sqrt{\eta}$. Define a new probability distribution q_i^x which is 0 whenever $i \in \mathcal{B}_x$ and equal to $\frac{p_i^x}{\sum_{i\in\mathcal{G}_x}}$ otherwise.

Define $s_i \stackrel{\text{def}}{=} \sum_x p'(x) q_i^x$. Note that by definition, $\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B)) < \mathrm{D}_{\max}(\sigma_i^x \| \mathcal{E}_i(\theta_B))$ for all $i \in \mathcal{G}_x$. Using Equation 2, we observe that for all $x \in \mathcal{G}$ it holds that

$$q_i^x < \frac{1}{1-\sqrt{\eta}} 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))}.$$

This implies

$$\begin{aligned} s_i &= \sum_x p'(x) q_i^x \\ &\leq \frac{1}{1-\sqrt{\eta}} \sum_x p'(x) 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))} \\ &\leq \frac{1}{(1-\sqrt{\eta})^2} \sum_x p(x) 2^{-\mathrm{D}_{\max}^{\eta}(\Psi^x \| \mathcal{E}_i(\theta_B))} \\ &= \frac{t_i}{(1-\sqrt{\eta})^2} < \frac{2^{-a}}{(1-\sqrt{\eta})^2} \end{aligned} \tag{3}$$

where in first inequality, we have used the fact that for $x \in \mathcal{B}$, $p'(x) = 0$.

4. **Lower bound on expected communication:**
Since $p_i^x > (1-\sqrt{\eta}) q_i^x$ for all pair (x, i) such that $x \in \mathcal{G}$, the expected communication cost is lower bounded by

$$\sum_x p(x) \sum_i p_i^x \log(i) > (1-\sqrt{\eta}) \sum_{x\in\mathcal{G}} p(x) \sum_i q_i^x \log(i) > (1-\sqrt{\eta})^2 \sum_x p'(x) \sum_i q_i^x \log(i).$$

From Equation 3, we have $s_i \leq \frac{2^{-a}}{(1-\eta)^2}$ and $\sum_i s_i = 1$. Thus, the quantity $\sum_i s_i \log(i)$ is minimized if $s_i = \frac{2^{-a}}{(1-\sqrt{\eta})^2}$ for all $i \leq 2^a(1-\sqrt{\eta})^2$. This gives following lower bound on expected communication cost

$$(1-\sqrt{\eta})^2 \cdot \frac{2^{-a}}{(1-\sqrt{\eta})^2} 2^a (1-\sqrt{\eta})^2 \log(2^a(1-\sqrt{\eta})^2/e) > (1-\sqrt{\eta})^2 \cdot a - 1. \qquad \blacktriangleleft$$

B Proof of Claims 14 and 15

Proof of Claim 14 . For a fixed i, let ρ_i be the state that achieves the infimum in the definition of $\mathrm{D}_{\max}^{\eta}(\Psi_i \| \sigma)$. It satisfies $\langle\Psi_i| \rho_i |\Psi_i\rangle \geq 1-\eta$. This means the largest eigenvalue of ρ_i is at least $1-\eta$. Thus, consider the eigen-decomposition $\rho_i = \lambda_1 |\lambda_1\rangle\langle\lambda_1| + \sum_{j>1} \lambda_j |\lambda_j\rangle\langle\lambda_j|$. We have $\lambda_1 > 1-\eta$ or equivalently $\sum_{j>1} \lambda_j < \eta$. Thus,

$$1-\eta < \langle\Psi_i| \rho_i |\Psi_i\rangle = \lambda_1 |\langle\Psi_i|\lambda_1\rangle|^2 + \sum_{j>1} \lambda_j |\langle\Psi_i|\lambda_j\rangle|^2 < |\langle\Psi_i|\lambda_1\rangle|^2 + \sum_{j>1} \lambda_j < |\langle\Psi_i|\lambda_1\rangle|^2 + \eta.$$

Hence, $|\langle\Psi_i|\lambda_1\rangle|^2 > 1-2\eta$. Moreover,

$$2^{\mathrm{D}_{\max}(\rho_i\|\sigma)} = \|\sigma^{-\frac{1}{2}} \rho \sigma^{-\frac{1}{2}}\|_\infty > (1-\eta) \|\sigma^{-\frac{1}{2}} |\lambda_1\rangle\langle\lambda_1| \sigma^{-\frac{1}{2}}\|_\infty = (1-\eta) \langle\lambda_1| \sigma^{-1} |\lambda_1\rangle,$$

where σ^{-1} is the pseudo-inverse of σ. From the definition of the projector Q^-, the following inequality easily follows:

$$\langle\lambda_1|\,\sigma^{-1}\,|\lambda_1\rangle \geq k\,\langle\lambda_1|\,Q^-\,|\lambda_1\rangle\,.$$

Thus we get

$$2^{\mathrm{D}_{\max}(\rho_i\|\sigma)} > k(1-\eta)\,\langle\lambda_1|\,Q^-\,|\lambda_1\rangle\,.$$

Inverting and using $|\langle\Psi_i|\lambda_1\rangle|^2 > 1-2\eta$, we have

$$2^{-\mathrm{D}_{\max}(\rho_i\|\sigma)} < \frac{1}{k(1-\eta)\,\langle\lambda_1|\,Q^-\,|\lambda_1\rangle} < \frac{1}{k(1-\eta)S^{2\eta}(\Psi_i||Q^-)}.$$

This proves the claim. ◀

Proof of Claim 15. Let $|\lambda_i\rangle$ be the state that achieves the infimum in the definition of $S^\eta(\Psi_i||Q^-)$. We know that $|\lambda_i\rangle$ has fidelity at least $1-\eta$ with $|\Psi_i\rangle$ and also minimizes the overlap with the subspace Q^-. Intuitively, this state must lie in the span of two vectors $\{Q^-\,|\Psi_i\rangle\,, Q^+\,|\Psi_i\rangle\}$. This we shall find to be true below.

Let us expand

$$|\lambda_i\rangle = aQ^-\,|\Psi_i\rangle + bQ^+\,|\Psi_i\rangle + c\,|\theta\rangle\,,$$

where $|\theta\rangle$ is normalized vector orthogonal to $\{Q^-\,|\Psi_i\rangle\,, Q^+\,|\Psi_i\rangle\}$. Then we have the conditions:

$$|a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + |b|^2\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle + |c|^2 = 1, \quad |a\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + b\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle\,| > \sqrt{1-\eta} \tag{4}$$

where the first condition is normalization condition and second condition says that overlap between $|\lambda_i\rangle$ and $|\Psi_i\rangle$ is at least $\sqrt{1-\eta}$. We would like to minimize the function

$$\langle\lambda_i|\,Q^-\,|\lambda_i\rangle = \langle\lambda_i|\,(aQ^-\,|\Psi_i\rangle + cQ^-\,|\theta\rangle) = |a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + |c|^2\,\langle\theta|\,Q^-\,|\theta\rangle \tag{5}$$

Note that $\langle\Psi_i|\,Q^-Q^-\,|\theta\rangle = 0$, hence the above expression.

First we shall show that a, b, c can be chosen to be real. Clearly c can be chosen real as it only appears as $|c|^2$. Only place where a, b appear as complex is in the constraint $|a\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + b\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle\,| > \sqrt{1-\eta}$. Let $a = a_R + ia_I, b = b_R + ib_I$. Then

$$\begin{aligned}
& |a\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + b\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle\,|^2 \\
= \;& (a_R\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + b_R\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle)^2 + (a_I\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + b_I\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle)^2 \\
= \;& |a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle^2 + |b|^2\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle^2 + 2(a_Rb_R + a_Ib_I)\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle \\
\leq \;& |a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle^2 + |b|^2\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle^2 + 2(\sqrt{a_R^2+a_I^2}\sqrt{a_I^2+b_I^2})\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle \\
= \;& |a|^2\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle^2 + |b|^2\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle^2 + 2|a||b|\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle \\
= \;& (|a|\,\langle\Psi_i|\,Q^-\,|\Psi_i\rangle + |b|\,\langle\Psi_i|\,Q^+\,|\Psi_i\rangle)^2
\end{aligned}$$

Thus, changing the complex coefficients a, b to $|a|, |b|$ does not change the objective function (Equation 5) and ensures that the constraints (Equation 4) are still satisfied. Thus, we can restrict ourselves to real variables a, b.

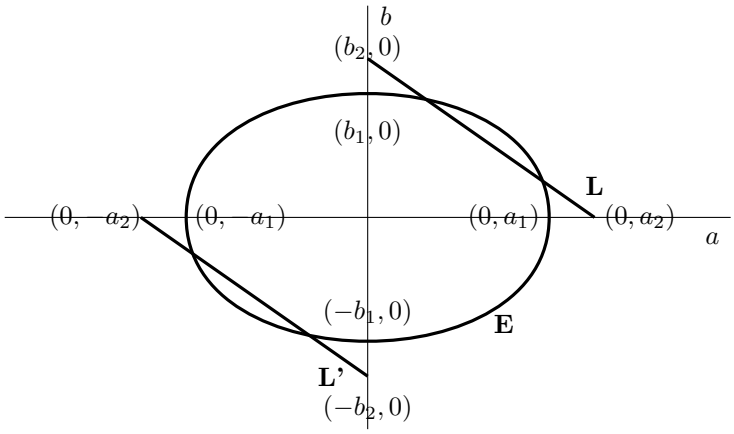

Figure 1 Plot of the constraints.

To find the optimal solution for equations 4 and 5, we fix a c and minimize a^2 with the constraints

$$a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle = 1 - c^2, \quad |a \langle\Psi_i| Q^- |\Psi_i\rangle + b \langle\Psi_i| Q^+ |\Psi_i\rangle| > \sqrt{1-\eta}.$$

We plot these constraints on (a, b) plane in Figure 1. The ellipse

$$E \stackrel{\text{def}}{=} a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle = 1 - c^2$$

intersects a-axis at $|a_1| = \sqrt{\frac{1-c^2}{\langle\Psi_i|Q^-|\Psi_i\rangle}}$ and intersects b-axis at $|b_1| = \sqrt{\frac{1-c^2}{\langle\Psi_i|Q^+|\Psi_i\rangle}}$. The lines

$$L \stackrel{\text{def}}{=} a \langle\Psi_i| Q^- |\Psi_i\rangle + b \langle\Psi_i| Q^+ |\Psi_i\rangle = \sqrt{1-\eta},$$
$$L' \stackrel{\text{def}}{=} a \langle\Psi_i| Q^- |\Psi_i\rangle + b \langle\Psi_i| Q^+ |\Psi_i\rangle = -\sqrt{1-\eta}$$

intersect a-axis at $|a_2| = \frac{\sqrt{1-\eta}}{\langle\Psi_i|Q^-|\Psi_i\rangle}$ and intersects b-axis at $|b_2| = \frac{\sqrt{1-\eta}}{\langle\Psi_i|Q^+|\Psi_i\rangle}$.

First note that if $c^2 > \eta$, then there is no solution. For this, consider

$$\begin{aligned} 1 - \eta &< (a \langle\Psi_i| Q^- |\Psi_i\rangle + b \langle\Psi_i| Q^+ |\Psi_i\rangle)^2 \\ &\leq (\langle\Psi_i| Q^- |\Psi_i\rangle + \langle\Psi_i| Q^+ |\Psi_i\rangle)(a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle) \\ &= (a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle) = 1 - c^2. \end{aligned}$$

So we assume that $c^2 \leq \eta$. Now lets focus on first quadrant. We can easily observe from the plot that we get $a = 0$ as minimum value of a^2 whenever ellipse E intersects b-axis above the line L. This occurs when

$$\sqrt{\frac{1-c^2}{\langle\Psi_i| Q^+ |\Psi_i\rangle}} > \frac{\sqrt{1-\eta}}{\langle\Psi_i| Q^+ |\Psi_i\rangle} \rightarrow \langle\Psi_i| Q^+ |\Psi_i\rangle > \frac{1-\eta}{1-c^2}.$$

But this is obvious, since the condition implies $\langle\Psi_i| Q^+ |\Psi_i\rangle > 1 - \eta$ in which case there is a vector in Q^+ with high overlap with $|\Psi_i\rangle$ and hence the objective function is 0.

So lets assume that $\langle\Psi_i| Q^+ |\Psi_i\rangle < 1 - \eta$, in which case, for all c, the ellipse E intersects b-axis below the line L. To find the point of intersection, we simultaneously solve the equations for line and ellipse, that is

$$a^2 \langle\Psi_i| Q^- |\Psi_i\rangle + b^2 \langle\Psi_i| Q^+ |\Psi_i\rangle = 1 - c^2, \quad a \langle\Psi_i| Q^- |\Psi_i\rangle + b \langle\Psi_i| Q^+ |\Psi_i\rangle = \sqrt{1-\eta}.$$

The value of a, b thus obtained is

$$a = \sqrt{1-\eta} - \sqrt{\frac{\langle\Psi_i| Q^+ |\Psi_i\rangle (\eta - c^2)}{\langle\Psi_i| Q^- |\Psi_i\rangle}}, \quad b = \sqrt{1-\eta} + \sqrt{\frac{\langle\Psi_i| Q^- |\Psi_i\rangle (\eta - c^2)}{\langle\Psi_i| Q^+ |\Psi_i\rangle}}.$$

It is easy to verify that the solution satisfies above equations. The other solution is with signs reversed.

Thus, we have the result that whenever $\langle\Psi_i| Q^+ |\Psi_i\rangle < 1-\eta$, the minimum $|a|^2 \langle\Psi_i| Q^- |\Psi_i\rangle + |c|^2 \langle\theta| Q^- |\theta\rangle$ is

$$(\sqrt{1-\eta} - \sqrt{\frac{\langle\Psi_i| Q^+ |\Psi_i\rangle (\eta - c^2)}{\langle\Psi_i| Q^- |\Psi_i\rangle}})^2 \langle\Psi_i| Q^- |\Psi_i\rangle + c^2 \langle\theta| Q^- |\theta\rangle .$$

This quantity is monotonically increasing with c. Hence above expression is minimized when $c = 0$. This justifies our intuition that the optimal vector lies in the plane $\{Q^+ |\Psi_i\rangle , Q^- |\Psi_i\rangle\}$. With this, we have found an overall minimum to be

$$(\sqrt{1-\eta} - \sqrt{\frac{\langle\Psi_i| Q^+ |\Psi_i\rangle \eta}{\langle\Psi_i| Q^- |\Psi_i\rangle}})^2 \langle\Psi_i| Q^- |\Psi_i\rangle = (\sqrt{(1-\eta)\langle\Psi_i| Q^- |\Psi_i\rangle} - \sqrt{\langle\Psi_i| Q^+ |\Psi_i\rangle \eta})^2.$$

This proves the claim. ◀

C Proof of Lemma 18

Proof. Proof shall proceed in two stages.

1. **Concentration of measure for Epsilon-nets:**

   From Claim 17, we infer that the Lipschitz constant of the function $f(|y\rangle) \stackrel{\text{def}}{=} \langle\Psi_y| Q^- |\Psi_y\rangle$ is upper bounded by $4\sqrt{\delta}$. From Lemma 16, we have that $\int \mu(y) dy f(y) = (1 - \delta - \frac{\delta}{d}) \langle 0| Q^- |0\rangle + \delta(1 - \frac{k-1}{d})$.

   Let α be a positive real to be chosen later. It now follows from Levy's concentration lemma (Fact 8) that

   $$\Pr_\mu(\langle\Psi_y| Q^- |\Psi_y\rangle < \int \mu(y) dy f(y) - \alpha) \leq e^{-\frac{d\alpha^2}{18\pi^3 \cdot 16\delta}} = e^{-\frac{d\alpha^2}{288\pi^3\delta}} \tag{6}$$

   In other words,

   $$\Pr_\mu(\langle\Psi_y| Q^- |\Psi_y\rangle < \delta(1 - \frac{k-1}{d}) - \alpha) \leq e^{-\frac{d\alpha^2}{288\pi^3\delta}}.$$

   Now, let S be the set of all $|y\rangle \in S^d$ for which $\langle\Psi_y| Q^- |\Psi_y\rangle > \delta(1 - \frac{k-1}{d}) - \alpha$. Let $\mathcal{G}$ be the set of all i such that S_i has an intersection with S. Let $T \stackrel{\text{def}}{=} \cup_{i\in\mathcal{G}} S_i$. Then T contains S, except for some points of measure zero, and furthermore from Claim 17, any $|z\rangle \in T$ satisfies

   $$\langle\Psi_z| Q^- |\Psi_z\rangle \geq \delta(1 - \frac{k-1}{d}) - \alpha - 4\sqrt{\delta}\varepsilon > \delta(1 - \frac{k-1}{d}) - \alpha - 2\varepsilon.$$

   Since $\mu(T) > (1 - e^{-\frac{d\alpha^2}{288\pi^3\delta}})$, and T is a union of S_i with $i \in \mathcal{G}$, it holds that for an i drawn according to $\lambda(i)$, probability that $i \in \mathcal{G}$ is equal to $\mu(T)$ and hence at least $(1 - e^{-\frac{d\alpha^2}{288\pi^3\delta}})$. Thus we have show the following inequality

   $$\Pr_\lambda(\langle\Psi_i| Q^- |\Psi_i\rangle \geq \delta(1 - \frac{k-1}{d}) - \alpha - 2\varepsilon) \geq 1 - 2e^{-\frac{d\alpha^2}{288\pi^3\delta}}, \tag{7}$$

2. Using concentration of measure for upper bound:
Now, to evaluate $\mathbb{E}_i 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)}$, we divide the expectation into two parts. For all i for which $\langle\Psi_i| Q^- |\Psi_i\rangle < \delta(1-\frac{k}{d}) - \alpha - 2\varepsilon$, we upper bound $2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} < 1$. For the rest of i, we use Lemma 13 to obtain

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} < \frac{1}{k(1-\eta)(\sqrt{(1-2\eta)(\delta(1-\frac{k}{d})-\alpha-2\varepsilon)} - \sqrt{2(1-\delta(1-\frac{k}{d})+\alpha+2\varepsilon)\eta})^2}$$

Note that for this to hold, we need $\delta(1-\frac{k}{d}) - \alpha - 2\varepsilon > 2\eta$. For this, we set $k = \frac{d}{4}, \alpha^2 = \frac{\delta}{\sqrt{d}} < \frac{\delta^2}{16}$ and we can upper bound $\delta(1-\frac{k}{d}) - \alpha - 2\varepsilon > \frac{\delta}{4} > 2\eta$, using $\varepsilon < \frac{\delta}{100}, \eta < \frac{\delta}{16}$ (assumptions of theorem). Then we have

$$2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} \leq \frac{4}{d(1-\eta)(\sqrt{(1-2\eta)(\delta/4)} - \sqrt{2(1-\delta/4)\eta})^2} \leq \frac{40}{d(1-\eta)\delta}.$$

Thus we get

$$\begin{aligned}
\mathbb{E}_i 2^{-\mathrm{D}^{\eta}_{\max}(\Psi_i\|\sigma)} &< 2e^{-\frac{d\alpha^2}{144\pi^3\delta}} + \frac{40}{d(1-2\eta)^2\delta} \\
&= 2e^{-\frac{\sqrt{d}}{288\pi^3}} + \frac{40}{d(1-2\eta)^2\delta} \\
&< \frac{50}{d\delta(1-2\eta)^2} = 2^{-\log(\frac{d\delta(1-2\eta)^2}{50})}
\end{aligned}$$

Last inequality holds for $d > 10^{12}$. This proves the theorem. ◀

Simple, Near-Optimal Quantum Protocols for Die-Rolling

Jamie Sikora[*†]

Centre for Quantum Technologies, National University of Singapore, Singapore; and
MajuLab, CNRS-UNS-NUS-NTU International Joint Research Unit, UMI 3654, Singapore
cqtjwjs@nus.edu.sg

Abstract

Die-rolling is the cryptographic task where two mistrustful, remote parties wish to generate a random D-sided die-roll over a communication channel. Optimal quantum protocols for this task have been given by Aharon and Silman (New Journal of Physics, 2010) but are based on optimal weak coin-flipping protocols which are currently very complicated and not very well understood. In this paper, we first present very simple classical protocols for die-rolling which have decent (and sometimes optimal) security which is in stark contrast to coin-flipping, bit-commitment, oblivious transfer, and many other two-party cryptographic primitives. We also present quantum protocols based on the idea of integer-commitment, a generalization of bit-commitment, where one wishes to commit to an integer. We analyze these protocols using semidefinite programming and finally give protocols which are very close to Kitaev's lower bound for any $D \geq 3$.

1998 ACM Subject Classification D.4.6 Security and Protection, G.1.6 Optimization

Keywords and phrases Quantum Cryptography, Semidefinite Programming, Die-Rolling, Integer-Commitment

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.4

1 Introduction

Die-rolling is the two-party cryptographic primitive in which two spatially separated parties, Alice and Bob, wish to agree upon an integer $d \in [D] := \{1, \ldots, D\}$, generated uniformly at random, over a communication channel. When designing die-rolling protocols, the security goals are:

1. *Completeness:* If both parties are honest, then their outcomes are the same, uniformly random, and neither party aborts.
2. *Soundness against cheating Bob:* If Alice is honest, then a dishonest (i.e., cheating) Bob cannot influence her protocol outcome away from uniform.
3. *Soundness against cheating Alice:* If Bob is honest, then a dishonest (i.e., cheating) Alice cannot influence his protocol outcome away from uniform.

* J.S. is supported in part by NSERC Canada.
Research at the Centre for Quantum Technologies at the National University of Singapore is partially funded by the Singapore Ministry of Education and the National Research Foundation, also through the Tier 3 Grant "Random numbers from quantum processes," (MOE2012-T3-1-009).
† A full version of the paper is available at https://arxiv.org/abs/1605.08156.

We note here that Alice and Bob start uncorrelated and unentangled. Otherwise, Alice and Bob could each start with half of the following maximally entangled state

$$\frac{1}{\sqrt{D}} \sum_{d \in [D]} |d\rangle_{\mathcal{A}} |d\rangle_{\mathcal{B}}$$

and measure in the computational basis to obtain a perfectly correlated, uniformly random die-roll. Thus, such a primitive would be trivial if they were allowed to start entangled.

Die-rolling is a generalization of a well-studied primitive known as *coin-flipping* [5] which is the special case of die-rolling when $D = 2$. In this paper, we analyze die-rolling protocols in a similar fashion that is widely adopted for coin-flipping protocols [3, 17, 13, 16, 8, 18, 19]. That is, we assume perfect completeness and calculate the soundness in terms of the *cheating probabilities*, as defined by the symbols:

$P^*_{B,d}$: The maximum probability with which a dishonest Bob can force an honest Alice to accept the outcome $d \in [D]$ by digressing from protocol.

$P^*_{A,d}$: The maximum probability with which dishonest Alice can force an honest Bob to accept the outcome $d \in [D]$ by digressing from protocol.

We are concerned with designing protocols which minimize the maximum of these $2D$ quantities since a protocol is only as good as its worst cheating probability. Coincidentally, all the protocols we consider in this paper have the property that all of Alice's cheating probabilities are equal and similarly for a cheating Bob. Therefore, for brevity, we introduce the following shorthand notation:

$$P^*_A := \max\{P^*_{A,1}, \ldots, P^*_{A,D}\} \quad \text{and} \quad P^*_B := \max\{P^*_{B,1}, \ldots, P^*_{B,D}\}.$$

When $D = 2$, the security definition for die-rolling above aligns with that of *strong* coin-flipping. For strong coin-flipping, it was shown by Kitaev [14] that any quantum protocol satisfies $P^*_{A,1}P^*_{B,1} \geq 1/2$ and $P^*_{A,2}P^*_{B,2} \geq 1/2$, implying that at least one party can cheat with probability at least $1/\sqrt{2}$. It was later shown by Chailloux and Kerenidis [8] that all four cheating probabilities can be made arbitrarily close to $1/\sqrt{2}$ by using optimal quantum protocols for *weak* coin-flipping as discovered by Mochon [16].

As pointed out in [1], Kitaev's proof for the lower bound on coin-flipping extends naturally to die-rolling; it can be shown that for any quantum die-rolling protocol, we have

$$P^*_{A,d}P^*_{B,d} \geq \frac{1}{D}$$

for any $d \in [D]$. This implies the lower bound $\max\{P^*_A, P^*_B\} \geq 1/\sqrt{D}$. In fact, extending the optimal coin-flipping protocol construction in [8], it was shown by Aharon and Silman [1] that for $D > 2$, it is possible to find quantum protocols where the maximum of the $2D$ probabilities is at most $1/\sqrt{D} + \delta$, for any $\delta > 0$.

The optimal protocols in [8] and [1] are not explicit as they rely on using Mochon's optimal weak coin-flipping protocols as subroutines. Moreover, Mochon's protocols are very complicated and not given explicitly, although they have been simplified [2].

The best known *explicit* quantum protocol for die-rolling[1] of which we are aware is given in [1]. It uses three messages and has cheating probabilities

$$P^*_A := \frac{D+1}{2D} \quad \text{and} \quad P^*_B := \frac{2D-1}{D^2}.$$

[1] The protocols considered in this paper have a much different form than these protocols.

These probabilities have the attractive property of approximating Kitaev's lower bound in the limit, but since $P_A^* \to 1/2$ as $D \to \infty$, the maximum cheating probability is quite large.

This motivates the work in this paper which is to find simple and explicit protocols for die-rolling that approximate Kitaev's lower bound on the maximum cheating probability

$$\max\{P_A^*, P_B^*\} \geq 1/\sqrt{D}.$$

1.1 Simple classical protocols

We first show that simple classical protocols exist with decent security.

▶ **Protocol 1** (Classical protocol)**.**

- *Bob chooses a subset $S \subseteq [D]$ with $|S| = m$, uniformly at random, and sends S to Alice. If $|S| \neq m$, Alice aborts.*
- *Alice selects $d \in S$ uniformly at random and tells Bob her selection. If $d \notin S$, Bob aborts.*
- *Both parties output d.*

We see that this is a valid die-rolling protocol as each party outputs the same value $d \in [D]$ and each value occurs with equal probability. As for the cheating probabilities, it is straightforward to see that

$$P_A^* = \frac{m}{D} \quad \text{and} \quad P_B^* = \frac{1}{m}.$$

Besides being extremely simple, this protocol has the following interesting properties:

- The product $P_{A,d}^* P_{B,d}^* = 1/D$, for any $d \in [D]$, saturates Kitaev's lower bound for every $d \in [D]$.
- For D square and $m = \sqrt{D}$, we have $P_A^* = P_B^* = 1/\sqrt{D}$, yielding an optimal protocol!
- If D is not square, then one party has a cheating advantage, i.e., $P_A^* \neq P_B^*$.

Note that to minimize $\max\{P_A^*, P_B^*\}$, it does not make sense to choose large m (greater than $\lceil\sqrt{D}\rceil$) or small m (less than $\lfloor\sqrt{D}\rfloor$). We can see that for $D = 3$, $D = 7$, or $D = 8$, for example, that choosing the ceiling is better while for $D = 5$ or $D = 10$ choosing the floor is better. Thus, we keep both the cases and summarize the overall security of the above protocol in the following lemma.

▶ **Lemma 2.** *For $D \geq 2$, there exists a classical die-rolling protocol satisfying*

$$\frac{1}{\sqrt{D}} \leq \max\{P_A^*, P_B^*\} = \min\left\{\frac{\lceil\sqrt{D}\rceil}{D}, \frac{1}{\lfloor\sqrt{D}\rfloor}\right\} \tag{1}$$

which is optimal when D is square.

Note that the special case of $D = 2$ has either Alice or Bob able to cheat perfectly, which is the case for all classical coin-flipping protocols. However, Kitaev's bound on the product of cheating probabilities is still (trivially) satisfied. For $D = 3$, we can choose $m = 2$ to obtain $\max\{P_A^*, P_B^*\} = 2/3$ proving that even classical protocols can have nontrivial security, which is vastly different than the $D = 2$ case. The values from (1) for $D \in \{2, \ldots, 10\}$ are later presented in Table 1.

We are not aware of other lower bounds for classical die-rolling protocols apart from those implied by Kitaev's bounds above. We see that sometimes classical protocols can be optimal, for example when D is square. We now consider how to design (simple) quantum protocols and see what levels of security they can offer.

1.2 Simple quantum protocols

Many of the best known explicit protocols for strong coin-flipping are based on the idea of *bit-commitment* [4, 20, 13, 19]. Optimal protocols are known for bit-commitment as well [9], but are again based on weak coin-flipping and are thus very complicated.

In this paper, we generalize the above simple, explicit protocols such that Alice commits to an *integer* instead of a bit. More precisely, our quantum protocols have the following form.

▶ **Protocol 3** (Quantum protocol)**.** *A quantum die-rolling protocol based on the idea of integer-commitment, denoted here as* DRIC*, is defined as follows:*

- *Alice chooses a random* $a \in [D]$ *and creates the state*

$$|\psi_a\rangle \in \mathcal{A} \otimes \mathcal{B}$$

  and sends the subsystem $\mathcal{B}$ *to Bob.*
- *Bob sends a uniformly random* $b \in [D]$ *to Alice.*
- *Alice reveals* a *to Bob and sends him the subsystem* $\mathcal{A}$*.*
- *Bob checks if* $\mathcal{A} \otimes \mathcal{B}$ *is in state* $|\psi_a\rangle$ *using the measurement*

$$\{\Pi_a := |\psi_a\rangle\langle\psi_a|, \Pi_{abort} := I - \Pi_a\}.$$

  Bob accepts/rejects a *based on his measurement outcome.*
- *If Bob does not abort, Alice and Bob output*

$$d := (a + b) \bmod D + 1 \in [D].$$

The special case of $D = 2$ yields the structure of the simple, explicit coin-flipping protocols mentioned above. Indeed, these protocols are very easy to describe, one needs only the knowledge of the D states $|\psi_a\rangle$ and, implicitly, the systems they act on, $\mathcal{A}$ and $\mathcal{B}$.

We start by formulating the cheating probabilities of a DRIC-protocol using semidefinite programming. Once we have established the semidefinite programming cheating strategy formulations, we are able to analyze the security of DRIC-protocols. Furthermore, we are able to analyze *modifications* to such protocols and the corresponding changes in security.

In this paper, we present a DRIC-protocol with near-optimal security. We develop this protocol in several steps described below.

The first step is to start with a protocol with decent security. To do this, we show how to create a DRIC-protocol with the same cheating probabilities as Protocol 1.

▶ **Proposition 4.** *There exists a* DRIC*-protocol with the same cheating probabilities as in Protocol 1.*

The second step is to give a process which (approximately) balances the maximum cheating probabilities of Alice and Bob. We accomplish this by modifying the protocol in order to decrease the overall maximum cheating probability (while possibly increasing lesser cheating probabilities).

▶ **Proposition 5.** *If there exists a* DRIC*-protocol with cheating probabilities* $P_A^* = \alpha$ *and* $P_B^* = \beta$*, then there exists a* DRIC*-protocol with maximum cheating probability*

$$\max\{P_A^*, P_B^*\} \leq \frac{D\max\{\beta,\alpha\} - \min\{\beta,\alpha\}}{D|\beta - \alpha| + D - 1} \leq \max\{\beta,\alpha\}.$$

Moreover, the last inequality is strict when $\alpha \neq \beta$ *yielding a strictly better protocol.*

Table 1 Values of our bounds (as truncated percentages) for various protocols and values of D. We see that the quantum protocol performs very well, even for D as small as 3.

D	2	3	4	5	6	7	8	9	10
Explicit Protocol in [1]	75%	66%	62%	60%	58%	57%	56%	55%	55%
Our Classical Protocol	100%	66%	50%	50%	50%	42%	37%	33%	33%
Our Quantum Protocol	75%	60%	50%	46%	44%	40%	36%	33%	32%
Kitaev's lower bound	70%	57%	50%	44%	40%	37%	35%	33%	31%

By combining the above two propositions, we are able to obtain the main result of this paper.

▶ **Theorem 6.** *For any $D \geq 2$, there exists a (quantum)* DRIC*-protocol satisfying*

$$\frac{1}{\sqrt{D}} \leq \max\{P_A^*, P_B^*\} \leq \min\left\{\frac{D+\lfloor\sqrt{D}\rfloor}{D(\lfloor\sqrt{D}\rfloor+1)}, \frac{1+\lceil\sqrt{D}\rceil}{D+\lceil\sqrt{D}\rceil}\right\}$$

which is strictly better than Protocol 1 when D is not square.

Since $\min\left\{\frac{D+\lfloor\sqrt{D}\rfloor}{D(\lfloor\sqrt{D}\rfloor+1)}, \frac{1+\lceil\sqrt{D}\rceil}{D+\lceil\sqrt{D}\rceil}\right\} \approx \frac{1}{\sqrt{D}}$ for large D, this bound is very close to optimal. To compare numbers, we list the values for $D \in \{2, \ldots, 10\}$, below.

1.3 Related literature

Quantum protocols for a closely related cryptographic task known as string-commitment have been considered [12, 21, 22, 7, 11]. Technically, this is the case of integer-commitment when $D = 2^n$ (if the string has n bits). It is worth noting that the quantum protocols considered in this paper are quite similar, but the security definitions are very different. Roughly speaking, the references above are concerned with quantum protocols where Alice is able to "cheat" on a bits and Bob is able to "learn" b bits of information about the n bit string. Multiple protocols and security trade-offs are given in the above references.

The use of semidefinite programming has been very valuable in the study of quantum cryptographic protocols, see for example [14, 15, 16, 10, 18, 19]. Roughly speaking, if one is able to formulate cheating probabilities as semidefinite programs, then the problem of analyzing cryptographic security can be translated into a concrete mathematical problem. Moreover, one then has the entire theory of semidefinite programming at their disposal. This is the approach taken in this work, to shine new light on a cryptographic task using the lens of semidefinite programming.

2 Semidefinite programming cheating strategy formulations

In this section, we use the theory of semidefinite programming to formulate Alice and Bob's maximum cheating probabilities for a DRIC-protocol. The formulations in this section are a generalization of those for bit-commitment, see [19] and the references therein for details about this special case.

2.1 Semidefinite programming

Semidefinite programming is the theory of optimizing a linear function over a positive semidefinite matrix variable subject to finitely many affine constraints. A semidefinite

program (SDP) can be written in the following form without loss of generality:

$$p^* := \sup\{\langle C, X\rangle : \mathcal{A}(X) = B,\ X \succeq 0\} \tag{2}$$

where $\mathcal{A}$ is a linear transformation, C and B are Hermitian, and $X \succeq Y$ means that $X - Y$ is (Hermitian) positive semidefinite.

Associated with every SDP is a dual SDP:

$$d^* := \inf\{\langle B, Y\rangle : \mathcal{A}^*(Y) = C + S,\ S \succeq 0,\ Y \text{ is Hermitian}\} \tag{3}$$

where $\mathcal{A}^*$ is the adjoint of $\mathcal{A}$.

We refer to the optimization problem (2) as the *primal* or *primal SDP* and to the optimization problem (3) as the *dual* or *dual SDP*. We say that the primal is *feasible* if there exists an X satisfying the (primal) constraints

$$\mathcal{A}(X) = B \quad \text{and} \quad X \succeq 0$$

and we say the dual is *feasible* if there exists (Y, S) satisfying the (dual) constraints

$$\mathcal{A}^*(Y) = C + S, \quad S \succeq 0, \quad \text{and} \quad Y \text{ is Hermitian.}$$

If further we have X positive definite, then the primal is said to be *strictly* feasible. If further we have S positive definite, then the dual is said to be *strictly* feasible.

Semidefinite programming has a rich and powerful duality theory. In particular, we use the following:

Weak duality:	If the primal and dual are both feasible, then $p^* \leq d^*$.
Strong duality:	If the primal and dual are both *strictly* feasible, then $p^* = d^*$ *and both attain an optimal solution.*

For more information about semidefinite programming and its duality theory, the reader is referred to [6].

2.2 Cheating strategy formulations

To study a fixed DRIC-protocol, it is convenient to define the following reduced states

$$\rho_a := \mathrm{Tr}_{\mathcal{A}}(|\psi_a\rangle \langle\psi_a|)$$

for all $a \in [D]$. We show that they appear in both the case of cheating Alice and cheating Bob.

2.2.1 Cheating Bob

To see how Bob can cheat, notice that he only has one message he sends to Alice. Thus, he must send $b \in [D]$ to force the outcome he wishes. For example, if he wishes to force the outcome d, he would send b such that $d = (a + b) \bmod D + 1$. Therefore, he must extract the value of a from $\mathcal{B}$ to accomplish this. Suppose he measures $\mathcal{B}$ with the measurement

$$\{M_1, \ldots, M_D\}$$

where the outcome of the measurement corresponds to Bob's guess for a. If Alice chose $a \in [D]$, he succeeds in cheating if his guess is correct, which happens with probability

$$\langle M_a, \rho_a\rangle.$$

Since the choice of Alice's integer a is uniformly random, we can calculate Bob's optimal cheating probability as

$$P_B^* = \max\left\{\frac{1}{D}\sum_{a\in[D]}\langle M_a, \rho_a\rangle : \sum_{a\in[D]} M_a = I_{\mathcal{B}},\ M_a \succeq 0, \forall a \in [D]\right\} \tag{4}$$

noting that the variables being optimized over correspond to a POVM measurement. Note that the maximum is attained since the set of feasible $(M_1, \ldots, M_D)$ forms a compact set.

Now that Bob's optimal cheating probability is stated in terms of an SDP, we can examine its dual as shown in the lemma below. Note that the lemma below follows from strong duality (details in the full version).

▶ **Lemma 7.** *For any* DRIC*-protocol, we have*

$$P_B^* = \min\left\{\mathrm{Tr}(X) : X \succeq \frac{1}{D}\rho_a, \forall a \in [D]\right\}. \tag{5}$$

We refer to the optimization problem (4) as *Bob's primal SDP* and to the optimization problem (5) as *Bob's dual SDP*. The utility of having dual SDP formulations is that any feasible solution yields an *upper bound* on the maximum cheating probability. Proving upper bounds on cheating probabilities would otherwise be a very hard task.

2.2.2 Cheating Alice

If Alice wishes to force Bob to accept outcome $d \in [D]$, she must convince him that the state in $\mathcal{A}\otimes\mathcal{B}$ is indeed $|\psi_a\rangle$ where a is such that $d = (a+b) \bmod D+1$. Note that this choice of a is determined after learning b from Bob, which occurs with uniform probability.

To quantify the extent to which Alice can cheat, we examine the states Bob has during the protocol. We know that Bob measures and accepts a with the measurement operator $\Pi_a := |\psi_a\rangle\langle\psi_a|$. Let $(a, \mathcal{A})$ be Alice's last message. Then Bob's state at the end of the protocol is given by a density operator σ_a acting on $\mathcal{A}\otimes\mathcal{B}$ which is accepted with probability $\langle\sigma_a, |\psi_a\rangle\langle\psi_a|\rangle$. Note that Alice's first message $\mathcal{B}$ is in state $\sigma := \mathrm{Tr}_{\mathcal{A}}(\sigma_a)$ which is independent of a (since Alice's first message does not depend on a when she cheats). Thus, the states under Bob's control are subject to the constraints

$$\mathrm{Tr}_{\mathcal{A}}(\sigma_a) = \sigma,\ \forall a \in [D], \quad \mathrm{Tr}(\sigma) = 1, \quad \sigma, \sigma_1 \ldots, \sigma_D \succeq 0. \tag{6}$$

(Note that $\mathrm{Tr}(\sigma_a) = 1$, for all $a \in [D]$, is implied by the constraints above, and is thus omitted.) On the other hand, if Alice maintains a purification of the states above, then using Uhlmann's Theorem [23] she can prepare any set of states satisfying conditions (6).

Thus, we have

$$P_A^* = \max\left\{\frac{1}{D}\sum_{a\in[D]}\langle\sigma_a, |\psi_a\rangle\langle\psi_a|\rangle : \mathrm{Tr}_{\mathcal{A}}(\sigma_a) = \sigma,\ \forall a \in [D],\ \mathrm{Tr}(\sigma) = 1,\ \sigma, \sigma_1 \ldots, \sigma_D \succeq 0\right\}. \tag{7}$$

Again, since the set of feasible $(\sigma, \sigma_1, \ldots, \sigma_D)$ is compact, the above SDP attains an optimal solution.

Similar to the case of cheating Bob, we can view the dual of Alice's cheating SDP above as shown in the lemma below. Again, the lemma below follows by strong duality (details in the full version).

▶ **Lemma 8.** *For any* DRIC*-protocol, we have*

$$P_A^* = \min\left\{ s : sI_{\mathcal{B}} \succeq \sum_{a\in[D]} Z_a,\ I_{\mathcal{A}} \otimes Z_a \succeq \frac{1}{D}\left|\psi_a\right\rangle\left\langle\psi_a\right|,\ \forall a \in [D],\ Z_a \text{ is Hermitian} \right\}. \quad (8)$$

We refer to the optimization problem (7) as *Alice's primal SDP* and the optimization problem (8) as *Alice's dual SDP*.

Note that every solution feasible in Alice's dual SDP has Z_a being positive semidefinite, for all $a \in [D]$. We can further assume that each Z_a is positive definite if we sacrifice the attainment of an optimal solution. This is because we can take an optimal solution $(s, Z_1, \ldots, Z_D)$ and consider $(s + \varepsilon D, Z_1 + \varepsilon I_{\mathcal{B}}, \ldots, Z_D + \varepsilon I_{\mathcal{B}})$ which is also feasible for any $\varepsilon > 0$, and $s + \varepsilon D$ approaches $s = P_A^*$ as ε decreases to 0.

Next, we use an analysis similar to one found in [15] and [24] to simplify the constraint

$$I_{\mathcal{A}} \otimes Z_a \succeq \left|\psi_a\right\rangle\left\langle\psi_a\right|$$

when Z_a is positive definite. Since $X \to ZXZ^{-1}$ is an automorphism of the set of positive semidefinite matrices for any fixed positive definite Z, we have

$$I_{\mathcal{A}} \otimes Z_a \succeq \frac{1}{D}\left|\psi_a\right\rangle\left\langle\psi_a\right| \iff I_{\mathcal{A}\otimes\mathcal{B}} \succeq (I_{\mathcal{A}} \otimes Z_a^{-1/2})\left(\frac{1}{D}\left|\psi_a\right\rangle\left\langle\psi_a\right|\right)(I_{\mathcal{A}} \otimes Z_a^{-1/2}). \quad (9)$$

Note that since the quantity on the right is positive semidefinite with rank at most 1, its largest eigenvalue is equal to its trace which is equal to

$$\frac{1}{D}\langle I_{\mathcal{A}} \otimes Z_a^{-1}, \left|\psi_a\right\rangle\left\langle\psi_a\right|\rangle = \frac{1}{D}\langle Z_a^{-1}, \mathrm{Tr}_{\mathcal{A}}(\left|\psi_a\right\rangle\left\langle\psi_a\right|)\rangle = \frac{1}{D}\langle Z_a^{-1}, \rho_a\rangle.$$

Thus, we can rewrite (9) as

$$I_{\mathcal{A}} \otimes Z_a \succeq \frac{1}{D}\left|\psi_a\right\rangle\left\langle\psi_a\right| \iff \frac{1}{D}\langle Z_a^{-1}, \rho_a\rangle \leq 1 \iff \langle Z_a^{-1}, \rho_a\rangle \leq D.$$

Therefore, we have the following lemma.

▶ **Lemma 9.** *For any* DRIC*-protocol, we have*

$$P_A^* = \inf\left\{ s : sI_{\mathcal{B}} \succeq \sum_{a\in[D]} Z_a,\ \langle Z_a^{-1}, \rho_a\rangle \leq D, \forall a \in [D],\ Z_a \text{ is positive definite}, \forall a \in [D] \right\}. \quad (10)$$

We also refer to the optimization problem (10) as Alice's dual SDP and we distinguish them by equation number.

3 Finding a decent DRIC-protocol

In this section, we exhibit a DRIC-protocol which has the same cheating probabilities as Protocol 1:

$$P_B^* = \frac{1}{m} \quad \text{and} \quad P_A^* = \frac{m}{D}.$$

To do this, define T_m to be the subsets of $[D]$ of cardinality m and note that $|T_m| = \binom{D}{m}$. Consider the following states

$$|\psi_a\rangle := \frac{1}{\sqrt{\binom{D-1}{m-1}}} \sum_{S \in T_m \,:\, a \in S} |S\rangle\,|S\rangle \in \mathcal{A} \otimes \mathcal{B},$$

for $a \in [D]$, where $\mathcal{A} = \mathcal{B} = \mathbb{C}^{|T_m|}$. Notice that

$$\rho_a := \mathrm{Tr}_{\mathcal{A}}\left(|\psi_a\rangle\langle\psi_a|\right) = \frac{1}{\binom{D-1}{m-1}} \sum_{S \in T_m \,:\, a \in S} |S\rangle\langle S|\,.$$

We now use the cheating SDPs developed in the previous section to analyze the cheating probabilities of this protocol.

3.1 Cheating Bob

To prove that Bob can cheat with probability at least $1/m$, suppose he measures his message from Alice in the computational basis. He then obtains a random subset $S \in T_m$ such that $a \in S$. He then guesses which integer is a and responds with the appropriate choice for b to get his desired outcome. He succeeds if and only if his guess for a (from the m choices in S) is correct. This strategy succeeds with probability $1/m$. Thus, $P_B^* \geq 1/m$.

To prove Bob cannot cheat with probability greater than $1/m$, notice that

$$X = \frac{1}{D\binom{D-1}{m-1}} I_{\mathcal{B}}$$

satisfies

$$X \succeq \frac{1}{D}\rho_a,\ \forall a \in [D],$$

and thus is feasible in Bob's dual (5). Therefore, $P_B^* \leq \mathrm{Tr}(X) = 1/m$, as desired.

3.2 Cheating Alice

Alice can cheat by creating the maximally entangled state

$$|T_m\rangle := \frac{1}{\sqrt{|T_m|}} \sum_{S \in T_m} |S\rangle\,|S\rangle \in \mathcal{A} \otimes \mathcal{B}$$

and sending $\mathcal{B}$ to Bob. After learning b, she sends a such that $(a+b) \bmod D+1$ is her desired outcome. She also sends $\mathcal{A}$ to Bob (without altering it in any way). Thus, her cheating probability is precisely the probability of her passing Bob's cheat detection which is

$$\langle \Pi_a, |T_m\rangle\langle T_m| \rangle = \langle |\psi_a\rangle\langle\psi_a|, |T_m\rangle\langle T_m| \rangle = |\langle T_m|\psi_a\rangle|^2 = \frac{m}{D}.$$

Therefore, this cheating strategy succeeds with probability m/D, proving $P_A^* \geq m/D$.

To prove this strategy is optimal, we use Alice's dual (10). Define

$$Z_a := \frac{1}{D} \sum_{S \in T_m \,:\, a \in S} |S\rangle\langle S| + \varepsilon \sum_{S \in T_m \,:\, a \notin S} |S\rangle\langle S|$$

where ε is a small positive constant. Z_a is invertible and we can write

$$Z_a^{-1} := D \sum_{S \in T_m \,:\, a \in S} |S\rangle\langle S| + \frac{1}{\varepsilon} \sum_{S \in T_m \,:\, a \notin S} |S\rangle\langle S| .$$

We see that each Z_a satisfies $\langle Z_a^{-1}, \rho_a \rangle = D$, for all $a \in [D]$. Also,

$$Z_a \preceq \frac{1}{D} \sum_{S \in T_m \,:\, a \in S} |S\rangle\langle S| + \varepsilon I_{\mathcal{B}}$$

thus

$$\sum_{a \in [D]} Z_a \preceq \frac{1}{D} \sum_{a \in [D]} \sum_{S \in T_m \,:\, a \in S} |S\rangle\langle S| + \varepsilon\, D\, I_{\mathcal{B}} = \left(\frac{m}{D} + \varepsilon D\right) I_{\mathcal{B}}.$$

Thus, $s = \dfrac{m}{D} + \varepsilon D$ satisfies

$$s\, I_{\mathcal{B}} \succeq \sum_{a \in [D]} Z_a$$

proving $P_A^* \leq s = \dfrac{m}{D} + \varepsilon D$, for all $\varepsilon > 0$. Therefore, $P_A^* = m/D$, as desired.

4 Balancing Alice and Bob's cheating probabilities

This section is comprised of two parts. We first focus on reducing Bob's cheating probabilities, then Alice's.

4.1 Building new protocols that reduce Bob's cheating

We start with a lemma.

▶ **Lemma 10.** *If there exists a* DRIC*-protocol with cheating probabilities $P_A^* = \alpha$ and $P_B^* = \beta$, then there exists another* DRIC*-protocol with cheating probabilities $P_A^* = \alpha'$ and $P_B^* = \beta'$ where*

$$\beta' \leq (1-t)\beta + \frac{t}{D} \quad \text{and} \quad \alpha' \leq (1-t)\alpha + t$$

for any $t \in (0,1)$.

We sketch the proof here. Fix a DRIC-protocol with cheating probabilities $P_A^* = \alpha$ and $P_B^* = \beta$ defined by the states $|\psi_a\rangle \in \mathcal{A} \otimes \mathcal{B}$, for $a \in [D]$. Extend each of the Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$ by another basis vector $|\perp\rangle$ and denote these Hilbert spaces by $\mathcal{A}'$ and $\mathcal{B}'$, respectively. In short, $\mathcal{A}' := \mathcal{A} \oplus \text{span}\{|\perp\rangle\}$ and $\mathcal{B}' := \mathcal{B} \oplus \text{span}\{|\perp\rangle\}$. Note that

$$\langle \perp, \perp | \psi_a \rangle = 0, \quad \text{for all } a \in [D].$$

We now analyze the cheating probabilities of Alice and Bob in the new DRIC-protocol defined by the states

$$|\psi'_a\rangle := \sqrt{1-t}\,|\psi_a\rangle + \sqrt{t}\,|\perp,\perp\rangle \in \mathcal{A}' \otimes \mathcal{B}', \quad \text{for all } a \in [D]$$

as a function of $t \in (0,1)$. For this, note that

$$\rho'_a := \text{Tr}_{\mathcal{A}}\left(|\psi'_a\rangle\langle\psi'_a|\right) = (1-t)\,\rho_a + t\,|\perp\rangle\langle\perp|\,,$$

where $\rho_a := \text{Tr}_{\mathcal{A}}\left(|\psi_a\rangle\langle\psi_a|\right)$.

To show how Bob's cheating probability changes, consider an optimal solution X to Bob's dual SDP (5) corresponding to the original protocol. Then one can show that

$$X' := (1-t)X + \frac{t}{D} |\perp\rangle\langle\perp|$$

is feasible in Bob's dual SDP after the protocol has been modified. This proves that

$$P_B^* \leq (1-t)\beta + t/D$$

for the new protocol.

Concerning cheating Alice, let $(s, Z_1, \ldots, Z_D)$ be a feasible solution for Alice's dual (10) for the original protocol. Then one can show that

$$\begin{aligned} s' &:= s(1-t)+t \\ Z_1' &:= ((1-t)+t/s)\, Z_1 + \left(\frac{s(1-t)+t}{D}\right) |\perp\rangle\langle\perp| \\ &\vdots \\ Z_D' &:= ((1-t)+t/s)\, Z_D + \left(\frac{s(1-t)+t}{D}\right) |\perp\rangle\langle\perp| \end{aligned}$$

is feasible for Alice's dual for the new protocol. Thus,

$$P_A^* \leq s' = s(1-t)+t$$

and since s can be taken arbitrarily close to α, the result follows.

Intuitively, Alice can cheat more if the states ρ_a are "close" to each other and Bob can cheat more if they are "far apart". What this protocol modification does is make all the states closer together to increase Alice's cheating probability but to decrease Bob's.

Note that this lemma is useful when $\beta > \alpha$. In this case, one can choose

$$t = \frac{\beta - \alpha}{(1-1/D) + (\beta - \alpha)} \in (0,1)$$

to equate the upper bounds. If $\alpha > \beta$, then no choice of $t \in (0,1)$ will make the two upper bounds in Lemma 10 equal. We summarize in the following corollary.

▶ **Corollary 11.** *If there exists a* DRIC*-protocol with cheating probabilities $P_A^* = \alpha$ and $P_B^* = \beta$,* with *$\beta > \alpha$, then there exists another* DRIC*-protocol with maximum cheating probability*

$$\max\{P_A^*, P_B^*\} \leq \frac{D\beta - \alpha}{D\beta - D\alpha + D - 1} < \beta.$$

4.2 Building new protocols that reduce Alice's cheating

In this subsection, we show how to reduce Alice's cheating probabilities in a DRIC-protocol.

▶ **Lemma 12.** *If there exists a* DRIC*-protocol with cheating probabilities $P_A^* = \alpha$ and $P_B^* = \beta$, then there exists another* DRIC*-protocol with cheating probabilities $P_A^* = \alpha'$ and $P_B^* = \beta'$ where*

$$\beta' \leq (1-t)\beta + t \quad \text{and} \quad \alpha' \leq (1-t)\alpha + \frac{t}{D},$$

for $t \in (0,1)$.

We sketch the proof here. Fix a DRIC-protocol with cheating probabilities $P_A^* = \alpha$ and $P_B^* = \beta$ defined by the states $|\psi_a\rangle \in \mathcal{A} \otimes \mathcal{B}$, for $a \in [D]$. Extend each of the Hilbert spaces $\mathcal{A}$ and $\mathcal{B}$ by the set of orthogonal basis vectors $\{|\perp_a\rangle : a \in [D]\}$, and denote these new Hilbert spaces by $\mathcal{A}'$ and $\mathcal{B}'$, respectively. In other words,

$$\mathcal{A}' := \mathcal{A} \oplus \text{span}\{|\perp_1\rangle, \ldots, |\perp_D\rangle\} \quad \text{and} \quad \mathcal{B}' := \mathcal{B} \oplus \text{span}\{|\perp_1\rangle, \ldots, |\perp_D\rangle\}.$$

Note that

$$\langle \perp_{a''}, \perp_{a'} | \psi_a \rangle = 0, \quad \text{for all } a, a', a'' \in [D].$$

Again, we analyze the cheating probabilities of Alice and Bob in the new DRIC-protocol defined by the states

$$|\psi'_a\rangle := \sqrt{1-t}\,|\psi_a\rangle + \sqrt{t}\,|\perp_a\rangle\,|\perp_a\rangle \in \mathcal{A}' \otimes \mathcal{B}'$$

for $a \in [D]$. The reduced states are

$$\rho'_a := (1-t)\,\rho_a + t\,|\perp_a\rangle\langle\perp_a|$$

for $a \in [D]$, recalling that $\rho_a := \text{Tr}_{\mathcal{A}}(|\psi_a\rangle\langle\psi_a|)$. We now analyze the cheating probabilities of this new protocol as a function of $t \in (0,1)$.

To show how Bob's cheating probability changes, we can use a similar argument. Consider an optimal solution X to Bob's dual (5) for the original protocol. Then one can show that

$$X' := (1-t)X + \frac{t}{D} \sum_{a \in [D]} |\perp_a\rangle\langle\perp_a|$$

is feasible for Bob's dual for the modified protocol. This shows that

$$P_B^* \leq (1-t)\beta + t.$$

Concerning cheating Alice, let $(s, Z_1, \ldots, Z_D)$ be a feasible solution for Alice's dual (10) for the original protocol. Then one can show that

$$\begin{aligned}
s' &:= (1-t)s + t/D + \zeta(D-1) \\
Z'_1 &:= \left((1-t) + \frac{t}{Ds}\right) Z_1 + \left((1-t)s + \frac{t}{D}\right) |\perp_1\rangle\langle\perp_1| + \zeta \sum_{c \in [D], c \neq 1} |\perp_c\rangle\langle\perp_c| \\
&\vdots \\
Z'_D &:= \left((1-t) + \frac{t}{Ds}\right) Z_D + \left((1-t)s + \frac{t}{D}\right) |\perp_D\rangle\langle\perp_D| + \zeta \sum_{c \in [D], c \neq D} |\perp_c\rangle\langle\perp_c|
\end{aligned}$$

is feasible for Alice's dual for the new protocol for $\zeta > 0$ a small constant. Thus,

$$P_A^* \leq (1-t)s + t/D$$

and since s can be taken arbitrarily close to α, the result follows.

Intuitively, this protocol modification works in the opposite manner of the last. Here, we are making the states farther apart as to decrease Alice's cheating at the expense of increasing Bob's.

As opposed to Lemma 10, the above lemma is useful when $\alpha > \beta$. Similarly, if $\beta > \alpha$, then no choice of $t \in (0,1)$ will make the two upper bounds in Lemma 12 equal.

By symmetry, we have the following corollary.

▶ **Corollary 13.** *If there exists a* DRIC*-protocol with cheating probabilities* $P_A^* = \alpha$ *and* $P_B^* = \beta$*, with* $\alpha > \beta$*, then there exists another* DRIC*-protocol with maximum cheating probability*

$$\max\{P_A^*, P_B^*\} \leq \frac{D\alpha - \beta}{D\alpha - D\beta + D - 1} < \alpha.$$

Note that if $\alpha = \beta$, the quantity $\frac{D\alpha-\beta}{D\alpha-D\beta+D-1}$ is equal to $\alpha(= \beta)$. Thus, we still have

$$\max\{P_A^*, P_B^*\} \leq \frac{D\alpha - \beta}{D\alpha - D\beta + D - 1}$$

holding, although no protocol modification is necessary. Therefore, Proposition 5 now follows from combining Corollaries 11 and 13 and the comment above.

Acknowledgements. I thank Sevag Gharibian for useful discussions.

References

1. N. Aharon and J. Silman. Quantum dice rolling: a multi-outcome generalization of quantum coin flipping. *New Journal of Physics*, 12(3):033027, 2010.
2. Dorit Aharonov, André Chailloux, Maor Ganz, Iordanis Kerenidis, and Loïck Magnin. A simpler proof of existence of quantum weak coin flipping with arbitrarily small bias. *SIAM Journal of Computing*, to appear, 2016.
3. Dorit Aharonov, Amnon Ta-Shma, Umesh Vazirani, and Andrew Chi-Chih Yao. Quantum bit escrow. In *Proceedings of 32nd Annual ACM Symposium on the Theory of Computing*, pages 705–714. ACM, 2000. doi:10.1145/335305.335404.
4. Andris Ambainis. A new protocol and lower bounds for quantum coin flipping. In *Proceedings of 33rd Annual ACM Symposium on the Theory of Computing*, pages 134–142. ACM, 2001. doi:10.1109/FOCS.2004.13.
5. Manuel Blum. Coin flipping by telephone. In Allen Gersho, editor, *Advances in Cryptology: A Report on CRYPTO 81, CRYPTO 81, IEEE Workshop on Communications Security, Santa Barbara, California, USA, August 24-26, 1981*, pages 11–15. U. C. Santa Barbara, Dept. of Elec. and Computer Eng., ECE Report No. 82-04, 1982, 1981.
6. Stephen Boyd and Lieven Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.
7. Harry Buhrman, Matthias Christandl, Patrick Hayden, Hoi-Kwong Lo, and Stephanie Wehner. Possibility, impossibility, and cheat-sensitivity of quantum bit string commitment. *Phys. Rev. A*, 78:022316, 2008.
8. André Chailloux and Iordanis Kerenidis. Optimal quantum strong coin flipping. In *Proceedings of 50th IEEE Symposium on Foundations of Computer Science*, pages 527–533. IEEE Computer Society, 2009.
9. André Chailloux and Iordanis Kerenidis. Optimal bounds for quantum bit commitment. In *Proceedings of the 52nd Annual IEEE Symposium on Foundations of Computer Science*, pages 354–362. IEEE Computer Society, 2011. doi:10.1109/FOCS.2011.42.
10. André Chailloux, Iordanis Kerenidis, and Jamie Sikora. Lower bounds for quantum oblivious transfer. *Quantum Information & Computation*, 13(1 & 2):158–177, 2013.
11. Rahul Jain. New binding-concealing trade-offs for quantum string commitment. *Journal of Cryptology*, 21:579–592, 2008.
12. Adrian Kent. Quantum bit string commitment. *Phys. Rev. Lett.*, 90:237901, 2003.
13. Iordanis Kerenidis and Ashwin Nayak. Weak coin flipping with small bias. *Information Processing Letters*, 89(3):131–135, 2004. doi:10.1016/j.ipl.2003.07.007.

14 Alexei Kitaev. Quantum coin-flipping. Unpublished result. Talk at the 6th Annual workshop on Quantum Information Processing (QIP 2003), 2002.

15 Carlos Mochon. A large family of quantum weak coin-flipping protocols. *Physical Review A*, 72:022341, 2005. URL: `http://arxiv.org/abs/quant-ph/0502068`, `doi:10.1103/PhysRevA.72.022341`.

16 Carlos Mochon. Quantum weak coin flipping with arbitrarily small bias. Available as arXiv.org e-Print quant-ph/0711.4114, 2007.

17 Ashwin Nayak and Peter W. Shor. Bit-commitment based quantum coin flipping. *Physical Review A*, 67:012304, 2003. `doi:10.1103/PhysRevA.67.012304`.

18 Ashwin Nayak, Jamie Sikora, and Levent Tunçel. Quantum and classical coin-flipping protocols based on bit-commitment and their point games. Available as arXiv.org e-Print quant-ph/1504.04217, 2015.

19 Ashwin Nayak, Jamie Sikora, and Levent Tunçel. A search for quantum coin-flipping protocols using optimization techniques. *Mathematical Programming*, 156(1):581–613, 2016.

20 Robert W. Spekkens and Terence Rudolph. Degrees of concealment and bindingness in quantum bit commitment protocols. *Physical Review A*, 65:012310, 2001. `doi:10.1103/PhysRevA.65.012310`.

21 Toyohiro Tsurumaru. Implementable quantum bit-string commitment protocol. *Phys. Rev. A*, 71:012313, 2005.

22 Toyohiro Tsurumaru. Group covariant protocols for quantum string commitment. *Phys. Rev. A*, 74:042307, 2006.

23 A. Uhlmann. The "transition probability" in the state space of a *-algebra. *Reports on Mathematical Physics*, 9(2):273–279, 1976.

24 John Watrous. Semidefinite programs for completely bounded norms. *Theory of Computing*, 5:217–238, 2009.

Robust Bell Inequalities from Communication Complexity*

Sophie Laplante[1], Mathieu Laurière[2], Alexandre Nolin[3], Jérémie Roland[4], and Gabriel Senno[5]

1 **IRIF, Université Paris-Diderot, Paris, France**
`laplante@liafa.univ-paris-diderot.fr`
2 **IRIF, Université Paris-Diderot, Paris, France**
`lauriere@liafa.univ-paris-diderot.fr`
3 **IRIF, Université Paris-Diderot, Paris, France**
`nolin@liafa.univ-paris-diderot.fr`
4 **Université Libre de Bruxelles, Brussels, Belgium**
`jroland@ulb.ac.be`
5 **CONICET & Departamento de Computación, FCEN, Universidad de Buenos Aires, Buenos Aires, Argentina**
`gsenno@dc.uba.ar`

Abstract

The question of how large Bell inequality violations can be, for quantum distributions, has been the object of much work in the past several years. We say a Bell inequality is normalized if its absolute value does not exceed 1 for any classical (i.e. local) distribution. Upper and (almost) tight lower bounds have been given in terms of number of outputs of the distribution, number of inputs, and the dimension of the shared quantum states. In this work, we revisit normalized Bell inequalities together with another family: inefficiency-resistant Bell inequalities. To be inefficiency-resistant, the Bell value must not exceed 1 for any local distribution, including those that can abort. Both these families of Bell inequalities are closely related to communication complexity lower bounds. We show how to derive large violations from any gap between classical and quantum communication complexity, provided the lower bound on classical communication is proven using these lower bounds. This leads to inefficiency-resistant violations that can be exponential in the size of the inputs. Finally, we study resistance to noise and inefficiency for these Bell inequalities.

1998 ACM Subject Classification F.1.3 Complexity Measures and Classes, F.2.3 Tradeoffs between Complexity Measures

Keywords and phrases Communication complexity, Bell inequalities, nonlocality, detector efficiency

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.5

1 Introduction

The question of achieving large Bell violations has been studied since Bell's seminal paper in 1964 [6]. In one line of investigation, proposals have been made to exhibit families of

* This work was partially supported by the European Union Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 600700 (QALGO), the Argentinian ANPCyT (PICT-2014-3711), the Laboratoire International Associée INFINIS, the Belgian ARC project COPHYMA and the Belgian *Fonds de la Recherche Scientifique* – FNRS under grant no. F.4515.16 (QUICTIME), and the French ANR Blanc grant RDAM ANR-12-BS02-005.

Table 1 Bounds on quantum violations of bipartite normalized Bell inequalities, in terms of the dimension d of the local Hilbert space, the number of settings (or inputs) N and the number of outcomes K (or outputs) per party. In the fourth column, we compare ad hoc results to the recent constructions of [10] (Theorem 7) which gives a lower bound of $\frac{\sqrt{c}}{q}$, where c (resp. q) stands for the classical (resp. quantum) communication complexity of simulating a distribution. We give upper bounds on their construction in terms of the parameters d, N, K.

Parameter	Upper bound	Ad hoc lower bounds	Best possible lower bound from [10]
Number of inputs N	$2^c \leq N$ [29, 12, 21]	$\frac{\sqrt{N}}{\log(N)}$ [19]	$\frac{\sqrt{c}}{q} \leq \log(N)$
Number of outputs K	$O(K)$ [19]	$\Omega\left(\frac{K}{(\log(K))^2}\right)$ [11]	$\leq \log(K)$
Dimension d	$O(d)$ [21]	$\Omega\left(\frac{d}{(\log(d))^2}\right)$ [11]	$\leq \log\log(d)$

distributions which admit unbounded violations [33, 28, 34, 36]. In another, various measures of nonlocality have been studied, such as the amount of communication necessary and sufficient to simulate quantum distributions classically [32, 7, 42, 43, 37, 12], or the resistance to detection inefficiencies and noise. More recently, focus has turned to giving upper and lower bounds on violations achievable, in terms of various parameters: number of players, number of inputs, number of outputs, dimension of the quantum state, and amount of entanglement [12, 21, 19].

Up until quite recently, violations were studied in the case of specific distributions (measuring Bell states), or families of distributions. Buhrman *et al.* [11] gave a construction that could be applied to several problems which had efficient quantum protocols (in terms of communication) and for which one could show a trade-off between communication and error in the classical setting. This still required an *ad hoc* analysis of communication problems. Recently Buhrman *et al.* [10] proposed the first general construction of quantum states along with Bell inequalities from any communication problem. The quantum states violate the Bell inequalities when there is a sufficiently large gap between quantum and classical communication complexity (a super-quadratic gap is necessary, unless a quantum protocol without local memory exists).

Table 1 summarizes the best known upper and lower bounds on quantum violations achievable with normalized Bell inequalities.

1.1 Our results

We revisit the question of achieving large Bell violations by exploiting known connections with communication complexity. Strong lower bounds in communication complexity, equivalent to the partition bound, amount to finding *inefficiency-resistant Bell inequalities* [27]. These are Bell functionals that are bounded above by 1 on all local distributions *that can abort*.

First, we study the resistance of normalized Bell inequalities to inefficiency. We show that, up to a constant factor in the value of the violation, any normalized Bell inequality can be made resistant to inefficiency while maintaining the normalization property (**Theorem 6**).

Second, we show how to derive large Bell violations from any communication problem for which the partition bound is bounded below and the quantum communication complexity is bounded above. The problems studied in communication complexity are far beyond the quantum set, but we show how to easily derive a quantum distribution from a quantum

Table 2 Comparison of the Bell violations obtained by the general construction of Buhrman *et al.* [10] for normalized Bell violations (second column) and this work, for inefficiency-resistant Bell violations (see Propositions 13, 14, 15, and 16). The parameter n is the size of the input (typically, $N = 2^n$.) Explicit Bell inequalities are given in the Appendix. The construction of Buhrman *et al.* only yields a violation when the gap between classical and quantum complexities is more than quadratic. In the case where the gap is too small to prove a violation, we indicate this with "N/A".

Problem	Normalized Bell violations [10]	Inefficiency-resistant Bell violations (this work)
VSP [38, 24]	$\Omega\left(\sqrt[6]{n}/\sqrt{\log n}\right)$ $d = 2^{\Theta(n \log n)}, K = 2^{\Theta(n)}$	$2^{\Omega(\sqrt[3]{n})-O(\log n)}$ $d = 2^{O(\log n)}, K = 3$
DISJ [39, 40, 1]	N/A	$2^{\Omega(n)-O(\sqrt{n})}$ $d = 2^{O(\sqrt{n})}, K = 3$
TRIBES [18, 9]	N/A	$2^{\Omega(n)-O(\sqrt{n}\log^2 n)}$ $d = 2^{O(\sqrt{n}\log^2 n)}, K = 3$
ORT [41, 9]	N/A	$2^{\Omega(n)-O(\sqrt{n}\log n)}$ $d = 2^{O(\sqrt{n}\log n)}, K = 3$

protocol. The Bell value we obtain is 2^{c-2q}, where c is the partition lower bound on the classical communication complexity of the problem considered, and q is an upper bound on its quantum communication complexity (**Theorem 8 and Corollary 9**). The quantum distribution has one extra output per player compared to the original distribution and uses the same amount of entanglement as the quantum protocol plus as many EPR pairs as needed to teleport the quantum communication in the protocol. We show that these Bell violations can be made noise-resistant, at the cost of a 2^{2q} factor in the number of outcomes per player (**Theorem 10**).

Finally, we provide tools to build Bell inequalities from communication lower bounds in the literature. Lower bounds used in practice to separate classical from quantum communication complexity are usually achieved using corruption bounds and its variants. In **Theorem 12**, we give an explicit construction which translates these bounds into a suitable Bell functional. Table 2 summarizes the new results or the improvements that we obtain in this work.

1.2 Related work

The study of the maximum violation of Bell inequalities began with Tsirelson [44], who showed that for two-outcome correlation Bell inequalities, the maximum violation is bounded above by Grothendieck's constant. Tsirelson also raised the question of whether one can have unbounded violations of Bell inequalities. More precisely, he asked whether there exist families of Bell inequalities for which the amount of the violation grows arbitrarily large.

The first answer to this question came from Mermin [33], who gave a family Bell inequalities for which a violation exponential in the number of parties is achieved. In the years that followed, several new constructions appeared for number of parties and number of inputs [3, 30, 28, 34, 36].

The study of upper bounds on violations of normalized Bell inequalities resumed in [12], where an upper bound of $O(K^2)$ (with K the number of outputs per player) and of $2^c \leq N$ (with c the communication complexity and N the number of inputs per player) were proven. In [21] the authors proved a bound of $O(d)$ in terms of the dimension d of the local Hilbert space, and in [19], the bound in terms of the number of outputs was improved to $O(K)$.

In [19], Bell inequalities are constructed for which a near optimal, but probabilistic, violation of order $\Omega(\sqrt{m}/\log m)$, with $N = K = d = m$, is proven. In [11], the same violation, although requiring $N = 2^m$ inputs, is achieved for a family of Bell inequalities and quantum distributions built using the quantum advantage in one-way communication complexity for the Hidden Matching problem (with $K = d = m$). In the same paper, a violation of order $\Omega(m/(\log m)^2)$, with $K = d = m$ and $N = 2^m/m$ is achieved with the Khot-Vishnoi game. Recently, an asymmetric version of that game was introduced to allow one of the parties to only make dichotomic measurements, with a smaller (although almost optimal for this scenario) violation $\Omega(\sqrt{m}/(\log m)^2)$ [35].

For *inefficiency-resistant* Bell inequalities, the bounds in [19] do not apply. In fact, Laplante *et al.* proved in [27] a violation exponential in the dimension and the number of outputs for this type of Bell functionals, achieved by a quantum distribution built, as in [11], from the Hidden Matching communication complexity problem.

The connection exhibited in [11] between Bell violations and communication complexity is generalized by Buhrman *et al.* in [10] where a fully general construction is given to go from a quantum communication protocol for a function f to a Bell inequality and a quantum distribution which achieves a violation of order $\Omega\left(\frac{\sqrt{R_{1/3}(f)}}{Q_{1/3}(f)}\right)$. The downside to this construction is that the quantum distribution has a double exponential (in the communication) number of outputs and the protocol to implement it uses an additional double exponential amount of entanglement. Also, this result does not apply for quantum advantages in a zero-error setting.

2 Preliminaries

2.1 Quantum nonlocality

Local, quantum, and nonsignaling distributions have been widely studied in quantum information theory since the seminal paper of Bell [6]. In an experimental setting, two players share an entangled state and each player is given a measurement to perform. The outcomes of the measurements are predicted by quantum mechanics and follow some probability distribution $p(a, b|x, y)$, where a is the outcome of Alice's measurement x, and b is the outcome of Bob's measurement y.

We consider bipartite distribution families of the form $\mathbf{p} = (p(\cdot, \cdot|x, y))_{(x,y)\in\mathcal{X}\times\mathcal{Y}}$ with inputs $(x, y) \in \mathcal{X} \times \mathcal{Y}$ determining a probability distribution $p(\cdot, \cdot|x, y)$ over the outcomes $(a, b) \in \mathcal{A} \times \mathcal{B}$, with the usual positivity and normalization constraints. The set of probability distribution families is denoted by $\mathcal{P}$. For simplicity, we call simply "distributions" such probability distribution families. The expression "Alice's marginal" refers to her marginal output distribution, that is $\sum_b p(\cdot, b|x, y)$ (and similarly for Bob).

The *local deterministic distributions*, denoted $\mathcal{L}_{\text{det}}$, are the ones where Alice outputs according to a deterministic strategy, i.e., a (deterministic) function of x, and Bob independently outputs as a function of y, without communicating. The *local distributions* $\mathcal{L}$ are obtained by taking distributions over the local deterministic strategies. Operationally, this corresponds to protocols with shared randomness and no communication. Geometrically, $\mathcal{L}$ is the convex hull of $\mathcal{L}_{\text{det}}$.

A *Bell test* [6] consists of estimating all the probabilities $p(a, b|x, y)$ and computing a *Bell functional*, or linear function, on these values. The Bell functional B is chosen together with a threshold τ so that any local classical distribution ℓ verifies the *Bell inequality* $B(\ell) \leq \tau$, but the chosen distribution $\mathbf{p}$ exhibits a *Bell violation*: $B(\mathbf{p}) > \tau$. By normalizing, we can assume without loss of generality that ℓ verifies $B(\ell) \leq 1$ for any $\ell \in \mathcal{L}$, and $B(\mathbf{p}) > 1$.

In this paper, we will also consider strategies that are allowed to abort the protocol with some probability. When they abort, they output the symbol $\perp$ ($\perp$ denotes a new symbol which is not in $\mathcal{A} \cup \mathcal{B}$). We will use the notation $\mathcal{L}^{\perp}_{\text{det}}$ and $\mathcal{L}^{\perp}$ to denote local strategies that can abort, where $\perp$ is added to the possible outputs for both players. When $\ell \in \mathcal{L}^{\perp}_{\text{det}}$ or $\mathcal{L}^{\perp}$, $\ell(a, b|x, y)$ is *not* conditioned on $a, b \neq \perp$ since $\perp$ is a valid output for such distributions.

The *quantum distributions*, denoted $\mathcal{Q}$, are the ones that result from applying measurements x, y to their part of a shared entangled bipartite state. Each player outputs his or her measurement outcome (a for Alice and b for Bob). In communication complexity terms, these are zero-communication protocols with shared entanglement. If the players are allowed to abort, then the corresponding set of distributions is denoted $\mathcal{Q}^{\perp}$.

Boolean (and other) functions can be cast as sampling problems. Consider a boolean function $f : \mathcal{X} \times \mathcal{Y} \to \{0, 1\}$ (non-boolean functions and relations can be handled similarly). First, we split the output so that if $f(x, y) = 0$, Alice and Bob are required to output the same bit, and if $f(x, y) = 1$, they output different bits. Let us further require Alice's marginal distribution to be uniform, likewise for Bob, so that the distribution is well defined. Call the resulting distribution $\mathbf{p}_f$, that is, for any $a, b \in \{0, 1\}$ and $(x, y) \in \mathcal{X} \times \mathcal{Y}$, we have $p_f(a, b|x, y) = 1/2$ if $a \oplus b = f(x, y)$, and $p_f(a, b|x, y) = 0$ otherwise, $\oplus$ being the 1-bit XOR.

If $\mathbf{p}_f$ were local, f could be computed with one bit of communication using shared randomness: Alice sends her output to Bob, and Bob XORs it with his output. If $\mathbf{p}_f$ were quantum, there would be a 1-bit protocol with shared entanglement for f. In communication complexity, we are usually interested in distributions having nontrivial communication complexity, and lie well beyond these sets.

Finally, a distribution is *nonsignaling* if for each player, its marginal output distributions, given by $p_A(a|x, y) = \sum_b p(a, b|x, y)$, for Alice, and $p_B(b|x, y) = \sum_a p(a, b|x, y)$, for Bob, do not depend on the other player's input. When this is the case, we write the marginals as $p_A(a|x)$ and $p_B(b|y)$. Operationally, this means that each player cannot influence the statistics of what the other player observes with his own choice of input. We note with $\mathcal{C}$ the set of nonsignaling distributions, also referred to as the *causal* set, and we note $\mathcal{C}^{\perp}$ when we allow aborting. The well-known inclusion relations between these sets are $\mathcal{L} \subset \mathcal{Q} \subset \mathcal{C} \subset \mathcal{P}$.

For any Boolean function f, the distribution $\mathbf{p}_f$ is nonsignaling since the marginals are uniform. A fundamental question of quantum mechanics has been to establish experimentally whether nature is truly *nonlocal*, as predicted by quantum mechanics, or whether there is a purely classical (i.e., *local*) explanation to the phenomena that have been predicted by quantum theory and observed in the lab.

2.2 Measures of nonlocality

We have described nonlocality as a yes/no property, but some distributions are somehow more nonlocal than others. To have a robust measure of nonlocality, it should verify some common sense properties: for a fixed distribution, the measure should be bounded; it should also be convex, since sampling from the convex combination of two distributions can be done by first picking randomly one of the two distributions using shared randomness, and then sampling from that distribution. We also expect such a measure of nonlocality to have various equivalent formulations. Several measures have been proposed and studied: resistance to noise [22, 2, 36, 20], resistance to inefficiency [30, 31, 27], amount of communication necessary to reproduce them [32, 7, 42, 43, 37, 12], information-theoretic measures [8, 14, 13], etc.

In the form studied in this paper, normalized Bell inequalities were first studied in [12], where they appeared as the dual of the linear program for a well-studied lower bound on communication complexity, known as the nuclear norm ν [29] (the definition is given in Section 2.3). There are many equivalent formulations of this bound. For distributions

arising from boolean functions, it has the mathematical properties of a norm, and it is related to winning probabilities of XOR games. It can also be viewed as a gauge, that is, a quantity measuring by how much the local set must be expanded in order to contain the distribution considered. For more general nonsignaling distributions, besides having a geometrical interpretation in terms of affine combinations of local distributions, it has also been shown to be equivalent to the amount of local noise that can be tolerated before the distribution becomes local [21].

A subsequent paper [27] studied equivalent formulations of the partition bound, one of the strongest lower bounds in communication complexity [17]. This bound also also has several formulations: the primal formulation can be viewed as resistance to detector inefficiency, and the dual formulation is given in terms of inefficiency-resistant Bell inequality violations.

In this paper, we show how to deduce large violations on quantum distributions from large violations on nonsignaling distributions, provided there are efficient quantum communication protocols for the latter.

2.3 Communication complexity and lower bounds

In classical communication complexity (introduced by [45]), two players each have a share of the input, and wish to compute a function on the full input. Communication complexity measures the number of bits they need to exchange to solve this problem in the worst case, over all inputs of a given size n. In this paper we consider a generalization of this model, where instead of computing a function, they each produce an output, say a and b, which should follow, for each (x, y), some prescribed distribution $p(a, b|x, y)$ (which depends on their inputs x, y). We assume that the order in which the players speak does not depend on the inputs. This is without loss of generality at a cost of a factor of 2 in the communication.

We use the following notation for communication complexity of distributions. $R_\epsilon(\mathbf{p})$ is the minimum number of bits exchanged in the worst case to output with the distribution $\mathbf{p}$, up to ϵ in total variation distance for all x, y. We call total variation distance between distributions the distance denoted by $|.|_1$, and defined as $|\mathbf{p}-\mathbf{p}'|_1 = \max_{x,y} \sum_{a,b} |p(a, b|x, y) - p'(a, b|x, y)|$.

We use Q to denote quantum communication complexity (see [47]), and we use the superscript $*$ to denote the presence of shared entanglement. For randomized communication, we assume shared randomness.

To give upper bounds on communication complexity it suffices to give a protocol and analyze its complexity. Proving lower bounds is often a more difficult task, and many techniques have been developed to achieve this. The methods we describe here are complexity measures which can be applied to any function. To prove a lower bound on communication, it suffices to give a lower bound on one of these complexity measures, which are bounded above by communication complexity for any function. We describe here most of the complexity measures relevant to this work.

The nuclear norm ν, given here in its dual formulation and extended to nonsignaling distributions, is expressed by the following linear program [29, 12]. (There is a quantum analogue, γ_2, which is not needed in this work. We refer the interested reader to the definition for distributions in [12]).

▶ **Definition 1** ([29, 12])**.** The nuclear norm ν of a nonsignaling distribution $\mathbf{p} \in \mathcal{C}$ is given by

$$\begin{aligned} \nu(\mathbf{p}) = \max_{B} \quad & B(\mathbf{p}) \\ \text{subject to} \quad & |B(\ell)| \leq 1 \quad \forall \ell \in \mathcal{L}_{det}. \end{aligned}$$

With error ϵ, $\nu_\epsilon(\mathbf{p}) = \min_{\mathbf{p}' \in \mathcal{C}: |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon} \nu(\mathbf{p}')$. We call any Bell functional that satisfies the constraint in the above linear program *normalized Bell functional*.

In this definition and in the rest of the paper, unless otherwise specified (in particular in Lemma 19), B ranges over vectors of real coefficients $B_{a,b,x,y}$ and $B(\mathbf{p})$ denotes $\sum_{a,b,x,y} B_{a,b,x,y} p(a,b|x,y)$, where a, b ranges over the non-abort outputs and x, y ranges over the inputs. So even when B and $\mathbf{p}$ have coefficients on the abort events, we do not count them. Table 1 summarizes the known upper and lower bounds on ν for various parameters. The (log of the) nuclear norm is a lower bound on classical communication complexity.

▶ **Proposition 2** ([29, 12]). *For any nonsignaling distribution* $\mathbf{p} \in \mathcal{C}$, $\mathrm{R}_\epsilon(\mathbf{p}) + 1 \geq \log(\nu_\epsilon(\mathbf{p}))$, *and for any boolean function* f, $\mathrm{R}_\epsilon(f) \geq \log(\nu_\epsilon(\mathbf{p}_f))$.

As lower bounds on communication complexity of Boolean functions go, ν is one of the weaker bounds, equivalent to the smooth discrepancy [17], and no larger than the approximate nonnegative rank and the smooth rectangle bounds [25]. More significantly for this work, up to small multiplicative constants, for boolean functions, (the log of) ν is a lower bound on quantum communication, so it is useless to establish gaps between classical and quantum communication complexity. (This limitation, with the upper bound in terms of the number of outputs on normalized Bell violations, is a consequence of Grothendieck's theorem [15].)

The classical and quantum efficiency measures, given here in their dual formulations, are expressed by the following two convex optimization programs. The classical bound is a generalization to distributions of the partition bound of communication complexity [17, 27]. This bound is one of the strongest lower bounds known, and can be exponentially larger than ν (an example is the Vector in Subspace problem). It is always as least as large as the relaxed partition bound which is in turn always at least as large as the smooth rectangle bound [17, 23]. Its weaker variants have been used to show exponential gaps between classical and quantum communication complexity. The definition we give here is a stronger formulation than the one given in [27]. We show they are equivalent in Appendix D.

▶ **Definition 3** ([27]). The ϵ-error efficiency bound of a distribution $\mathbf{p} \in \mathcal{P}$ is given by

$$\begin{aligned} eff_\epsilon(\mathbf{p}) = \max_{B,\beta} \quad & \beta & & \\ \text{subject to} \quad & B(\mathbf{p}') \geq \beta & & \forall \mathbf{p}' \in \mathcal{P} \text{ s.t. } |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon, \\ & B(\ell) \leq 1 & & \forall \ell \in \mathcal{L}_{det}^{\perp}. \end{aligned}$$

We call any Bell functional that satisfies the second constraint in the above linear program *inefficiency-resistant Bell functional*. The ϵ-error quantum efficiency bound of a $\mathbf{p} \in \mathcal{P}$ is

$$\begin{aligned} eff_\epsilon^*(\mathbf{p}) = \max_{B,\beta} \quad & \beta & & \\ \text{subject to} \quad & B(\mathbf{p}') \geq \beta & & \forall \mathbf{p}' \in \mathcal{P} \text{ s.t. } |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon, \\ & B(\mathbf{q}) \leq 1 & & \forall \mathbf{q} \in \mathcal{Q}^{\perp}. \end{aligned}$$

We denote $eff = eff_0$ and $eff^* = eff_0^*$ the 0-error bounds.

For any given distribution $\mathbf{p}$, its classical communication complexity is bounded below by the (log of the) efficiency. For randomized communication complexity with error ϵ, the bound is $\log(eff_\epsilon)$ and for quantum communication complexity, the bound is $\log(eff_\epsilon^*)$. Note that for any $\mathbf{p} \in \mathcal{Q}$, the quantum communication complexity is 0 and the eff^* bound is 1. For any function f, the efficiency bound $eff_\epsilon(\mathbf{p}_f)$ is equivalent to the partition bound [17, 27].

▶ **Proposition 4** ([27]). *For any* $\mathbf{p} \in \mathcal{P}$ *and any* $0 \leq \epsilon < 1/2$, $\mathrm{R}_\epsilon(\mathbf{p}) \geq \log(eff_\epsilon(\mathbf{p}))$ *and* $Q_\epsilon(\mathbf{p}) \geq \frac{1}{2}\log(eff_\epsilon^*(\mathbf{p}))$. *For any* $\mathbf{p} \in \mathcal{C}$ *and any* $0 \leq \epsilon \leq 1$, $\nu_\epsilon(\mathbf{p}) \leq 2eff_\epsilon(\mathbf{p})$.

Theorem 8 below involves upper bounds on the quantum efficiency bound. To give an upper bound on the quantum efficiency of a distribution $\mathbf{p}$, it is more convenient to use the primal formulation, and upper bounds can be given by exhibiting a local (or quantum) distribution with abort which satisfies the following two properties: the probability of aborting should be the same on all inputs x, y, and conditioned on not aborting, the outputs of the protocol should reproduce the distribution $\mathbf{p}$. The efficiency is inverse proportional to the probability of not aborting, so the goal is to abort as little as possible.

▶ **Proposition 5** ([27]). *For any distribution $\mathbf{p} \in \mathcal{P}$, $eff^*(\mathbf{p}) = 1/\eta^*$, with η^* the optimal value of the following optimization problem (non-linear, because $\mathcal{Q}^\perp$ is not a polytope).*

$$\begin{aligned} &\max_{\zeta, \mathbf{q} \in \mathcal{Q}^\perp} && \zeta \\ &\textit{subject to} && q(a,b|x,y) = \zeta p(a,b|x,y) \quad \forall x,y,a,b \in \mathcal{X}\times\mathcal{Y}\times\mathcal{A}\times\mathcal{B} \end{aligned}$$

Moreover, for any $0 \leq \epsilon \leq 1$, $eff^*_\epsilon(\mathbf{p}) = \min_{\mathbf{p}' \in \mathcal{P} : |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon} eff^*(\mathbf{p}')$.

3 Properties of Bell inequalities

Syntactically, there are two differences between the normalized Bell functionals (Definition 1) and the inefficiency-resistant ones (Definition 3). The first difference is that the normalization constraint is relaxed: for inefficiency-resistant functionals, the lower bound on the Bell value for local distributions is removed. Since this is a maximization problem, this relaxation allows for larger violations. This difference alone would not lead to a satisfactory measure of nonlocality, since one could obtain unbounded violations by shifting and dilating the Bell functional. The second difference prevents this. The upper bound is required to hold not only for local distributions, but also those that can abort. This is a much stronger condition. Notice that a local distribution can selectively abort on configurations that would otherwise tend to keep the Bell value small, making it harder to satisfy the constraint.

In this section, we show that normalized Bell violations can be modified to be resistant to local distributions that abort, while preserving the violation on any nonsignaling distribution, up to a factor of 3. This means that we can add the stronger constraint of resistance to local distributions that abort to Definition 1, incurring a loss of just a factor of 3, and the only remaining difference between the resulting linear programs is the relaxation of the lower bound (dropping the absolute value) for local distributions that abort.

▶ **Theorem 6.** *Let B be a normalized Bell functional on $\mathcal{A} \times \mathcal{B} \times \mathcal{X} \times \mathcal{Y}$ and $\mathbf{p} \in \mathcal{C}$ a nonsignaling distribution such that $B(\mathbf{p}) \geq 1$. Then there exists a normalized Bell functional B^* on $(\mathcal{A} \cup \{\perp\}) \times (\mathcal{B} \cup \{\perp\}) \times \mathcal{X} \times \mathcal{Y}$ with 0 coefficients on the $\perp$ outputs such that : $\forall \mathbf{p} \in \mathcal{C}$, $B^*(\mathbf{p}) \geq \frac{1}{3} B(\mathbf{p}) - \frac{2}{3}$, and $\forall \ell \in \mathcal{L}^\perp_{\text{det}}$, $|B^*(\ell)| \leq 1$.*

The formal proof of Theorem 6 is deferred to Appendix A, and we will only give its high-level structure in this part of the paper. First, we show (see Observation 17) how to rescale a normalized Bell functional so that it saturates its normalization constraint. Then, Definition 18 adds weights to abort events to make the Bell functional resistant to inefficiency. Finally, Lemma 19 removes the weights on the abort events of a Bell functional while keeping it bounded on the local set with abort, without dramatically changing the values it takes on the nonsignaling set. Our techniques are similar to the ones used in [31].

4 Exponential violations from communication bounds

Recently, Buhrman *et al.* gave a general construction to derive normalized Bell inequalities from any sufficiently large gap between classical and quantum communication complexity.

▶ **Theorem 7** ([10]). *For any function f for which there is a quantum protocol using q qubits of communication but no prior shared entanglement, there exists a quantum distribution $\mathbf{q} \in \mathcal{Q}$ and a normalized Bell functional B such that $B(\mathbf{q}) \geq \frac{\sqrt{R_{1/3}(f)}}{6\sqrt{3q}}(1-2^{-q})^{2q}$.*

Their construction is quite involved, requiring protocols to be memoryless, which they show how to achieve in general, and uses multiport teleportation to construct a quantum distribution. The Bell inequality they construct expresses a correctness constraint.

In this section, we show how to obtain large inefficiency-resistant Bell violations for quantum distributions from gaps between quantum communication and classical communication lower bounds. We first prove the stronger of two statements, which gives violations of $\frac{eff_\epsilon(\mathbf{p})}{eff^\star_{\epsilon'}(\mathbf{p})}$. For any problem for which a classical lower bound c is given using the efficiency or partition bound or any weaker method (including the rectangle bound and its variants), and any upper bound q on quantum communication complexity, it implies a violation of 2^{c-2q}.

▶ **Theorem 8.** *For any distribution $\mathbf{p} \in \mathcal{P}$ and any $0 \leq \epsilon' \leq \epsilon \leq 1$, if (B, β) is a feasible solution to the dual of $eff_\epsilon(\mathbf{p})$ and $(\zeta, \mathbf{q})$ is a feasible solution to the primal for $eff^\star_{\epsilon'}(\mathbf{p})$, then there is a quantum distribution $\overline{\mathbf{q}} \in \mathcal{Q}$ such that $B(\overline{\mathbf{q}}) \geq \zeta\beta$ and $B(\ell) \leq 1, \forall \ell \in \mathcal{L}^\perp_{det}$, and in particular, if both are optimal solutions, then $B(\overline{\mathbf{q}}) \geq \frac{eff_\epsilon(\mathbf{p})}{eff^\star_{\epsilon'}(\mathbf{p})}$. The distribution $\overline{\mathbf{q}}$ has one additional output per player compared to the distribution $\mathbf{p}$.*

Proof. Let (B, β) be a feasible solution to the dual of $eff_\epsilon(\mathbf{p})$, $\mathbf{p}'$ be such that $eff^\star_{\epsilon'}(\mathbf{p}) = eff^\star(\mathbf{p}')$ with $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon'$, and $(\zeta, \mathbf{q})$ be a feasible solution to the primal for $eff^\star(\mathbf{p}')$. From the constraints, we have $\mathbf{q} \in \mathcal{Q}^\perp$, $q(a,b|x,y) = \zeta p'(a,b|x,y)$ for all $(a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y}$, $B(\ell) \leq 1$ for all $\ell \in \mathcal{L}^\perp_{det}$, and $B(\mathbf{p}'') \geq \beta$ for all $\mathbf{p}''$ s.t. $|\mathbf{p}'' - \mathbf{p}|_1 \leq \epsilon$. Then $B(\mathbf{q}) = \zeta B(\mathbf{p}') \geq \zeta\beta$. However, $\mathbf{q} \in \mathcal{Q}^\perp$ but technically we want a distribution in $\mathcal{Q}$ (not one that aborts). So we add a new (valid) output 'A' to the set of outputs of each player, and they should output 'A' instead of aborting whenever $\mathbf{q}$ aborts. The resulting distribution, say $\overline{\mathbf{q}} \in \mathcal{Q}$ (with additional outcomes 'A' on both sides), is such that $B(\overline{\mathbf{q}}) = B(\mathbf{q})$ (since the Bell functional B does not have any weight on $\perp$ or on 'A'). ◀

Theorem 7 and Theorem 8 are both general constructions, but there are a few significant differences. Firstly, Theorem 8 requires a lower bound on the partition bound in the numerator, whereas Theorem 7 only requires a lower bound on communication complexity (which could be exponentially larger). Secondly, Theorem 7 requires a quantum communication protocol in the denominator, whereas our theorem only requires an upper bound on the quantum efficiency (which could be exponentially smaller). Thirdly, our bound is exponentially larger than Buhrman *et al.*'s for most problems considered here, and applies to subquadratic gaps, but their bounds are of the more restricted class of normalized Bell inequalities.

Theorem 8 gives an explicit Bell functional provided an explicit solution to the efficiency (partition) bound is given and the quantum distribution is obtained from a solution to the primal of $eff^\star$ (Proposition 5). Recall that a solution to the primal of $eff^\star$ is provided by a quantum zero-communication protocol that can abort, which conditioned on not aborting, outputs following $\mathbf{p}$. We can also start from a quantum protocol, as we show below. From the quantum protocol, we derive a quantum distribution using standard techniques.

▶ **Corollary 9.** *For any distribution* $\mathbf{p} \in \mathcal{P}$ *and any* $0 \leq \epsilon' \leq \epsilon \leq 1$ *such that* $R_\epsilon(\mathbf{p}) \geq \log(eff_\epsilon(\mathbf{p})) \geq c$ *and* $Q_{\epsilon'}(\mathbf{p}) \leq q$*, there exists an explicit inefficiency-resistant* B *derived from the efficiency lower bound, and an explicit quantum distribution* $\overline{\mathbf{q}} \in \mathcal{Q}$ *derived from the quantum protocol such that* $B(\overline{\mathbf{q}}) \geq 2^{c-2q}$*.*

Proof. Let (B, β) be an optimal solution to $eff_\epsilon(\mathbf{p})$ and let c be such that $eff_\epsilon(\mathbf{p}) = \beta \geq 2^c$. By optimality of B, we have $B(\mathbf{p}') \geq 2^c$ for any $\mathbf{p}'$ such that $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon$. Since $Q_{\epsilon'}(\mathbf{p}) \leq q$, there exists a q-qbit quantum protocol for some distribution $\mathbf{p}'$ with $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon$. Then, we can use teleportation to obtain a $2q$ classical bit, entanglement-assisted protocol for $\mathbf{p}'$. We can simulate it without communication by picking a shared $2q$-bit random string and running the protocol but without sending any messages. If the measurements do not match the string, output a new symbol 'A' (not in the output set of the quantum protocol and different from $\perp$). We obtain a quantum distribution $\overline{\mathbf{q}}$ such that $B(\overline{\mathbf{q}}) = B(\mathbf{p}')/2^{2q} \geq 2^{c-2q}$. ◀

Most often, communication lower bounds are not given as efficiency or partition bounds, but rather using variants of the corruption bound. We show in Section 6.1 how to map a corruption bound to explicit Bell coefficients.

5 Noise-resistant violations from communication bounds

Normalized Bell inequalities are naturally resistant to any local noise: if the observed distribution is $\tilde{\mathbf{p}} = (1-\varepsilon)\mathbf{p} + \varepsilon\ell$ for some $\ell \in \mathcal{L}$, then $B(\tilde{\mathbf{p}}) \geq (1-\varepsilon)B(\mathbf{p}) - \varepsilon$ since $|B(\ell)| \leq 1$. In inefficiency-resistant Bell inequalities, relaxing the absolute value leads to the possibility that $B(\ell)$ has a large negative value for some local ℓ. (Indeed, such large negative values are inherent to large gaps between ν and eff.) If this distribution were used as adversarial noise, the observed distribution, $(1-\varepsilon)\mathbf{p} + \varepsilon\ell$, could have a Bell value much smaller than 1. This makes inefficiency-resistant Bell inequalities susceptible to adversarial local noise.

Our construction from Theorem 8 is susceptible to uniform noise since most of the time, the output is 'A'. Uniform noise will disproportionately hit the non-'A' outputs, destroying the structure of the distribution. In Theorem 10, we show that our construction can be made resistant to uniform noise, by including a (possible) transcript from the protocol in the outputs. (Notice that this leads to a much larger output set.) Since the transcripts in our construction are teleportation measurements, they follow a uniform distribution, making the modified distribution resistant to uniform noise. The tolerance to noise comes from the error parameter in the classical communication lower bound.

Let $N_\varepsilon(\mathbf{p}) = \{(1-\delta)\mathbf{p} + \delta\mathbf{u},\ \delta \in [0, \varepsilon]\} \subseteq \mathcal{P}$ be the ε-noisy neighbourhood of $\mathbf{p}$, where $\mathbf{u}$ the uniform noise distribution, that is: $u(a, b|x, y) = \frac{1}{|\mathcal{A}|\cdot|\mathcal{B}|}$ for all $(a, b) \in \mathcal{A} \times \mathcal{B}$.

▶ **Theorem 10.** *For any distribution* $\mathbf{p} \in \mathcal{P}$ *and any* $0 \leq \epsilon' \leq \epsilon \leq 1$ *such that* $R_\epsilon(\mathbf{p}) \geq \log(eff_\epsilon(\mathbf{p})) \geq c$ *and* $Q_{\epsilon'}(\mathbf{p}) \leq q$*, there exists an explicit inefficiency-resistant* $\tilde{B}$ *derived from the efficiency lower bound, and an explicit quantum distribution* $\overline{\mathbf{q}} \in \mathcal{Q}$ *derived from the quantum protocol such that* $\tilde{B}(\mathbf{q}') \geq 2^{c-2q}$ *for any* $\mathbf{q}' \in N_{\varepsilon-\varepsilon'}(\overline{\mathbf{q}})$*.*

Proof. Let $\mathcal{A}$ (resp. $\mathcal{B}$) be Alice's (resp. Bob's) possible outputs for $\mathbf{p}$. From a quantum communication protocol for $\mathbf{p}'$ with $|\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon'$ using q qubits of communication, we construct an entanglement-assisted protocol using $2q$ bits of communication and teleportation. Let $\mathcal{M}_A$ (resp. $\mathcal{M}_B$) be the set of possible transcripts for Alice (resp. Bob), with $|\mathcal{M}_A| = M_A$ (resp. $|\mathcal{M}_B| = M_B$), and note that $\log M_A + \log M_B = 2q$.

We define the quantum distribution $\overline{\mathbf{q}}$ where Alice's possible outputs are $\mathcal{A} \times \mathcal{M}_A$ and Bob's possible outputs are $\mathcal{B} \times \mathcal{M}_B$. Alice proceeds as follows (Bob proceeds similarly):

1. She runs the quantum protocol for $\mathbf{p}'$ as if all bits received from Bob were 0.
2. She outputs (a, μ_A), where μ_A is the transcript of the messages she would have sent to Bob and a is the output she would have produced in the original protocol.

By definition, this distribution is such that, for all a, b, x, y, $\overline{q}(a, 0, b, 0|x, y) = \frac{1}{2^{2q}} p'(a, b|x, y)$.

Let $eff_\varepsilon(\mathbf{p}) \geq 2^c$ be achieved by the Bell functional B. By definition, we have $B(\ell) \leq 1$ for all $\ell \in \mathcal{L}_{det}^{\perp}$, and $B(\mathbf{p}'') \geq 2^c$ for all $\mathbf{p}''$ such that $|\mathbf{p}'' - \mathbf{p}|_1 \leq \varepsilon$. In particular for any $\mathbf{p}'' \in N_{\varepsilon-\varepsilon'}(\mathbf{p})$, that is, $\mathbf{p}'' = (1-\delta)\mathbf{p} + \delta\mathbf{u}$ for some $\delta \in [0, \varepsilon - \varepsilon']$, we have $|\mathbf{p}'' - \mathbf{p}|_1 \leq \varepsilon$ and therefore $B(\mathbf{p}'') = (1-\delta)B(\mathbf{p}') + \delta B(\mathbf{u}) \geq 2^c$, where $B(\mathbf{u}) = \frac{1}{AB}\sum_{a,b,x,y} B_{a,b,x,y}$.

Let the Bell functional $\tilde{B}$ for distributions over $(\mathcal{A} \times \mathcal{M}_A) \times (\mathcal{B} \times \mathcal{M}_B)$ be defined as follows: $\tilde{B}_{(a,\mu_A),(b,\mu_B),x,y} = B_{a,b,x,y}$ if $\mu_A = \mu_B = 0$, and $\tilde{B}_{(a,\mu_A),(b,\mu_B),x,y} = 0$ otherwise.

Let $\tilde{\mathcal{L}}_{det}^{\perp}$ be the local set for distributions over $(\mathcal{A} \times \mathcal{M}_A) \times (\mathcal{B} \times \mathcal{M}_B)$. Then $\tilde{B}$ satisfies $\tilde{B}(\ell) \leq 1$ for all $\ell \in \tilde{\mathcal{L}}_{det}^{\perp}$ (by assimilating any event with $\mu_A \neq 0$ or $\mu_B \neq 0$ to a $\perp$ event), as well as $\tilde{B}(\overline{\mathbf{q}}) = \frac{1}{2^{2q}} B(\mathbf{p}')$. Hence, $\forall \delta \in [0, \varepsilon - \varepsilon']$, we also have $(1-\delta)\tilde{B}(\overline{\mathbf{q}}) + \delta\tilde{B}(\mathbf{u}) =$
$= (1-\delta)\frac{1}{2^{2q}} B(\mathbf{p}') + \delta \frac{1}{ABM_AM_B} \sum_{a,\mu_A,b,\mu_B,x,y} \tilde{B}_{(a,\mu_A),(b,\mu_B),x,y}$
$= \frac{1}{2^{2q}} \left[(1-\delta)B(\mathbf{p}') + \delta \frac{1}{AB} \sum_{a,b,x,y} B_{a,b,x,y}\right] = \frac{1}{2^{2q}} \left[(1-\delta)B(\mathbf{p}') + \delta B(\mathbf{u})\right].$

Therefore, for all $\mathbf{q}' \in N_{\varepsilon-\varepsilon'}(\overline{\mathbf{q}})$, $\tilde{B}(\mathbf{q}') \geq 2^{c-2q}$, as claimed. ◀

6 Explicit constructions

6.1 From corruption bound to Bell inequality violation

We now explain how to construct an explicit Bell inequality violation from the corruption bound. The corruption bound, introduced by Yao in [46], is a very useful lower bound technique. It has been used for instance in [39] to get a tight $\Omega(n)$ lower bound on the randomized communication complexity of Disjointness (whereas the approximate rank, for example, can only show a lower bound of $\Theta(\sqrt{n})$). Let us recall that a rectangle R of $\mathcal{X} \times \mathcal{Y}$ is a subset of that set of the form $R_A \times R_B$, where $R_A \subseteq \mathcal{X}$ and $R_B \subseteq \mathcal{Y}$.

▶ **Theorem 11** (Corruption bound [46, 4, 26]). *Let f be a (possibly partial) Boolean function on $\mathcal{X} \times \mathcal{Y}$. Given $\gamma, \delta \in (0,1)$, suppose that there is a distribution μ on $\mathcal{X} \times \mathcal{Y}$ such that for every rectangle $R \subseteq \mathcal{X} \times \mathcal{Y}$*

$$\mu(R \cap f^{-1}(1)) > \gamma\mu(R \cap f^{-1}(0)) - \delta$$

Then, for every $\epsilon \in (0,1)$, $2^{R_\epsilon(f)} \geq \frac{1}{\delta}\left(\mu(f^{-1}(0)) - \frac{\epsilon}{\gamma}\right)$.

See, e.g., Lemma 3.5 in [5] for a rigorous treatment. For several problems, such a μ is already known. In Theorem 12 below, whose proof we defer to Appendix B, we show how to construct a Bell inequality violation from this type of bound.

▶ **Theorem 12.** *Let f be a (possibly partial) Boolean function on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$. Fix $z \in \{0,1\}$. Let μ be an input distribution, and $(U_i)_{i\in I}$ (resp. $(V_j)_{j\in J}$) be a family of pairwise nonoverlapping subsets of $f^{-1}(\bar{z})$ (resp. of $f^{-1}(z)$). Assume that there exists $g : \mathbb{N} \to (0, +\infty)$ such that, for any rectangle $R \subseteq \mathcal{X} \times \mathcal{Y}$*

$$\sum_{i\in I} u_i \mu(R \cap U_i) \geq \sum_{j\in J} v_j \mu(R \cap V_j) - g(n). \tag{1}$$

Then, the Bell functional B given by the following coefficients: for all $a, b, x, y \in \{0,1\} \times$

$\{0,1\} \times \mathcal{X} \times \mathcal{Y}$,

$$B_{a,b,x,y} = \begin{cases} 1/2(-u_i g(n)^{-1} \mu(x,y)) & \text{if } (x,y) \in U_i \text{ and } a \oplus b = z, \\ 1/2(v_j g(n)^{-1} \mu(x,y)) & \text{if } (x,y) \in V_j \text{ and } a \oplus b = z, \\ 0 & \text{otherwise.} \end{cases} \tag{2}$$

satisfies

$$B(\ell) \leq 1, \quad \forall \ell \in \mathcal{L}_{det}^{\perp}, \tag{3}$$

$$B(\mathbf{p}_f) = \frac{1}{2 \cdot g(n)} \sum_j v_j \mu(V_j) \tag{4}$$

and for any $\mathbf{p}' \in \mathcal{P}$ *such that* $|\mathbf{p}' - \mathbf{p}_f|_1 \leq \epsilon$ *:*

$$B(\mathbf{p}') \geq \frac{1}{2 \cdot g(n)} \left[\sum_j v_j \mu(V_j) - \epsilon \left(\sum_j |v_j| \mu(V_j) + \sum_i |u_i| \mu(U_i) \right) \right]. \tag{5}$$

For many other problems in the literature, such as Vector in Subspace and Tribes, stronger variants of the corruption bound are needed to obtain good lower bounds. These stronger variants have been shown to be no stronger than the partition bound (more specifically, the relaxed partition bound) [23]. The generalization in Theorem 12 of the hypothesis of Theorem 11, which the reader might have notice, allow us to construct explicit Bell functionals also for these problems.

6.2 Some specific examples

Using Corollary 9 and the construction to go from a corruption bound (or its variants) to a Bell inequality (Theorem 12), we give explicit Bell inequalities and violations for several problems studied in the literature. Since our techniques also apply to small gaps, we include problems for which the gap between classical and quantum communication complexity is polynomial.

Vector in Subspace

In the Vector in Subspace Problem $\mathrm{VSP}_{0,n}$, Alice is given an $n/2$ dimensional subspace of an n dimensional space over $\mathbb{R}$, and Bob is given a vector. This is a partial function, and the promise is that either Bob's vector lies in the subspace, in which case the function evaluates to 1, or it lies in the orthogonal subspace, in which case the function evaluates to 0. Note that the input set of $\mathrm{VSP}_{0,n}$ is continuous, but it can be discretized by rounding, which leads to the problem $\widetilde{\mathrm{VSP}}_{\theta,n}$ (see [24] for details). Klartag and Regev [24] show that the VSP can be solved with an $O(\log n)$ quantum protocol, but the randomized communication complexity of this problem is $\Omega(n^{1/3})$. As shown in [23], this is also a lower bound on the relaxed partition bound. Hence Corollary 9 yields the following.

▶ **Proposition 13.** *There exists a Bell inequality* B *and a quantum distribution* $\overline{\mathbf{q}}_{VSP} \in \mathcal{Q}$ *such that* $B\left(\overline{\mathbf{q}}_{VSP}\right) \in 2^{\Omega(n^{1/3}) - O(\log n)}$ *and for all* $\ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$.

Note that the result of [24] (Lemma 4.3) is not of the form needed to apply Theorem 12. It is yet possible to obtain an explicit Bell functional following the proof of Lemma 5.1 in [23].

Disjointness

In the Disjointness problem, the players receive two sets and have to determine whether they are disjoint or not. More formally, the Disjointness predicate is defined over $\mathcal{X} = \mathcal{Y} = \mathcal{P}([n])$ by $\mathrm{DISJ}_n(x, y) = 1$ iff x and y are disjoint. It is also convenient to see this predicate as defined over length n inputs, where $\mathrm{DISJ}_n(x, y) = 1$ for $x, y \in \{0,1\}^n$ if and only if $|\{i : x_i = 1 = y_i\}| = 0$. The communication complexity for DISJ_n is $\Omega(n)$ using a corruption bound [39] and there is a quantum protocol using $O(\sqrt{n})$ communication [1]. Combining these results with ours, we obtain the following.

▶ **Proposition 14.** *There is a quantum distribution $\overline{\mathbf{q}}_{\mathrm{DISJ}} \in \mathcal{Q}$ and an explicit Bell inequality B satisfying: $B(\overline{\mathbf{q}}_{\mathrm{DISJ}}) = 2^{\Omega(n)-O(\sqrt{n})}$, and for all $\ell \in \mathcal{L}^{\perp}_{det}$, $B(\ell) \leq 1$.*

The proof is deferred to the Appendix (see Section C.1).

Tribes

Let $r \geq 2$, $n = (2r+1)^2$. Let $\mathrm{TRIBES}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be defined as: $\mathrm{TRIBES}_n(x, y) := \bigwedge_{i=1}^{\sqrt{n}} \left(\bigvee_{j=1}^{\sqrt{n}} (x_{(i-1)\sqrt{n}+j} \wedge y_{(i-1)\sqrt{n}+j}) \right)$. The Tribes function has an $\Omega(n)$ classical lower bound [16] using the smooth rectangle bound and a $O(\sqrt{n}(\log n)^2)$ quantum protocol [9]. Combining these results with ours, we obtain the following.

▶ **Proposition 15.** *There is a quantum distribution $\overline{\mathbf{q}}_{\mathrm{TRIBES}} \in \mathcal{Q}$ and an explicit Bell inequality B satisfying: $B(\overline{\mathbf{q}}_{\mathrm{TRIBES}}) = 2^{\Omega(n)-O(\sqrt{n}(\log n)^2)}$, and for all $\ell \in \mathcal{L}^{\perp}_{det}$, $B(\ell) \leq 1$.*

The proof is deferred to the Appendix (see Section C.2).

Gap Orthogonality

The Gap Orthogonality (ORT) problem was introduced by Sherstov as an intermediate step to prove a lower bound for the Gap Hamming Distance (GHD) problem [41]. We derive an explicit Bell inequality for ORT from Sherstov's lower bound of $\Omega(n)$, shown in [23] to be a relaxed partition bound. (Applying Corollary 9 also gives a (non-explicit) violation for GHD.) The quantum upper bound is $O(\sqrt{n}\log n)$ by the general result of [9]. In the ORT problem, the players receive vectors and need to tell whether they are nearly orthogonal or far from orthogonal. More formally, we consider the input space $\{-1,+1\}^n$ (to stick to the usual notations for this problem), and we denote $\langle \cdot, \cdot \rangle$ the scalar product on $\{-1,+1\}^n$. Let $\mathrm{ORT}_n : \{-1,+1\}^n \times \{-1,+1\}^n \to \{-1,+1\}$ be the partial function defined as in [41] by: $\mathrm{ORT}_n(x, y) = -1$ if $|\langle x, y \rangle| \leq \sqrt{n}$, and $\mathrm{ORT}_n(x, y) = +1$ if $|\langle x, y \rangle| \geq 2\sqrt{n}$. Combining the results mentioned above with ours, we obtain the following.

▶ **Proposition 16.** *There is a quantum distribution $\overline{\mathbf{q}}_{\mathrm{ORT}} \in \mathcal{Q}$ and an explicit Bell inequality B satisfying: $B(\overline{\mathbf{q}}_{\mathrm{ORT}}) = 2^{\Omega(n)-O(\sqrt{n}\log n)}$, and for all $\ell \in \mathcal{L}^{\perp}_{det}$, $B(\ell) \leq 1$.*

The proof is deferred to the Appendix (see Section C.3).

7 Discussion

We have given three main results. First, we showed that normalized Bell inequalities can be modified to be bounded in absolute value on the larger set of local distributions that can abort without significantly changing the value of the violations achievable with nonsignaling

distributions. Then, we showed how to derive large inefficiency-resistant Bell violations from any gap between the partition bound and the quantum communication complexity of some given distribution $\mathbf{p}$. The distributions $\mathbf{q}$ achieving the large violations are relatively simple (only 3 outputs for boolean distributions $\mathbf{p}$) and can be made resistant to uniform noise at the expense of an increase in the number of outputs exponential in $Q(\mathbf{p})$. Finally, we showed how to construct explicit Bell inequalities when the separation between classical and quantum communication complexity is proven via the corruption bound.

From a practical standpoint, the specific Bell violations we have studied are probably not feasible to implement, because the parameters needed are still impractical or the quantum states are infeasible to implement. However, our results suggest that we could consider functions with small gaps in communication complexity, in order to find practical Bell inequalities that are robust against uniform noise and detector inefficiency. Let us consider an experimental setup with non-abort probability η per side, and ε uniform noise. Suppose we have a Boolean function with a lower bound of $c > 3\log(1/\eta^2)$ on classical communication with ε' error, and an $(\varepsilon'-\varepsilon)$-correct quantum protocol, with $\varepsilon' > \varepsilon$, using $q = \log(1/\eta^2)$ qubits. Our construction gives an inefficiency-resistant Bell violation of $2^{c-2q} > 1/\eta^2$, which is robust against ε uniform noise. (The number of outcomes per side increases to $\frac{2}{\eta^2}$.) Factoring in the inefficiency, the observed violation would still be $\eta^2 2^{c-2q} > 1$.

Regarding upper bounds, since (the log of) efficiency is a lower bound on communication complexity, inefficiency-resistant Bell violations are bounded above by the number of inputs per side. For dimension d and number of outcomes K, we obtain the upper bound $eff_\epsilon(\mathbf{q}) \leq 2^{O((\frac{Kd}{\epsilon})^2 \log^2(K))}$ for quantum distributions, by combining known bounds. Indeed, we know that $R_\epsilon(\mathbf{p}) \leq O((\frac{K\nu(\mathbf{p})}{\epsilon})^2 \log^2(K))$ for any $\mathbf{p} \in \mathcal{C}$ (see [12]). Combining this with the bounds $eff_\epsilon(\mathbf{p}) \leq 2^{R_\epsilon(\mathbf{p})}$ (Proposition 4), and $\nu(\mathbf{q}) \leq O(d)$ for any $\mathbf{q} \in \mathcal{Q}$ (see [21]), gives the desired upper bound. Hence unbounded violations are possible for $K = 3$ outputs per side.

References

1. S. Aaronson and A. Ambainis. Quantum search of spatial regions. *Theory of Computing*, 1:47–79, 2005.
2. A. Acín, T. Durt, N. Gisin, and J. I. Latorre. Quantum nonlocality in two three-level systems. *Physical Review A*, 65:052325, 2002.
3. M. Ardehali. Bell inequalities with a magnitude of violation that grows exponentially with the number of particles. *Physical Review A*, 46:5375–5378, 1992.
4. L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *Proc. 27th FOCS*, pages 337–347. IEEE, 1986.
5. P. Beame, T. Pitassi, N. Segerlind, and A. Wigderson. A strong direct product theorem for corruption and the multiparty communication complexity of disjointness. *Computational Complexity*, 15(4):391–432, 2006.
6. J. S. Bell. On the Einstein Podolsky Rosen paradox. *Physics*, 1:195, 1964.
7. G. Brassard, R. Cleve, and A. Tapp. Cost of exactly simulating quantum entanglement with classical communication. *Physical Review Letters*, 83(9):1874, 1999.
8. N. Brunner, D. Cavalcanti, A. Salles, and P. Skrzypczyk. Bound nonlocality and activation. *Physical Review Letters*, 106:020402, 2011.
9. H. Buhrman, R. Cleve, and A. Wigderson. Quantum vs classical communication and computation. In *Proc. 30th STOC*, pages 63–68, 1998.
10. H. Buhrman, Ł. Czekaj, A. Grudka, Mi. Horodecki, P. Horodecki, M. Markiewicz, F. Speelman, and S. Strelchuk. Quantum communication complexity advantage implies violation of a Bell inequality. *Proceedings of the National Academy of Sciences*, 113(12):3191–3196, 2016.

11 H. Buhrman, O. Regev, G. Scarpa, and R. de Wolf. Near-optimal and explicit Bell inequality violations. *Theory of Computing*, 8(1):623–645, 2012.

12 J. Degorre, M. Kaplan, S. Laplante, and J. Roland. The communication complexity of non-signaling distributions. *Quantum information & computation*, 11(7-8):649–676, 2011.

13 M. Forster, S. Winkler, and S. Wolf. Distilling nonlocality. *Physical Review Letters*, 102:120401, 2009.

14 R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués. Operational framework for nonlocality. *Physical Review Letters*, 109:070401, 2012.

15 A. Grothendieck. Résumé de la théorie métrique des produits tensoriels topologiques. *Boletim da Sociedade de Matemática de São Paulo*, 8:1–79, 1953.

16 P. Harsha and R. Jain. A Strong Direct Product Theorem for the Tribes Function via the Smooth-Rectangle Bound. In *Procs. 33rd FSTTCS*, volume 24, pages 141–152, 2013.

17 R. Jain and H. Klauck. The partition bound for classical communication complexity and query complexity. In *Proc. 25th CCC*, pages 247–258, 2010.

18 T. S. Jayram, R. Kumar, and D. Sivakumar. Two applications of information complexity. In *Proc. 35th STOC*, pages 673–682, 2003.

19 M. Junge and C. Palazuelos. Large violation of Bell inequalities with low entanglement. *Communications in Mathematical Physics*, 306(3):695–746, 2011.

20 M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf. Operator space theory: A natural framework for Bell inequalities. *Physical Review Letters*, 104:170405, 2010.

21 M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf. Unbounded violations of bipartite Bell inequalities via operator space theory. *Communications in Mathematical Physics*, 300(3):715–739, 2010.

22 D. Kaszlikowski, P. Gnaciński, M. Żukowski, W. Miklaszewski, and A. Zeilinger. Violations of local realism by two entangled N-dimensional systems are stronger than for two qubits. *Physical Review Letters*, 85:4418–4421, 2000.

23 I. Kerenidis, S. Laplante, V. Lerays, J. Roland, and D. Xiao. Lower bounds on information complexity via zero-communication protocols and applications. *SIAM Journal on Computing*, 44(5):1550–1572, 2015.

24 B. Klartag and O. Regev. Quantum one-way communication can be exponentially stronger than classical communication. In *Proc. 43th STOC*, pages 31–40, 2011.

25 G. Kol, S. Moran, A. Shpilka, and A. Yehudayoff. Approximate nonnegative rank is equivalent to the smooth rectangle bound. In *Automata, Languages, and Programming*, pages 701–712. Springer, 2014.

26 E. Kushilevitz and N. Nisan. *Communication complexity*. Cambridge University Press, 1997.

27 S. Laplante, V. Lerays, and J. Roland. Classical and quantum partition bound and detector inefficiency. In *Proc. 39th ICALP*, pages 617–628, 2012.

28 W. Laskowski, T. Paterek, M. Żukowski, and Č Brukner. Tight multipartite Bell's inequalities involving many measurement settings. *Physical Review Letters*, 93(20):200401, 2004.

29 N. Linial and A. Shraibman. Lower bounds in communication complexity based on factorization norms. *Random Structures & Algorithms*, 34(3):368–394, 2009.

30 S. Massar. Nonlocality, closing the detection loophole, and communication complexity. *Physical Review A*, 65:032121, 2002.

31 S. Massar, S. Pironio, J. Roland, and B. Gisin. Bell inequalities resistant to detector inefficiency. *Physical Review A*, 66:052112, 2002.

32 T. Maudlin. Bell's inequality, information transmission, and prism models. In *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association*, pages 404–417. JSTOR, 1992.

33 D. N. Mermin. Extreme quantum entanglement in a superposition of macroscopically distinct states. *Physical Review Letters*, 65(15):1838, 1990.

34 K. Nagata, W. Laskowski, and T. Paterek. Bell inequality with an arbitrary number of settings and its applications. *Physical Review A*, 74(6):062109, 2006.

35 C. Palazuelos and Z. Yin. Large bipartite Bell violations with dichotomic measurements. *Physical Review A*, 92:052313, 2015.

36 D. Pérez-García, M. M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge. Unbounded violation of tripartite Bell inequalities. *Communications in Mathematical Physics*, 279(2):455–486, 2008.

37 S. Pironio. Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations. *Physical Review A*, 68(6):062102, 2003.

38 R. Raz. Exponential separation of quantum and classical communication complexity. In *Proc. 31th STOC*, pages 358–367, 1999.

39 A. A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

40 A. A. Razborov. Quantum communication complexity of symmetric predicates. *Izvestiya: Mathematics*, 67(1):145, 2003.

41 A. A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.

42 M. Steiner. Towards quantifying non-local information transfer: finite-bit non-locality. *Physics Letters A*, 270(5):239–244, 2000.

43 B. F. Toner and D. Bacon. Communication cost of simulating Bell correlations. *Physical Review Letters*, 91(18):187904, 2003.

44 B. S. Tsirel'son. Quantum analogues of the Bell inequalities. the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36(4):557–570, 1987.

45 A. C. C. Yao. Some complexity questions related to distributed computing. In *Proc. 11th STOC*, pages 209–213, 1979.

46 A. C. C. Yao. Lower bounds by probabilistic arguments. In *Proc. 24th FOCS*, pages 420–428. IEEE, 1983.

47 A. C. C. Yao. Quantum circuit complexity. In *Proc. 34th FOCS*, pages 352–361. IEEE, 1993.

A Proof of Theorem 6

▶ Observation 17. Let B be a non-constant normalized Bell functional and $\mathbf{p} \in \mathcal{C}$ such that $B(\mathbf{p}) \geq 1$. Consider $\ell^- \in \mathcal{L}_{\text{det}}$ such that $B(\ell^-) = m = \min\{B(\ell) | \ell \in \mathcal{L}_{\text{det}}\}$ and $\ell^+ \in \mathcal{L}_{\text{det}}$ such that $B(\ell^+) = M = \max\{B(\ell) | \ell \in \mathcal{L}_{\text{det}}\}$. We have $m < M$ because B is non-constant. The Bell functional $\tilde{B}$ defined by $\tilde{B}(\cdot) = \frac{1}{M-m}(2B(\cdot) - M - m)$, is such that $\tilde{B}(\ell^+) = 1$, $\tilde{B}(\ell^-) = -1$, $|\tilde{B}(\ell)| \leq 1$ for all $\ell \in \mathcal{L}_{\text{det}}^{\perp}$, and $\tilde{B}(\mathbf{p}) \geq B(\mathbf{p})$.

▶ **Definition 18.** For any two families of distributions, $\mathbf{m}_A = (m_A(\cdot|x))_{x \in \mathcal{X}}$ over outcomes in $\mathcal{A}$ for Alice and $\mathbf{m}_B = (m_B(\cdot|y))_{y \in \mathcal{Y}}$ over outcomes in $\mathcal{B}$ for Bob, $f_{\mathbf{m}_A, \mathbf{m}_B} : \mathcal{C}^{\perp} \to \mathcal{C}$ replaces abort events on Alice's (resp. Bob's) side by a sample from $\mathbf{m}_A$ (resp. $\mathbf{m}_B$).

For B a normalized Bell functional with coefficients only on non-abort events, the Bell functional $B^{\perp}_{\mathbf{m}_A,\mathbf{m}_B}$ on $(\mathcal{A}\cup\{\perp\})\times(\mathcal{B}\cup\{\perp\})\times\mathcal{X}\times\mathcal{Y}$ is given by

$$(B^{\perp}_{\mathbf{m}_A,\mathbf{m}_B})_{a,b,x,y} = B_{a,b,x,y} + \chi_{\{\perp\}}(a)\sum_{a'\neq\perp} m_A(a'|x)B_{a',b,x,y}$$
$$+\chi_{\{\perp\}}(b)\sum_{b'\neq\perp} m_B(b'|y)B_{a,b',x,y} + \chi_{\{\perp\}}(a)\chi_{\{\perp\}}(b)\sum_{a',b'\neq\perp} m_A(a'|x)m_B(b'|y)B_{a',b',x,y}$$

where $\chi_{\mathcal{S}}$ is the indicator function for set $\mathcal{S}$ taking value 1 on $\mathcal{S}$ and 0 everywhere else.

Note that $f_{\mathbf{m}_A,\mathbf{m}_B}$ preserves locality, and $B^{\perp}_{\mathbf{m}_A,\mathbf{m}_B}(\mathbf{p}) = B(f_{\mathbf{m}_A,\mathbf{m}_B}(\mathbf{p}))$, $\forall\mathbf{p}\in\mathcal{C}^{\perp}$, so $B^{\perp}_{\mathbf{m}_A,\mathbf{m}_B}(\mathbf{p}) = B(\mathbf{p})$, for all $\mathbf{p}\in\mathcal{C}$, and $|B^{\perp}_{\mathbf{m}_A,\mathbf{m}_B}(\ell)|\leq 1$, for all $\ell\in\mathcal{L}^{\perp}$.

▶ **Lemma 19.** *Let B' be a normalized Bell functional on $(\mathcal{A}\cup\{\perp\})\times(\mathcal{B}\cup\{\perp\})\times\mathcal{X}\times\mathcal{Y}$. (possibly with weights on $\perp$.) Then the Bell functional B'' on the same set defined by*

$$B''_{a,b,x,y} = B'_{a,b,x,y} - B'_{a,\perp,x,y} - B'_{\perp,b,x,y} + B'_{\perp,\perp,x,y}, \tag{6}$$

for all $(a,b,x,y)\in(\mathcal{A}\cup\{\perp\})\times(\mathcal{B}\cup\{\perp\})\times\mathcal{X}\times\mathcal{Y}$ satisfies :
1. *If $B''_{a,b,x,y}=0$, then $a=\perp$ or $b=\perp$,*
2. *$\forall\mathbf{p}\in\mathcal{C}$, $B''(\mathbf{p}) = B'(\mathbf{p}) - B'(\mathbf{p}_{A,\perp}) - B'(\mathbf{p}_{\perp,B}) + B'(\mathbf{p}_{\perp,\perp})$,*
where $\mathbf{p}_{A,\perp}\in\mathcal{L}^{\perp}$ (resp. $\mathbf{p}_{\perp,B}\in\mathcal{L}^{\perp}$) is the local distribution obtained from $\mathbf{p}$ if Bob (resp. Alice) replaces any of his (resp. her) outputs by $\perp$, and $\mathbf{p}_{\perp,\perp}\in\mathcal{L}^{\perp}$ is the local distribution where Alice and Bob always output $\perp$. In Item 2 above, for any $\mathbf{p}'$, $B'(\mathbf{p}') = \sum_{a,b,x,y} B'_{a,b,x,y}\mathbf{p}'(a,b|x,y)$ where the sum is also over the abort events.

Proof. Item 1 follows from (6). We prove Item 2. For $\mathbf{p}\in\mathcal{C}^{\perp}$ with marginals $\mathbf{p}_A$ and $\mathbf{p}_B$, we have: for all $y\in Y$, $p_A(a|x) = \sum_{b\in\mathcal{B}\cup\{\perp\}} p(a,b|x,y)$, and for all $x\in X$, $p_B(b|y) = \sum_{a\in\mathcal{A}\cup\{\perp\}} p(a,b|x,y)$. For the remainder of this proof, summations involving a (resp. b) are over $a\in\mathcal{A}\cup\{\perp\}$ (resp. $b\in\mathcal{B}\cup\{\perp\}$). By definition, $p_{A,\perp}(a,b|x,y) = p_A(a|x)\chi_{\{\perp\}}(b)$, $p_{\perp,B}(a,b|x,y) = \chi_{\{\perp\}}(a)p_B(b|y)$, and $p_{\perp,\perp}(a,b|x,y) = \chi_{\{\perp\}}(a)\chi_{\{\perp\}}(b)$. We have:

$$\begin{aligned}
B''(\mathbf{p}) &= \sum_{a,b,x,y}\left[B'_{a,b,x,y} - B'_{a,\perp,x,y} - B'_{\perp,b,x,y} + B'_{\perp,\perp,x,y}\right]p(a,b|x,y)\\
&= \sum_{a,b,x,y} B'_{a,b,x,y}p(a,b|x,y) - \sum_{a,x,y} B'_{a,\perp,x,y}\sum_{b} p(a,b|x,y)\\
&\quad - \sum_{b,x,y} B'_{\perp,b,x,y}\sum_{a} p(a,b|x,y) + \sum_{x,y} B'_{\perp,\perp,x,y}\sum_{a,b} p(a,b|x,y)\\
&= B'(\mathbf{p}) - \sum_{a,x,y} B'_{a,\perp,x,y}p_A(a|x) - \sum_{b,x,y} B'_{\perp,b,x,y}p_B(b|y) + \sum_{x,y} B'_{\perp,\perp,x,y}\\
&= B'(\mathbf{p}) - B'(\mathbf{p}_{A,\perp}) - B'(\mathbf{p}_{\perp,B}) + B'(\mathbf{p}_{\perp,\perp}).
\end{aligned}$$

◀

We are now ready to prove Theorem 6.

Proof of Theorem 6. From Observation 17, we can assume that there exists $\ell^-,\ell^+\in\mathcal{L}_{\text{det}}$ such that $B(\ell^-)=-1$ and $B(\ell^+)=1$ (otherwise, we replace B by its saturated version $\tilde{B}$). Since ℓ^- and ℓ^+ are deterministic distributions, we have: $\ell^- = \ell^-_A\otimes\ell^-_B$ and $\ell^+ = \ell^+_A\otimes\ell^+_B$, for some marginals $\ell^-_A,\ell^-_B,\ell^+_A$, and ℓ^+_B. We consider the two replacing Bell functionals from Definition 18 constructed from (B,ℓ^-_A,ℓ^-_B) on one hand, and (B,ℓ^+_A,ℓ^+_B) on the other hand. Taking $B' = \frac{1}{2}(B^{\perp}_{\ell^-_A,\ell^-_B} + B^{\perp}_{\ell^+_A,\ell^+_B})$, we have $|B'(\ell)|\leq 1$, $\forall\ell\in\mathcal{L}^{\perp}$, and therefore we can apply

Lemma 19 to B' to get B''. Since $B'(\mathbf{p}_{\perp,\perp}) = \frac{1}{2}(B^{\perp}_{\ell_A^-,\ell_B^-}(\mathbf{p}_{\perp,\perp}) + B^{\perp}_{\ell_A^+,\ell_B^+}(\mathbf{p}_{\perp,\perp})) = \frac{1}{2}(B(\ell^-) + B(\ell^+)) = 0$, we have for all $\mathbf{p} \in \mathcal{C}^{\perp}$, $B''(\mathbf{p}) = B'(\mathbf{p}) - B'(\mathbf{p}_{A,\perp}) - B'(\mathbf{p}_{\perp,B})$. Therefore $B^* = \frac{1}{3}B''$ satisfies all the required properties. In particular, since $|B'(\ell)| \leq 1\ \forall \ell \in \mathcal{L}^{\perp}$, we have for any $\mathbf{p} \in \mathcal{C}$, $B^*(\mathbf{p}) \geq \frac{1}{3}B'(\mathbf{p}) - \frac{1}{3}|B'(\mathbf{p}_{A,\perp})| - \frac{1}{3}|B'(\mathbf{p}_{\perp,B})| \geq \frac{1}{3}B'(\mathbf{p}) - \frac{2}{3}$, and for any $\ell \in \mathcal{L}^{\perp}$, $|B^*(\ell)| \leq \frac{1}{3}|B'(\ell)| + \frac{1}{3}|B'(\ell_{A,\perp})| + \frac{1}{3}|B'(\ell_{\perp,B})| \leq 1$. ◀

B Proof of Theorem 12

Proof. Let us first set $B_{z,x,y} = B_{a,b,x,y}$ for all $a \oplus b = z$. Let $\ell \in \mathcal{L}^{\perp}_{det}$. Then, we have:

$$B(\ell) = \sum_{(x,y)\in R} B_{z,x,y} + \sum_{(x,y)\in S} B_{z,x,y}$$

where R and S are the two rectangles where ℓ outputs z. Let us take a rectangle R. Then :

$$\sum_{(x,y)\in R} B_{z,x,y} = \frac{1}{2 \cdot g(n)} \left(\sum_j v_j \mu(V_j \cap R) - \sum_i u_i \mu(U_i \cap R) \right) \leq 1/2$$

with the inequality following from (1). This proves (3).

Let us now compute $B(\mathbf{p}_f)$. By linearity of B and the definition of its coefficients, we have:

$$\begin{aligned} B(\mathbf{p}_f) &= \sum_{a,b,x,y} B_{a,b,x,y} \mathbf{p}_f(a,b|x,y) \\ &= \frac{1}{2} \sum_{(x,y)\in f^{-1}(z),a,b} B_{a,b,x,y} \chi_{\{z\}}(a \oplus b) + \frac{1}{2} \sum_{(x,y)\in f^{-1}(\bar{z}),a,b} B_{a,b,x,y} \chi_{\{\bar{z}\}}(a \oplus b) \\ &= 1/2 \sum_j \sum_{(x,y)\in V_j} v_j g(n)^{-1} \mu(x,y) \\ &= \frac{1}{2 \cdot g(n)} \sum_j v_j \mu(V_j) \end{aligned}$$

(for the third equality we used the fact that $B_{a,b,x,y} = 0$ when $a \oplus b = \bar{z}$). This proves (4).

Moreover, for any family of additive error terms $\Delta(a,b|x,y) \in [-1,1]$ such that

$$\sum_{a,b} |\Delta(a,b|x,y)| \leq \epsilon \qquad \forall x,y \in \mathcal{X} \times \mathcal{Y},$$

denoted collectively as $\mathbf{\Delta}$, we have

$$\begin{aligned} |B(\mathbf{\Delta})| &= \left| \sum_{a,b,x,y} B_{a,b,x,y} \Delta(a,b|x,y) \right| \\ &= \frac{1}{2 \cdot g(n)} \left| \sum_{a,b\,:\,a\oplus b=z} \left[\sum_i \sum_{(x,y)\in U_i} (-u_i) \mu(x,y) \Delta(a,b|x,y) + \right. \right. \\ &\qquad \left. \left. \sum_j \sum_{(x,y)\in V_j} v_j \mu(x,y) \Delta(a,b|x,y) \right] \right| \end{aligned}$$

$$\leq \frac{1}{2 \cdot g(n)} \left[\sum_i \sum_{(x,y) \in U_i} |u_i| \mu(x,y) \left(\sum_{a,b} |\Delta(a,b|x,y)| \right) + \right.$$
$$\left. \sum_j \sum_{(x,y) \in V_j} |v_j| \mu(x,y) \left(\sum_{a,b} |\Delta(a,b|x,y)| \right) \right]$$
$$\leq \frac{\epsilon}{2 \cdot g(n)} \left[\sum_i |u_i| \mu(U_i) + \sum_j |v_j| \mu(V_j) \right]$$

From this calculation and (4), we obtain, for $\mathbf{p}' = \mathbf{p}_f + \mathbf{\Delta}$:

$$B(\mathbf{p}') = B(\mathbf{p}_f) + B(\mathbf{\Delta}) \geq \frac{1}{2 \cdot g(n)} \left[\sum_j v_j \mu(V_j) - \epsilon \left(\sum_j |v_j| \mu(V_j) + \sum_i |u_i| \mu(U_i) \right) \right],$$

which proves (5). ◀

C Explicit examples

Let us formulate a special case of Theorem 12 that will be useful in the examples. Here there is just one subset in $f^{-1}(0)$ and one in $f^{-1}(1)$.

▶ **Corollary 20.** *Let f be a (possibly partial) Boolean function on $\mathcal{X} \times \mathcal{Y}$, where $\mathcal{X}, \mathcal{Y} \subseteq \{0,1\}^n$. Given $\gamma \in (0,1)$ and $g : \mathbb{N} \to (0,1)$, suppose that there is a distribution μ on $\mathcal{X} \times \mathcal{Y}$ such that: for any rectangle $R \subseteq \mathcal{X} \times \mathcal{Y}$,*

$$\mu(R \cap f^{-1}(1)) > \gamma \mu(R \cap f^{-1}(0)) - g(n). \tag{7}$$

Then μ satisfies (1) with $z = 0$, $i = j = 1$, $U_1 = f^{-1}(1)$, $V_1 = f^{-1}(0)$, $u_1 = 1$, $v_1 = \gamma$. Let B be defined by (2), that is: for all $a, b, x, y \in \{0,1\} \times \{0,1\} \times \mathcal{X} \times \mathcal{Y}$,

$$B_{a,b,x,y} = \begin{cases} -\frac{1}{2 \cdot g(n)} \mu(x,y) & \text{if } f(x,y) = 1 \text{ and } a \oplus b = 0 \\ \frac{\gamma}{2 \cdot g(n)} \mu(x,y) & \text{if } f(x,y) = 0 \text{ and } a \oplus b = 0 \\ 0 & \text{otherwise.} \end{cases}$$

Then, B satisfies

$$B(\ell) \leq 1, \quad \forall \ell \in \mathcal{L}_{det}^{\perp},$$
$$B(\mathbf{p}_f) = \frac{\gamma}{2 \cdot g(n)} \mu(f^{-1}(0))$$

and for any $\mathbf{p}' \in \mathcal{P}$ such that $|\mathbf{p}' - \mathbf{p}_f|_1 \leq \epsilon$:

$$B(\mathbf{p}') \geq \frac{1}{2 \cdot g(n)} \Big[\gamma \mu(f^{-1}(0)) - \epsilon \big(\gamma \mu(f^{-1}(0)) + \mu(f^{-1}(1)) \big) \Big].$$

C.1 Disjointness

In [39], Razborov proved the following.

▶ **Lemma 21** ([39]). *There exist two distributions μ_0 and μ_1 with $\mathrm{supp}(\mu_0) \subseteq \mathrm{DISJ}_n^{-1}(1)$ and $\mathrm{supp}(\mu_1) \subseteq \mathrm{DISJ}_n^{-1}(0)$, such that: for any rectangle R in the input space,*

$$\mu_1(R) \geq \Omega(\mu_0(R)) - 2^{\Omega(n)}.$$

Following his proof, one can check that we actually have:

$$\mu_1(R) \geq \frac{1}{45}\mu_0(R) - 2^{-\epsilon n + \log_2(2/9)}.$$

So, letting $\mu := (\mu_0 + \mu_1)/2$,

$$\mu(R \cap f^{-1}(0)) \geq \frac{1}{45}\mu(R \cap f^{-1}(1)) - 2^{-\epsilon n + \log_2(4/9)}. \tag{8}$$

▶ Remark. Actually, $\mathrm{supp}(\mu_1) = A_1 := \{(x,y) : |x| = |y| = m, |x \cap y| = 1\} \subseteq \mathrm{DISJ}_n^{-1}(0)$.

Note that by this construction, $\mu(f^{-1}(0)) = \mu(f^{-1}(1)) = 1/2$. Combining (8) with Corollary 20 (with $g(n) = 2^{-\epsilon n + \log_2(4/9)}$), we obtain:

▶ **Corollary 22.** *There exists a Bell inequality B satisfying: $\forall \ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$,*

$$B(\mathbf{p}_{\mathrm{DISJ}_n}) = \frac{1}{90}2^{\epsilon n - \log_2(4/9)},$$

and for any distribution $\mathbf{p}' \in \mathcal{P}$ such that $|\mathbf{p}' - \mathbf{p}_{\mathrm{DISJ}_n}|_1 \leq \varepsilon$,

$$B(\mathbf{p}') \geq 2^{\epsilon n - \log_2(4/9)}\frac{1 - 46\epsilon}{90}.$$

More precisely, Theorem 12 gives an explicit construction of such a Bell inequality: we can define B as:

$$B_{a,b,x,y} = \begin{cases} -2^{\epsilon n - \log_2(4/9)}\mu(x,y) & \text{if } \mathrm{DISJ}_n(x,y) = 0 \text{ and } a \oplus b = 1 \\ \frac{1}{45}2^{\epsilon n - \log_2(4/9)}\mu(x,y) & \text{if } \mathrm{DISJ}_n(x,y) = 1 \text{ and } a \oplus b = 1 \\ 0 & \text{otherwise.} \end{cases}$$

To obtain Proposition 14, we use Corollary 9 together with the fact that $Q_{\varepsilon'}(\mathrm{DISJ}_n) = O(\sqrt{n})$.

C.2 Tribes

Let $n = (2r+1)^2$ with $r \geq 2$ and let $\mathrm{TRIBES}_n : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be defined as:

$$\mathrm{TRIBES}_n(x,y) := \bigwedge_{i=1}^{\sqrt{n}} \left(\bigvee_{j=1}^{\sqrt{n}} (x_{(i-1)\sqrt{n}+j} \text{ and } y_{(i-1)\sqrt{n}+j}) \right).$$

In [16][Sec. 3] the following is proven:

▶ **Lemma 23.** *There exists a probability distribution μ on $\{0,1\}^n \times \{0,1\}^n$ for which there exist numbers $\alpha, \lambda, \gamma, \delta > 0$ such that for sufficiently large n and for any rectangle R in the input space:*

$$\gamma\mu(U_1 \cap R) \geq \alpha\mu(V_1 \cap R) - \lambda\mu(V_2 \cap R) - 2^{-\delta n/2+1}$$

where $U_1 = \mathrm{TRIBES}_n^{-1}(0)$, $\{V_1, V_2\}$ forms a partition of $\mathrm{TRIBES}_n^{-1}(1)$ and $\mu(U_1) = 1 - 7\beta^2/16$, $\mu(V_1) = 6\beta^2/16$, $\mu(V_2) = \beta^2/16$ with $\beta = \frac{r+2}{r+1}$.

In [16], the coefficients are $\alpha = 0.99, \lambda = \frac{16}{3(0.99)^2}$ and $\gamma = \frac{16}{(0.99)^2}$ (the authors say these values have not been optimized).

Combining this result with our Theorem 12 (taking $z = 1, i = 1, j = 2$, U_1, V_1, V_2 as in Lemma 23, $u_1 = \gamma, v_1 = \alpha, v_2 = -\lambda$, and $g(n) = 2^{-\delta n/2+1}$), we obtain:

▶ **Corollary 24.** *There exists a Bell inequality satisfying:* $\forall \ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$,

$$B(\mathbf{p}_{\mathrm{TRIBES}_n}) = 2^{\delta n/2-1}\frac{\beta^2}{16}(6\alpha - \lambda),$$

and for any distribution $\mathbf{p}' \in \mathcal{P}$ *such that* $|\mathbf{p}' - \mathbf{p}_{\mathrm{TRIBES}_n}|_1 \leq \varepsilon$,

$$B(\mathbf{p}') \geq 2^{\delta n/2-1}\left[\frac{\beta^2}{16}(6\alpha-\lambda) - \epsilon(\gamma(1-7\beta^2/16) + \lambda\beta^2/16 + \alpha 6\beta^2/16)\right].$$

More precisely, Theorem 12 provides a Bell inequality B yielding this bound, defined as:

$$B_{a,b,x,y} = \begin{cases} -\gamma 2^{\delta n/2-1}\mu(x,y) & \text{if } (x,y) \in U_1 \text{ and } a \oplus b = 1 \\ \alpha 2^{\delta n/2-1}\mu(x,y) & \text{if } (x,y) \in V_1 \text{ and } a \oplus b = 1 \\ -\lambda 2^{\delta n/2-1}\mu(x,y) & \text{if } (x,y) \in V_2 \text{ and } a \oplus b = 1 \\ 0 & \text{otherwise.} \end{cases}$$

To obtain Proposition 15, we use Corollary 9 together with the fact that $Q_{\varepsilon'}(\mathrm{TRIBES}_n) = O(\sqrt{n}(\log n)^2)$.

C.3 Gap Orthogonality

Let f_n be the partial functions over $\{-1,+1\}^n \times \{-1,+1\}^n$ by $f_n(x,y) = \mathrm{ORT}_{64n}(x^{64}, y^{64})$, that is:

$$f_n(x,y) = \begin{cases} -1 & \text{if } |\langle x,y\rangle| \leq \sqrt{n}/8 \\ +1 & \text{if } |\langle x,y\rangle| \geq \sqrt{n}/4. \end{cases}$$

In [41], Sherstov proves the following result.

▶ **Lemma 25** ([41])**.** *Let* $\delta > 0$ *be a sufficiently small constant and* μ *the uniform measure over* $\{0,1\}^n \times \{0,1\}^n$. *Then,* $\mu(f_n^{-1}(+1)) = \Theta(1)$ *and for all rectangle* R *in* $\{0,1\}^n \times \{0,1\}^n$ *such that* $\mu(R) > 2^{-\delta n}$,

$$\mu(R \cap f_n^{-1}(+1)) \geq \delta\mu(R \cap f_n^{-1}(-1)).$$

This implies that if we put uniform weight on inputs of ORT_{64n} of the form (x^{64}, y^{64}) and put 0 weight on the others, we get a distribution μ' satisfying the constraints of Corollary 20 for ORT_{64n} together with $\gamma = \delta$ from Lemma 4 and $g(64n) = 2^{\delta n}$.

To get a distribution satisfying the constraints of Corollary 20 on inputs of ORT_{64n+l} for all $0 \leq l \leq 63$ we extend μ' as follows:

$$\tilde{\mu}(xu, yv) = \begin{cases} \mu'(x,y) & \text{if } u = +1^l,\ v = -1^l \text{ and } \left(\langle x,y\rangle < -\sqrt{64n} \text{ or } 0 \leq \langle x,y\rangle \leq \sqrt{64n}\right) \\ \mu'(x,y) & \text{if } u = +1^l,\ v = +1^l \text{ and } \left(-\sqrt{64n} \leq \langle x,y\rangle < 0 \text{ or } \langle x,y\rangle > \sqrt{64n}\right) \\ 0 & \text{otherwise} \end{cases}$$

Using this distribution $\tilde{\mu}$ together with $\gamma = \delta$ from Lemma 25 and with $g(n) = 2^{-\delta n}$ we obtain, from Corollary 20, a Bell inequality violation for ORT_{64n+l} for all $0 \leq l \leq 63$:

▶ **Corollary 26.** *There exists a Bell inequality* B *satisfying:* $\forall \ell \in \mathcal{L}_{det}^{\perp}$, $B(\ell) \leq 1$,

$$B(\mathbf{p}_{\mathrm{ORT}_{64n+l}}) = 2^{\delta n}\delta\tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(-1)),$$

and for any distribution $\mathbf{p}' \in \mathcal{P}$ *such that* $|\mathbf{p}' - \mathbf{p}_{\mathrm{ORT}_{64n+l}}|_1 \leq \varepsilon$,

$$B(\mathbf{p}') \geq 2^{\delta n}\left(\delta\tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(-1)) - \epsilon\left[\delta\tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(-1)) + \tilde{\mu}(\mathrm{ORT}_{64n+l}^{-1}(+1))\right]\right).$$

More precisely, Theorem 12 gives an explicit construction of such a Bell inequality: we can define B as:

$$B_{a,b,x,y} = \begin{cases} -2^{\delta n}\tilde{\mu}(x,y) & \text{if } (x,y) \in \mathrm{ORT}^{-1}_{64n+l}(+1) \text{ and } a \oplus b = -1 \\ \delta 2^{\delta n}\tilde{\mu}(x,y) & \text{if } (x,y) \in \mathrm{ORT}^{-1}_{64n+l}(-1) \text{ and } a \oplus b = -1 \\ 0 & \text{otherwise.} \end{cases}$$

To obtain Proposition 16, we use Corollary 9 together with the fact that $Q_{\varepsilon'}(\mathrm{ORT}_n) = O(\sqrt{n}\log n)$.

D Equivalent formulations of the efficiency bounds

In [27], the zero-error efficiency bound was defined in its primal and dual forms as follows

▶ **Definition 27** ([27])**.** The efficiency bound of a distribution $\mathbf{p} \in \mathcal{P}$ is given by

$$\begin{aligned} eff(\mathbf{p}) = \min_{\zeta,\mu_\ell \geq 0} \quad & \frac{1}{\zeta} \\ \text{subject to} \quad & \sum_{\ell \in \mathcal{L}^{\perp}_{det}} \mu_\ell \ell(a,b|x,y) = \zeta p(a,b|x,y) \quad \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\ & \sum_{\ell \in \mathcal{L}^{\perp}_{det}} \mu_\ell = 1 \\ = \max_{B} \quad & B(\mathbf{p}) \\ \text{subject to} \quad & B(\ell) \leq 1 \qquad \forall \ell \in \mathcal{L}^{\perp}_{det} \end{aligned}$$

The ϵ-error efficiency bound was in turn defined as $\min_{\mathbf{p}' \in \mathcal{P} | \mathbf{p}' - \mathbf{p}|_1 \leq \epsilon} eff(\mathbf{p}')$. In this appendix, we show that this is equivalent to the definition used in the present article (Definition 3). In the original definition, the Bell functional could depend on the particular $\mathbf{p}'$. We show that it is always possible to satisfy the constraint with the same Bell functional for all $\mathbf{p}'$ close to $\mathbf{p}$.

In order to prove this, we will need the following notions.

▶ **Definition 28.** A *distribution error* $\mathbf{\Delta}$ is a family of additive error terms $\Delta(a,b|x,y) \in [-1,1]$ for all $(a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y}$ such that

$$\sum_{a,b} \Delta(a,b|x,y) = 0 \qquad \forall (x,y) \in \mathcal{X}\times\mathcal{Y}.$$

For any $0 \leq \epsilon \leq 1$, the set Δ_ϵ is the set of distribution errors $\mathbf{\Delta}$ such that

$$\sum_{a,b} |\Delta(a,b|x,y)| \leq \epsilon \qquad \forall (x,y) \in \mathcal{X}\times\mathcal{Y}.$$

This set is a polytope, so it admits a finite set of extremal points. We denote this set by Δ^{ext}_ϵ.

We will use the following properties of Δ_ε.

▶ Fact 29. For any distribution $\mathbf{p} \in \mathcal{P}$, we have

$$\{\mathbf{p}' \in \mathcal{P} |\ |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon\} \subseteq \{\mathbf{p} + \mathbf{\Delta} |\ \mathbf{\Delta} \in \Delta_\varepsilon\}$$

The reason why the set on the right hand side might be larger is that $\mathbf{p}+\mathbf{\Delta}$ might not be a valid distribution. In order to ensure that this is the case, it is sufficient to impose that all obtained purposed probabilities are nonnegative, leading to the following property.

▶ Fact 30. For any distribution $\mathbf{p} \in \mathcal{P}$, we have

$$\{\mathbf{p}' \in \mathcal{P} |\ |\mathbf{p}' - \mathbf{p}|_1 \leq \epsilon\} = \{\mathbf{p}+\mathbf{\Delta} |\ \mathbf{\Delta} \in \Delta_\varepsilon \ \&\ p(a,b|x,y) + \Delta(a,b|x,y) \geq 0\ \forall a,b,x,y\}$$

We are now ready to prove the following theorem.

▶ **Theorem 31.** *Let $\mathbf{p} \in \mathcal{P}$ be a distribution, $eff_\varepsilon(\mathbf{p})$ be defined as in Definition 3 and $eff(\mathbf{p})$ be defined as in Definition 27. Then, we have*

$$eff_\varepsilon(\mathbf{p}) = \min_{\mathbf{p}' \in \mathcal{P} : |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon} eff(\mathbf{p}').$$

Proof. Let $\overline{eff}_\varepsilon(\mathbf{p}) = \min_{\mathbf{p}' \in \mathcal{P} : |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon} eff(\mathbf{p}')$. We first show that $eff_\varepsilon(\mathbf{p}) \leq \overline{eff}_\varepsilon(\mathbf{p})$. Let (B,β) be an optimal feasible point for $eff_\varepsilon(\mathbf{p})$, so that

$$\begin{aligned} eff_\varepsilon(\mathbf{p}) &= \beta, \\ B(\mathbf{p}') &\geq \beta && \forall \mathbf{p}' \text{ s.t. } |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon, \\ B(\ell) &\leq 1 && \forall \ell \in \mathcal{L}_{det}^{\perp}. \end{aligned}$$

Therefore (B,β) is also a feasible point for $eff(\mathbf{p}')$ for all $\mathbf{p}' \in \mathcal{P}$ such that $|\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon$, so that $eff(\mathbf{p}') \geq \beta$ for all such $\mathbf{p}'$, and $\overline{eff}_\varepsilon(\mathbf{p}) \geq \beta = eff_\varepsilon(\mathbf{p})$.

It remains to show that $eff_\varepsilon(\mathbf{p}) \geq \overline{eff}_\varepsilon(\mathbf{p})$. In order to do so, we first use the primal form of $eff(\mathbf{p}')$ in Definition 27 to express $\overline{eff}_\varepsilon(\mathbf{p})$ as follows

$$\begin{aligned} \overline{eff}_\varepsilon(\mathbf{p}) = \min_{\substack{\mathbf{p}' \in \mathcal{P} \\ \text{s.t } |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon}} & \quad eff(\mathbf{p}') \\ = \min_{\zeta, \mu_\ell \geq 0, \mathbf{p}' \in \mathcal{P}} & \quad \frac{1}{\zeta} \\ \text{subject to} & \quad \sum_{\ell \in \mathcal{L}_{det}^{\perp}} \mu_\ell \ell(a,b|x,y) = \zeta p'(a,b|x,y) \quad \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\ & \quad \sum_{\ell \in \mathcal{L}_{det}^{\perp}} \mu_\ell = 1, \quad |\mathbf{p}'-\mathbf{p}|_1 \leq \epsilon \\ = \min_{\zeta, \mu_\ell \geq 0, \mathbf{\Delta} \in \Delta_\varepsilon} & \quad \frac{1}{\zeta} \\ \text{subject to} & \quad \sum_{\ell \in \mathcal{L}_{det}^{\perp}} \mu_\ell \ell(a,b|x,y) = \\ & \qquad \zeta[p(a,b|x,y) + \Delta(a,b|x,y)] \qquad \forall (a,b,x,y) \in \mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\ & \quad \sum_{\ell \in \mathcal{L}_{det}^{\perp}} \mu_\ell = 1, \end{aligned}$$

where the last equality follows from Fact 30 and the fact that the first condition of the program imposes that $p(a,b|x,y) + \Delta(a,b|x,y)$ is nonnegative (since $\sum_\ell \mu_\ell \ell(a,b|x,y)$ is nonnegative).

Since Δ_ε is a polytope, $\overline{eff}_\varepsilon(\mathbf{p})$ can be expressed as the following linear program

$$\begin{aligned}
\overline{eff}_\varepsilon(\mathbf{p}) = \min_{\zeta,\mu_\ell\geq 0,\nu_\Delta\geq 0} \quad & \frac{1}{\zeta} \\
\text{subject to} \quad & \sum_{\ell\in\mathcal{L}_{det}^{\perp}} \mu_\ell \ell(a,b|x,y) = \zeta[p(a,b|x,y) + \\
& \qquad \sum_{\mathbf{\Delta}\in\Delta_\varepsilon^{ext}} \nu_\Delta \Delta(a,b|x,y)] \qquad\qquad \forall (a,b,x,y)\in\mathcal{A}\times\mathcal{B}\times\mathcal{X}\times\mathcal{Y} \\
& \sum_{\ell\in\mathcal{L}_{det}^{\perp}} \mu_\ell = 1, \quad \sum_{\mathbf{\Delta}\in\Delta_\varepsilon^{ext}} \nu_\Delta = 1.
\end{aligned}$$

Note that this can be written in standard LP form via the change of variables $\mu_\ell = \zeta w_\ell$. By LP duality, we then obtain

$$\begin{aligned}
\overline{eff}_\varepsilon(\mathbf{p}) = \max_{B,\beta} \quad & \beta \\
\text{subject to} \quad & B(\mathbf{p}+\mathbf{\Delta}) \geq \beta \qquad \forall \mathbf{\Delta}\in\Delta_\varepsilon, \\
& B(\ell) \leq 1 \qquad\qquad \forall \ell\in\mathcal{L}_{det}^{\perp}.
\end{aligned}$$

Comparing this to the definition of $eff_\varepsilon(\mathbf{p})$ (Definition 3) and together with Fact 29, we therefore have $\overline{eff}_\varepsilon(\mathbf{p}) \leq eff_\varepsilon(\mathbf{p})$. ◀

How Hard Is Deciding Trivial Versus Nontrivial in the Dihedral Coset Problem?

Nai-Hui Chia[*1] **and Sean Hallgren**[†2]

1 **Department of Computer Science and Engineering, The Pennsylvania State University, University Park, PA 16802, USA**
nxc233@cse.psu.edu

2 **Department of Computer Science and Engineering, The Pennsylvania State University University Park, PA 16802, USA**
hallgren@cse.psu.edu

Abstract

We study the hardness of the dihedral hidden subgroup problem. It is known that lattice problems reduce to it, and that it reduces to random subset sum with density > 1 and also to quantum sampling subset sum solutions. We examine a decision version of the problem where the question asks whether the hidden subgroup is trivial or order two. The decision problem essentially asks if a given vector is in the span of all coset states. We approach this by first computing an explicit basis for the coset space and the perpendicular space. We then look at the consequences of having efficient unitaries that use this basis. We show that if a unitary maps the basis to the standard basis in any way, then that unitary can be used to solve random subset sum with constant density > 1. We also show that if a unitary can exactly decide membership in the coset subspace, then the collision problem for subset sum can be solved for density > 1 but approaching 1 as the problem size increases. This strengthens the previous hardness result that implementing the optimal POVM in a specific way is as hard as quantum sampling subset sum solutions.

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

Keywords and phrases Quantum algorithms, hidden subgroup problem, random subset sum problem

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.6

1 Introduction

The dihedral coset problem is an important open problem in quantum algorithms. It comes from the hidden subgroup problem, which is defined as: given a function on a group G that is constant and distinct on cosets a subgroup H, find H. Here we will focus on the case when G is the dihedral group of order $2N$. It is known that this problem reduces to the case when the subgroup is order two [4]. All known approaches for solving the hidden subgroup problem over the dihedral group start by evaluating the function in superposition and measuring the function value. The result is a random coset state $\frac{1}{\sqrt{2}}(|0, x\rangle + |1, x + d\rangle)$, where $d \in \mathbb{Z}_N$ is a fixed label of the subgroup and x is a coset representative uniformly chosen in $\mathbb{Z}_N$. For

* Partially supported by National Science Foundation award CCF-1218721, and by the National Security Agency (NSA) under Army Research Office (ARO) contract number W911NF-12-1-0522.

† Partially supported by National Science Foundation awards CCF-1218721 and CCF-1618287, and by the National Security Agency (NSA) under Army Research Office (ARO) contract number W911NF-12-1-0522.

our purposes, it is more convenient to have the following quantum problem rather than the hidden subgroup problem.

The *dihedral coset problem* [13] is: given a tensor product of k coset states

$$|c^{(d)}_{x_1,x_2,\cdots,x_k}\rangle = \frac{1}{\sqrt{2}}(|0,x_1\rangle + |1,x_1+d\rangle) \otimes \cdots \otimes \frac{1}{\sqrt{2}}(|0,x_k\rangle + |1,x_k+d\rangle),$$

where $x_1,\ldots,x_k$ are randomly chosen in $\mathbb{Z}_N$, compute d. The first register of each state is mod 2, and the second register is mod N.

This is a natural problem to consider after the successes with abelian groups such as $\mathbb{Z}_N$. The dihedral group with $2N$ elements has $\mathbb{Z}_N$ as a normal subgroup. The representations are mostly two dimensional, so it does not have obvious problems like the symmetric group, where we know large entangled measurements are required to get information from the states [7]. Furthermore, Regev [13] showed that the unique shortest vector problem reduces to the dihedral coset problem, so it could provide a pathway for finding a quantum algorithm for lattice problems.

Much is known about the dihedral coset problem, at least compared to most other nonabelian groups (although there are groups with efficient algorithms, e.g. [6, 9, 3]). Ettinger and Hoyer [4] showed that a polynomial number of measurements in the Fourier basis has enough classical information to determine d, but the best known algorithm takes exponential time to compute it. Kuperberg found subexponential time algorithms [10, 11] for the problem. He also showed that computing one bit of d was sufficient to compute all of d. This algorithm was a big step, although it should be noted that it seems difficult to combine this with Regev's uSVP to dihedral group HSP reduction to get a subexponential time algorithm for the uSVP, partly due to the fact that the coset states created in the reduction have errors with some probability.

The dihedral coset problem also has some connections to the subset sum problem. Bacon, Childs, and van Dam analyzed how well a "pretty good measurement" performs [1]. This type of measurement maximizes the probability of computing d correctly. It is unknown how to compute the measurement they find without quantum sampling subset sum solutions. A unitary implementing this can be used to solve the worst case subset sum, which is NP-complete. Regev showed how to reduce the dihedral coset problem to the random subset sum problem density $\rho > 1$ where ρ also approaches 1 as the problem size increases. Density 1 is the hardest case for the random subset sum problem as shown in Proposition 1.2 in [8]. But is solving the dihedral coset problem as hard as subset sum, and if so, for what parameters? The only connection we are aware of is to compose two known reductions. First, random subset sum with density $\rho = 1/\log k$ reduces to uSVP. Then uSVP reduces to the dihedral coset problem. It is open if an efficient quantum algorithm exists for random subset sum, and density $1/\log k$ may not be as hard to solve as constant density.

1.1 New approach

In this paper we focus on distinguishing trivial from order two subgroups. Instead of trying to compute d, we define a problem which asks if the state is an order two coset state, or is the trivial subgroup case. We define this problem as the *dihedral coset space problem (DCSP)*: either an order two coset state is given, or a random standard basis vector is given, decide which. The random standard basis vector corresponds to the trivial subgroup case in the hidden subgroup problem. This problem is a special case of the decision version of the HSP defined by Fenner and Zhou [5] since we are restricting to order two subgroups. In

their paper, they found a search to decision reduction when N is a power of two. So it turns out that the problem is not computationally easier in that case.

We start by finding a set of vectors that span C and $C^{\perp}$. Let $\vec{l} \in \mathbb{Z}_N^k$, and $p \in \mathbb{Z}_N$. The vectors have the form

$$|S_{\vec{l},p}^{m}\rangle = \frac{1}{\sqrt{|T_{\vec{l},p}|}} \sum_{j=0}^{|T_{\vec{l},p}|-1} \omega_{|T_{\vec{l},p}|}^{mj} |\vec{b}_{\vec{l},p}^{(j)}\rangle |\chi_{\vec{l}}\rangle,$$

where $T_{\vec{l},p}$ contains the subset sum solutions for $(\vec{l}, p)$, and the vectors $\vec{b}$ are an ordered set of the subset sum solutions. We call this set of orthonormal vectors the *subset sum* basis. We prove that the $m = 0$ subset of vectors span C and the remaining ones, which have $m \geq 1$, span $C^{\perp}$.

Ideally we would like to reduce subset sum to the DCSP. Since this is still out of reach, we prove a weaker relationship. Instead, we assume there is an algorithm that uses the subset sum basis to solve the DCSP and examine the consequences. Such an algorithm needs to decide if $m = 0$ or $m \geq 1$ to distinguish if the vector is in C or $C^{\perp}$. In this paper we consider two main types of unitaries that use this basis. We show that in one case such a unitary can be used to solve random subset sum and in the other case it can be used to solve the random collision problem. This may indicate that the unitaries are difficult to implement.

The first type of unitary we consider maps the subset sum basis to the standard basis. An example would be one that maps each vector $|S_{\vec{l},p}^{m}\rangle$ to the corresponding standard basis vector $|m, p, \vec{l}\rangle$, identifying the vector. This unitary can be used to solve a subset sum instance $(\vec{l}, p)$ by taking $|0, p, \vec{l}\rangle$, applying U^{-1} to get $|S_{\vec{l},p}^{0}\rangle$ and measuring, since $|S_{\vec{l},p}^{0}\rangle$ is a uniform superposition of solutions. The ability to identify the basis vector in this way is very strong because it can solve an NP-complete problem, but we show the connection for a wider range of unitaries. In particular, we show that any unitary that maps the subset sum basis to the standard basis in some way can be used to solve the random subset sum problem in the cryptographic range of constant density $\rho > 1$. This can be view as generalizing the connection to quantum sampling in [1].

The proof for this case works by showing that such a unitary can be used to solve worst case collision for the subset sum function. That is, given a subset sum instance $(\vec{l}, p)$ and a solution vector $\vec{b}$, the goal is to compute a second solution $\vec{b}'$ if one exists. Then we use the fact that random subset sum reduces to random collision for density a constant greater than one [8].

The second type of unitary we consider maps the subset sum basis to vectors where the first bit is zero if the vector is in C, and is one if the vector is from $C^{\perp}$. This type of unitary can be used to solve the DCSP by computing the unitary on the input vector and measuring the first bit. It is a relaxation of the first type of unitary because it could be followed by another unitary mapping to the standard basis. We show that this type of unitary can be used to solve the random collision problem for subset sum with density $\rho = 1 + c \log\log N / \log N$. This collision problem for this density appears to be less well understood than for constant density.

The proof for this case uses the unitary that can solve the DCSP to solve the random collision problem for subset sum. The problem in this case has an arbitrary solution vector $\vec{b}$ fixed, and then a vector $\vec{l}$ is chosen at random. The goal is again to find a second solution $\vec{b}' \neq \vec{b}$ such that $\vec{b}' \cdot \vec{l} = \vec{b} \cdot \vec{l} \bmod N$ on input $\vec{b}$ and $\vec{l}$.

In addition to these two main types of unitaries we show that a small generalization of the form of the subset sum bases has similar results.

The hardness of random subset sum depends on the density and the same is true for the collision problem. But for collision the definition is important also. There are four definitions of finding collisions of hash functions [14]. Our definition of random subset sum collision is based on the universal one-way hash-function family. That is, for any point in the domain, given the hash function uniformly at random from the family, the goal is to find another point in the domain having the same hash value. Impagliazzo and Naor have shown that random subset sum collision is at least as hard as the random subset sum problem when the density is a constant greater than 1 [8]. However, the density of the random subset sum collision problem we consider has density $\rho \leq 1 + c \log N \log N / \log N$. This density is between the density used for subset sum in [13] and the cryptographic one. The hardness of densities for collision in this range is not known, but it can be contrasted with random subset sum, where the problem gets harder as the density approaches one [8].

There are several open questions. Can the second type of unitary above also be used to solve random subset sum? Consider unitaries which decide membership of C with small error, e.g., $1/poly$. Can these unitaries be implemented efficiently or solve some hard problems? Is it possible to implement a unitary efficiently distinguishing C from $C^{\perp}$, with the subset sum basis, or some other basis? If a space has a basis that seems hard to be implemented for some reason, does that mean that no basis for that space is efficient? Is it possible that a larger space C' containing C exists where it is easier to test C' vs. $C'^{\perp}$? Deciding membership in a subspace or its complement is a generalization of classical languages to quantum languages. Are there other examples?

2 Background

In this section, we give the background of the dihedral coset problem and the random subset sum problem.

The dihedral coset problem comes from the dihedral hidden subgroup problem which is:

▶ **Definition 2.1** (Dihedral Hidden Subgroup Problem)**.** Given the dihedral group D_{2N} and a function f that maps D_{2N} to some finite set such that f hides a subgroup H (f takes same value within each coset of H and takes distinct value on different cosets), the problem is to find a set of generators for H.

Ettinger and Hoyer showed that the problem reduces to the case when the subgroup is order two [4]. Hence, we can assume H is an order two subgroup, which can be represented as $\{1, d\}$ for $d \in \mathbb{Z}_N$. All known approaches for solving this problem start by evaluating the function in superposition to get $\sum_{g \in D_{2N}} |g, f(g)\rangle$, and then measuring the function value. This results in an order two coset state $\frac{|0,x\rangle + |1,x+d\rangle}{\sqrt{2}}$, where $x \in \mathbb{Z}_N$ is a random coset representative. Then the problem becomes to find d when given many random order two coset states. This problem is defined as follows:

▶ **Definition 2.2** (Dihedral Coset Problem (DCP))**.** Given a random k-register order two coset state

$$|c^{(d)}_{x_1,x_2,\ldots,x_k}\rangle = \frac{1}{\sqrt{2}}(|0, x_1\rangle + |1, x_1 + d\rangle) \otimes \cdots \otimes \frac{1}{\sqrt{2}}(|0, x_k\rangle + |1, x_k + d\rangle).$$

The problem is to find d.

The hardness of the DCP has been studied by reducing to the random subset sum problem [13] which is defined as follows:

▶ **Definition 2.3** (Random Subset Sum Problem). Given a vector of positive integers $\vec{l} = [l_1, l_2, \ldots, l_k]^T$ uniformly distributed in $\mathbb{Z}_N^k$ and $s = \vec{b} \cdot \vec{l} \pmod N$ where $\vec{b} \in \mathbb{Z}_2^k$ is chosen uniformly at random, find a vector $\vec{b'} \in \mathbb{Z}_2^k$ such that $\vec{l} \cdot \vec{b'} = s \pmod N$. The density is defined as $\rho = \frac{k}{\log(N)}$.

Although the worst-case subset sum problem is NP-hard, the random subset sum problem can be solved in polynomial time when the density is in a certain range. There is no known polynomial-time algorithm for solving the case when ρ is $\Omega(1/k)$ and $O(\frac{k}{\log^2(k)})$. Regev [13] showed that a solution to the random subset sum problem with $\rho > 1$ implies an efficient quantum algorithm for solving the DCP. Moreover, we note that one can reduce the random subset sum problem with $\rho = O(1/\log k)$ to a lattice problem [12, 2], and then to the DCP [13]. Since these two ranges are generally believed not equivalent, it is still not clear if the DCP is equivalent to random subset sum with ρ in a hard range.

In the rest of this section, we define one more problem which will be used in the section 5.

▶ **Definition 2.4** (Random Subset Sum Collision Problem). Let $\vec{b} \in \mathbb{Z}_2^k$ be an arbitrary fixed vector. Given $\vec{b}$, and a vector $\vec{l} \in \mathbb{Z}_N^k$ chosen uniformly at random, the problem is to find a solution $\vec{b'} \in \mathbb{Z}_2^k$ such that $\vec{b} \cdot \vec{l} \equiv \vec{b'} \cdot \vec{l} \pmod N$ and $\vec{b'} \neq \vec{b}$.

The worst-case version of this problem is to find $\vec{b'}$ for arbitrary $\vec{b}$ and $\vec{l}$ which are given. For simplicity, we will call this problem the random collision problem and the worst-case version as the worst-case collision problem in the rest of the paper.

Impagliazzo and Naor showed a relationship between random collision problem and the random subset sum problem. The input in their notation has n numbers modulo $2^{\ell(n)}$ plus the target value.

▶ **Theorem 2.5** (Theorem 3.1 in [8]). *Let $\ell(n) = (1-c)n$ for $c > 0$. If the subset sum function for length $\ell(n)$ is one-way, then it is also a family of universal one-way hash functions.*

The subset sum function for length $\ell(n)$ can be represented by n integers each of which is $\ell(n)$-bits long. The input is an n-bit binary string $\vec{b}$ which indicates a subset of the n integers and the function outputs an integer s which is the sum of the subsets of integers indexed by $\vec{b}$. A family of universal one-way hash functions is the set of functions $\mathcal{F} = \{f\}$ which satisfies the property that if for all x, when f is chosen randomly from $\mathcal{F}$, then finding a collision (i.e., $y \neq x$ and $f(x) = f(y)$) is hard. Note that the random subset sum problem can be viewed as inverting a random subset sum function and the random collision problem is as finding a collision for a random subset sum function.

In the proof of Theorem 2.5 [8], Impagliazzo and Naor showed that finding a collision for a random subset sum function is at least as hard as inverting a random subset sum function. Therefore, we can give the following corollary:

▶ **Corollary 2.6.** *The random subset sum problem with N a power of 2 and ρ a constant > 1 reduces to the random collision problem with the same N and ρ.*

This corollary will be used in the Section 5.

3 The Dihedral Coset Space Problem

In this section we set up our approach. We first define the dihedral coset space problem and show how to use it to solve the dihedral coset problem. Then we define the coset space which we wish to understand.

▶ **Definition 3.1** (Dihedral Coset Space Problem (DCSP))**.** Given a state $|\tau\rangle$ which is promised to be a random order-two coset state $|c^{(d)}_{x_1,x_2,\ldots,x_k}\rangle$ or a random standard basis state $|\vec{b},x\rangle$ where $\vec{b}\in\mathbb{Z}_2^k$ and $x\in\mathbb{Z}_N^k$, the problem is to decide if $|\tau\rangle$ is a k-register order two coset state or not.

A solution to the DCSP implies a polynomial-time algorithm for solving the DCP with N a power of 2 as shown in [5]. We include a proof of our special case here.

▶ **Claim 3.2.** *The dihedral coset problem (DCP) with N a power of 2 reduces to the dihedral coset space problem (DCSP).*

Proof. Suppose we are given the input of the DCP with subgroup d, we first show how to get the least significant bit of d.

Since N is a power of 2, the least significant bit of x and $x+d \pmod N$ are equal for $x\in\mathbb{Z}_N$ if and only if d is even. Therefore by measuring the least significant bit of the state $\frac{|0,x\rangle+|1,x+d\rangle}{\sqrt{2}}$, we get the same state if d is even and get either $|0,x\rangle$ or $|1,x+d\rangle$ (which are standard-basis states) otherwise.

According to the observation above, the least significant bit of d can be computed by the following algorithm. First, we measure the least significant bit of each register. Then all the registers do not change or collapse to a standard-basis state. Finally, apply the algorithm for the DCSP. If the result is an order-two coset state, the least significant bit is 0; otherwise, the least significant bit is 1.

To get bit $(i+1)$, one subtracts d by the least significant i bits computed and measure the $I+1$-th least significant bit of the state. Repeat the process above until all bits of d are known. ◀

It is worth noting that this fact also implies that the lattice problem can be reduced to the DCSP due to the known reduction from the lattice problem to the DCP with N a power of 2 [13].

The main objects we want to understand are the coset space and its complement.

▶ **Definition 3.3.** The coset space $C=\mathrm{span}(\{|c^{(d)}_{x_1,\ldots,x_k}\rangle : d,x_1,\ldots,x_k\in\mathbb{Z}_N\})$ and the orthogonal complement of C is $C^\perp$.

Note that a test for a vector being in C or $C^\perp$ is sufficient to solve the DCSP if k is big enough. This follows from counting the number of k-register order two coset states. There are at most N subgroups, and at most N^k coset representatives, so the number of k-register order two coset states is at most $N(N)^k$. The dimension of the whole space is $(2N)^k$. Hence, the subspace spanned by k-register order-two coset states is at most $1/2$ of the whole space when $k\geq\log 2N$.

▶ **Claim 3.4.** *Let $k=\log 2N+k'$. Let Π_C be a projector onto C and $\Pi_{C^\perp}$ be a projector onto $C^\perp$. If the input is an order two coset state, the measurement $\{\Pi_C,\Pi_{C^\perp}\}$ outputs C always. Otherwise, if the input is a random standard basis state, then this measurement outputs $C^\perp$ with probability at least $1-1/2^{k'+1}$.*

4 The Subset Sum Basis

In this section, we start by finding an orthonormal basis for C and one for $C^\perp$. Note that if we can give a unitary which distinguishes which of the two subspaces we are in (C or $C^\perp$) efficiently, we can solve the DCSP efficiently as in Claim 3.4.

In order to make the basis easier to understand, we permute the subsystems so that the first bit of all registers are on the left, and the integers mod N are on the right. That is, write the original basis state $|b_1, x_1, b_2, x_2, \ldots, b_k, x_k\rangle$ as $|b_1, b_2, \cdots, b_k, x_1, x_2, \cdots, x_k\rangle$. In this notation the coset state is written as

$$|c^{(d)}_{x_1,x_2,\ldots,x_k}\rangle = \frac{1}{2^{k/2}} \sum_{b_1,\ldots,b_k=0}^{1} |b_1, \ldots, b_k, x_1 + b_1 d, \ldots, x_k + b_k d\rangle = \frac{1}{2^{k/2}} \sum_{\vec{b}\in\{0,1\}^k} |\vec{b}, \vec{x} + \vec{b}d\rangle.$$

The subset sum basis is defined as follows:

▶ **Definition 4.1** (The Subset Sum Basis)**.** Let $\vec{l} = (l_1, l_2, \cdots, l_k)^T \in \mathbb{Z}_N^k$, and $p \in \mathbb{Z}_N$. Let $T_{\vec{l},p} = \{\vec{b} : \vec{b}\cdot\vec{l} = p, \vec{b} \in \mathbb{Z}_2^k\}$ contain subset sum solutions for input $\vec{l}$, p. If $|T_{\vec{l},p}| = 0$ then define $|S^m_{\vec{l},p}\rangle = 0$. If $|T_{\vec{l},p}| \geq 1$, then let $m \in \{0, \ldots, |T_{\vec{l},p}| - 1\}$ and pick an ordering $\{\vec{b}^{(j)}_{\vec{l},p}\}$ of the solutions in $T_{\vec{l},p}$. Define the vector

$$|S^m_{\vec{l},p}\rangle = \frac{1}{\sqrt{|T_{\vec{l},p}|}} \sum_{j=0}^{|T_{\vec{l},p}|-1} \omega^{mj}_{|T_{\vec{l},p}|} |\vec{b}^{(j)}_{\vec{l},p}\rangle. \tag{1}$$

For $N, k \in \mathbb{Z}$, define two sets

$$\mathcal{B}^{\perp} = \mathcal{B}^{\perp}_{k,N} = \{|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle : \vec{l} \in \mathbb{Z}_N^k,\ p \in \mathbb{Z}_N,\ m \in \{1, \ldots, |T_{\vec{l},p}| - 1\},\ |T_{\vec{l},p}| \geq 2\} \tag{2}$$

and

$$\mathcal{B}^{0} = \mathcal{B}^{0}_{k,N} = \{|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle : \vec{l} \in \mathbb{Z}_N^k,\ p \in \mathbb{Z}_N,\ m = 0,\ |T_{\vec{l},p}| \geq 1\}. \tag{3}$$

The set $\mathcal{B} = \mathcal{B}^0 \bigcup \mathcal{B}^{\perp}$ is called the *subset sum basis* of $\mathbb{C}^{(2N)^k}$.

In this definition, $|\chi_j\rangle$ is the Fourier basis state $|\chi_j\rangle = \frac{1}{\sqrt{N}} \sum_i \omega^{ij}_N |i\rangle$, and $|\chi_{\vec{l}}\rangle = |\chi_{l_1}\rangle \cdots |\chi_{l_k}\rangle$. Note that $\mathcal{B}^0 \cup \mathcal{B}^{\perp}$ is an orthonormal basis for the whole space and the two sets are disjoint. The vector $|S^m_{\vec{l},p}\rangle$ is a superposition of solution vectors $\vec{b}$ to the equation $\vec{l}\cdot\vec{b} = p$. If no such $\vec{b}$ exists then there is no corresponding $|S^m_{\vec{l},p}\rangle$. If at least one solution $\vec{b}$ exists then $|S^0_{\vec{l},p}\rangle$ is in $\mathcal{B}^0$. If at least two solutions $\vec{b}$ exist then vectors appear in $\mathcal{B}^{\perp}$. Varying m gives orthogonal superpositions of the solutions. Ranging over all $\vec{l} \in \mathbb{Z}_N^k$ and $p \in \mathbb{Z}$ covers all possible bit vectors. Furthermore, these vectors are tensored with every possible Fourier basis state over $\mathbb{Z}_N$.

Next we show that $\mathcal{B}^{\perp}$ forms an orthonormal basis for $C^{\perp}$.

▶ **Claim 4.2.** *The vectors in the set $\mathcal{B}^{\perp}$ form an orthonormal basis of a space that is orthogonal to the k-register order two coset space.*

Proof. As noted, the vectors form an orthonormal basis of the whole space. We will show that an arbitrary state in $\mathcal{B}^{\perp}$ is orthogonal to all k-register order two coset states. Fix $\vec{l}$ and p, and let

$$|\psi\rangle = |S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|}} \sum_{j=0}^{|T|-1} \omega^{mj}_{|T|} |\vec{b}^{(j)}\rangle|\chi_{\vec{l}}\rangle$$

be a state in $\mathcal{B}^\perp$ where $T = T_{\vec{l},p}$ and $\vec{b}^{(j)} = \vec{b}^{(j)}_{\vec{l},p}$ to simplify notation. Then for an arbitrary order-two coset state $|c^d_{x_1,x_2,\cdots,x_k}\rangle$, the inner product $\langle c^{(d)}_{x_1,x_2,\cdots,x_k}|\psi\rangle$ is

$$\begin{aligned}
\frac{1}{\sqrt{2^k|T|}} \sum_{\vec{b}\in\{0,1\}^k} \langle\vec{b}|\langle\vec{x}+\vec{b}d| \sum_{j=0}^{|T|-1} \omega_{|T|}^{mj} |\vec{b}^{(j)}\rangle|\chi_{\vec{l}}\rangle &= \frac{1}{\sqrt{2^k N^k|T|}} \sum_{j=0}^{|T|-1} \omega_{|T|}^{mj} \omega_N^{\vec{l}\cdot(\vec{x}+d\vec{b}^{(j)})} \\
&= \frac{\omega_N^{\vec{l}\cdot\vec{x}}}{\sqrt{2^k N^k|T|}} \sum_{j=0}^{|T|-1} \omega_{|T|}^{mj} \omega_N^{dp} \qquad (4) \\
&= \frac{\omega_N^{\vec{l}\cdot\vec{x}+dp}}{\sqrt{2^k N^k|T|}} \sum_{j=0}^{|T|-1} \omega_{|T|}^{mj} = 0. \qquad (5)
\end{aligned}$$

Eq. 4 is true because $\vec{b}^{(j)} \cdot \vec{l} = p$ iff $\vec{b}^{(j)} \in T$ by the definition of T. Then since $m \geq 1$ and $|T| \geq 2$ by the definition of $\mathcal{B}^\perp_{k,N}$, Eq. 5 is true. ◀

According to Claim 4.2, $\text{span}(\mathcal{B}^\perp) \subseteq C^\perp$. Next we show that $\mathcal{B}^0$ exactly spans the subspace C (and thus $\text{span}(\mathcal{B}^\perp) = C^\perp$).

▶ **Lemma 4.3.** *The set $\mathcal{B}^0$ is an orthonormal basis for the subspace spanned by the order-two coset states.*

Proof. Because C is orthogonal to $\text{span}(\mathcal{B}^\perp)$ by Claim 4.2, $C \subseteq \text{span}(\mathcal{B}^0)$. We want to show equality. Suppose for contradiction that $C \subset \text{span}(\mathcal{B}^0)$. Then there is a vector $|\alpha\rangle \in C^\perp$ that is orthogonal to $\text{span}(\mathcal{B}^\perp)$, so $|\alpha\rangle \in C^\perp \cap \text{span}(\mathcal{B}^0)$. We show that there is no non-zero linear combination of states in $\mathcal{B}^0$ whose inner product with all order-two coset states is zero.

Suppose the state

$$|\alpha\rangle = \sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \alpha_{\vec{l},p} |S^0_{\vec{l},p}\rangle|\chi_{l_1},\ldots,\chi_{l_k}\rangle$$

is orthogonal to all order-two coset states, i.e., $\langle c^{(d)}_{x_1,x_2,\ldots,x_k}|\alpha\rangle = 0$ for $x_1,\ldots,x_k,d \in \mathbb{Z}_N$, for some nonzero vector $|\alpha\rangle \in \text{span}(B^0)$. This inner product is

$$\begin{aligned}
&\frac{1}{\sqrt{2^k}} \sum_{\vec{b}\in\{0,1\}^k} \langle\vec{b},\vec{x}+\vec{b}d| \sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \alpha_{\vec{l},p} |S^0_{\vec{l},p}\rangle|\chi_{l_1},\ldots,\chi_{l_k}\rangle \\
= &\frac{1}{\sqrt{2^k}} \sum_{\vec{b}\in\{0,1\}^k} \langle\vec{b},\vec{x}+\vec{b}d| \sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \alpha_{\vec{l},p} \frac{1}{\sqrt{|T_{\vec{l},p}|}} \sum_{j=0}^{|T_{\vec{l},p}|-1} |\vec{b}^{(j)}_{\vec{l},p}\rangle|\chi_{l_1},\ldots,\chi_{l_k}\rangle \\
= &\frac{1}{\sqrt{2^k N^k}} \sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \alpha_{\vec{l},p} \frac{1}{\sqrt{|T_{\vec{l},p}|}} \sum_{\vec{b}\in\{0,1\}^k} \sum_{j=0}^{|T_{\vec{l},p}|-1} \langle\vec{b}|b^{(j)}_{\vec{l},p}\rangle \omega_N^{\vec{l}\cdot(\vec{x}+\vec{b}d)} \\
= &\frac{1}{\sqrt{2^k N^k}} \sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \alpha_{\vec{l},p} \frac{\omega_N^{\vec{l}\cdot\vec{x}+pd}}{\sqrt{|T_{\vec{l},p}|}} \sum_{\vec{b}:\vec{b}\cdot\vec{l}=p} \sum_{j=0}^{|T_{\vec{l},p}|-1} \langle\vec{b}|b^{(j)}_{\vec{l},p}\rangle \\
= &\frac{1}{\sqrt{2^k N^k}} \sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \alpha_{\vec{l},p} \omega_N^{\vec{l}\cdot\vec{x}+pd} \sqrt{|T_{\vec{l},p}|}.
\end{aligned}$$

Then we have the following equations:

$$\sum_{\vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N} \sqrt{\frac{|T_{\vec{l},p}|}{2^k N^k}}\alpha_{\vec{l},p}\cdot\omega_N^{x_1\cdot l_1+\cdots+x_k\cdot l_k+d\cdot p}=0,\ \forall x_1,\ldots,x_k,d\in\mathbb{Z}_N.$$

Define $\vec{v}$ as an $N^{k+1}\times 1$ vector such that $(\vec{v})_{\vec{l},p}=\sqrt{\frac{|T_{\vec{l},p}|}{2^k N^k}}\alpha_{\vec{l},p}$. The sums above can be represented as follows:

$$A^{\otimes(k+1)}\cdot\vec{v}=\vec{0}, \tag{6}$$

where A is an $N\times N$ Fourier matrices with the (i,j)-th entry as $A_{i,j}=\omega_N^{ij}$ and $\vec{0}$ is an $N^{k+1}\times 1$ vector with all entries as 0. Note that the column of $A^{\otimes(k+1)}$ is indexed by $\vec{l}$ and p and the the row is indexed by $\vec{x}$ and d.

The determinant of $A^{\otimes(k+1)}$ is not zero, so the only vector $\vec{v}$ satisfying Equation 6 is $\vec{v}=\vec{0}$. When $|T_{\vec{l},p}|\geq 1$ this forces $\alpha_{\vec{l},p}=0$ for every coefficient used in $|\alpha\rangle$. When $|T_{\vec{l},p}|=0$, $\alpha_{\vec{l},p}$ is not used in the sum because $|S^m_{\vec{l},p}\rangle=0$. Therefore, these facts contradict the hypothesis that there exists a nonzero vector $|\alpha\rangle\in\text{span}(\mathcal{B}^0)$ which is orthogonal to all order-two coset states. ◀

Now, it is easy to see that a unitary which can efficiently distinguish $\text{span}(\mathcal{B}^0)$ from $\text{span}(\mathcal{B}^\perp)$ also distinguishes C from $C^\perp$ by Claim 3.4 and Lemma 4.3. The next question we address is whether any unitaries that use this basis can be implemented efficiently or not.

5 The hardness results

In general we would like to understand unitaries that can be used to decide if a state is in the coset space C or in $C^\perp$. In this section we look at two types of unitaries using the subset sum basis, plus an extension of each one:

1. A unitary U_S that maps every basis vector $|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle$ to a standard basis state. Note that if these standard basis states specify p and $\vec{l}$, then this can be used to solve the worst case subset sum, but we are allowing a more general type of unitary here.
2. A unitary U_C that maps every basis vector $|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle$ to $|m=0?\rangle|\phi^m_{\vec{l},p}\rangle$, indicating whether or not the state is in the coset space.
3. A unitary $U=\tilde{U}_S$ that satisfies condition (1) or $U=\tilde{U}_C$ that satisfies (2), but U uses a slightly more general basis, where any basis can be chosen for each $(\vec{l},p)$ subspace $\text{span}\{|S^m_{\vec{l},p},\chi_{\vec{l}}\rangle : m\geq 1\}$.

For the last type we use any basis satisfying the following definition.

▶ **Definition 5.1.** Let $\tilde{\mathcal{B}}^0=\mathcal{B}^0=\{|S^0_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle : \vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N, m=0, |T_{\vec{l},p}|\geq 1\}$ be as in Definition 4.1, and let $\tilde{\mathcal{B}}^\perp=\{|\tilde{S}^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle : \vec{l}\in\mathbb{Z}_N^k, p\in\mathbb{Z}_N, m\in\{1,\ldots,|T_{\vec{l},p}|-1\}, |T_{\vec{l},p}|\geq 2\}$ be an orthogonal basis such that $\text{span}(\{|\tilde{S}^m_{\vec{l},p}\rangle : m\in\{1,\ldots,|T_{\vec{l},p}|-1\}\})=\text{span}(\{|S^m_{\vec{l},p}\rangle : m\in\{1,\ldots,|T_{\vec{l},p}|-1\}\})$ for all $\vec{l},p$.

We show that unitaries of type 1 above can be used to solve random subset sum for the cryptographic density ρ a constant greater than 1, indicating that such a unitary may be hard to implement. This strengthens the result in [1] which is a special case where the unitary must perform quantum sampling, i.e., map an input $|\vec{l},p\rangle$ to a superposition of solutions $|\vec{l},S^0_{\vec{l},p}\rangle$. Such a unitary implementing quantum sampling can used to solve worst-case subset

sum by taking an input $|0, p, \vec{l}\rangle$, applying U inverse to get $|S^0_{\vec{l},p}\rangle|\vec{l}\rangle$ and measuring, since this is a uniform superposition of solutions.

An algorithm that uses the subset sum basis to solve the DCSP needs to decide if $m = 0$ (for C), or $m > 0$ (for $C^\perp$). The second type of unitary above allows an arbitrary unitary that writes the answer in the first bit. We show that such a unitary can solve random collision for density $\rho = 1 + c \log\log N / \log N$. This may indicate that no such unitary can be efficiently implemented, although we are less clear on the difficulty of the random collision problem.

The third type of unitary allows an arbitrary basis within each subspace of solutions, but does not mix solutions of different inputs $\vec{l}, p$. Note that $|S^0_{\vec{l},p}\rangle$ cannot change in this case, since it is one dimension in $\mathcal{B}^0$. Let $\tilde{\mathcal{B}} = \tilde{\mathcal{B}}^0 \cup \tilde{\mathcal{B}}^\perp$ be the basis used by the unitary.

The proofs work by using U_S to solve the worst-case collision problem, or U_C, $\tilde{U}_S$, or $\tilde{U}_C$ to solve the random collision problem.

5.1 Unitary mapping to a standard basis

First we give an algorithm that finds a solution to the worst-case collision problem when given a unitary U_S that maps the subset sum basis $\mathcal{B}$ to the standard basis in an arbitrary way. Given $\vec{b}$ and $\vec{l}$ where $\vec{b} \in \mathbb{Z}_2^k$ and $\vec{l} \in \mathbb{Z}_N^k$, the task in the worst-case collision problem is to find $\vec{b}' \neq \vec{b}$ such that $\vec{b}' \cdot \vec{l} = \vec{b} \cdot \vec{l}$.

▶ **Algorithm 1.** *On input $\vec{l} \in \mathbb{Z}_N^k$ and $\vec{b} \in \mathbb{Z}_2^k$:*
1. *Prepare the quantum state $|\vec{b}, \vec{l}\rangle$.*
2. *Apply QFT_N^k on $\vec{l}$, then the state becomes $|\vec{b}\rangle|\chi_{\vec{l}}\rangle$.*
3. *Apply U_S to $|\vec{b}\rangle|\chi_{\vec{l}}\rangle$.*
4. *Measure $U_S(|\vec{b}\rangle|\chi_{\vec{l}}\rangle)$ in the standard basis.*
5. *Apply $U_S^\dagger$.*
6. *Measure value $\vec{b}'$ in the first register.*

Here QFT_N^k is the quantum Fourier transform over $\mathbb{Z}_N^k$.

▶ **Theorem 5.2.** *If there exists an efficient unitary operator U_S, where U_S is a bijection between the subset sum basis and the standard basis, then the worst-case collision problem can be solved efficiently by a quantum algorithm. Therefore random subset sum with density a constant greater than 1 can also be solved.*

Proof. Given $\vec{l}$ and $\vec{b}$ as input, let $p = \vec{l} \cdot \vec{b}$ and $T = T_{\vec{l},p}$. For $\vec{b} = \vec{b}^{(j_0)} \in T$, after computing the Fourier transform of the second register, the resulting state $|\vec{b}^{(j_0)}, \chi_{\vec{l}}\rangle$ can be wrtten in the subset sum basis as

$$|\vec{b}^{(j_0)}, \chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|}} \sum_{m=0}^{|T|-1} \omega^{-j_0 m} |S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle.$$

Applying U_S to this state gives the state

$$\frac{1}{\sqrt{|T|}} \sum_{m=0}^{|T|-1} \omega^{-j_0 m} |D^m_{\vec{l},p}\rangle, \tag{7}$$

where $|D^m_{\vec{l},p}\rangle := U_S(|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle)$ is a standard basis vector by assumption on U_S. Measuring the state in Equation (7) in the standard basis gives $|D^m_{\vec{l},p}\rangle$ for some $m \in [0 : |T| - 1]$. Applying

$U_S^\dagger$ to $|D^m_{\vec{l},p}\rangle$ gives $|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle$, where the first register is $|S^m_{\vec{l},p}\rangle = \frac{1}{\sqrt{|T|}}\sum_{j=0}^{|T|-1}\omega^{jm}|\vec{b}^{(j)}\rangle$ in the standard basis. Measuring this gives a vector $\vec{b}' \neq \vec{b}$ with probability $\frac{|T|-1}{|T|}$.

Theorem 2.5 reduces random subset sum to solving the random collision problem for constant density greater than one, so random subset sum also reduces to the worst case collision problem. ◀

The proof that Algorithm 1 works used a special property of the subset sum basis, which is that every basis vector $|S^m_{\vec{l},p}\rangle$ spreads the solutions with equal magnitude. When this is not the case then the algorithm does not work for the worst case collision problem. However, we will later show that it solves the random collision problem, as long as the number of solutions is not too large.

First we describe an example basis where the algorithm fails. The idea is that the unitary can map a solution vector $|\vec{b},\chi_\ell\rangle$ to a vector that is very close to itself. In that case the algorithm will measure the same value $\vec{b}$ that it started with and not solve the collision problem, which can be seen as follows. Let $\vec{b} = \vec{b}^{(0)}$, let

$$|\hat{S}^1\rangle|\chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|-1}}\sum_{m=1}^{|T|-1}|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle,$$

and pick arbitrary orthonormal vectors $|\hat{S}^2\rangle,\ldots,|\hat{S}^{|T|-1}\rangle$ to form a basis for the subspace $\text{span}(\{|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle : m \in [1:|T|-1]\})$. Note that $|\langle\vec{b},\chi_{\vec{l}}|\hat{S}^1,\chi_{\vec{l}}\rangle|^2 = \frac{|T|-1}{|T|}$, which implies that one gets $U_S(|\hat{S}^1\rangle|\chi_{\vec{l}}\rangle)$ with probability $\frac{|T|-1}{|T|}$ after applying U_S and measuring in the standard basis. In that case, applying $U_S^\dagger$ results in the input vector $\vec{b}$. Therefore, given a unitary mapping this new basis $\{|S^0,\chi_{\vec{l}}\rangle,|\hat{S}^1,\chi_{\vec{l}}\rangle,\ldots,|\hat{S}^{|T|-1},\chi_{\vec{l}}\rangle\}$ to standard basis, the algorithm returns an answer $\vec{b}' \neq \vec{b}$ happens with probability $1/|T|$. The number of solutions $|T|$ can be very large for larger densities.

Next we show that if we limit the size of T, then random collision can be solved.

▶ **Corollary 5.3.** *Suppose Algorithm 1 is run with $\tilde{U}_S$. If $\tilde{U}_S$ is an efficient unitary operator which maps every state in $\tilde{\mathcal{B}}$ to an arbitrary state in the standard basis, then on input $\vec{l},\vec{b}$, the algorithm solves the collision problem with probability at least $\frac{1}{|T_{\vec{l},p}|}(1-\frac{1}{|T_{\vec{l},p}|})$, where $p = \vec{l}\cdot\vec{b}$. In particular, when $k \leq \log N + c\log\log N$, the random collision problem can be solved in quantum polynomial time.*

Proof. Similar to the proof for Theorem 5.2, first represent $|\vec{b},\chi_{\vec{l}}\rangle$ as a linear combination of states in $\tilde{\mathcal{B}}$ as follows:

$$|\vec{b},\chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|}}|S^0\rangle|\chi_{\vec{l}}\rangle + \sqrt{\frac{|T|-1}{|T|}}\Big(\sum_{m=1}^{|T|-1}c_m|\tilde{S}^m\rangle\Big)|\chi_{\vec{l}}\rangle,$$

where $T = T_{\vec{l},p}$ and $\tilde{S}^m = \tilde{S}^m_{\vec{l},p}$.

After applying the unitary $\tilde{U}_S$, the state is

$$\tilde{U}_S|\vec{b},\chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|}}|D^0\rangle + \sqrt{\frac{|T|-1}{|T|}}\Big(\sum_{m=1}^{|T|-1}c_m|D^m\rangle\Big), \tag{8}$$

where D^m for $m \in [0:|T|-1]$ are arbitrary distinct states in the standard basis. By measuring the state in the Equation (8) in the standard basis, $|D^0\rangle$ is measured with probability $1/|T|$. Then applying $\tilde{U}_S^\dagger$ and measuring the output state in the standard basis gives $\vec{b}' \neq \vec{b}$ with

probability $\frac{|T|-1}{|T|}$. Based on the Claim 1.1, $|T| = poly(k)$ with high probability. Thus, given a random input $\vec{b}$ and $\vec{l}$, the probability to get $\vec{b}' \neq \vec{b}$ using $\tilde{U}_S$ in Algorithm 1 is at least $\frac{|T|-1}{|T|^2} = 1/poly(k)$. ◀

5.2 Deciding membership in C

The unitary U_S illustrated how our algorithm works and used the subset sum basis, but U_S may not be useful for distinguishing C from $C^\perp$ in general. Next we consider a unitary U_C that can distinguish C from $C^\perp$. Suppose U_C works on a larger Hilbert space to have work space and exactly distinguishes $\mathcal{B}^0$ from $\mathcal{B}^\perp$ in the first qubit as follows:

▶ **Definition 5.4.** Let U_C be a unitary operator such that

$$U_C(|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle|0\rangle) = \begin{cases} |0\rangle|\psi_{\vec{l},p,0}\rangle & \text{if } m = 0 \\ |1\rangle|\psi_{\vec{l},p,m}\rangle & \text{otherwise} \end{cases}$$

where $\{|\psi_{\vec{l},p,m}\rangle : \vec{l} \in \mathbb{Z}_N^k,\ p \in \mathbb{Z}_N,\ m \in [0 : |T_{\vec{l},p}| - 1]\}$ are states resulting from applying U_C and the third register is a workspace initialized to $|0\rangle$.

We modify Algorithm 1 so that only the first bit is measured in step four, and U_C is used instead of U_S.

▶ **Algorithm 2.** *On input $\vec{l} \in \mathbb{Z}_N^k$ and $\vec{b} \in \mathbb{Z}_2^k$:*
1. *Prepare the quantum state $|\vec{b}, \vec{l}\rangle$.*
2. *Apply QFT_N^k on $\vec{l}$, then the state becomes $|\vec{b}\rangle|\chi_{\vec{l}}\rangle$.*
3. *Apply U_C to $|\vec{b}\rangle|\chi_{\vec{l}}\rangle$.*
4. *Measure the first qubit of $U_C(|\vec{b}\rangle|\chi_{\vec{l}}\rangle)$ in the standard basis.*
5. *Apply $U_C^\dagger$.*
6. *Measure the first register in the standard basis.*

▶ **Theorem 5.5.** *If U_C can be implemented efficiently, then Algorithm 2 solves the collision problem on input $\vec{l}, \vec{b}$ with probability $\frac{2}{|T_{\vec{l},p}|}(1 - \frac{1}{|T_{\vec{l},p}|})$, where $p = \vec{l} \cdot \vec{b}$. In particular, when $k \leq \log N + c \log\log N$ the random collision problem can be solved in quantum polynomial time.*

Proof. Given $\vec{l}$ and $\vec{b}$ as input, let $p = \vec{l} \cdot \vec{b}$ and $T = T_{\vec{l},p}$. For $\vec{b} = \vec{b}^{(j_0)} \in T$, after computing the Fourier transform of the second register, we can write the resulting state $|\vec{b}^{(j_0)}, \chi_{\vec{l}}\rangle$ in the subset sum basis as follows:

$$|\vec{b}^{(j_0)}, \chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|}} \sum_{m=0}^{|T|-1} \omega^{-j_0 m} |S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle.$$

Applying U_C to this state plus a work register results in

$$\frac{1}{\sqrt{|T|}}(|0\rangle|\psi_{\vec{l},p,0}\rangle + \sum_{m=1}^{|T|-1} \omega^{-j_0 m}|1\rangle|\psi_{\vec{l},p,m}\rangle). \tag{9}$$

Measuring the first qubit of the state in Equation (9) gives $|0\rangle|\psi_{\vec{l},p,m}\rangle$ with probability $1/|T|$ and $\frac{1}{\sqrt{|T|-1}} \sum_{m=1}^{|T|-1} \omega^{-j_0 m}|1\rangle|\psi_{\vec{l},p,m}\rangle$ with probability $1 - 1/|T|$.

Applying $U_C^\dagger$ to the result gives $|S^0_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle$ in the first case and

$$\frac{1}{\sqrt{|T|-1}}\sum_{m=1}^{|T|-1}\omega^{-j_0 m}|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle$$

in the second case.

Finally, the state is measured in the standard basis. In the first case, when a zero is measured in the first bit, which happens with probability $1/|T|$, a vector $\vec{b}' \neq \vec{b}$ is measured with probability $1-1/|T|$ in the last step. In the second case when a one is measured the amplitude of $|\vec{b}^{(j_0)}, \chi_{\vec{l}}\rangle$ in $\frac{1}{\sqrt{|T|-1}}\sum_{m=1}^{|T|-1}\omega^{-j_0 m}|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle$ is

$$\frac{1}{\sqrt{|T|-1}}\sum_{m=1}^{|T|-1}\omega^{-j_0 m}\langle\vec{b}^{(j_0)}, \chi_{\vec{l}}|S^m_{\vec{l},p}\rangle|\chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|-1}}\sum_{m=1}^{|T|-1}\omega^{-j_0 m}\langle\vec{b}^{(j_0)}|S^m_{\vec{l},p}\rangle$$

$$= \frac{1}{\sqrt{(|T|-1)|T|}}\sum_{m=1}^{|T|-1}\omega^{-j_0 m}\omega^{j_0 m} = \frac{|T|-1}{\sqrt{(|T|-1)|T|}}.$$

Thus, the probability that the measurement gives $\vec{b}' \neq \vec{b}$ is $1-\frac{(|T|-1)^2}{(|T|-1)|T|} = 1/|T|$. Therefore, the probability the algorithm returns $\vec{b}' \neq \vec{b}$ is $\frac{2}{|T|}(1-\frac{1}{|T|})$.

By Claim 1.1 the probability that a randomly chosen $\vec{l}$ satisfies $|T_{\vec{l},p}| \leq poly(k)$ is at least $1/poly(k)$ when $k = \log N + c\log\log N$. Thus, the random collision problem can be solved by repeating the algorithm $poly(k)$ times. ◀

Now we consider the case where an arbitrary basis can be used within each subspace spanned by solutions of a given subset sum instance $\vec{l}$, p as in Definition 5.1. Let $\tilde{U}_C$ be a unitary that maps every state in $\tilde{\mathcal{B}}$ to quantum state whose first qubit indicates if the state is in $\tilde{\mathcal{B}}^0$ or $\tilde{\mathcal{B}}^\perp$

▶ **Corollary 5.6.** *If Algorithm 2 is run with $\tilde{U}_C$ on input $\vec{l}, \vec{b}$, then it solves the collision problem with probability at least $\frac{1}{|T_{\vec{l},p}|}(1-\frac{1}{|T_{\vec{l},p}|})$, where $p = \vec{l}\cdot\vec{b}$. In particular, if $k \leq \log N + c\log\log N$ then it solves the random collision problem in quantum polynomial time.*

Proof. Suppose $\tilde{U}_C$ maps $|S^0_{\vec{l},p}\rangle$ to a state $|0\rangle|\psi_{\vec{l},p,0}\rangle$ and maps $|\tilde{S}^m_{\vec{l},p}\rangle$ to $|1\rangle|\psi_{\vec{l},p,m}\rangle$ for $m \in [1:|T|-1]$, where the set of vectors $\{|\psi_{\vec{l},p,m}\rangle : m \in [1:|T|-1]\}$ are an arbitrary orthonormal set of quantum states. The analysis is similar to the proof above, but we only consider the case when the state collapses to $m=0$. Specifically, after applying $\tilde{U}_C$ to $|\vec{b}, \chi_{\vec{l}}\rangle$, the state is

$$\tilde{U}_C|\vec{b}, \chi_{\vec{l}}\rangle = \frac{1}{\sqrt{|T|}}|0\rangle|\psi_{\vec{l},p,0}\rangle + \sqrt{\frac{|T|-1}{|T|}}\Big(\sum_{m=1}^{|T|-1} c_m|1\rangle|\psi_{\vec{l},p,m}\rangle\Big). \tag{10}$$

The probability the state collapses to $|0\rangle|\psi_{\vec{l},p,0}\rangle$ after measuring the first qubit is $1/|T|$. After applying $\tilde{U}_C^\dagger$ and measuring the state a vector $\vec{b}' \neq \vec{b}$ is measured with probability $1-1/|T|$. In total the probability of success is at least $\frac{|T|-1}{|T|^2}$. For the choice of k given, this is at least $1/poly(k)$ with probability $1/poly(k)$ by Claim 1.1. ◀

References

1 D. Bacon, A. M. Childs, and Wim van Dam. Optimal measurements for the dihedral hidden subgroup problem. *Chicago J. Theor. Comput. Sci.*, 2006.

2 Matthijs J. Coster, Antoine Joux, Brian A. LaMacchia, Andrew M. Odlyzko, Claus-Peter Schnorr, and Jacques Stern. Improved low-density subset sum algorithms. *Computational Complexity*, 2(2):111–128, 1992.

3 Thomas Decker, Gábor Ivanyos, Raghav Kulkarni, Youming Qiao, and Miklos Santha. An efficient quantum algorithm for finding hidden parabolic subgroups in the general linear group. In Erzsébet Csuhaj-Varjú, Martin Dietzfelbinger, and Zoltán Ésik, editors, *Mathematical Foundations of Computer Science 2014: 39th International Symposium, MFCS 2014, Budapest, Hungary, August 25-29, 2014. Proceedings, Part II*, pages 226–238, Berlin, Heidelberg, 2014. Springer Berlin Heidelberg. doi:10.1007/978-3-662-44465-8_20.

4 Mark Ettinger and Peter Høyer. On quantum algorithms for noncommutative hidden subgroups. *Advances in Applied Mathematics*, 25(3):239–251, 2000.

5 Stephen Fenner and Yong Zhang. *Theory and Applications of Models of Computation: 5th International Conference, TAMC 2008, Xi'an, China, April 25-29, 2008. Proceedings*, chapter On the Complexity of the Hidden Subgroup Problem, pages 70–81. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi:10.1007/978-3-540-79228-4_6.

6 Katalin Friedl, Gábor Ivanyos, Frédéric Magniez, Miklos Santha, and Pranab Sen. Hidden translation and translating coset in quantum computing. *SIAM J. Comput.*, 43(1):1–24, 2014. doi:10.1137/130907203.

7 Sean Hallgren, Cristopher Moore, Martin Rötteler, Alexander Russell, and Pranab Sen. Limitations of quantum coset states for graph isomorphism. *J. ACM*, 57:34:1–34:33, November 2010. doi:10.1145/1857914.1857918.

8 Russell Impagliazzo and Moni Naor. Efficient cryptographic schemes provably as secure as subset sum. *Journal of Cryptology*, 9:236–241, 1996.

9 Gábor Ivanyos, Luc Sanselme, and Miklos Santha. An efficient quantum algorithm for the hidden subgroup problem in nil-2 groups. In Eduardo Sany Laber, Claudson Bornstein, Loana Tito Nogueira, and Luerbio Faria, editors, *LATIN 2008: Theoretical Informatics: 8th Latin American Symposium, Búzios, Brazil, April 7-11, 2008. Proceedings*, pages 759–771, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. doi:10.1007/978-3-540-78773-0_65.

10 Greg Kuperberg. A subexponential-time quantum algorithm for the dihedral hidden subgroup problem. *SIAM J. Comput.*, 35(1):170–188, 2005.

11 Greg Kuperberg. Another subexponential-time quantum algorithm for the dihedral hidden subgroup problem. In *8th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2013, May 21-23, 2013, Guelph, Canada*, pages 20–34, 2013.

12 J. C. Lagarias and A. M. Odlyzko. Solving low-density subset sum problems. *Journal of the Association for Computing Machinery*, 32(1):229–246, 1985.

13 Oded Regev. Quantum computation and lattice problems. *SIAM J. Comput.*, 33(3):738–760, 2004.

14 Phillip Rogaway and Thomas Shrimpton. *Cryptographic Hash-Function Basics: Definitions, Implications, and Separations for Preimage Resistance, Second-Preimage Resistance, and Collision Resistance*, pages 371–388. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004. doi:10.1007/978-3-540-25937-4_24.

A Appendix

▶ **Claim 1.1.** *Let $\vec{b} \in \mathbb{Z}_2^k$ be an arbitrary fixed vector with $k = \log N + c \log\log N$ for some constant c. Then over random choices of $\vec{l} \in \mathbb{Z}_N^k$, the probability that $|T_{\vec{l},\vec{b}\cdot\vec{l}}| \leq poly(k)$ is at least $\frac{1}{poly(k)}$.*

Proof. Fix $\vec{b} \in \mathbb{Z}_2^k$ and let $X_{\vec{b}'}$ be a random variable over $\vec{l}$ such that $X_{\vec{b}'} = 1$ if $\vec{b}' \cdot \vec{l} = \vec{b} \cdot \vec{l}$ and $X_{\vec{b}'} = 0$ otherwise. Then $|T_{\vec{l},\vec{l}\cdot\vec{b}}| = \sum_{\vec{b}' \in \mathbb{Z}_2^k} X_{\vec{b}'} = \sum_{\vec{b}' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} X_{\vec{b}'} + 1$.

For $\vec{b}' \neq \vec{b}$, the expected value of $X_{\vec{b}'}$ is $\mathrm{E}[X_{\vec{b}'}] = \mathrm{Prob}_{\vec{l}}(X_{\vec{b}'} = 1) = \mathrm{Prob}_{\vec{l}}(\vec{l}\cdot\vec{b}' = \vec{l}\cdot\vec{b}) = \frac{1}{N}$. The last equality can be seen by choosing i such that $b'_i = 1$ and $b_i = 0$ without loss generality ($\vec{b}$ and $\vec{b}'$ can be swapped if needed). Then by fixing l_j for $j \neq i$, and choosing l_i uniformly, $\vec{b} \cdot \vec{l}$ is fixed while $\vec{b}' \cdot \vec{l}$ is uniformly distributed. The variance of $X_{\vec{b}'}$ is $\mathrm{Var}(X_{\vec{b}'}) = \frac{1}{N} - \frac{1}{N^2}$.

Therefore, the expected value of $|T_{\vec{l},\vec{l}\cdot\vec{b}}| - 1$ is $\mathrm{E}[\sum_{\vec{b}' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} X_{\vec{b}'}] = \sum_{\vec{b}' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} \mathrm{E}[X_{\vec{b}'}] = \frac{2^k - 1}{N}$, and the variance of $|T_{\vec{l},\vec{l}\cdot\vec{b}}| - 1$ is

$$\begin{aligned}
\mathrm{Var}\Big(\sum_{\vec{b}' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} X_{\vec{b}'}\Big) &= \sum_{\vec{b}' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} \mathrm{Var}(X_{\vec{b}'}) + \sum_{\vec{b}' \neq \vec{b}'', \vec{b}', \vec{b}'' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} \mathrm{Cov}(X_{\vec{b}'}, X_{\vec{b}''}) \\
&\leq \frac{2^k - 1}{N} + \sum_{\vec{b}'' \neq \vec{b}', \vec{b}', \vec{b}'' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} \mathrm{Cov}(X_{\vec{b}'}, X_{\vec{b}''}). \qquad (11)
\end{aligned}$$

This results in $\mathrm{Var}(\sum_{\vec{b}' \in \mathbb{Z}_2^k \setminus \{\vec{b}\}} X_{\vec{b}'}) \leq \frac{2^k - 1}{N}$ provided that the covariences are all zero, which we show below. First we finish proving the claim by applying Chebyshev's inequality to get

$$\mathrm{Prob}(|T_{\vec{l},\vec{b}\cdot\vec{l}}| \geq poly(k)) \leq \frac{2^k - 1}{N} \frac{1}{poly(k)} = \frac{1}{poly(k)},$$

when $k \leq \log N + c \log\log N$.

In the following, we show that $X_{\vec{b}'}$ and $X_{\vec{b}''}$ are independent when $\vec{b}$, $\vec{b}'$, and $\vec{b}''$ are all different values, which implies $\mathrm{Cov}(X_{\vec{b}'}, X_{\vec{b}''}) = 0$. To see this let 1 be a coordinate such that $b'_1 = 1$ and $b''_1 = 0$ without loss of generality (b'_j and b''_j can be swapped). If $b_1 = 0$, then

$$\begin{aligned}
& \mathrm{Prob}_{\vec{l}}(X_{\vec{b}'} = 1, X_{\vec{b}''} = 1) \\
= & \sum_{l_2,\ldots,l_k=0:}^{N-1} \mathrm{Prob}_{l_1}(X_{\vec{b}'} = 1, X_{\vec{b}''} = 1 | l_2, \ldots, l_k) \cdot \mathrm{Prob}(l_2, \ldots, l_k) \\
= & \frac{1}{N^{k-1}} \sum_{l_2,\ldots,l_k=0}^{N-1} \mathrm{Prob}_{l_1}(X_{\vec{b}'} = 1 | l_2, \ldots, l_k) \cdot \mathrm{Prob}_{l_1}(X_{\vec{b}''} = 1 | l_2, \ldots, l_k) \qquad (12) \\
= & \frac{1}{N} \cdot \frac{1}{N^{k-1}} \sum_{l_2,\ldots,l_k=0}^{N-1} \mathrm{Prob}_{l_1}(X_{\vec{b}''} = 1 | l_2, \ldots, l_k) \\
= & \frac{1}{N} \cdot \frac{1}{N^{k-1}} \cdot N^{k-2} = \frac{1}{N^2}. \qquad (13)
\end{aligned}$$

Equation 12 is true because $X_{b''}$ is fixed after fixing $l_2, \ldots, l_k$. For Equation 13 note that $\vec{b}$ and $\vec{b}''$ differ in at least one bit besides position $i = 1$. Therefore a $1/N$ fraction of the N^{k-1} choices for $l_2, \ldots, l_k$ satisfy $\vec{l} \cdot \vec{b} = \vec{l} \cdot \vec{b}''$.

In the case where $b_1 = 1$ the properties of $X_{\vec{b}'}$ and $X_{\vec{b}''}$ are reversed:

$$\begin{aligned}
& \mathrm{Prob}_{\vec{l}}(X_{\vec{b}'} = 1, X_{\vec{b}''} = 1) \\
= & \sum_{l_2,\ldots,l_k=0}^{N-1} \mathrm{Prob}_{l_1}(X_{\vec{b}'} = 1, X_{\vec{b}''} = 1 | l_2, \ldots, l_k) \cdot \mathrm{Prob}(l_2, \ldots, l_k) \\
= & \frac{1}{N^{k-1}} \sum_{l_2,\ldots,l_k=0}^{N-1} \mathrm{Prob}_{l_1}(X_{\vec{b}'} = 1 | l_2, \ldots, l_k) \cdot \mathrm{Prob}_{l_1}(X_{\vec{b}''} = 1 | l_2, \ldots, l_k) \qquad (14) \\
= & \frac{1}{N} \frac{1}{N^{k-1}} \sum_{l_2,\ldots,l_k=0}^{N-1} \mathrm{Prob}_{l_1}(X_{\vec{b}'} = 1 | l_2, \ldots, l_k) \\
= & \frac{1}{N} \cdot \frac{1}{N^{k-1}} \cdot N^{k-2} = \frac{1}{N^2}. \qquad (15)
\end{aligned}$$

Equation 14 is true because $X_{\vec{b}'}$ is fixed to 0 or 1 for all l_1. Equation 15 is true because $\vec{b}$ and $\vec{b}'$ differ in at least one bit besides $i = 1$.

Therefore, the covariance of $X_{\vec{b}'}$ and $X_{\vec{b}''}$ is 0. ◀

The Structure of Promises in Quantum Speedups*

Shalev Ben-David

Massachusetts Institute of Technology, Cambridge, MA, USA
shalev@mit.edu

Abstract

In 1998, Beals, Buhrman, Cleve, Mosca, and de Wolf showed that no super-polynomial quantum speedup is possible in the query complexity setting unless there is a promise on the input. We examine several types of "unstructured" promises, and show that they also are not compatible with super-polynomial quantum speedups. We conclude that such speedups are only possible when the input is known to have some structure.

Specifically, we show that there is a polynomial relationship of degree 18 between $\mathrm{D}(f)$ and $\mathrm{Q}(f)$ for any Boolean function f defined on permutations (elements of $[n]^n$ in which each alphabet element occurs exactly once). More generally, this holds for all f defined on orbits of the symmetric group action (which acts on an element of $[M]^n$ by permuting its entries). We also show that any Boolean function f defined on a "symmetric" subset of the Boolean hypercube has a polynomial relationship between $\mathrm{R}(f)$ and $\mathrm{Q}(f)$ – although in that setting, $\mathrm{D}(f)$ may be exponentially larger.

1998 ACM Subject Classification F.1.2 Modes of Computation

Keywords and phrases Quantum computing, quantum query complexity, decision tree complexity, lower bounds, quantum adversary method

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.7

1 Introduction

When can quantum computers provide super-polynomial speedups over classical computers? This has been one of the central questions of quantum computing research since its inception. On one hand, Shor [10] showed that quantum computers can be used to factor an n-bit integer in $O(n^3)$ time – exponentially faster than the best known classical algorithm (which is only conjectured to achieve $e^{O(n^{1/3}\log^{2/3} n)}$ time [6]). On the other hand, quantum algorithms are not believed be able to solve NP-complete problems efficiently, which heavily restricts the set of problems for which they may offer such a speedup. The intuition, then, is that quantum algorithms help only for certain "structured" problems, but not for unstructured ones.

In the query complexity model, we can hope to formalize this intuition. To this end, in 1998, Beals, Buhrman, Cleve, Mosca, and de Wolf [5] showed that the classical and quantum query complexities of any total Boolean function are polynomially related. On the other hand, *partial* functions – functions that assume the input satisfies some promise – can exhibit exponential quantum speedups [11, 8, 2]. However, we still do not have an understanding of *which* partial functions should be expected to provide such speedups.

▶ **Open Problem 1.** Can we characterize the partial functions f for which $\mathrm{Q}(f) = \mathrm{R}(f)^{o(1)}$?

* This work is partially supported by the National Science Foundation and by the Natural Sciences and Engineering Council of Canada. I thank Scott Aaronson for many helpful discussions.

Although we are currently far from such a characterization, a natural first step would be to find *any* type of promise for which we can show a polynomial relationship between $\mathrm{R}(f)$ and $\mathrm{Q}(f)$ (similar to the Beals et al. result for total functions). In this work, we give the first such relationship. We show that when the promise is "the input is a permutation of $\{1, 2, \ldots, n\}$," there is a power 18 relationship between quantum and deterministic query complexities. We also show that when the promise has the form "the input has Hamming weight in the set S" (with $S \subseteq \mathbb{N}$), there is a power 18 relationship between quantum and randomized query complexities (though it's possible for the deterministic query complexity to be exponentially larger). We generalize these results to other classes of promises.

1.1 Previous Work

In 1998, Beals, Buhrman, Cleve, Mosca, and de Wolf [5] proved the following theorem.

▶ **Theorem 1** ([5]). *Let $f : \{0,1\}^n \to \{0,1\}$ be a total function. Then $\mathrm{Q}(f) = \Omega(\mathrm{D}(f)^{1/6})$.*

Their result easily extends to larger alphabets:

▶ **Theorem 2.** *Let $f : [M]^n \to \{0,1\}$ be a total function. Then $\mathrm{Q}(f) = \Omega(\mathrm{D}(f)^{1/6})$.*

This tells us that there is never a super-polynomial quantum speedup for total functions. Note that these results compare quantum query complexity to deterministic query complexity, which is stronger than comparing to randomized query complexity. However, no better relationship is known, even between $\mathrm{Q}(f)$ and $\mathrm{R}(f)$. For more information, see [7].

Another interesting result was proved by Aaronson and Ambainis [1]. They defined a function f to be *permutation-invariant* if

$$f(x_1, x_2, \ldots, x_n) = f(\tau(x_{\sigma(1)}), \tau(x_{\sigma(2)}), \ldots, \tau(x_{\sigma(n)})) \tag{1}$$

for all inputs x and all permutations $\sigma \in S_n$ and $\tau \in S_M$. Here f may be a partial function, but the domain of f must itself be invariant under these permutations. As an example, if $M = 2$, the domain of f might contain all binary strings of Hamming weight in $\{1, 2, n-2, n-1\}$, and $f(x)$ will depend only on the Hamming weight of x (with the value of f being equal on Hamming weights k and $n-k$). Note in particular that the Collision problem – in which we're promised that the input either contains no repeated alphabet symbol, or else repeats each symbol exactly twice, and must discern which is the case – is a permutation-invariant function.

Aaronson and Ambainis [1] proved the following theorem.

▶ **Theorem 3.** *Let f be permutation-invariant. Then $\mathrm{Q}(f) = \tilde{\Omega}(\mathrm{R}(f)^{1/7})$.*

This theorem means that if f is unstructured in a way that looks like the Collision problem, $\mathrm{Q}(f)$ and $\mathrm{R}(f)$ are polynomially related. However, the property of "looking like Collision" places strong constraints on both the function and its promise. In this work, we will show a relationship that holds for *all functions defined defined over a fixed promise* P: we will not assume anything about the structure of f. However, our results will not generalize Theorem 3 (we will provide a generalization of Theorem 1 instead).

Recently, Aaronson and Ben-David [3] characterized the total Boolean functions f that can be "sculpted" to give an exponential quantum speedup; that is, the functions f that can be restricted to a promise P on which quantum algorithms provide an exponential advantage. They showed that the sculptable total functions are those with a large number of large certificates. In particular, this means most total functions are scultable. One interpretation

of this is that *quantum speedups are all about the promise* – if we can carefully chose the promise, we can make almost any function exhibit an enormous quantum speedup over classical algorithms.

The question we study here flips the quantifiers: on which promises does there exist a function that exhibits a large quantum advantage? Plausibly, there are very few such promises, which means that characterizing them is a useful way to approach Open Problem 1. Unfortunately, proving the non-existence of quantum-friendly functions can be difficult.

1.2 Our Results

Our first result is a polynomial relationship between $\mathrm{Q}(f)$ and $\mathrm{D}(f)$ for all functions whose domain is the set of permutations.

▶ **Theorem 4.** *Let $M = n$, and let $P \subseteq [M]^n$ be the set of permutations. Then for all $f : P \to \{0,1\}$, $\mathrm{Q}(f) = \Omega(\mathrm{D}(f)^{1/18})$.*

We prove this result as a special case of a more general theorem. To state the general version, we need a few definitions.

Given $x \in [M]^n$, the *orbit* $\mathrm{orb}(x)$ of x is the set of all strings in $[M]^n$ that can be reached by permuting the characters of x (in other words, the orbit under the symmetric group action acting on the entries of the input). Note that each orbit is uniquely identified by a the multiset $\{x_1, x_2, \ldots, x_n\}$ of characters that appear in the strings of that orbit. We will use $\tau(x)$ to refer to this multiset. If $T \subseteq [M]^n$ is an orbit, we will also use $\tau(T)$ to refer to $\tau(x)$ for any $x \in T$. For example, the orbit of $x = (1,1,2)$ is $\mathrm{orb}(x) = \{(1,1,2),(1,2,1),(2,1,1)\}$, and corresponds to the multiset $\tau(x) = \{1,1,2\}$.

We prove the following generalization of Theorem 4.

▶ **Theorem 5.** *Let $T \subseteq [M]^n$ be an orbit, and let $f : T \to \{0,1\}$. Then $\mathrm{Q}(f) = \Omega(\mathrm{D}(f)^{1/18})$.*

Note that this is a relationship between quantum query complexity and deterministic (not randomized) query complexity. In this sense, the result is similar to Theorem 2, and indeed we use some similar tools in its proof. However, unlike for total functions, a function defined on an orbit might have certificate complexity exponentially smaller than $\mathrm{D}(f)$, which prevents the techniques of [5] from directly applying. We develop some new tools to get around this.

Our second result extends the previous theorem from promises that are orbits to promises that are unions of orbits; that is, the promise may be any "symmetric" set. Here we are only able to prove a polynomial relationship when M is constant.

▶ **Theorem 6.** *Let M be constant. If $f : X \to \{0,1\}$ is a function on any symmetric promise $X \subseteq [M]^n$ (that is, a set X satisfying $x \in X \Rightarrow \tau(x) \subseteq X$), then $\mathrm{Q}(f) = \Omega\left(\mathrm{R}(f)^{1/(18(M-1))}\right)$. In particular, when $M = 2$, we have $\mathrm{Q}(f) = \Omega\left(\mathrm{R}(f)^{1/18}\right)$, so any function defined on a symmetric subset of the Boolean hypercube does not exhibit a super-polynomial quantum speedup.*

Unlike the previous theorem, this one only relates quantum query complexity to randomized (rather than deterministic) query complexity. This is necessary; indeed, if X is the set of binary strings of Hamming weight 0 or $\lfloor n/2 \rfloor$ and f is defined to be 0 on 0^n and 1 elsewhere, then $\mathrm{D}(f) = \lfloor n/2 \rfloor + 1$ but $\mathrm{R}(f)$ is constant.

Notice that this last theorem applies even to the promise $X = [M]^n$ (for constant M), so it can be viewed as a generalization of Theorem 1 (although our polynomial relationship has higher degree, and our generalization replaces $\mathrm{D}(f)$ with $\mathrm{R}(f)$).

As a final note, we remark that our results are incomparable with the Aaronson-Ambainis result (Theorem 3). When M is constant, our Theorem 6 is *much* more general (since it doesn't place restrictions on the function). However, when M is constant, Theorem 3 is not very difficult in the first place; most of the work in [1] went towards dealing with the fact that M may be large (as it is in the Collision problem).

2 Preliminaries

In query complexity, there is a known (possibly partial) function $f : [M]^n \to \{0,1\}$ and an unknown string x in the domain of f. The goal is to determine the value of $f(x)$ using as few queries to the entries of x as possible. Here $[M] := \{0, 1, \ldots, M-1\}$ is the input alphabet; often we set $M = 2$, so the domain is $\{0,1\}^n$.

The query complexity achieved by an algorithm A is defined to be the number of queries used by A over the worst-case choice of x. The query complexity of the function f is then defined to be the minimum query complexity achieved by any algorithm A.

When A is a deterministic algorithm, we denote the query complexity of f by $\mathrm{D}(f)$; when A is a bounded-error randomized algorithm, we denote it by $\mathrm{R}(f)$; and when A is a bounded-error quantum algorithm, we denote it by $\mathrm{Q}(f)$. We also define the zero-error randomized query complexity $\mathrm{R}_0(f)$ to be the expected number of queries used by the best zero-error randomized algorithm (over the worst-case choice of input x). As expected, we have the relationship $\mathrm{D}(f) \geq \mathrm{R}_0(f) \geq \mathrm{R}(f) \geq \mathrm{Q}(f)$ for every function f. We denote the domain of f by $\mathrm{Dom}(f)$. We sometimes refer to the domain as the *promise* of f.

A partial assignment is a string $p \in ([M] \cup \{*\})^n$ that represents partial knowledge of a string in $[M]^n$. An input $x \in [M]^n$ is consistent with a partial assignment p if for all indices i, either $p_i = x_i$ or $p_i = *$. The size $|p|$ of p is the number of non-star entries in p.

A partial assignment is called a 0-certificate for f if the only strings in $\mathrm{Dom}(f)$ it is consistent with are 0-inputs to f. 1-certificates are defined similarly. A partial assignment is a certificate if it is a 0- or 1-certificate. The certificate complexity $\mathrm{C}_x(f)$ of an input x is the minimum size of a certificate for f consistent with x. The maximum of $\mathrm{C}_x(f)$ over all $x \in \mathrm{Dom}(f)$ is the certificate complexity $\mathrm{C}(f)$ of f.

A block is a set of indices in $\{1, 2, \ldots n\}$. We say that a block B is sensitive for a string $x \in \mathrm{Dom}(f)$ if there is a string $y \in \mathrm{Dom}(f)$ that agrees with x outside of B, and satisfies $f(y) \neq f(x)$. The maximum number of disjoint sensitive blocks of x is the block sensitivity of x, denoted by $\mathrm{bs}_x(f)$. The maximum block sensitivity of x over all $x \in \mathrm{Dom}(f)$ is called the block sensitivity of f, denoted by $\mathrm{bs}(f)$.

If $f : [M]^n \to \{0,1\}$ is a total function, we can also define the sensitivity $\mathrm{s}_x(f)$ of a string x as the maximum number of disjoint sensitive blocks of size 1, and the sensitivity $\mathrm{s}(f)$ of f as the maximum value of $\mathrm{s}_x(f)$ over all $x \in [M]^n$. However, since we will be dealing with non-total functions, we will define sensitivity slightly differently in the next section.

It is not hard to see that $\mathrm{s}_x(f) \leq \mathrm{bs}_x(f) \leq \mathrm{C}_x(f)$ for all $x \in \mathrm{Dom}(f)$. Also, since a zero-error algorithm always finds a certificate, we have $\mathrm{s}(f) \leq \mathrm{bs}(f) \leq \mathrm{C}(f) \leq \mathrm{R}_0(f) \leq \mathrm{D}(f)$. The lower bound on Grover search implies $\mathrm{Q}(f) = \Omega(\sqrt{\mathrm{bs}(f)})$ and $\mathrm{R}(f) = \Omega(\mathrm{bs}(f))$. For total functions, we have $\mathrm{D}(f) \leq \mathrm{C}(f)\,\mathrm{bs}(f)$ and $\mathrm{C}(f) \leq \mathrm{s}(f)\,\mathrm{bs}(f)$ [5], so

$$\mathrm{D}(f) \leq \mathrm{C}(f)\,\mathrm{bs}(f) \leq \mathrm{s}(f)\,\mathrm{bs}(f)^2 \leq \mathrm{bs}(f)^3 = O(\mathrm{Q}(f)^6). \tag{2}$$

For a nice survey of query complexity, see [7].

3 Orbit Promises

In this section, we show that the deterministic and quantum query complexity measures are polynomially related when the promise is exactly an orbit, proving Theorem 5.

One particular case which will motivate a lot of our analysis is the case where $M = n$ and T is the orbit corresponding to the multiset $\{0, 1, \ldots, n-1\}$ (i.e. the case where the inputs are all permutations), together with the function f satisfying $f(x) = 0$ if and only if 0 occurs in the first $\lfloor \frac{n}{2} \rfloor$ entries of x. This is sometimes called the permutation inversion problem.

Informally, in this problem we are promised that the input x is a permutation of the elements $0, 1, \ldots, n-1$, and the task is to find the 0 element using as few queries as possible (to turn this into a decision problem, we only ask whether the 0 occurs in the first half of the entries of x). The permutation inversion problem has been shown to require $\Omega(\sqrt{n})$ quantum queries using a variety of methods [9]; our approach uses Ambainis's adversary method [4].

3.1 Sensitivity on Orbit Promises

We start by attempting to mimic the proof that $\mathrm{D}(f) = O(\mathrm{Q}(f)^6)$ for total functions. There are two missing pieces that don't immediately work for partial functions. One is the relationship $\mathrm{C}(f) \leq \mathrm{bs}(f)\,\mathrm{s}(f)$; as defined, $\mathrm{s}(f) = 0$ for permutation inversion, since it's impossible to change only one bit and stay in the promise. The other is the relationship $\mathrm{D}(f) \leq \mathrm{C}(f)\,\mathrm{bs}(f)$; for permutation inversion, we have $\mathrm{D}(f) = \lfloor n/2 \rfloor$ but $\mathrm{C}(f) = \mathrm{bs}(f) = 1$.

We fix the former by changing the definition of $\mathrm{s}(f)$ for orbit promises. The latter problem is harder to handle, and does not have an elementary solution. In the next section, we will attack it by showing that the permutation inversion problem – in which we are looking for a hidden marked item that's promised to be unique – is essentially the only difficult case.

▶ **Definition 7.** Let $T \subseteq [M]^n$ be an orbit, let $f : T \to \{0, 1\}$, and let $x \in T$. We define the sensitivity $\mathrm{s}_{2,x}(f)$ of x is the maximum number of disjoint sensitive blocks of size 2 (instead of size 1). The sensitivity $\mathrm{s}_2(f)$ of f is the maximum value of $\mathrm{s}_{2,x}(f)$ out of all $x \in T$.

Note that letting blocks have size 2 allows two entries to be swapped, maintaining the promise. It is also clear that we still have $\mathrm{s}_{2,x}(f) \leq \mathrm{bs}_x(f)$ for all $x \in T$.

▶ **Theorem 8.** *For all $f : T \to \{0, 1\}$ with $T \subseteq [M]^n$ an orbit, we have $\mathrm{C}(f) \leq 3\,\mathrm{bs}(f)\,\mathrm{s}_2(f)$.*

Proof. Let $x \in T$. Then x has $\mathrm{bs}_x(f)$ disjoint sensitive blocks; let them be $b_1, b_2, \ldots, b_{\mathrm{bs}_x(f)}$, and assume each b_i is minimal (under subsets). Then $\bigcup b_i$ is a certificate consistent with x (for otherwise, x would have more than $\mathrm{bs}_x(f)$ disjoint sensitive blocks). We claim that the size of a sensitive block b_i is at most $3\,\mathrm{s}_2(f)$. This gives us the desired result, because we then have a certificate of size at most $3\,\mathrm{bs}_x(f)\,\mathrm{s}_2(f)$.

Let $y \in T$ agree with x outside b_i with $f(y) \neq f(x)$. Since x and y have the same orbit, the difference between them must be a permutation on the entries of b_i. In other words, there is some permutation σ on b_i such that for $j \in b_i$, we have $y_j = x_{\sigma(j)}$.

Consider the cycle decomposition $c_1 c_2 \ldots c_k$ of σ. Let $c_j = (a_1, a_2, \ldots, a_m)$ be any cycle in it. We claim that switching a_s and a_{s+1} for $s \in \{1, 2, \ldots, m-1\}$ gives a sensitive block for y of size 2. Indeed, if this was not a sensitive block, then block b_i would not be minimal, since $(a_s, a_{s+1})\sigma$ would be a permutation corresponding to a smaller sensitive block (with a_s removed). Note that the number of disjoint sensitive blocks of size 2 we can form this way is at least $\frac{|b_i|}{3}$, since for each cycle c_j we can form $\lfloor \frac{|c_j|}{2} \rfloor \geq \frac{|c_j|}{3}$ of them. Thus $\mathrm{s}_2(f) \geq \frac{1}{3}|b_i|$, as desired. ◀

▶ **Corollary 9.** *Let $f : T \to \{0,1\}$ with $T \subseteq [M]^n$ an orbit. Then $\mathrm{R}(f) = \Omega(\mathrm{C}(f)^{1/2})$ and $\mathrm{Q}(f) = \Omega(\mathrm{C}(f)^{1/4})$.*

Proof. We have $\mathrm{C}(f) \leq 3\,\mathrm{bs}(f)\,\mathrm{s}_2(f) \leq 3\,\mathrm{bs}(f)^2$, so $\mathrm{bs}(f) = \Omega(\sqrt{\mathrm{C}(f)})$. Combined with $\mathrm{Q}(f) = \Omega(\sqrt{\mathrm{bs}(f)})$ and $\mathrm{R}(f) = \Omega(\mathrm{bs}(f))$, this gives the desired result. ◀

3.2 The Structure of Small Certificates

The previous section showed a lower bound on $\mathrm{Q}(f)$ in terms of $\mathrm{C}(f)$ on orbit promises. However, this result by itself cannot be used to relate $\mathrm{Q}(f)$ to $\mathrm{D}(f)$ or $\mathrm{R}(f)$, because the certificate complexity of a function on an orbit promise may be much smaller than the query complexities (an example of this is given by permutation inversion, in which $\mathrm{C}(f) = 1$).

In this section, we prove the following technical lemma, which will be the main tool for handling functions for which the certificate complexity is much smaller than the deterministic query complexity.

▶ **Lemma 10.** *Let $f : T \to \{0,1\}$ with $T \subseteq [M]^n$ an orbit. Fix any $k \leq \frac{1}{2}\sqrt{\mathrm{D}(f)}$. If $k \geq \mathrm{C}(f)$, then there is*

- *a partial assignment p, consistent with some input in T, of size at most $4k^2$, and*
- *a set of alphabet elements $S \subseteq [M]$, of size at most $4k^2$, whose elements each occur less than $2k$ times in $\tau(T) - \tau(p)$*

such that for any $x \in T$ consistent with p and any certificate c consistent with x of size at most k, at least one of the alphabet elements of $c - p$ is in S.

Some clarifications are in order. By $\tau(T) - \tau(p)$, we mean multiset subtraction between the alphabet elements in T and those occurring in p (multiset subtraction is defined analogously to set subtraction; the frequency count of an element in $\tau(T) - \tau(p)$ is the difference of frequency counts in $\tau(T)$ and $\tau(p)$, or 0 if this difference is negative). By $c - p$, we mean the string d with $d_i = c_i$ when $p_i = *$ and $d_i = *$ otherwise.

Intuitively, this lemma is saying that if we fix a few input coordinates p and restrict to inputs consistent with p, then there is a small set $S \subseteq [M]$ of alphabet elements such that an element of S must exist in any small certificate. For example, for the problem of inverting a permutation, we can choose $p = *^n$, $S = \{0\}$, and $k = \lfloor n/2 \rfloor - 1$; then any certificate of size less than k must include the alphabet element 0. The intuition, then, is that solving the function quickly requires searching for an alphabet symbol in S, which will be a difficult task since there are few of them and each occurs a small number of times.

The proof of this lemma is motivated by the proof that $\mathrm{D}(f) \leq \mathrm{C}(f)^2$ for total boolean functions. That proof describes a deterministic algorithm for computing $f(x)$: repeatedly pick 0-certificates consistent with the entries of x seen so far, and query the entries of x corresponding to the non-$*$ entries of that certificate. Since each 0-certificate conflicts with all 1-certificates, each time we do this we reveal a new entry of every 1-certificate. Therefore, after $\mathrm{C}(f)$ iterations, a certificate has been revealed and the value of $f(x)$ has been determined.

Our proof works similarly, except that it is no longer true that each 0-certificate must contradict every 1-certificate on some entry. Instead, it might be possible that a 0-certificate and a 1-certificate disagree on the location of an alphabet element. However, in that case we can conclude that there are a few alphabet elements that are included in all small certificates.

Proof. Fix such T, f, and k. The proof is based on the following algorithm, which either generates the desired p and S or else computes $f(x)$ for a given input x. We will proceed by arguing that the algorithm always generates p and S after at most $4k^2$ queries, which

must happen before it computes $f(x)$ when x is the worst-case input (as guaranteed by the requirement that $k \leq \frac{1}{2}\sqrt{\mathrm{D}(f)}$). The algorithm is as follows.

Get input x
Set $p = *^n$, $S = \emptyset$, $R = \emptyset$
loop
    Find any certificate c (consistent with a legal input) that
- has size at most k
- is consistent with p
- has the property that $c - p$ has no alphabet elements in S.

    If there are no such certificates, output p and S and halt.
    Add all the alphabet elements of c to R.
    Set S to be the set of elements i of R that occur less than $2k$ times in $\tau(T) - \tau(p)$.
    Query x on all domain elements of c and add the results to p.
    If p is a 0-certificate, output "$f(x) = 0$" and halt; if it's a 1-certificate, output "$f(x) = 1$" and halt.

We claim that this algorithm will go through the loop at most $4k$ times. Indeed, each iteration through the loop selects a certificate. A 0-certificate must conflict with all 1-certificates, and vice versa. There are two ways for certificates to conflict: either they disagree on the value of an entry, or else there is some alphabet element i that they claim to find in different places (and in addition, there must be few unrevealed instances of i in x).

This motivates the following definition: for a certificate c, let $h_{p,S}(c)$ be $|c - p| + |\,\mathrm{alphabet}(c) - S|$ if c is consistent with p, and zero otherwise (here $\mathrm{alphabet}(c)$ denotes the set of alphabet elements occurring in c). Note that at the beginning of the algorithm, $h_{p,S}(c) \leq 2|c| \leq 2k$ for all certificates c of size at most k. Now, whenever the algorithm considers a 0-certificate c_0, the value of $h_{p,S}(c_1)$ decreases for all 1-certificates c_1 of size at most k (unless it is already 0). This is because either c_0 and c_1 conflict on an input, in which case an entry of c_1 is revealed and included in p, decreasing $|c_1 - p|$ (or contradicting c_1), or else c_0 and c_1 both include an alphabet element i which has less than $2k$ occurrences left to be revealed (if it had at least $2k$ unrevealed occurrences, it wouldn't be the source of a conflict between c_0 and c_1, since they each have size at most k). In the latter case, i is added to S, which decreases $|\,\mathrm{alphabet}(c_1) - S|$.

We have shown that each iteration of the algorithm decreases $h_{p,S}(c)$ either for all 0-certificates or for all 1-certificates (of size at most k). This means that unless the loop is terminated, one of the two values will reach 0 in less than $4k$ iterations. We claim this cannot happen, implying the loop terminates in less than $4k$ iterations.

Suppose by contradiction that $h_{p,S}(c)$ reaches 0 for all 0-certificates. This means p is either a certificate – in which case the value of $f(x)$ was determined, which is a contradiction – or else p is not a certificate, and conflicts with all 0-certificates of size at most k. In the latter case, there is some input y consistent with p such that $f(y) = 0$, and there are no certificates consistent with y of size at most k. Thus $C(f) > k$, contradicting the assumption in the lemma.

This shows the loop always terminates in less than $4k$ iterations, which means it cannot calculate $f(x)$, and must instead output p and S. This gives the desired result, since all certificates of size at most k that are consistent with p have the property that $c - p$ has an alphabet element in S. ◀

Note that if we restrict to inputs consistent with p, then the lemma asserts that finding a small certificate requires finding an element of S. This gives us the following corollary.

▶ **Corollary 11.** *If $f : T \to \{0,1\}$ with $T \subseteq [M]^n$ an orbit, then we have*

$$\mathrm{R}_0(f) = \Omega(\min(\mathrm{D}(f)^{1/2}, n^{1/4})) = \Omega(\mathrm{D}(f)^{1/4}). \tag{3}$$

When $M = n$ and T is the set of permutations, $\mathrm{R}_0(f) = \Omega(\min(\mathrm{D}(f)^{1/2}, n^{1/3})) = \Omega(\mathrm{D}(f)^{1/3})$.

Proof. Fix T and f, and let $k = \lfloor \min(\frac{1}{2}\sqrt{\mathrm{D}(f)}, \frac{1}{4}n^{1/4}) \rfloor - 1$ (in the case of permutations, let $k = \lfloor \min(\frac{1}{2}\sqrt{\mathrm{D}(f)}, \frac{1}{4}n^{1/3}) \rfloor - 1$). Since a zero-error randomized algorithm must find a certificate, if $k < \mathrm{C}(f)$, the result follows. It remains to treat the case where $k \geq \mathrm{C}(f)$.

In this case, let p and S be as in the lemma. We restrict to inputs consistent with p. Any zero-error randomized algorithm A must find a certificate on such inputs. If A uses $\mathrm{R}_0(f)$ expected queries, then it finds a certificate with probability at least $1/2$ after $2\,\mathrm{R}_0(f)$ queries.

If $2\,\mathrm{R}_0(f) \leq k$, then A must find a certificate of size at most k with probability $1/2$. But this means that on all inputs x, A finds an element of S in x outside p with probability at least $\frac{1}{2}$. However, there are at most $2k|S| = 8k^3$ such elements in the entries of x outside p (in the case of permutations, at most $|S| = 4k^2$ such elements), and the size of the domain is $n - |p| \geq n - 4k^2 \geq \frac{n}{2}$. If x is generated by fixing p and permuting the remaining entries randomly, the chance of a query finding an element of S is at most $\frac{16k^3}{n}$, so by the union bound, the chance of finding such an element after k queries is at most $\frac{16k^4}{n}$ (in the case of permutations, this becomes $\frac{8k^3}{n}$). Since $k < \frac{1}{2}n^{1/4}$ (or $k < \frac{1}{2}n^{1/3}$ in the case of permutations) gives the desired contradiction. ◀

3.3 Lower bounds on R(f) and Q(f)

We now put everything together to prove lower bounds on $\mathrm{R}(f)$ and $\mathrm{Q}(f)$ in terms of $\mathrm{D}(f)$, proving Theorem 5.

▶ **Theorem 12.** *Let $f : T \to \{0,1\}$ with $T \subseteq [M]^n$ an orbit. Then:*

$$\mathrm{R}(f) = \Omega(\min\{\mathrm{D}(f)^{1/6}, n^{1/9}\}) = \Omega(\mathrm{D}(f)^{1/9}),$$
$$\mathrm{Q}(f) = \Omega(\min\{\mathrm{D}(f)^{1/12}, n^{1/18}\}) = \Omega(\mathrm{D}(f)^{1/18}).$$

Throughout this proof, we will identify a partial assignment $q \in ([M] \cup \{*\})^n$ with the set $\{(i, q_i) : i \in \mathbb{N}, q_i \neq *\}$. This will allow us to use set theoretic notation to manipulate partial assignments; for example, if q and q' are consistent partial assignments, then $q \cup q'$, $q \cap q'$, and $q - q'$ are all also a partial assignments.

Proof of Theorem 12. Apply Lemma 10 with $k = \min\{\frac{1}{4}\sqrt{\mathrm{D}(f)}, \frac{1}{3}n^{1/3}\}$. If $\mathrm{C}(f) > k$, then we're done by Corollary 9. Otherwise, we get p and S from the lemma, with all certificates of size at most k that are consistent with p having alphabet elements in S.

We use Ambainis's adversary method (see Appendix A) to get a lower bound for $\mathrm{Q}(f)$, which will look very similar to the lower bound for permutation inversion found in [4]. In order to apply the adversary method, we use p and S to find some specific inputs to our function f.

First, let Y be the multiset of alphabet symbols in $\tau(T) - \tau(p)$, except for the alphabet symbols found in S. Note that there are at most $4k^2$ alphabet symbols in S, and each occurs at most $2k$ times outside of p. Since $|\tau(T)| = n$ and $|\tau(p)| \leq 4k^2$, we have $|Y| \geq n - 4k^2 - 8k^3 \geq n - 12k^3 \geq n/2$.

We now run the following procedure.

We now describe the modification in step 5. We make c consistent with p in two sub-steps: in step A, we expand c by setting $c_i = \gamma$ for various choices of $i \in [n]$ and $\gamma \in [M]$; and in

1: Initialize the partial assignment $r = *^n$.
2: Initialize the multiset $Z = Y$.
3: **while** $|r| \leq k$ **do**
4: Pick any 0-certificate c consistent with r of size at most $\mathrm{C}(f)$.
5: Modify c to get c' consistent with p and r, as described below.
6: Replace any alphabet symbols of $c' - p$ that are in S by symbols from Z to get c''.
7: Update Z by removing the used symbols from it.
8: Add the entries of c'' to r.
9: If $|r| > k$, stop. Otherwise, repeat steps 4-8 for a 1-certificate.

step B, we permute the non-$*$ entries of $c - r$. The choices of i in step A will always be entries with $c_i = r_i = *$, and the choices of γ will always be alphabet symbols from either Z or $\tau(p - r)$. When we use symbols from Z, we also update Z to remove those symbols.

Explicitly, we do the following. First, note that c is consistent with r and r is consistent with p. Let $d = c - r$ and let $q = p - r$. We will modify d to make it consistent with q. First, for each symbol γ in the multiset $\tau(d) \cap \tau(q)$, we ensure there is a distinct entry i such that $d_i \neq *$ and $q_i = \gamma$. If there isn't one, we pick i with $q_i = \gamma$ and $d_i = *$ and set c_i to an element of Z (and remove that element from Z). This step ensures that all alphabet elements of d that "must be" part of q can be placed inside q by permuting the non-$*$ entries of d. Next, for each i such that $d_i \neq *$ and $q_i \neq *$, we ensure there is a distinct j such that $d_j = q_i$. If there isn't one, we pick j such that $r_j = p_j = q_j = d_j = c_j = *$ and set $c_j = q_i$.

It is not hard to see that after these additions to c, we can permute the non-$*$ entries of $c - r$ to make it consistent with $p - r$: we can ensure the intersection of non-$*$ entries of $p - r$ and $c - r$ get filled with the correct alphabet symbols, and doing this also ensures that the only alphabet symbols used in the remainder of the partial assignment are not necessary for p. Hence we get c' consistent with p and r. c' was formed by increasing the size of c' by at most $2|c|$, so $|c'| \leq 3|c| \leq 3\,\mathrm{C}(f)$.

We note a few invariants of this algorithm. The first invariant is that any alphabet symbols of $c' - p$ that are in S do not occur in $\tau(p - c')$. This means that after the current iteration, these symbols will not occur in $\tau(p - r)$, so they will be swapped for elements of Z whenever they occur outside of r.

The second invariant is that r is consistent with p and that $r - p$ uses no alphabet elements in S. By Lemma 10, r is not a certificate as long as $|r| \leq k$. Hence step 4 of the algorithm (where a certificate is chosen) never fails. Moreover, each iteration of the algorithm increases $|r|$ by at most $6\,\mathrm{C}(f)$. Thus the loop repeats at least $\frac{k}{6\,\mathrm{C}(f)}$ times.

Consider the entries of r that were added by the selection of 0-certificates. Let the first α of them be the partial assignments $a_1^{(0)}, a_2^{(0)}, \ldots, a_\alpha^{(0)}$, with $\alpha = \lfloor \frac{k}{6\,\mathrm{C}(f)} \rfloor$. Each $a_i^{(0)}$ is equal to $c'' - r$ at the i-th round of the algorithm. Similarly, let the subsets of r that were added by 1-certificates be $a_1^{(1)}, a_2^{(1)}, \ldots, a_\alpha^{(1)}$.

Let W be the multiset $\tau(T) - \tau(p)$ restricted to S; that is, the collection of alphabet symbols in S outside of p. Note that if some of the non-$*$ alphabet symbols in $a_i^{(0)}$ were replaced by some symbols from S and the non-star entries of $a_i^{(0)}$ were permuted, we would get a 0-certificate, and similarly for $a_i^{(1)}$. By the first invariant, it is actually sufficient to replace the alphabet symbols of $a_i^{(0)}$ or $a_i^{(1)}$ by those from the multiset W. We use this fact to construct the sets for the adversary method.

Let A be the sub-multiset of W consisting of the symbols in W that actually need to be swapped in for some $a_i^{(0)}$ or $a_i^{(1)}$. Since the total size of the $a_i^{(0)}$ and $a_i^{(1)}$ sets is $|r| \leq k$, we have $|A| \leq k$.

To each $a_i^{(j)}$, we add an arbitrary block of $|A|$ entries outside r with alphabet symbols from Y. To be able to do this, we require that $2|A|\alpha \leq n - |r| - 2k|A|$ (the $2k|A|$ term appears because each alphabet element in A may occur up to $2k$ times), and also that $|Y| - |r| \geq 2|A|\alpha$. Since $|r| \leq k$, $|A| \leq k$, $\alpha \leq k/6$, $|p| \leq 4k^2$, and $|Y| \geq n/2$, and since $k \leq n^{1/3}/3$, it is easy to check that these conditions hold.

Now, for each $j \in \{0,1\}$ and each $i = 1, 2, \ldots, \alpha$, we can place all the alphabet elements of A inside $a_i^{(j)}$ and permute its non-$*$ entries in a way that restores the j-certificate. We can thus generate 2α inputs, α of which have value 0 and α of which have value 1, such that the only difference between the inputs is which of the 2α disjoint bins has the alphabet elements of A (and has been shuffled). This is essentially a version of permutation inversion.

It's clear that a classical randomized algorithm must make $\Omega(\alpha)$ queries, since it must find the special bin containing the alphabet elements of A. For the quantum lower bound, we use Theorem 17. Let X be the set of inputs in which the elements of A were placed for a 0-certificate bin, and let Z be the set of inputs in which the elements of A were placed for a 1-certificate bin. Our relation R will simply be $X \times Z$. Then each element of X has α neighbors in Z, and vice versa. However, for each entry t and $(x, y) \in R$, we have $l_{x,t} = 1$ or $l_{y,t} = 1$, so $l_{x,t} l_{y,t} \leq \alpha$. Thus we get a quantum lower bound of $\Omega(\sqrt{\alpha})$.

Finally, to complete the proof, we note that $\alpha = \Omega(\min(\frac{n}{k}, \frac{k}{\mathrm{C}(f)})) = \Omega(\frac{k}{\mathrm{C}(f)})$ (since $n \geq k^2$), so that, combining with Corollary 9, $\mathrm{R}(f) = \Omega(\beta)$ and $\mathrm{Q}(f) = \Omega(\sqrt{\beta})$ with $\beta = \max(\sqrt{\mathrm{C}(f)}, \frac{k}{\mathrm{C}(f)})$. Note that this satisfies $\beta = \Omega(k^{1/3})$. This gives:

$$\mathrm{R}(f) = \Omega(\min\{\mathrm{D}(f)^{1/6}, n^{1/9}\}),$$
$$\mathrm{Q}(f) = \Omega(\min\{\mathrm{D}(f)^{1/12}, n^{1/18}\}).$$ ◀

4 Symmetric Promises with Small Alphabets

In this section, we show a polynomial relationship between $\mathrm{Q}(f)$ and $\mathrm{R}(f)$ for any function on a symmetric promise whose alphabet size is constant, proving Theorem 6. We will use the term symmetric to refer to invariance under permutation of the indices of the inputs. That is, a promise X is symmetric if $x_\sigma \in X$ for all $x \in X$ and $\sigma \in S_n$, where $x_\sigma = (x_{\sigma(1)}, x_{\sigma(2)}, \ldots, x_{\sigma(n)})$, and a function $f : X \to \{0,1\}$ is symmetric if X is symmetric and $f(x) = f(x_\sigma)$ for all $x \in X$ and $\sigma \in S_n$.

4.1 The case of Symmetric Functions

We start by dealing with the case where the function f is itself symmetric.

▶ **Theorem 13.** *Let $f : X \to \{0,1\}$ be a symmetric function. Then $\mathrm{Q}(f) = \Omega\left(\frac{\mathrm{R}(f)^{1/8}}{M \log^{1/8} M}\right)$.*

To prove this theorem, we relate $\mathrm{Q}(f)$ and $\mathrm{R}(f)$ to $g(f)$, which we now define.

▶ **Definition 14.** If T_1, T_2 are orbits on alphabet $[M]$, the distance $d(T_1, T_2)$ between T_1 and T_2 is the maximum over all $i \in [M]$ of the difference between the multiplicity of i in T_1 and the multiplicity of i in T_2. If $f : X \to \{0,1\}$ is a symmetric function, let $d(f)$ be the minimum of $d(T_1, T_2)$ for orbits $T_1, T_2 \subseteq X$ with different values under f. Define $g(f) := \frac{n}{d(f)}$.

We proceed to prove lemmas relating $g(f)$ to $\mathrm{R}(f)$ and $\mathrm{Q}(f)$ to $g(f)$.

▶ **Lemma 15.** *For any $x \in [M]^n$, $O(\frac{n^2 \log M}{d^2})$ queries suffice to find an orbit T such that $d(T, \tau(x)) < d$ with probability at least $\frac{2}{3}$ (where $\tau(x)$ denotes the orbit of x). Hence, if $f : X \to \{0,1\}$ is symmetric, $\mathrm{R}(f) = O(g(f)^2 \log M)$.*

Proof. We describe a classical randomized algorithm for estimating the orbit of input x. The algorithm is simply the basic sampling procedure that queries random entries of x and keeps track of the number r_i of times each alphabet element i was observed. The orbit T is then formed by $T(i) = \text{round}\left(\frac{r_i}{\sum_{i\in[M]} r_i} n\right)$, where the rounding operation sometimes rounds up and sometimes down, in order to preserve the sum of $T(i)$ being n. Note that the ratios $T(i)/n$ are within $1/n$ of the observed frequency of i.

Let the orbit of x be $\tau(x) = (t_1, t_2, \ldots, t_M)$, so that the multiplicity of alphabet element i in $\tau(x)$ is t_i. A version of the Chernoff bound states that if we have $k \geq \frac{3}{\epsilon^2} \ln \frac{2}{\delta}$ samples estimating the proportion p of the population with some property, the proportion of the sample with that property is in $(p-\epsilon, p+\epsilon)$ with probability at least $1-\delta$.

Suppose $d \geq 2$. Setting $\epsilon = \frac{d-1}{n}$ and $\delta = \frac{1}{3M}$, we see that $O(\frac{n^2 \log(M)}{d^2})$ samples suffice for $\frac{T(i)}{n}$ to be within $\frac{d}{n}$ of $\frac{t_i}{n}$ with probability at least $1 - \frac{1}{3M}$. In other words, we have $|T(i) - t_i| < d$ with probability $1 - \frac{1}{3M}$ for each i. The union bound then gives us $|T(i) - t_i| < d$ for all i with probability at least $\frac{2}{3}$. This shows that $d(T, \tau(x)) < d$, as desired.

When $d = 1$, we set $\epsilon = \frac{1}{2n}$ and $\delta = \frac{1}{3M}$ to get frequency ratios within $\frac{1}{2n}$ of the true values $\frac{t_i}{n}$ with probability at least $2/3$. The closest integer orbit to the observed frequency ratios will then be exactly correct with probability at least $2/3$.

To compute $f(x)$ for symmetric f, a randomized algorithm can estimate the orbit of x to within $\frac{d(f)}{2}$, and then just output the value of f on any input of the orbit within $\frac{d(f)}{2}$ of the estimated orbit T. Since $g(f) = \frac{n}{d(f)}$, we get $\mathrm{R}(f) = O(g(f)^2 \log M)$. ◀

▶ **Lemma 16.** *If $f : X \to \{0,1\}$ is symmetric, then $\mathrm{Q}(f) = \Omega(g(f)^{1/4}/M^2)$.*

Proof. Let S and T be orbits with distance $d(f)$ such that if x has orbit S and y has orbit T then $f(x) \neq f(y)$. We claim that a quantum algorithm cannot distinguish between these orbits in less than the desired number of queries.

We proceed by a hybrid argument. We form a sequence of orbits $\{S_i\}_{i=0}^k$ with $k \leq M$ such that $S_0 = S$, $S_k = T$, and for all $i = 0, 1, \ldots, k-1$, the orbits S_i and S_{i+1} differ in the multiplicity of at most 2 alphabet elements and have distance at most $d(f)$.

We do this as follows. Set $S_0 = S$. Let A be the set of alphabet elements on whose multiplicities the current S_i agrees with T; at the beginning, A is the set of alphabet elements on which S and T have the same multiplicity, which may be empty. To construct S_{i+1} given S_i, we simply pick an alphabet element r for which S_i has a larger multiplicity than T and an alphabet element r' for which S_i has a smaller multiplicity than T. We then set S_{i+1} to have the same multiplicities as S_i, except that the multiplicity of r is reduced to that in T and the multiplicity of r' is increased to make up the difference. Note that the multiplicity of r is then equal in S_i and T, so r gets added to A. Moreover, note that $d(S_i, S_{i+1}) \leq d(S_i, T)$, and also $d(S_{i+1}, T) \leq d(S_i, T)$. Since this is true for all i, it follows that $d(S_i, S_{i+1}) \leq d(S, T) = d(f)$.

Since an alphabet element gets added to A each time and the elements are never removed, this procedure is terminated with $S_k = T$ after at most M steps. Thus $k \leq M$. Also, consecutive orbits differ in the multiplicities of 2 elements and have distance at most $d(f)$.

We now give a lower bound on the quantum query complexity of distinguishing S_i from S_{i+1}. Without loss of generality, let the alphabet elements for which S_i and S_{i+1} differ be 0 and 1, with 0 having a smaller multiplicity in S_i. Let a be the multiplicity of 0 in S_i, and let b be the multiplicity of 1 in S_i, with $0 < b - a \leq d(f)$. Let c and d be the multiplicities of 0 and 1 in S_{i+1}, respectively. Then $c + d = a + b$. Let $e = a + b = c + d$.

We prove two lower bounds using Ambainis's adversary method, corresponding to e being either large or small. For the small case, consider an input x of orbit S_i split into $2\alpha = \lfloor \frac{n}{e} \rfloor$

blocks $B_1, B_2, \ldots, B_{2\alpha}$ of size e each, such that all the 0 and 1 elements lie in block B_1. To change the input from orbit S_i to S_{i+1}, we must simply change the first block. Also, note that rearranging the blocks does not change the orbit. Let X be the set of inputs given by rearranging the blocks of x so that the block B_1 ends up in the first α blocks, and let Y be the set of inputs given by replacing B_1 to get orbit S_{i+1} and then rearranging the blocks so that B_1 ends up in the last α blocks. We now have a reduction from permutation inversion, so using Ambainis's adversary method, we get a lower bound of $\Omega(\sqrt{\alpha}) = \Omega(\sqrt{\frac{n}{e}})$.

For the case when e is big, we restrict to inputs in which all elements are fixed except for those with value 0 or 1. A lower bound of $\Omega(\sqrt{e/d(f)})$ then follows from the appendix of [1].

If $e \leq \sqrt{nd(f)}$, the former bound gives a lower bound of $\Omega((\frac{n}{d(f)})^{1/4})$ for distinguishing S_i from S_{i+1} by quantum queries. If $e \geq \sqrt{nd(f)}$, the latter bound gives the same. Thus we have a lower bound of $\Omega(g(f)^{1/4})$ in all cases.

Finally, note that if a quantum algorithm could compute $f(x)$ in $\mathrm{Q}(f)$ queries, then for some i it could distinguish S_i from S_{i+1} with probability $\Omega(\frac{1}{M})$. Thus we could use $M^2\,\mathrm{Q}(f)$ queries to distinguish S_i from S_{i+1} with constant probability, so $\mathrm{Q}(f) = \Omega(g(f)^{1/4}/M^2)$. ◀

These two lemmas combine to prove Theorem 13.

4.2 The General Case

We now prove Theorem 6. The proof proceeds by describing a classical algorithm that doesn't use too many more queries than the best quantum algorithm. An interesting observation is that this classical algorithm is mostly deterministic, and uses only $O(\mathrm{Q}(f)^8 M^{16} \log M)$ randomized queries at the beginning (to estimate the orbit of the input).

For this proof, we will often deal with certificates c for f that only work on inputs of some specific orbit S; that is, all inputs $x \in X$ of orbit S that are consistent with c have the same value under f. We will say c is a certificate for the orbit S.

Proof. Let f be a function. We describe a classical algorithm for computing f on an input x, and argue that a quantum algorithm cannot do much better.

As a first step, the algorithm will estimate the orbit of x using $O(\mathrm{Q}(f)^8 M^{16} \log M)$ queries. By Lemma 15, this will provide an orbit T such that $d(T, \tau(x)) < \frac{n}{CM^8\,\mathrm{Q}(f)^4}$ with high probability, where we choose the constant C to be larger than twice the asymptotic constant in Lemma 16. We restrict our attention to orbits that are within $\frac{n}{CM^8\,\mathrm{Q}(f)^4}$ of T.

Now, notice that if we fix an orbit S and assume that x has this orbit, then there is a deterministic algorithm that determines the value of $f(x)$ in at most α steps, where $\alpha = O(\mathrm{Q}(f)^{18})$. Since this is a deterministic algorithm, it must find a certificate of size at most α for the orbit S. The only other possibility is that the deterministic algorithm finds a partial assignment that contradicts the orbit S, in which case it cannot proceed. Running this deterministic algorithm on orbit S will be called *examining* S.

Note further that we can assume we never find a 0-certificate c_0 for some orbit S_0 and a 1-certificate c_1 for some other orbit S_1 without the certificates contradicting either orbit. This is because if we found such certificates, then we can lower bound $\mathrm{Q}(f)$ restricting the function to inputs that agree with c_0 and c_1 and have orbit equal to either S_0 or S_1. The quantum algorithm for f would then have to distinguish between these orbits, which is equivalent to distinguishing between orbits with multisets $S_0 - (\tau(c_0) \cup \tau(c_1))$ and $S_1 - (\tau(c_0) \cup \tau(c_1))$ on inputs of size $n - |c_0 \cup c_1|$. These orbits have distance at most $\frac{n}{CM^8\,\mathrm{Q}(f)^4}$. Since $n > 4\alpha$ (or else $\mathrm{R}(f) = O(n) = O(\mathrm{Q}(f)^{18})$), we have $n - |c_0 \cup c_1| > \frac{n}{2}$, and $(\frac{n}{2})/(\frac{n}{CM^8\,\mathrm{Q}(f)^4}) = \frac{CM^8\,\mathrm{Q}(f)^4}{2}$;

then Lemma 16 together with the choice of c imply that a quantum algorithm takes more than $Q(f)$ queries to distinguish these orbits, a contradiction.

For an orbit S, we now define $v(S) \in [2\alpha+1]^M$ to be the vector with $v(S)_i = \min(S(i), 2\alpha)$ for all i, where $S(i)$ is the multiplicity of i in the orbit S. If an input has orbit S, we call $v(S)$ the simplified orbit of the input. We consider the partial order on simplified orbits given by $v(S) \geq v(R)$ if and only if $v(S)_i \geq v(R)_i$ for all $i = 1, 2, \ldots M$. We say a simplified orbit $v(S)$ is maximal if it is maximal in this partial order.

The algorithm proceeds by finding the set of maximal simplified orbits, and selecting a representative orbit S for each maximal simplified orbit v so that $v(S) = v$. Let the orbits selected this way be $S_1, S_2, \ldots, S_\beta$. For each S_i, we then run the deterministic algorithm that uses α queries assuming orbit S_i. Let c_i be the set of queries made by this algorithm for orbit S_i. Note that the total number of queries made this way is at most $\alpha\beta$.

For each S_i, the partial assignment c_i is either a certificate for S_i or a disproof of the orbit S_i. Consider the pairwise unions $c_i \cup c_j$. We restrict our attention to the orbits S_i that are consistent with $c_i \cup c_j$ for all j. We claim that there is at least one such orbit. Indeed, if T is the true orbit of the input, then $v \geq v(T)$ for some maximal simplified orbit v, and $v(S_k) = v$ for some k. Then S_k cannot be disproven in 2α queries, as that would disprove v and therefore $v(T)$ as well.

Now, let S_i and S_j be any two orbits remaining. Then they are both consistent with $c_i \cup c_j$. As we saw earlier, we cannot have c_i be a 0-certificate for S_i and c_j be a 1-certificate for S_j (or vice versa); the certificates c_i and c_j must agree. We conclude that the certificates c_i for the remaining orbits are either all 0-certificates (for their respective orbits) or all 1-certificates. Our algorithm will then output 0 in the former case and 1 in the latter.

To see that the algorithm is correct, recall that S_k is one of the remaining orbits, with $v(S_k) = v \geq v(T)$. Without loss of generality, suppose the algorithm output 0, so that c_k is a 0-certificate. Suppose by contradiction that $f(x) = 1$ for the our input. Let c be a 1-certificate consistent with x of size at most α. Then c is a 1-certificate for the orbit T. Now, $c \cup c_k$ cannot disprove $v(T)$ (since it has size at most 2α), so $c \cup c_k$ cannot disprove T. Since $c \cup c_k$ cannot disprove $v(T)$, it also cannot disprove v, so it cannot disprove S_k. This means T and S_k are not disproven by their 0- and 1-certificates, which we've shown is a contradiction. Thus if the algorithm outputs 0, we must have $f(x) = 0$ as well.

The total number of queries required is $O(\mathrm{Q}(f)^8 M^8 \log M) + \alpha\beta$, where $\alpha = O(\mathrm{Q}(f)^{18})$. We must estimate β, the number of maximal simplified orbits. This is at most the number of maximal elements in $[2\alpha + 1]^M$ in our partial order. We can show by induction that this is at most $(2\alpha + 1)^{M-1}$: in the base case of $M = 1$, the value is 1, and when M increases by 1 the number of maximal elements can increase by at most a factor of $(2\alpha + 1)$. This gives a final bound of $O(\mathrm{Q}(f)^{18M})$ on the number of queries when M is constant.

To reduce this to $O(\mathrm{Q}(f)^{18(M-1)})$, we note that some alphabet element a must occur at least n/M times in T, by the pigeonhole principle. We could then use $O(M\alpha)$ queries to find 2α instances of a with high probability. Then each simplified orbit v will have $v_a = 2\alpha$, so the simplified orbits are effectively elements of $[2\alpha + 1]^{M-1}$ instead of $[2\alpha + 1]^M$. This decreases β to $(2\alpha+1)^{M-2}$, so the total number of queries decreases to $O(\mathrm{Q}(f)^{18(M-1)})$. ◀

References

1 Scott Aaronson and Andris Ambainis. The need for structure in quantum speedups. *arXiv preprint* `arXiv:0911.0996`, 2009.

2 Scott Aaronson and Andris Ambainis. Forrelation: A problem that optimally separates quantum from classical computing. In *Proceedings of the 47th ACM Symposium on Theory of Computing (STOC 2015)*, pages 307–316, 2015. doi:10.1145/2746539.2746547.

3 Scott Aaronson and Shalev Ben-David. Sculpting quantum speedups. *arXiv preprint* arXiv:1512.04016, 2015.

4 Andris Ambainis. Quantum lower bounds by quantum arguments. *Journal of Computer and System Sciences*, 64(4):750–767, 2002. doi:10.1006/jcss.2002.1826.

5 Robert Beals, Harry Buhrman, Richard Cleve, Michele Mosca, and Ronald de Wolf. Quantum lower bounds by polynomials. *Journal of the ACM*, 48(4):778–797, 2001. arXiv:arXiv:quant-ph/9802049, doi:10.1145/502090.502097.

6 Joe P Buhler, Hendrik W Lenstra Jr, and Carl Pomerance. Factoring integers with the number field sieve. In *The development of the number field sieve*, pages 50–94. Springer, 1993.

7 Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002. doi:10.1016/S0304-3975(01)00144-X.

8 Richard Cleve. The query complexity of order-finding. *Information and Computation*, 192(2):162–171, 2004.

9 Ashwin Nayak. Inverting a permutation is as hard as unordered search. *arXiv preprint* arXiv:1007.2899, 2010.

10 Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. arXiv:quant-ph/9508027.

11 Daniel R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997. doi:10.1137/S0097539796298637.

A The Quantum Adversary Method

▶ **Theorem 17.** *Let $f : X \to \{0,1\}$ with $X \subseteq [M]^n$. Let $A, B \subseteq X$ be such that $f(a) = 0$ for all $a \in A$ and $f(b) = 1$ for all $b \in B$. Let $R \subseteq A \times B$ be such that*

1. *For each $a \in A$, there exist at least m different $b \in B$ such that $(a,b) \in R$.*

2. *For each $b \in B$, there exist at least m' different $a \in A$ such that $(a,b) \in R$.*

Let $l_{a,i}$ be the number of $b \in B$ such that $(a,b) \in R$ and $a_i \neq b_i$. Let $l_{b,i}$ be the number of $a \in A$ such that $(a,b) \in R$ and $a_i \neq b_i$. Let l_{max} be the maximum of $l_{a,i} l_{b,i}$ over all $(a,b) \in R$ and $i \in \{1,2,\ldots,n\}$ such that $a_i \neq b_i$. Then $Q(f) = \Omega\left(\sqrt{\frac{mm'}{l_{max}}}\right)$.

Quantum Algorithms for Abelian Difference Sets and Applications to Dihedral Hidden Subgroups

Martin Roetteler

Microsoft Research, Quantum Architectures and Computation Group, One Microsoft Way, Redmond, WA 98052, USA
martinro@microsoft.com

Abstract

Difference sets are basic combinatorial structures that have applications in signal processing, coding theory, and cryptography. We consider the problem of identifying a shifted version of the characteristic function of a (known) difference set and present a general algorithm that can be used to tackle any hidden shift problem for any difference set in any abelian group. We discuss special cases of this framework which include (a) Paley difference sets based on quadratic residues in finite fields which allow to recover the shifted Legendre function quantum algorithm, (b) Hadamard difference sets which allow to recover the shifted bent function quantum algorithm, and (c) Singer difference sets which allow us to define instances of the dihedral hidden subgroup problem which can be efficiently solved on a quantum computer.

1998 ACM Subject Classification F.1.1 Models of Computation, F.2 Analysis of Algorithms and Problem Complexity

Keywords and phrases Quantum algorithms, hidden shift problem, hidden subgroup problem, difference sets, Fourier transforms

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.8

1 Introduction

Many exponential speedups in quantum computing are the result of solving problems that belong to either the class of hidden subgroup problems (HSPs) or the class of hidden shift problems. For instance, the problems of factoring integers and of computing discrete logarithms in abelian groups [44] can be reformulated as solving instances of hidden subgroup problems in abelian groups [34, 26, 27, 6, 22, 23]: given a function f from an abelian group A to a set, so that f is constant on the cosets of some subgroup $H \leq A$ and takes distinct values on different cosets, the task is to find generators of H.

Successes of the hidden subgroup framework include period finding over the reals which was used by Hallgren to construct an efficient quantum algorithm for solving Pell's equation [18, 24] and more recently to the discovery of a quantum algorithm for computing unit groups of number fields of arbitrary degree [12]. Moreover, the hidden subgroup problem over symmetric and dihedral groups are related to the graph isomorphism problem [5, 2, 19, 13] and some computational lattice problems [40]. Constructing efficient algorithms for these problems are two major open questions in quantum algorithms.

As far as hidden shift problems are concerned, the shifted Legendre function problem [47], shifted sphere problems and shifts of other non-linear structures [8], problems of finding shifts of non-linear Boolean functions [43, 7] can be reformulated as solving instances of hidden shift problems in abelian groups: given a pair (f, g) of functions from an abelian group A to a set so that g is obtained from f by shifting the argument by an unknown shift

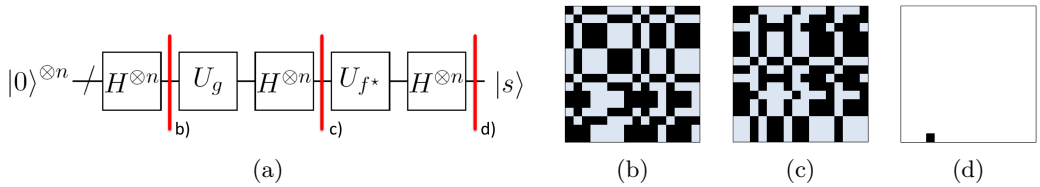

Figure 1 An example for hidden shift problem $g(x) = f(x \oplus s)$ over $A = \mathbb{Z}_2^{256}$ where the instance f is given by a bent function and $f^\star$ is the dual bent function of f. Shown in (a) is the circuit for the correlation-based algorithm from [43], where the red marker denotes the state at the respective point in time during the algorithm's execution. Shown in (b), (c), and (d) are visualizations of three stages during the algorithm's execution: (b) is the state after the shifted function has been computed into the ± 1-valued phase. Here black and light blue color stand for (re-normalized) values of $+1$ and -1, respectively. Shown in (c) is the state after the Fourier transform. As bent functions have a flat spectrum in absolute value, the state is again two valued at this point. Finally, in (d) the state after the final Hadamard transform is shown, after which all amplitude is supported on the shift $|s\rangle$. Here black denotes an amplitude of 1 and white an amplitude of 0. The main contribution of this paper, Algorithm 11, can be considered a generalization of this picture to more general classes of hiding functions f. These are obtained from difference sets which, in a precise sense, generalize the notion of bent functions.

$s \in A$, the task is to find this shift. For further background on hidden subgroup and hidden shift problems see [35, 27, 25, 9, 31].

An intriguing connection exists between *injective* instances of the hidden shift problem over abelian groups A and the hidden subgroup problem for semidirect products of the form $A \rtimes \mathbb{Z}_2$ where the action of $\mathbb{Z}_2$ is given by inversion. This connection includes the special case of the hidden shift problem over the cyclic groups $A = \mathbb{Z}_N$, where N is a large integer which are related to the dihedral groups $D_N = \mathbb{Z}_N \rtimes \mathbb{Z}_2$ via this connection. Despite much effort, a fully polynomial-time quantum algorithm for the hidden subgroup problem over the dihedral groups has remained elusive. In this paper we make a step toward solving the hidden subgroup problem over the dihedral groups by exhibiting some instances that can be solved efficiently on a quantum computer. By efficient we mean that the run-time of the quantum part of the computation is bounded polynomially in the input size, which is generally assumed to be $\log A$, and the run-time of the classical post-processing part of the computation is also bounded polynomially in the input size.

1.1 Our results

Based on the combinatorial structure of difference sets we derive a class of functions that have a two-level Fourier power spectrum. We then consider the hidden shift problem for these functions, following a general algorithm principle that was used earlier to solve the hidden Legendre symbol [47] and the hidden bent function problem [43].

The basic idea underlying all these algorithms is to use the fact that the quantum computer can perform quantum Fourier transforms efficiently. This is used in correlation-based techniques which try to identify a shift by first transforming the function into frequency (Fourier) domain, then performing a point-wise multiplication with the desired target correlator, followed by an inverse Fourier transform and a measurement in the computational basis. After these steps the shift might be obtained, even without further post-processing. An example for this approach is shown in Figure 1 where the underlying group is the Boolean hypercube and the shifted function is a so-called bent function.

A rich theory of difference sets exists and many explicit constructions are known. Furthermore, several applications of difference sets exist in signal processing [38], coding theory

[32], cryptography [37], see also [4] for further examples. In this paper we focus on the case of difference sets in abelian groups and show that a correlation-based approach can be successfully applied to several families of difference sets. In one application we consider so-called Singer difference sets, which are difference sets in cyclic groups. These difference sets have parameters

$$(v,k,\lambda) = \left(\frac{q^{d+1}-1}{q-1}, \frac{q^d-1}{q-1}, \frac{q^{d-1}-1}{q-1}\right),$$

where q is a constant and d is a parameter that defines the input size of the problem, and v, k, and λ are characteristic parameters of the difference set. We construct instances of dihedral hidden subgroup problems that can be (query) efficiently solved on a quantum computer. There is one step in our algorithm that requires the implementation of a diagonal operator whose diagonal elements are certain generalized Gauss sums whose flatness follows from a classic result due to Turyn [46]. In general, we do not known how to implement these diagonal operators efficiently and it seems that the actual computational cost has to be determined on a case-by-case basis. However, for the case of $N = 2^n - 1$ we can leverage a result by van Dam and Seroussi [48] to implement a quantum algorithm that is fully efficient in terms of its quantum complexity as well as classical complexity. Classically, the underlying problem of this white-box problem is at least as hard as the discrete logarithm problem over a finite field.

1.2 Related work

Several papers study the dihedral hidden subgroup, however, it is an open whether quantum computers can solve this problem efficiently. There is a quantum algorithm which is fully efficient in its quantum part, which however requires an exponential-time classical post-processing [14]. Furthermore, a subexponential-time quantum algorithm for the dihedral subgroup problem is known [28, 41, 29] based on a sieving idea. The dihedral hidden subgroup problem for adversarially chosen hiding functions is believed to be intractable on a quantum computer, even we are not aware of any evidence stronger for this intuition than reductions from lattice problems [40] and subset sum type problems [40, 1]. The connection between hidden shift problems over abelian groups and hidden subgroup problems over semidirect groups of the mentioned special form is well-known and was one of the reasons why the hidden shift problem has been studied for various groups [14, 47, 15, 33, 10, 21, 9].

The study of hidden shift problems has resulted in quantum algorithms that are of independent interest and have even inspired cryptographic schemes that might be candidates for post-quantum cryptography [39]. Besides the mentioned works, problems of hidden shift type were also studied in [42, 16, 17], in the rejection sampling [36] framework, and in the context of multiregister PGM algorithms for Boolean hidden shift problems [7]. The main result of this paper is Theorem 17 which asserts that there exist instances of the hidden subgroup problem over the dihedral groups D_N that can be solved in $O(\log N)$ queries to the hiding function, $O(polylog(N))$ quantum time, $O(\log N)$ quantum space, and trivial classical post-processing. Moreover, for $N = 2^n - 1$, where $n \geq 2$, there exist instances of the hidden subgroup problem over the dihedral group D_{2^n-1} for which the hiding function is white-box and for which the entire quantum computation can be performed in $O(poly(n))$ quantum time, $O(n)$ quantum space, and trivial classical post-processing. Moreover, the classical complexity of solving these instances is at least as hard as solving the discrete logarithm problem over finite fields. To the best of our knowledge this is the first exponential size family of instances of the dihedral hidden subgroup problem that can be solved efficiently on

a quantum computer, whereas for the same class of instances no efficient classical algorithm is known[1].

In Corollary 18 we show that for D_N where $N = 2^n - 1$ this theorem implies that there are an expected number of $O(2^{n^2})$ instances of the dihedral HSP (where the hidden subgroup is a reflection) that can be solved efficiently on a quantum computer. This is a small fraction of the set of all instances of such hidden subgroup problems as the number of all instances scales doubly exponential as $O(2^{n2^n})$. In particular, it seems unlikely that the set of such constructed instances has a non-trivial intersection with the set of instances that can be obtain via Regev's reduction from gapped unique-SVP lattice problems.

The rest of this paper is organized as follows. First, in Section 2 we introduce some notation and basic definitions such as Fourier transform, convolution, and the basic combinatorial object of study in this paper, namely difference sets in finite abelian groups. Next, in Section 3 we present a quantum algorithm that can be applied to any shifted difference set problem, albeit sometimes with low probability of success. We exhibit some instances of shifted difference set problems that can be solved efficiently. These special cases include the so-called class of Singer difference sets which are then used in Section 4 to construct instances of the dihedral hidden subgroup problem that can be solved efficiently on a quantum computer. Finally, in Section 5 we offer conclusions and end with some open problems.

2 Background

2.1 Quantum Fourier transforms over abelian groups

The main tool we will use are Fourier transforms over abelian groups. In the following we state some basic definitions and properties. Recall that for any abelian group A the character group $\widehat{A} = \mathrm{Hom}(A, \mathbb{C}^\times)$ is isomorphic to A. We denote the irreducible characters of A by $\chi : A \to \mathbb{C}^\times$.

▶ **Definition 1.** The *quantum Fourier transform* on $\mathbb{C}^d$ is a unitary transformation defined as $\mathrm{QFT}_A := \frac{1}{\sqrt{|A|}} \sum_{a\in A} \sum_{\chi\in\widehat{A}} \chi(a)|\chi\rangle\langle a|$.

▶ **Example 2.** For $A = \mathbb{Z}_2$ the QFT_A is given by the Hadamard transform $H := \frac{1}{\sqrt{2}}\left(\begin{smallmatrix} 1 & 1 \\ 1 & -1 \end{smallmatrix}\right)$.

▶ **Definition 3.** The *Fourier transform* of a (complex-valued) function $F : A \to \mathbb{C}$ is a function $\hat{F} : \widehat{A} \to \mathbb{C}$ defined as $\hat{F}(\chi) := \langle\chi|\mathrm{QFT}_A|F\rangle$ where $|F\rangle := \sum_{x\in A} F(x)|x\rangle$. Here $\hat{F}(\chi)$ is called the *Fourier coefficient* of F at $\chi \in \widehat{A}$. We can write it explicitly as $\hat{F}(\chi) = \frac{1}{\sqrt{|A|}} \sum_{x\in A} \chi(x)F(x)$. The set $\{\hat{F}(\chi) : \chi \in \widehat{A}\}$ is called the *Fourier spectrum* of F.

▶ **Definition 4.** The *convolution* of functions $F, G : A \to \mathbb{C}$ is a function $(F * G) : A \to \mathbb{C}$ defined as $(F * G)(x) = \sum_{y\in A} F(y)G(x - y)$.

▶ **Fact 5.** Let $F, G, H : A \to \mathbb{C}$ denote arbitrary functions. The Fourier transform and convolution have the following basic properties:

1. The Fourier transform is linear: $\widehat{F + G} = \hat{F} + \hat{G}$.

[1] It is easy to see that it is possible to find instances of the dihedral hidden subgroup problem that can be solved efficiently on a classical computer, e.g., functions that identify the points of a regular N-gon that are opposites along a symmetry axis in a linear increasing fashion. On these "taco"-like instances the hidden symmetry axis, and thereby the hidden subgroup, can be found simply by a binary search.

2. When applied twice, the Fourier transform satisfies $\hat{\hat{F}}(z) = F(-z)$. In particular, for $A = \mathbb{Z}_2$ the Fourier transform is self-inverse: $\hat{\hat{F}} = F$. From this property also follows that when the Fourier transform QFT_A is applied four times then the result is the identity.
3. QFT is unitary, so the Plancherel identity $\sum_{\chi\in A} |\hat{F}(\chi)|^2 = \sum_{x\in A} |F(x)|^2$ holds.
4. The convolution is commutative: $F*G = G*F$, and associative: $(F*G)*H = F*(G*H)$.
5. The Fourier transform and convolution are related through the following identities: $(\hat{F} * \hat{G})/\sqrt{|A|} = \widehat{FG}$ and $(\widehat{F*G})/\sqrt{|A|} = \hat{F}\hat{G}$, where $FG : A \to \mathbb{C}$ is the entry-wise product of functions F and G: $(FG)(x) := F(x)G(x)$.
6. A shift of a function in time domain leads to a point-wise multiplication with a "linear phase" in Fourier domain: If there exists $s \in A$ such that for all $x \in A$ it holds $G(x) = F(x-s)$, then for all $\chi \in \widehat{A}$ we have that $\hat{G}(\chi) = \chi(s)\hat{F}(\chi)$. This latter property will be crucial for the hidden shift algorithm presented later in this paper.

2.2 Difference sets

We recall the definition of difference sets in finite groups. We focus on the case of abelian groups in this paper. See also [3, 45, 30] for further information, in particular about the treatment for general, non-abelian groups.

Let A be a finite abelian group whose group operation we write additively and whose neutral element we denote with 0_A. Denote the pairwise inequivalent irreducible characters of A by $\widehat{A}$. For a subset $D \subseteq A$ of A we introduce the notation $D^- := \{-d : d \in D\}$ for the set of all inverses and $\Delta D := D + D^- = \{x - y : x, y \in D\}$ for the set of all differences of pairs of elements of D.

▶ **Definition 6** (Difference set)**.** Let A be a finite abelian group of size $v = |A|$. A subset $D \subseteq A$ of size $k = |D|$ is called a (v, k, λ)-difference set, where $\lambda \geq 1$, if the following equality holds in the group algebra $\mathbb{C}[A]$ of A:

$$\Delta D = \lambda(A \setminus \{0_A\}) + k0_A. \tag{1}$$

This means that the set of all differences covers each element the same number λ of times, except for the neutral element, which is covered precisely k times. A nice feature of difference sets in abelian group is that they allow to construction functions with almost flat spectrum: the following theorem [46] asserts that all Fourier coefficients of the characteristic function of a difference set in an abelian group have the same absolute value, with a possible exception of a peak at the zero frequency:

▶ **Theorem 7** (Turyn, 1965)**.** *Let A be an abelian group of order v and D be an (v, k, λ)-difference set in A. Let $\chi \in \widehat{A}$ be a non-trivial character. Then*

$$|\chi(D)| := \left|\sum_{d\in D} \chi(d)\right| = \sqrt{k - \lambda} \tag{2}$$

holds. For the trivial character χ_0 we have that $|\chi_0(D)| = k$.

Proof. We include a proof as it is instructive to see how the difference set condition can be used when interpreted as the identity (1) in the group ring $\mathbb{C}[A]$. Indeed, when identifying D with $\sum_{d\in D} d \in \mathbb{C}[A]$, we obtain from eq. (1) that

$$\left(\sum_{d\in D} d\right)\left(\sum_{d\in D} -d\right) = \lambda\left(\sum_{g\in A} g\right) + (k-\lambda)0_A. \tag{3}$$

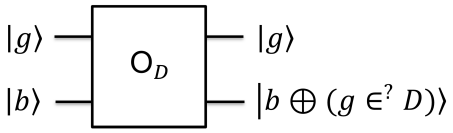

Figure 2 The oracle for the shifted difference set problem considered in this paper. The oracle allows to test membership of a given test input $g \in A$. The value of the test $g \in^? D$ is XORed onto a bit $b \in \{0,1\}$.

Let $\chi \in \widehat{A}$ be non-trivial. Then clearly $\chi(A) = 0$ holds which implies—by applying χ to both sides of eq. (3)—that $\chi(D)\overline{\chi(D)} = \chi(A) + (k-\lambda)\chi(0_A) = k - \lambda$. From this we obtain that $|\chi(D)| = \sqrt{k-\lambda}$ as claimed. ◀

With each difference set D we can canonically associate an incidence structure called the *development* of D, and denoted by $Dev(D)$.

▶ **Definition 8.** Let D be a (v,k,λ)-difference set in an abelian group A. Then the points of $Dev(D)$ are given by the elements of A and the blocks of $Dev(D)$ are given by $v + D := \{v + a : a \in D\}$, where $v \in A$.

It is well-known that $Dev(D)$ is a symmetric design. More precisely, we have the following result (for a proof see, e.g., [3], [45] or [30]):

▶ **Theorem 9.** *Let D be a (v,k,λ)-difference set in an abelian group A. Then $Dev(D)$ is a symmetric balanced-incomplete block design with parameters (v,k,λ).*

Theorem 9 implies that there are $|A|$ blocks, that each block has $|D|$ elements, that any two elements have precisely λ blocks in common and that in addition any two blocks intersect in precisely λ points. Also, it holds that $\lambda = k(k-1)/(v-1)$, see e.g. [30, Prop. 1.1], implying that λ is determined by the group order v of A and the size k of D. This equality allows us also to do a consistency check that the normalized state vector $\frac{1}{\sqrt{k}}\sum_{d\in D}|d\rangle$ is indeed mapped to a normalized vector under the Fourier transform QFT_A for the group A: using Theorem 7 we find that the length of the transformed vector is given by

$$\begin{aligned}\frac{1}{\sqrt{vk}^2}\left((v-1)|\chi(D)|^2 + |\chi_0(D)|^2\right) &= ((v-1)(k-\lambda)+k^2)/(vk)\\ &= ((v-1)k - k(k-1) + k^2)/(vk) = 1,\end{aligned}$$

as desired.

3 Quantum algorithm for shifted difference sets

▶ **Problem 10** (Shifted difference set problem)**.** Let A be an abelian group and let $s \in A$. Let $D \subseteq A$ be a (known) difference set and let $s + D$ be given by a membership oracle. The problem is to find s.

Similar to [43] we can modify Problem 10 by hiding not only the characteristic function of $s + D$ via a membership oracle but also the characteristic function of D itself. In this case we assume that we have access to membership oracles for both D and $s + D$. The following quantum algorithm is a general recipe to tackle instances of the shifted difference set problem specified in Problem 10. As we will show in the following, Algorithm 11 can be

used to find the hidden shift s efficiently in several cases of difference sets for various abelian groups A. It should be noted, however, that the probability of success crucially depends on the instance (A, D) of the problem and there are instances for which the algorithm recovers s successfully is only exponentially small. The algorithm can be seen as a generalization of correlation-based algorithms for solving hidden shift problems, e.g., [47], [43], and [7].

▶ **Algorithm 11.** The input to the algorithm is a membership oracle as in Problem 10.
Step 1: Prepare the input superposition:

$$|0\rangle \mapsto \frac{1}{\sqrt{|A|}} \sum_{g\in A} |g\rangle.$$

Step 2: Query the shifted difference set. This maps the state to:

$$\frac{1}{\sqrt{|A|}} \sum_{g\in A} (-1)^{(g\in^? s+D)} |g\rangle = \frac{1}{\sqrt{|A|}} \sum_{g\in A} |g\rangle - \frac{2}{\sqrt{|A|}} \sum_{d\in(s+D)} |d\rangle.$$

Step 3: Apply the quantum Fourier transform for A. This maps the state to:

$$|\chi_0\rangle - \frac{2}{|A|} \sum_{\chi\in\widehat{A}} \chi(s+D)|\chi\rangle = \left(1 - \frac{2k}{|A|}\right)|\chi_0\rangle - \frac{2}{|A|} \sum_{\chi\neq\chi_0} \chi(D)\chi(s)|\chi\rangle.$$

Step 4: Compute $\mathrm{diag}(1, \overline{\chi(D)}/\sqrt{k-\lambda} : \chi \neq \chi_0)$ into the phase. This maps the state to:

$$\left(1 - \frac{2k}{|A|}\right)|\chi_0\rangle - \frac{2(k-\lambda)}{|A|} \sum_{\chi\neq\chi_0} \chi(s)|\chi\rangle.$$

Step 5: Apply the inverse quantum Fourier transform for A. This maps the state to:

$$\frac{1}{\sqrt{|A|}} \left(1 - \frac{2(k-\sqrt{k-\lambda})}{|A|}\right) \sum_{g\in A} |g\rangle - \frac{2\sqrt{k-\lambda}}{\sqrt{|A|}} |-s\rangle$$

Step 5: Measure in the standard basis. Obtain $-s$ with probability $p := \frac{4(k-\lambda)}{|A|}$ and all other group elements uniformly with probability $(1-p)/|A|$.

3.1 Examples

3.1.1 Paley difference sets and shifted Legendre functions

Let A be the additive group of the finite field $\mathbb{F}_q$, where $q = p^n$ is a prime power such that $q \equiv 3 \pmod 4$. Define $D := \{x : x \text{ is a non-zero square in } \mathbb{F}_q\}$. It is well-known [3, 45] that D is then a difference set in A. These difference sets are also known as Paley difference sets. The parameters of D are as follows:

$$(v, k, \lambda) = \left(q, \frac{q-1}{2}, \frac{q-3}{4}\right).$$

▶ **Example 12.** Let $q = 27$ and consider the irreducible polynomial $f(x) = x^3 + x^2 + x + 2 \in \mathbb{F}_3[x]$, defining the finite field $\mathbb{F}_{27} \cong \mathbb{F}_3[x]/(f(x))$. Denote by $\{1, \alpha, \alpha^2\}$ an $\mathbb{F}_3$-basis of $\mathbb{F}_{27}$ where $\alpha := x \bmod f(x)$, the image of x under the canonical projection. Then D given by the following 13 elements

$$\begin{aligned} D \;=\; \{&1, \quad \alpha, \quad 2\alpha^2+2\alpha+1, \quad 2\alpha+2, \quad \alpha+2, \quad \alpha^2+2\alpha, \quad \alpha^2+1, \quad 2\alpha^2+1, \\ &\alpha^2+\alpha+1, \quad \alpha^2, \quad 2\alpha^2+2\alpha, \quad \alpha^2+2\alpha+1, \quad \alpha^2+2\alpha+2\} \end{aligned}$$

defines a $(27, 13, 6)$-difference set in $\mathbb{Z}_3^3$.

▶ Remark. Applying Algorithm 11 finds the hidden shift with probability of success

$$p_{success} = \left| \frac{2\left(\frac{q-1}{2} - \frac{q-3}{4}\right)^{1/2}}{q^{1/2}} \right|^2 \approx 1 - O(1/q).$$

This means that for large q, we can efficiently recover the hidden shift. In this case, the Algorithm 11 specializes to the algorithm given in [47]. We recover the result that an unknown shift of the Legendre symbol can be reconstructed with high probability using 1 query.

3.1.2 Hadamard difference sets and shifted bent functions

Let A be the elementary abelian 2-group $A = \mathbb{Z}_2^{2n}$, where $n \in \mathbb{N}$. Let $f : \mathbb{Z}_2^{2n} \to \mathbb{Z}_2$ be a bent function. Define $D := \{x \in \mathbb{Z}_2^{2n} : f(x) = 1\}$. It is well-known [3, 45] that D is a difference set in A. These difference sets are also known as Hadamard difference sets. The parameters of D are as follows:

$$(v, k, \lambda) = \left(2^{2n},\ 2^{2n-1} - 2^{n-1},\ 2^{2n-2} - 2^{n-1}\right).$$

▶ **Example 13.** Let $n = 4$ and let $f(x_1, x_2, x_3, x_4) = x_1x_2 \oplus x_3x_4 \oplus x_1 \in \mathbb{F}_2[x_1, x_2, x_3, x_4]$ be a bent function from the Maiorana-McFarland family [11]. Then $D = \{x \in \mathbb{F}_2^4 : f(x) = 1\}$ given by the following 6 elements

$$D = \{(1,1,0,0),\quad (1,1,1,0),\quad (1,1,0,1),\quad (0,0,1,1),\quad (1,0,1,1),\quad (0,1,1,1)\}$$

defines a $(16, 6, 2)$-difference set in $\mathbb{Z}_2^4$. The blocks of the development $Dev(D)$ of D are obtained by taking the characteristic function of f and shifting it under all elements of $A = \mathbb{Z}_2^4$. Hence, the incidence matrix of the $(16, 6, 2)$-design $Dev(D)$ is given by

$$\begin{bmatrix}
0&0&0&1&0&0&0&1&0&0&0&1&1&1&1&0\\
0&0&1&0&0&0&1&0&0&0&1&0&1&1&0&1\\
0&1&0&0&0&1&0&0&0&1&0&0&1&0&1&1\\
1&0&0&0&1&0&0&0&1&0&0&0&0&1&1&1\\
0&0&0&1&0&0&0&1&1&1&1&0&0&0&0&1\\
0&0&1&0&0&0&1&0&1&1&0&1&0&0&1&0\\
0&1&0&0&0&1&0&0&1&0&1&1&0&1&0&0\\
1&0&0&0&1&0&0&0&0&1&1&1&1&0&0&0\\
0&0&0&1&1&1&1&0&0&0&0&1&0&0&0&1\\
0&0&1&0&1&1&0&1&0&0&1&0&0&0&1&0\\
0&1&0&0&1&0&1&1&0&1&0&0&0&1&0&0\\
1&0&0&0&0&1&1&1&1&0&0&0&1&0&0&0\\
1&1&1&0&0&0&0&1&0&0&0&1&0&0&0&1\\
1&1&0&1&0&0&1&0&0&0&1&0&0&0&1&0\\
1&0&1&1&0&1&0&0&0&1&0&0&0&1&0&0\\
0&1&1&1&1&0&0&0&1&0&0&0&1&0&0&0
\end{bmatrix}.$$

Applying Algorithm 11 to the shifted difference problem for a Hadamard difference set finds the hidden shift with probability of success

$$p_{success} = \left| \frac{2\left(2^{2n-1} - 2^{2n-2}\right)^{1/2}}{2^{2n}{}^{1/2}} \right|^2 = 1.$$

This means that we always recover the hidden shift s with probability 1. In this case, the Algorithm 11 specializes to the algorithm given in [43]. We recover the result that an unknown shift of a bent function can be reconstructed using 1 query.

3.1.3 Singer difference sets and shifted hyperplanes

Let q be a prime power, let $d \geq 1$ and let $\mathbb{F}_{q^{d+1}}$ be the finite field with q^{d+1} elements. The Singer difference sets are constructed from d-dimensional projective spaces over $\mathbb{F}_q$ as follows: consider the trace map tr from $\mathbb{F}_{q^{d+1}}$ to $\mathbb{F}_q$. Let T be a transversal of $\mathbb{F}_q^*$ in $\mathbb{F}_{q^{d+1}}^*$ that is chosen in such a way that tr maps T onto the values 0 and 1 in $\mathbb{F}_q$ only. We can then define a group $A := \mathbb{F}_{q^{d+1}}^\times / \mathbb{F}_q^\times$ which turns out to be cyclic. Furthermore, we can define a subset $D := \{x : x \in A | \mathrm{tr}(x) = 0\}$. It turns out [3] that D is then a difference set in $\mathbb{Z}_N$, where $N = \frac{q^{d+1}-1}{q-1}$. This difference set has parameters

$$(v, k, \lambda) = \left(\frac{q^{d+1}-1}{q-1}, \frac{q^d-1}{q-1}, \frac{q^{d-1}-1}{q-1} \right). \tag{4}$$

▶ **Example 14.** Let $P = PG(2, 3)$ be the two-dimensional projective space over $\mathbb{F}_3$. Then $|P| = (3^3 - 1)/(3 - 1) = 13$. By choosing an $\mathbb{F}_3$-basis of $\mathbb{F}_{27}$ we obtain an embedding of $\mathbb{F}_{27}^\times$ into $\mathrm{GL}(3, \mathbb{F}_3)$. If $\alpha \in \mathbb{F}_{27}$ is a primitive element for $\mathbb{F}_{27}/\mathbb{F}_3$, then the corresponding matrix has order 26 and therefore generates a cyclic subgroup C of $\mathrm{GL}(3, \mathbb{F}_3)$ order 26. Under the canonical projection $\pi : \mathrm{GL}(3, \mathbb{F}_3) \to \mathrm{PGL}(3, \mathbb{F}_3)$, the subgroup C is mapped to a subgroup $\overline{C} = \langle \sigma \rangle$ of $\mathrm{PGL}(3, \mathbb{F}_3)$ of order 13 (see also [20, Kapitel II, Satz 7.3]). This subgroup is sometimes also called the "Singler cycle." The Singer cycle operates transitively on the points $\{(x : y : z) : x, y, z \in \mathbb{F}_3\}$ of the projective space P. By picking the particular order $[\sigma^i p_0 : i = 0, \ldots, 12]$, where p_0 is the point $(0 : 0 : 1)$, we obtain points that we can identify with $[0, 1, \ldots, 12]$. The image of the hyperplane given by all points $p \in \mathbb{F}_{27}$ with $\mathrm{tr}(p) = 0$ is given by the set $D := \{0, 1, 3, 9\}$. Then D is a $(13, 4, 1)$-difference set in the cyclic group $\mathbb{Z}_{13}$. The development $Dev(D)$ is given by:

$$\begin{bmatrix}
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 \\
1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 1
\end{bmatrix}.$$

If in eq. (4) we consider q to be constant and d be a parameter that corresponds to the input size of a hidden shift problem over $\mathbb{Z}_N$, we can use Algorithm 11 to solve the hidden shift problem over $\mathbb{Z}_N$ with probability of success

$$p_{success} = \left| \frac{4(q^d - q^{d-1})^{1/2}}{(q^{d+1}-1)^{1/2}} \right|^2 = \frac{2}{q} + O(1/q^2).$$

This means that for constant q, we can efficiently recover the hidden shift from a constant number of trials. In Section 4 we show how we can use the instances of hidden difference problems of Singer type to construct efficiently solvable instances of the dihedral hidden subgroup problem.

▶ Remark. We note that not all shifted difference set problems can be solved efficiently by using Algorithm 11. An example is given by the projective planes $(q^2+q+1, q+1, 1)$ of order q. In this case the input size is given by $\log q$ and the probability of success can be computed to be $p_{success} = |\frac{2q^{1/2}}{(q^2+q+1)^{1/2}}|^2 \approx \frac{2}{q} + O(1/q^2)$, i.e., the probability of success is exponentially small in this case. It is an open problem if cases like this can be tackled, e.g., by considering multi-register algorithms.

3.2 Injectivization

As mentioned in the introduction, it is well known that the hidden subgroup problem over semidirect products of the form $A \rtimes \mathbb{Z}_2$, where the action of $\mathbb{Z}_2$ is given by inversion, and the hidden shift problem over A are closely related. More precisely, there is a one-to-one correspondence between instances of the hidden subgroup problem in which the subgroup is a conjugate of the order 2 subgroup $H = \langle (0,1) \rangle$ and instances of *injective* hidden shift problems over A.

This leads to the question whether it is possible to relate instances of hidden shift problems where the hiding function $f : A \to S$ is not injective to the injective case. Thankfully, as shown in [17] such a connection indeed exists. We briefly review this construction.

For given $f : A \to S$, and a set $V := \{v_1, \ldots, v_m\} \subseteq A$ of m elements of A we define a new function $f_V(x) := (f(x+v_1), \ldots, f(x+v_m))$. Gharibi showed in [17] that if the set V is chosen uniformly at random, then the probability that the function f_V is not injective can be upper bounded as

$$Pr_V(f_V \text{ not injective}) \leq |A|^2(1-\gamma_{min})^m, \tag{5}$$

where $\gamma_{min} := \min_{v \neq 0}(\gamma_v(f))$ and for all $v \in A$ the so-called influences $\gamma_v(f)$ of f at v are defined as $\gamma_v(f) := Pr_x(f(x) \neq f(x+v))$, i.e., the probability that f changes its value when the input is toggled by v.

We now show that for instances of shifted difference set problems these influences can be bounded by the parameters of the difference set alone. This in turn allows to establish a bound on the overall number of copies m that are needed to make the hiding function injective, namely a bound that grows proportional to $\log |A|$.

▶ **Lemma 15.** *Let $f : A \to \{0,1\}$ be a hiding function corresponding to the characteristic function a (v,k,λ)-difference set in an abelian group A. Then for all $v \in A \setminus \{0\}$ we have that $\gamma_v(f) = \frac{2(k-\lambda)}{|A|}$.*

Proof. Note that

$$\begin{aligned} Pr_x(f(x) \neq f(x+v)) &= \frac{1}{|A|} \sum_{x \in A} (f(x) - f(x+v))^2 & (6) \\ &= \frac{1}{|A|} (|D| + |v+D| - 2|D \cap (v+D)|) & (7) \\ &= \frac{1}{|A|}(2|D| - 2\lambda) = \frac{2(k-\lambda)}{|A|}, & (8) \end{aligned}$$

where in the second equation we used the fact that only elements in the intersection contribute to $f(x)f(x+v)$ and in the third equation we used Theorem 9 which implies that $|D \cap (v+D)| = \lambda$ for all $v \neq 0$. ◀

We can now establish the claimed result that the number of copies only grows with the log of the group size.

▶ **Theorem 16.** *Let D be a (v,k,λ)-difference set in an abelian group A and $f : A \to \{0,1\}$ an instance of a hidden difference set problem for D. Then $m = O(\log|A|)$ copies are enough to obtain an injective instance f_V with probability greater than $1 - \frac{1}{64}$.*

Proof. From the cited bound (5) we obtain that

$$Pr(f \text{ injective}) \geq 1 - |A|^2(1-\gamma_{min}(f))^m$$

It is easy to see that lower bounding the right hand side in this expression by $1 - \frac{1}{64}$ is equivalent to choosing $m \geq \frac{1}{\log(1-\gamma_{min}(f))}(-6 - 2\log_2(|A|))$. Now, from Lemma 15 we have that $\gamma_{min}(f) = \frac{2(k-\lambda)}{|A|}$ from which we can conclude that in particular $|A| \geq 2(k-\lambda)$ holds. Using the fact that $\log_2(1-x) \leq -x$ holds for $x \in [0,1)$, this implies that

$$m \geq \left\lceil \frac{1}{\log_2\left(1 - \frac{2(k-\lambda)}{|A|}\right)}(-6 - 2\log_2|A|)\right\rceil \tag{9}$$

$$\geq \left\lceil -\frac{|A|}{2(k-\lambda)}(-6 - 2\log_2|A|)\right\rceil \geq 2\log_2|A| + 6. \tag{10}$$

Hence $m = O(\log|A|)$ copies are enough to guarantee that for $V = \{v_1, \ldots v_m\}$ chosen uniformly at random, the probability of f_V being injective is at least $1 - \frac{1}{64}$. ◀

4 Efficiently solvable dihedral hidden subgroup problems

▶ **Theorem 17.** *There exist instances of the hidden subgroup problem over the dihedral groups D_N that can be solved in $O(\log N)$ queries to the hiding function, $O(polylog(N))$ quantum time, $O(\log N)$ quantum space, and trivial classical post-processing. Moreover, for $N = 2^n - 1$, where $n \geq 2$, there exist instances of the hidden subgroup problem over the dihedral group D_{2^n-1} for which the hiding function is white-box and for which the entire quantum computation can be performed in $O(poly(n))$ quantum time, $O(n)$ quantum space, and trivial classical post-processing. Moreover, the classical complexity of solving these instances is at least as hard as solving the discrete logarithm problem over finite fields.*

Proof. To construct the instances that can be solved efficiently we proceed in three steps: (i) first, we show that a particular set of hidden shift problems over $\mathbb{Z}_N$ can be obtained from hiding functions that are indicator functions of hyperplanes and that these indicator functions can be implemented efficiently, (ii) next we show that Algorithm 11 is query, time, and space efficient for these instances; (iii) finally, we show that it is possible to construct instances of the hidden subgroup problem in D_N from the hidden shift instances constructed in (i) and that these instances are unlikely to be solvable on a classical computer, unless computing finite field discrete logarithms is possible in polynomial-time.

Step (i): We instantiate the abelian difference set quantum algorithm for the case of the cyclic group $A = \mathbb{Z}_N$, where $N = (q^{d+1}-1)/(q-1) = q^d + q^{q-1} + \ldots + 1$. Here q is constant and d is a parameter that corresponds to the input size of the problem. We use the explicitly (white-box) description of the function $f(x) = \mathrm{tr}(\alpha^x)$, where tr denotes the trace map from $\mathbb{F}_q^{d+1}$ to $\mathbb{F}_q$ and where α is a primitive element in $\mathbb{F}_q$. Now, the instance of the shifted difference set problem is defined by the hiding function $g(x) = \mathrm{tr}(\alpha^{x+s})$, where $s \in \mathbb{Z}_N$. This function can be given as a white-box function by providing the element $\beta := \alpha^s \in \mathbb{F}_{q^{d+1}}$ so that g can then be evaluated as $g(x) = \mathrm{tr}(\alpha^x\beta)$. Note that the set $\{x \in A : \mathrm{tr}(x) = 0\}$ defines a hyperplane and therefore a difference set D of Singer type.

Step (ii): We now go through each step of Algorithm 11 and check that the steps are time- and space-efficient. In the first step, a Fourier transform is applied to create the equal

superposition of all elements of A. As A is abelian, this can clearly be done efficiently. In the second step, we have to evaluate the function g in superposition. Again, as there is an explicit description of the trace which can be computed as sum of powers of the relative Frobenius from $\mathbb{F}_{q^{d+1}}$ to $\mathbb{F}_q$ as follows $\mathrm{tr}(x) = x + x^q + \ldots + x^{q^d}$ we can evaluate $g(x) = \mathrm{tr}(\alpha^x \beta)$ by first constructing a circuit for exponentiation $x \mapsto \alpha^x \in \mathbb{F}_{q^{d+1}}$ followed by scalar multiplication with β, followed by the application of the trace map. Clearly, all these operations can be efficiently implemented by means of a classical Boolean circuit whose size and depth are polynomial in d. Hence, by applying standard techniques from reversible computing, we can derive quantum circuits for the evaluation of f and g. Therefore we can compute Step 2 efficiently on a quantum computer.

Step 3 is another application of a quantum Fourier transform over the abelian group A which as in Step 1 can be done efficiently. Step 4 is the most challenging step in the entire algorithm. If we were just interested in the query complexity of the problem we would be done as we could simply apply the diagonal unitary operator $\Delta := \mathrm{diag}(1, \overline{\chi_1(D)}, \ldots, \overline{\chi_{N-1}(D)})$, where $\chi_1, \ldots, \chi_{N-1}$ runs through all non-trivial characters of $\mathbb{Z}_N$. This argument is sufficient to establish the first claimed statement in the theorem, i.e., the query complexity result.

For the white-box statement, we are interested in the time- and space-efficiency of the algorithm, i.e., we have to show that Δ can be implemented efficiently. For this we have to assume $N = 2^n - 1$ as required by one of the subsequent steps (and we highlight where). First we use a result due to van Dam and Seroussi [48] establishing that finite field Gauss sums can be approximated efficiently on a quantum computer. The connection to our situation is that the elements of Δ are Gauss sums. We briefly review the van Dam/Seroussi algorithm and then argue that we can apply it in superposition in order to compute Δ.

Let $\mathbb{F}_q$ be a finite field where $q = p^{d+1}$ and p prime. Let $\psi := \mathbb{F}_p \to \mathbb{C}^\times$ be a non-trivial additive character and let $\chi : \mathbb{F}_q^\times \to C^\times$ be a non-trivial multiplicative character. Then the Gauss sum $G(\psi, \chi)$ is defined as

$$G(\psi, \chi) = \sum_{x \in \mathbb{F}_q^\times} \chi(x)\psi(\mathrm{tr}(x)).$$

The additive and multiplicative characters of $\mathbb{F}_q$ have a simple description: For $n \in \mathbb{N}$ denote a primitive n-th root of unity in $\mathbb{C}^\times$ with ω_n. Then the additive characters take the form $\psi_\mu(x) := \omega_p^{\mathrm{tr}(\mu x)}$, where $\mu \in \mathbb{F}_q$ runs through all elements of $\mathbb{F}_q$. The multiplicative characters can be described using a primitive elements $\alpha \in \mathbb{F}_{p^{d+1}}$ as follows: $\chi_\beta(\alpha^i) := \omega_{p^{d+1}-1}^{\beta i}$, where β runs through all non-zero elements of $\mathbb{F}_{p^{d+1}}$. This means that evaluation $\chi_\beta(x) = \omega^{\beta \log_\alpha(x)}$ requires the computation of a discrete log over the multiplicative group of the field.

It is known that for non-trivial ψ and χ, the absolute value of the Gauss sum $G(\psi, \chi)$ evaluates to $|G(\psi, \chi)| = \sqrt{q}$, i.e., $G(\psi, \chi) = \sqrt{q}e^{i\theta}$, where $\theta \in [0, 2\pi)$. The paper [48] established that θ can be approximated with precision ε by a quantum algorithm in time $O(\frac{1}{\varepsilon} polylog(q))$. As we are overall only looking for a quantum algorithm that can solve the hidden shift problem over $\mathbb{Z}_N$ with bounded probability of success, it will be enough to approximate the diagonal elements of Δ with constant precision, i.e., we can use the van Dam/Seroussi algorithm to estimate $G(\psi, \chi)$. A minor complication is the fact that in [48] only the case of known character χ is considered, however, by making all steps of the algorithm conditioned on the character χ it can be easily seen that the transformation $|\chi\rangle \mapsto G(\chi, \psi)/\sqrt{q}|\chi\rangle$ can also be implemented coherently, i.e., on superposition of inputs χ. The final step is to show how to relate $\chi(D)$ and $G(\chi, \psi)$. For this we make the restriction

that $p = 2$ so that our parameters always take the form $N = 2^{d+1} - 1$. We then obtain that

$$\begin{aligned}\chi(D) &= \sum_{x:\mathrm{tr}(x)=0} \chi(x) = \sum_{x\in\mathbb{F}_q^\times} \chi(x)(1+(-1)^{\mathrm{tr}(x)}) = \sum_{x\in\mathbb{F}_q^\times} \chi(x) + \sum_{x\in\mathbb{F}_q^\times} \chi(x)(-1)^{\mathrm{tr}(x)} \\ &= \sum_{x\in\mathbb{F}_q^\times} \chi(x)(-1)^{\mathrm{tr}(x)} = G(\psi,\chi),\end{aligned}$$

where ψ denotes the additive character $\psi(x) := (-1)^{\mathrm{tr}(x)}$ of $\mathbb{F}_{2^n}$ and $\chi(x)$ denotes a multiplicative character of $\mathbb{F}_{2^n}^\times$. This argument establishes that we can approximate the operator Δ efficiently on a quantum computer with constant precision ε.

The final two steps of the algorithm are easy to do: Step 5 is just another Fourier transform and Step 6 a measurement in the computational basis, both of which can be done efficiently.

Step (iii): To construct the desired instances of the hidden subgroup problem from the hidden shift problem, we apply the results from Subsection 3.2 and specialize them to the case of the Singer difference sets. We pick $m = 2(d+1)+6$ random elements $v_1,\ldots,v_m \in \mathbb{F}_{q^{d+1}}$ and construct the hiding function $g_{v_1,\ldots,v_m}(x) := (g(x+v_1),\ldots,g(x+v_m))$ which according to Theorem 16 is injective with probability greater than $1-\frac{1}{64}$. We then apply another standard construction [15, 28] which allows to turn an instance of an injective hidden shift problem into a hidden subgroup problem. Indeed, if $f,g : A \to S$ is an injective instance of a hidden shift problem with shift $s \in A$, then the corresponding hidden subgroup problem over $A \rtimes \mathbb{Z}_2$ is given by the hiding function $F((a,0)) := f(a)$ and $F((a,1)) := g(a)$, where (a,t) is an encoding of the elements, i.e., $a \in A$ and $t \in \mathbb{Z}_2$. Conversely, if $F : A \rtimes \mathbb{Z}_2 \to S$ is a defining function of a hidden subgroup problem with hidden subgroup $H = \langle (a,1)\rangle$ of order 2, then $f(x) := F(x,0)$ and $g(x) := F(x,1)$ defines a hidden shift problem over A.

Overall, we established the claimed result of the existence of an efficient quantum algorithm to solve the hidden subgroup problem. The classical complexity of finding the shift s from β clearly is as least as hard as solving the discrete logarithm over a finite field. ◀

▶ **Corollary 18.** *Let $N = 2^n - 1$, where $n \geq 2$, there there exist an expected number of 2^{n^2} instances of hidden subgroup problems over D_N that can be solved efficiently on a quantum computer.*

Proof. From the proof of Theorem 17 we see that in step (iii) for each random choice of $m = O(\log|A|)$ elements, where $|A| = |\mathbb{Z}_{2^n-1}| = 2^n - 1$, we obtain a valid injectivization of the hidden shift function. There are an expected number of $O(|A|^m) = O((2^{\log_2(|A|)})^m) = O(2^{n^2})$ such functions. ◀

5 Conclusions

We showed that the property of difference sets to give rise to functions with two level Fourier (power) spectrum which makes them useful for classical applications also allows to define hidden shift problems which can then be tackled on a quantum computer. While a solution to general hidden shift problems for arbitrary difference sets remains elusive, we showed that several interesting special cases can indeed be solved efficiently on a quantum computer. This includes the known cases of the Legendre symbol which we show to be an instantiation of our framework for the case of a Paley difference set. Furthermore, it includes the case of hidden bent functions which we show to be special cases of Hadamard difference sets. The case of Singer difference sets appears to be new and allows us to construct white-box

instances of dihedral hidden subgroup problems that can be solved fully efficiently on a quantum computer, both in the quantum and in the classical parts of the algorithm.

Open problems include whether these findings have any consequence for more general classes of instances of the dihedral hidden subgroup problem and the hidden subgroup problem in other semidirect products of a similar form. Other open problems include whether it is possible to solve the shifted difference set problem for projective planes which we mentioned cannot be solved by our main algorithm with better than exponentially small probability of success. One possible avenue for future research is to consider multi-register algorithms to tackle this problem. Another open problem is the case of hidden shift problems over abelian groups for functions that have approximately constant spectra, possibly with the exception of the zero frequency as in case of the functions arising from difference sets considered in this paper.

Acknowledgments. The author would like to thank Schloss Dagstuhl for hosting Seminar 15371, during which part of this research was carried out.

References

1 Dave Bacon, Andrew M. Childs, and Wim van Dam. Optimal measurements for the dihedral hidden subgroup problem. *Chicago Journal of Theoretical Computer Science*, 2006(2), Oct 2006. quant-ph/0501044. URL: `http://cjtcs.cs.uchicago.edu/articles/2006/2/contents.html`.

2 Robert Beals. Quantum computation of Fourier transforms over symmetric groups. In *Proceedings of the twenty-ninth annual ACM symposium on Theory of computing*, STOC'97, pages 48–53, New York, NY, USA, 1997. ACM. `doi:10.1145/258533.258548`.

3 Thomas Beth, Dieter Jungnickel, and Hanfried Lenz. *Design Theory*, volume I. Cambridge University Press, 2nd edition, 1999.

4 Thomas Beth, Dieter Jungnickel, and Hanfried Lenz. *Design Theory*, volume II. Cambridge University Press, 2nd edition, 1999.

5 Dan Boneh and Richard Lipton. Quantum cryptanalysis of hidden linear functions. In *Advances in Cryptology — CRYPT0' 95*, volume 963 of *Lecture Notes in Computer Science*, pages 424–437. Springer, 1995. `doi:10.1007/3-540-44750-4_34`.

6 Gilles Brassard and Peter Høyer. An exact polynomial–time algorithm for Simon's problem. In *Proceedings of Fifth Israeli Symposium on Theory of Computing and Systems*, pages 12–33. ISTCS, IEEE Computer Society Press, 1997. arXiv quant–ph/9704027.

7 Andrew M. Childs, Robin Kothari, Maris Ozols, and Martin Roetteler. Easy and hard functions for the Boolean hidden shift problem. In *8th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2013, May 21-23, 2013, Guelph, Canada*, pages 50–79, 2013.

8 Andrew M. Childs, Leonard J. Schulman, and Umesh V. Vazirani. Quantum algorithms for hidden nonlinear structures. In *48th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2007), October 20-23, 2007, Providence, RI, USA, Proceedings*, pages 395–404, 2007.

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

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

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

12 Kirsten Eisenträger, Sean Hallgren, Alexei Kitaev, and Fang Song. A quantum algorithm for computing the unit group of an arbitrary degree number field. In *Symposium on Theory of Computing, STOC 2014, New York, NY, USA, May 31 – June 03, 2014*, pages 293–302, 2014.

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

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

15 Katalin Friedl, Gábor Ivanyos, Frédéric Magniez, Miklos Santha, and Pranab Sen. Hidden translation and orbit coset in quantum computing. In *Proceedings of the 35fth Annual ACM Symposium on Theory of Computing (STOC'03)*, pages 1–9. ACM, 2002. quant-ph/0211091. doi:10.1145/780542.780544.

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

17 Mirmojtaba Gharibi. Reduction from non-injective hidden shift problem to injective hidden shift problem. *Quantum Information and Computation*, 13(3&4):212–230, 2013.

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

19 Peter Høyer. Efficient quantum transforms. quant-ph/9702028, 1997.

20 Bertram Huppert. *Endliche Gruppen I*. Springer, 1967.

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

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

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

24 Richard Jozsa. Quantum computation in algebraic number theory: Hallgren's efficient quantum algorithm for solving Pell's equation. *Annals of Physics*, 306(2):241–279, 2003. quant-ph/0302134. doi:10.1016/S0003-4916(03)00067-8.

25 Phillip Kaye, Raymond Laflamme, and Michele Mosca. *An Introduction to Quantum Computing*. Oxford University Press, 2007.

26 Alexei Yu. Kitaev. Quantum measurements and the abelian stabilizer problem, 1995. quant-ph/9511026.

27 Alexei Yu. Kitaev, Alexander Shen, and Mikhail N. Vyalyi. *Classical and Quantum Computation*, volume 47 of *Graduate Studies in Mathematics*. American Mathematical Society, 2002.

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

29 Greg Kuperberg. Another subexponential-time quantum algorithm for the dihedral hidden subgroup problem. arXiv: 1112.3333, 2011. URL: http://arxiv.org/abs/1112.3333.

30 Eric S. Ladner. *Symmetric Designs: An Algebraic Approach*, volume 74 of *London Mathematical Society Lecture Note Series*. Cambridge University Press, 1983.

31 Chris Lomont. The hidden subgroup problem – review and open problems. quant-ph/0411037, 2004.

32 F. J. MacWilliams and N. J. A. Sloane. *The theory of error–correcting codes*. North–Holland, Amsterdam, 1977.

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

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

35 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.

36 Maris Ozols, Martin Roetteler, and Jérémie Roland. Quantum rejection sampling. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, ITCS'12, pages 290–308, New York, NY, USA, 2012. ACM. arXiv: 1103.2774.

37 Josef Pieprzyk, Thomas Hardjono, and Jennifer Seberry. *Fundamentals of computer security*. Springer, 2003.

38 Alexander Pott, Vijay Kumar, Tor Helleseth, and Dieter Jungnickel, editors. *Difference sets, sequences, and their correlations*, volume 542 of *NATO Science Series*. Kluwer, 1998.

39 Oded Regev. New lattice-based cryptographic constructions. *J. ACM*, 51(6):899–942, 2004.

40 Oded Regev. Quantum computation and lattice problems. *SIAM Journal on Computing*, 33(3):738–760, 2004.

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

42 Martin Roetteler. Quantum algorithms to solve the hidden shift problem for quadratics and for functions of large Gowers norm. In *Proceedings of the 34st International Symposium on Mathematical Foundations of Computer Science (MFCS'09)*, volume 5734 of *Lecture Notes in Computer Science*, pages 663–674. Springer, 2009. arXiv: 0911.4724. doi:10.1007/978-3-642-03816-7_56.

43 Martin Roetteler. Quantum algorithms for highly non-linear Boolean functions. In *Proceedings of the 21st Annual ACM-SIAM Symposium on Discrete Algorithms (SODA'10)*, pages 448–457, 2010. arXiv: 0811.3208. URL: http://arxiv.org/abs/0811.3208.

44 Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26(5):1484–1509, 1997. Preliminary version in FOCS 1994. doi:10.1137/S0097539795293172.

45 Douglas R. Stinson. *Combinatorial Designs: Constructions and Analysis*. Springer, 2003.

46 Richard J. Turyn. Character sums and difference sets. *Pacific Journal of Mathematics*, 15(1):319–346, 1965.

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

48 Wim van Dam and Gadiel Seroussi. Quantum algorithms for estimating Gauss sums. quant-ph/0207131, 2002.

Instantaneous Non-Local Computation of Low T-Depth Quantum Circuits

Florian Speelman*

Centrum Wiskunde & Informatica, Amsterdam, the Netherlands
f.speelman@cwi.nl

Abstract

Instantaneous non-local quantum computation requires multiple parties to jointly perform a quantum operation, using pre-shared entanglement and a single round of simultaneous communication. We study this task for its close connection to position-based quantum cryptography, but it also has natural applications in the context of foundations of quantum physics and in distributed computing. The best known general construction for instantaneous non-local quantum computation requires a pre-shared state which is exponentially large in the number of qubits involved in the operation, while efficient constructions are known for very specific cases only.

We partially close this gap by presenting new schemes for efficient instantaneous non-local computation of several classes of quantum circuits, using the Clifford+T gate set. Our main result is a protocol which uses entanglement exponential in the T-depth of a quantum circuit, able to perform non-local computation of quantum circuits with a (poly-)logarithmic number of layers of T gates with quasi-polynomial entanglement. Our proofs combine ideas from blind and delegated quantum computation with the garden-hose model, a combinatorial model of communication complexity which was recently introduced as a tool for studying certain schemes for quantum position verification. As an application of our results, we also present an efficient attack on a recently-proposed scheme for position verification by Chakraborty and Leverrier.

1998 ACM Subject Classification F.1.1 Models of Computation

Keywords and phrases Quantum Cryptography, Quantum Communication

Digital Object Identifier 10.4230/LIPIcs.TQC.2016.9

1 Introduction

We study the task of instantaneous non-local quantum computation, and present new protocols to efficiently perform this task for specific classes of quantum circuits. Our main motivation comes from position-based quantum cryptography, where previous attacks on schemes for position-based quantum cryptography have taken either of two forms:

First results on quantum position-based cryptography involved attacks on specific proposals for schemes, such as the attacks by Lau and Lo [31], those by Kent, Munro and Spiller [28], and the attack on Beigi and König's scheme using mutually-unbiased-bases [37]. A certain family of efficient attacks on a concrete class of single-qubit schemes [13] was formalized by the garden-hose model. Described as 'fast protocols for bipartite unitary operators', Yu, Griffiths and Cohen [40, 39] give protocols that, although not directly inspired by position-based quantum cryptography, can be translated to our setting.

On the other hand Buhrman et al. [12] constructed a general attack which treats the quantum functionality of the protocol to be attacked as a black box. For a protocol which uses

* The author is supported by the EU projects SIQS and QALGO.

a message of n qubits, the entanglement consumption of this attack is around $2^{\log(\frac{1}{\varepsilon})2^{4n}}$ EPR pairs, doubly exponential in n. Here ε represents the probability that the attack does not succeed. The construction of Buhrman et al. was based on a protocol for 'instantaneous non-local measurement' by Vaidman [38, 16]. Beigi and König [5] later constructed a more efficient general attack, using port-based teleportation – a new teleportation method introduced by Ishizaka and Hiroshima [25, 26]. The improved attack uses $O(n\frac{2^{8n}}{\varepsilon^2})$ EPR pairs, still an exponential dependence on n.

These protocols were able to solve the following task. Given a constant $\varepsilon \geq 0$ and an n-qubit quantum operation[1] U, where n is a natural number. Two players, Alice and Bob, receive an arbitrary input state ρ_{AB} of n qubits, with the players receiving $n/2$ qubits each. After a single round of simultaneous quantum[2] communication, the players must output a state ε-close to $U\rho_{AB}U^\dagger$. Alice outputs the first $n/2$ qubits of the state and Bob outputs the other $n/2$ qubits. We define $\mathrm{INQC}_\varepsilon(U)$ as the smallest number of EPR pairs that the players have to share at the start of a protocol which performs this task. $\mathrm{INQC}(U)$ is used as a shorthand for $\mathrm{INQC}_0(U)$, a protocol which works with no error. We present a more precise definition of INQC is presented in Appendix A.

In this work we partially bridge the gap between efficient specific constructions for instantaneous non-local computation and expensive general ones, by constructing a protocol for non-local computation of a unitary transformation U such that the entanglement use of the protocol depends on the quantum circuit which describes U.

In particular, writing quantum circuits over the Clifford+T gate set, we create a protocol using entanglement exponential in the *T-count*. We also present a protocol that uses an amount of entanglement which scales as the number of qubits n raised to the power of the *T-depth* of the circuit. Even though this is a quickly-growing dependence, for circuits of constant T-depth this amounts to a polynomial dependence on n, unlike any earlier construction. For circuits of polylogarithmic T-depth we obtain an amount of entanglement which is quasi-polynomial in n, i.e. a dependence of the form $2^{(\log n)^c}$ for some constant c. Note that the depth and size of the quantum circuit can be much higher than its T-depth: we allow an arbitrary number of gates from the Clifford group in addition to the limited number of T gates. Our results imply new efficient attacks on any scheme for position-verification where the action of the honest party can be written as a low T-depth quantum circuit.

Linking blind quantum computation and instantaneous non-local quantum computation was first considered by Broadbent[3] [8], who considered a setting where the parties have access to non-local boxes – correlations even stronger than those allowed by quantum mechanics. The techniques we use are also based on delegated and blind quantum computation [15, 4, 18, 19, 7] and results on computation via teleportation [24], but we combine them with new ideas from the *garden-hose model* [13, 29] – a recently-introduced combinatorial model for communication complexity with close links to a specific class of schemes for position verification.

We prove two main theorems, each improving on the entanglement consumption of the best-known previous constructions for non-local instantaneous quantum computation

[1] Our constructions only consider unitaries given by quantum circuits, but the task naturally extends to more general quantum operations. The motivation for Vaidman's original scheme [38], which formed the basis of Buhrman et al.'s construction, was to instantaneously perform a non-local measurement. Our constructions can also be applied to that case, by writing the measurement as a unitary operation followed by a measurement in the computational basis.

[2] Since restriction to classical communication is not necessarily dictated by the application in position-based quantum cryptography, we allow quantum communication. All presented protocols work equally well when all messages are classical instead.

[3] These results were first available as privately-circulated notes in December 2011, and were made available online in December 2015.

for specific circuits[4]. Additionally, we use our proof method to construct a new attack on a scheme for position verification which was recently proposed by Chakraborty and Leverrier [14].

▶ **Theorem 3.** *Any n-qubit Clifford+T quantum circuit C which has at most k T gates has a protocol for instantaneous non-local computation using $O(n2^k)$ EPR pairs.*

▶ **Theorem 5.** *Any n-qubit quantum circuit C using the Clifford+T gate set which has T-depth d, has a protocol for instantaneous non-local computation using $O((68n)^d)$ EPR pairs.*

The main technical tool we use in the proof of our depth-dependent construction is the following lemma, which is able to remove a conditionally-applied gate from the Clifford group without any communication – at an entanglement cost which scales with the garden-hose complexity of the function which describes the condition.

▶ **Lemma 4.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ be a function known to all parties, and let $GH(f)$ be the garden-hose complexity of the function f. Assume Alice has a single qubit with state $\mathrm{P}^{f(x,y)}|\psi\rangle$, for binary strings $x, y \in \{0,1\}^n$, where Alice knows the string x and Bob knows y. The following two statements hold:*

1. *There exists an instantaneous protocol without any communication which uses $2GH(f)$ pre-shared EPR pairs after which a chosen qubit of Alice is in the state $\mathrm{X}^{g(\hat{x},\hat{y})}\mathrm{Y}^{h(\hat{x},\hat{y})}|\psi\rangle$. Here $\hat{x}$ depends only on x and the $2GH(f)$ bits that describe the measurement outcomes of Alice, and $\hat{y}$ depends on y and the measurement outcomes of Bob.*
2. *The garden-hose complexities of the functions g and h are at most linear in the garden-hose complexity of the function f. More precisely, $GH(g) \leq 4GH(f) + 1$ and $GH(h) \leq 11GH(f) + 2$.*

Chakraborty and Leverrier [14] recently proposed a protocol for quantum position verification on the interleaved multiplication of unitaries. They show that all known attacks, applied to this protocol, require entanglement exponential in the number of terms t in the product. As an application of Lemma 4, we present an attack on their proposed protocol which has entanglement cost polynomial in t and the number of qubits n. The new attack requires an amount of entanglement which scales as $(\frac{t}{\varepsilon})^{O(1)}$ per qubit, and for each qubit succeeds with probability at least $1 - \varepsilon$.

2 Preliminaries

2.1 The Pauli matrices and the Clifford group

The single-qubit *Pauli matrices* are $\mathrm{X} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, $\mathrm{Y} = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}$, $\mathrm{Z} = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and the identity $\mathrm{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$. A *Pauli operator* on an n-qubit state is the tensor product of n one-qubit Pauli matrices, the group of n qubit Pauli operators[5] is $\mathcal{P} = \{\sigma_1 \otimes \cdots \otimes \sigma_n \mid$

[4] From now on, whenever we write 'quantum circuit', we will always mean a quantum circuit that only uses the Clifford group generators, together with T gates.

[5] The given definition includes a global phase, which is not important when viewing the elements as quantum gates.

$\forall j : \sigma_j \in \{I, X, Y, Z\}\} \times \{\pm 1, \pm i\}$. These are some of the simplest quantum operations and appear, for example, as corrections for standard quantum teleportation.

The *Clifford group* can be defined as those operations that take elements of the Pauli group to other elements of the Pauli group under conjugation – the *normalizer* of the Pauli group. We consider the Clifford group on n qubits, for some natural number n.

$$\mathcal{C} = \{U \in \mathcal{U}(2^n) \mid \forall \sigma : \sigma \in \mathcal{P} \implies U\sigma U^\dagger \in \mathcal{P}\} \tag{1}$$

Notable elements of the Clifford group are the single-qubit gates given by the Hadamard matrix $\mathrm{H} = \frac{1}{\sqrt{2}}\begin{pmatrix}1 & 1\\ 1 & -1\end{pmatrix}$ and the phase gate $\mathrm{P} = \begin{pmatrix}1 & 0\\ 0 & i\end{pmatrix}$, and the two-qubit CNOT gate given by $\mathrm{CNOT} = \begin{pmatrix}1 & 0 & 0 & 0\\ 0 & 1 & 0 & 0\\ 0 & 0 & 0 & 1\\ 0 & 0 & 1 & 0\end{pmatrix}$.

The set $\{\mathrm{H}, \mathrm{P}, \mathrm{CNOT}\}$ generates the Clifford group up to a global phase when applied to arbitrary qubits, see e.g. [23]. For all these gates, we will use subscripts to indicate the qubits or wires to which they are applied; e.g. H_j is a Hadamard gate applied to the j-th wire, and $\mathrm{CNOT}_{j,k}$ is a CNOT that has wire j as control and k as target.

Even though there exist interesting quantum circuits that use only gates from the Clifford group, it is not a universal set of gates. Indeed, the Gottesman–Knill states that such a circuit can be efficiently simulated by a classical computer, something which is not known to be true for general quantum circuits [22, 1]. By extending $\mathcal{C}$ with *any* gate, we do obtain a gate-set which is universal for quantum computation [32].

The gate we will use to extend the Clifford gates to a universal set is the T gate, sometimes called $\pi/8$-gate or R, defined by $\mathrm{T} = \begin{pmatrix}1 & 0\\ 0 & e^{i\pi/4}\end{pmatrix}$. We will write all circuits using gates from the set $\{\mathrm{X}, \mathrm{Z}, \mathrm{H}, \mathrm{P}, \mathrm{CNOT}, \mathrm{T}\}$. Technically X, P, and Z are redundant here, since they can be formed by the others as $\mathrm{P} = \mathrm{T}^2$, $\mathrm{Z} = \mathrm{P}^2$ and $\mathrm{X} = \mathrm{ZHZ}$, but we include them for convenience.

In our protocols for instantaneous non-local computation, we will alternate teleportation steps with gate operations, and therefore the interaction between the Pauli matrices and the other gates are especially important. We will make much use of the following identities, which can all be easily checked[6].

$$\begin{array}{llll} \mathrm{XZ} = \mathrm{ZX} & \mathrm{HX} = \mathrm{ZH} & \mathrm{CNOT}_{1,2}(\mathrm{X}\otimes\mathrm{I}) = (\mathrm{X}\otimes\mathrm{X})\mathrm{CNOT}_{1,2} \\ \mathrm{PZ} = \mathrm{ZP} & \mathrm{HZ} = \mathrm{XH} & \mathrm{CNOT}_{1,2}(\mathrm{I}\otimes\mathrm{X}) = (\mathrm{I}\otimes\mathrm{X})\mathrm{CNOT}_{1,2} \\ \mathrm{PX} = \mathrm{XZP} & \mathrm{TX} = \mathrm{PXT} & \mathrm{CNOT}_{1,2}(\mathrm{Z}\otimes\mathrm{I}) = (\mathrm{Z}\otimes\mathrm{I})\mathrm{CNOT}_{1,2} \\ & & \mathrm{CNOT}_{1,2}(\mathrm{I}\otimes\mathrm{Z}) = (\mathrm{Z}\otimes\mathrm{Z})\mathrm{CNOT}_{1,2} \end{array} \tag{2}$$

2.2 Key transformations from Clifford circuits

For a 0/1 vector v of length n and for any single-qubit operation U, we write $U^v = \bigotimes_{j=1}^n U^{v_j}$, i.e., U^v is the application of U on all qubits $j \in [n]$ for which $v_j = 1$. When Alice teleports a state $|\psi\rangle$ of n qubits to Bob, the uncorrected state at Bob's side can be written as $\mathrm{X}^{a_x}\mathrm{Z}^{a_z}|\psi\rangle$. Here we let a_x and a_z be the vectors representing the outcomes of the Bell measurements of Alice. In analogy with the the literature on assisted and blind quantum computation, we will call the teleportation measurement outcomes a_x and a_z the *key* needed to decode $|\psi\rangle$.

[6] Here equality is up to a global phase – which we will ignore from now on for simplicity.

The specific entries of these keys will often depend on several different measurement outcomes, given by earlier steps in the protocol, and we will therefore occasionally describe them as *polynomials* over $\mathbb{F}_2$. Viewing the keys as polynomials is especially helpful in the description of the more-complicated protocol of Section 5.

For any gate from the Clifford group $U \in \mathcal{C}$, if we apply U on the encoded state, we can describe the resulting state as $U|\psi\rangle$ with a new key. That is, $UX^{a_x}Z^{a_z}|\psi\rangle = X^{\hat{a}_x}Z^{\hat{a}_z}U|\psi\rangle$ for some new 0/1 keys $\hat{a}_x, \hat{a}_z$. The transformations of the keys will have a particularly simple form. (See for example [11] for a characterization of these transformations and a different application of Clifford circuit computation.)

For example, we can write the identities of Equation 2 in terms of key transformations. The transformations that occur when a bigger Pauli operator is applied, can then be easily found by writing the Pauli operator in terms of its generators $\{\mathrm{H}, \mathrm{P}, \mathrm{CNOT}\}$, and applying these rules one-by-one. We will write $(x_1, x_2 \mid z_1, z_2)$ as a shorthand for, respectively, the X key on the first and second qubit, and the Z key on the first and second qubit – this is a convenient notation[7] for the pair of vectors a_x and a_z that represent these keys. All addition of these keys will be over $\mathbb{F}_2$, i.e., the $+$ represents the binary exclusive or.

$$\begin{aligned} \mathrm{P}(x \mid z) &\to (x \mid x+z)\mathrm{P} \\ \mathrm{H}(x \mid z) &\to (z \mid x)\mathrm{H} \\ \mathrm{CNOT}_{1,2}(x_1, x_2 \mid z_1, z_2) &\to (x_1, x_1+x_2 \mid z_1+z_2, z_2)\mathrm{CNOT}_{1,2} \end{aligned}$$

2.3 Clifford+T quantum circuits, T-count and T-depth

In several different areas of quantum information, gates from the Clifford group are 'well-behaved' or 'easy', while the other non-Clifford gates are hard – an observation which was also made, with several examples, in the recent [10].

The *T-count* of a quantum circuit is defined as the number of T gates in the entire quantum circuit. The *T-depth* is the number of layers of T gates, when viewing the circuit as alternating between Clifford gates and a layer of simultaneous T gates. See for example Figure 5.

Given a quantum operation, it is not always obvious what is the best circuit in terms of T-count or T-depth. Recent work gave algorithms for finding circuits that are optimized in terms of T-depth [3, 21, 35, 2] and optimal constructions for arbitrary single-qubit unitaries have also been found [30, 34, 36]. These constructions sometimes increase the number of qubits involved by adding ancillas – the use of which can greatly decrease the T-depth of the resulting circuit.

2.4 The garden-hose model

The garden-hose model is a combinatorial model of communication complexity, first introduced by Buhrman, Fehr, Schaffner and Speelman [13]. The recent work by Klauck and Podder [29] further investigated the notion, proving several follow-up results. Here we repeat the basic definitions of the garden-hose model and its link to attacks on schemes for position-based quantum cryptography.

Alice has an input $x \in \{0,1\}^n$, Bob has an input $y \in \{0,1\}^n$, and the players want to compute a function $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ in the following way. Between the two

[7] This mapping is called the symplectic notation when used in the stabilizer formalism, although we won't need to introduce the associated symplectic inner product for our construction.

players are s pipes, and, in a manner depending on their respective inputs, the players link up these pipes one-to-one with hoses. Alice also has a water tap, which she can connect to one of these pipes. When $f(x, y) = 0$, the water should exit on Alice's side, and when $f(x, y) = 1$ we want the water to exit at Bob's side. The garden-hose complexity of a function f, written $GH(f)$, then is the least number s of pre-shared pipes the players need to compute the function in this manner.

There is a natural translation from strategies of the garden-hose game to a quantum protocol that routes a qubit to either Alice or Bob depending on their local inputs, up to teleportation corrections. Consider the following quantum task, again dependent on a function f like in the previous paragraph. Alice now receives a quantum state $|\psi\rangle$ and a classical input x, Bob receives input y, and the players are allowed one round of simultaneous communication. If $f(x, y) = 0$, Alice must output $|\psi\rangle$ after this round of communication, and otherwise Bob must output $|\psi\rangle$. We would like to analyze how much pre-shared entanglement the players need to perform this task.

From the garden-hose protocol for f, the players can come up with a strategy for this quantum task that needs at most $GH(f)$ EPR pairs pre-shared. Every pipe corresponds to an EPR pair. If a player's garden-hose strategy dictates a hose between some pipe j and another pipe k, then that player performs a Bell measurement of EPR-halves labeled j and k. Alice's connection of the water tap to a pipe corresponds to a Bell measurement between her input state $|\psi\rangle$ and the local half of an EPR pair. After their measurements, the correct player will hold the state $|\psi\rangle$, up to Pauli corrections incurred by the teleportations. The corrections can be performed after a step of simultaneous communication containing the outcomes of all measurements.

We will describe some of the logic in terms of the garden-hose model, as an abstraction away from the qubits involved. When we refer to a quantum implementation of a garden-hose strategy, we always mean the back-and-forth teleportation as described above.

The following lemma will prove to be useful. Let the number of *spilling pipes* of a garden-hose protocol for a player be the number of possible places the water could possibly exit. That is, the number of spilling pipes for Alice for a specific x, is the number of different places the water could exit on her side over all Bob's inputs y. The number of spilling pipes for Alice is then the maximum number of spilling pipes over all x. To be able to chain different parts of a garden-hose protocol together, it can be very convenient to only have a single spilling pipe for each player.

▶ **Lemma 1** (Lemma 11 of [29]). *A garden-hose protocol P for any function f with multiple spilling pipes can be converted to another garden-hose protocol P' for f that has only one spilling pipe on Alice's side and one spilling pipe on Bob's side. The size of P' is at most 3 times the size of P plus 1.*

Klauck and Podder also showed that computing the binary XOR of several protocols is possible with only a linear overhead in total garden-hose complexity [29, Theorem 18]. We give an explicit construction for this statement in AppendixC – the result already follows from the similar construction of [29, Lemma 12], except that we obtain a constant which is slightly better than unfolding their (more general) proof.

▶ **Lemma 2.** *Let $(f_1, f_2, \ldots, f_k)$ be functions, where each function f_i has garden-hose complexity $GH(f_i)$. Let $c \in \{0, 1\}$ be an arbitrary bit. Then,*

$$GH\left(c \oplus \bigoplus_{i=1}^{k} f_i\right) \leq 4 \sum_{i=1}^{k} GH(f_i) + 1\,.$$

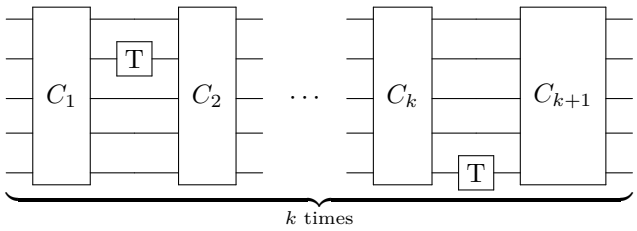

Figure 1 A circuit with T-count k. The C_i gates represent subcircuits consisting only of operation from the Clifford group $\mathcal{C}$.

3 Low T-count quantum circuits

▶ **Theorem 3.** *Let C be an n-qubit quantum circuit with gates from the Clifford+T gate set, and let C contain k T-gates in total. Then* $\mathrm{INQC}(C) \leq O(n2^k)$*, i.e., there exists a protocol for two-party instantaneous non-local computation of C which uses a pre-shared entangled state of $O(n2^k)$ EPR pairs.*

Proof. Let Alice's input state be some arbitrary quantum state $|\psi_0\rangle$. We will write the quantum state at step $t \in \{0, \ldots, k\}$, as intermediate result of executing the circuit C for t steps, as $|\psi_t\rangle$. Let C_t be the subcircuit, consisting only of Clifford gates, between the $(t-1)$th and tth T gates. At step t, the circuit alternates between the Clifford subcircuit C_t and a T-gate on some wire w_t which we write as T_{w_t}, that is, we define $T_{w_t} = \mathrm{I}^{\otimes w_t - 1} \otimes \mathrm{T} \otimes \mathrm{I}^{\otimes n - w_t - 1}$.

Because of the nature of the setting, all steps are done instantaneously unless noted otherwise, without waiting for a message of the other party. For example, if the description mentions that one party teleports a qubit, we can instantly describe the qubit as 'being on the other side', but the other party will act on the uncorrected qubit, since the communication will only happen afterwards and simultaneously.

We first give a high-level description of the protocol. Bob teleports his part of the state to Alice, who holds the entire state – up to teleportation corrections. Alice will now apply the first set of Clifford gates, followed by a single T gate. The teleportation corrections (all known to Bob) determine whether the T gate that Alice performs creates an unwanted extra P gate on the state. The extra P gate is created whenever an X correction is present, because of the relation $\mathrm{TX} = \mathrm{PXT}$. Therefore, even though Alice holds the state, only Bob knows whether the state has an extra unwanted P gate or not.

To remove the unwanted gate, Alice teleports all n qubits back to Bob, who corrects the phase gate (if present). The players then perform a garden-hose-like trick to keep the form of the key simple, at the cost of doubling the total size at each step.

Now we will give the precise description of the players' actions:

Step 0. Bob performs a Bell measurement to teleport all his $n/2$ qubits to Alice, where we write the needed X-corrections as $b^0_{x,i}$ and Z-corrections $b^0_{z,i}$, for $i = n/2+1, \ldots, n$. Now, since the qubits Alice already started with don't need a correction, we have $b^0_{x,i} = b^0_{z,i} = 0$ for $i = 1, \ldots, n/2$. Then we write b^0_x and b^0_z for the 0/1 vector containing the X corrections and Z correction respectively. The complete state is $\mathrm{X}^{b^0_x}\mathrm{Z}^{b^0_z}|\psi_0\rangle$, where all qubits are at Alice's side while Bob knows the key.

Step 1.a. Alice executes C_1 on the (uncorrected) qubits, so that the state now equals

$$C_1 \mathrm{X}^{b^0_x}\mathrm{Z}^{b^0_z}|\psi_0\rangle = \mathrm{X}^{\hat{b}^1_x}\mathrm{Z}^{\hat{b}^1_z} C_1|\psi_0\rangle\,,$$

where $(\hat{b}_x^1, \hat{b}_z^1) = f_1(b_x^0, b_z^0)$, with $f_1 : \mathbb{F}_2^n \times \mathbb{F}_2^n \to \mathbb{F}_2^n \times \mathbb{F}_2^n$ is a formula that consists of relabeling and addition over $\mathbb{F}_2$, and that is known to all parties. Bob knows all the entries of the vectors $\hat{b}_x^1$ and $\hat{b}_z^1$ that contain the new teleportation corrections.

Step 1.b. Alice executes the T gate on the correct wire $w_1 \in \{1, \ldots, n\}$ of the uncorrected qubits. Define $\mathbf{b}^1 = \hat{b}^1_{x,w_1}$, the w_1 entry of the vector $\hat{b}_x^1$. The state in Alice's possession is now

$$\mathrm{T}_{w_1} \mathrm{X}^{\hat{b}_x^1} \mathrm{Z}^{\hat{b}_z^1} C_1 |\psi_0\rangle = \mathrm{P}_{w_1}^{\mathbf{b}^1} \mathrm{X}^{\hat{b}_x^1} \mathrm{Z}^{\hat{b}_z^1} \mathrm{T}_{w_1} C_1 |\psi_0\rangle = \mathrm{P}_{w_1}^{\mathbf{b}^1} \mathrm{X}^{\hat{b}_x^1} \mathrm{Z}^{\hat{b}_z^1} |\psi_1\rangle \,.$$

That is, besides the presence of the Pauli gates, depending on the teleportation measurements, the w_1 qubit possibly has an extra phase gate that needs to be corrected before the protocol can continue.

Step 1.c. Alice teleports all qubits to Bob, with teleportation outcomes $a_x^1, a_z^1 \in \mathbb{F}_2^n$. We will define the $\mathbf{a}^1$ as the w_1 entry of a_x^1. Bob then has the state

$$\mathrm{X}^{a_x^1} \mathrm{Z}^{a_z^1} \mathrm{P}_{w_1}^{\mathbf{b}^1} \mathrm{X}^{\hat{b}_x^1} \mathrm{Z}^{\hat{b}_z^1} |\psi_1\rangle = \mathrm{P}_{w_1}^{\mathbf{b}^1} \mathrm{X}^{\hat{b}_x^1} \mathrm{Z}^{\hat{b}_z^1} \mathrm{Z}^{\mathbf{a}^1 \mathbf{b}^1} \mathrm{X}^{a_x^1} \mathrm{Z}^{a_z^1} |\psi_1\rangle \,.$$

Knowing the relevant variables from his measurement outcomes in the previous steps, Bob performs the operation $\mathrm{X}^{\hat{b}_x^1} \mathrm{Z}^{\hat{b}_z^1} (\mathrm{P}_{w_1}^{\mathbf{b}^1})^\dagger$ to transform the state to $\mathrm{Z}^{\mathbf{a}^1 \mathbf{b}^1} \mathrm{X}^{a_x^1} \mathrm{Z}^{a_z^1} |\psi_1\rangle$.

Step 1.d. For this step the players share two sets of n EPR pairs, one set labeled "$\mathbf{b}^1 = 0$", the other set labeled "$\mathbf{b}^1 = 1$". Bob teleports the state to Alice using the set corresponding to the value of $\mathbf{b}^1$, with teleportation outcomes b_x^2 and b_z^2.

Step 1.e. The set of qubits corresponding to the correct value of $\mathbf{b}^1$ are in the state

$$\mathrm{X}^{b_x^2} \mathrm{Z}^{b_x^2} \mathrm{Z}^{\mathbf{a}^1 \mathbf{b}^1} \mathrm{X}^{a_x^1} \mathrm{Z}^{a_z^1} |\psi_1\rangle \,.$$

On the set labeled "$\mathbf{b}^1 = 0$", Alice applies $\mathrm{X}^{a_x^1} \mathrm{Z}^{a_z^1}$, and on the set labeled "$\mathbf{b}^1 = 1$" Alice applies $\mathrm{X}^{a_x^1} \mathrm{Z}^{a_z^1} \mathrm{Z}^{\mathbf{a}^1}_{w_1}$, so that the state (at the correct set of qubits) equals $\mathrm{X}^{b_x^2} \mathrm{Z}^{b_z^2} |\psi_1\rangle$. We are now in almost the same situation as before the first step: Alice is in possession of a state for which Bob completely knows the needed teleportation corrections – with the difference that Alice does not know which of the two sets that is.

Steps 2. . . k. The players repeat the protocol from Step 1, but Alice performs all steps in parallel for *all* sets of states. The needed resources then double with each step: two sets for step 2, four for step 3, etc.

Step k+1, final step. When having executed this protocol for the entire circuit, Alice only teleports Bob's qubits back to him, i.e. the qubits corresponding to the last $n/2$ wires, instead of the entire state, so that in the correct groups, Alice and Bob are in possession of the state $|\psi_k\rangle$ up to simple teleportation corrections. Then, in their step of simultaneous communication, the players exchange all teleportation measurement outcomes. After receiving these measurement outcomes, the players discard the qubits that did not contain the state, and perform the Pauli corrections on the correct qubits.

The needed EPR pairs for this protocol consist of $n/2$ for Step 0. Then every set uses at most $3n$ pairs: n for the teleportation of Alice to Bob, and $2n$ for the teleportation back. The t-th step of the circuit starts with 2^{t-1} sets of parallel executions, therefore the total entanglement is upper bounded by $n/2 + \sum_{t=1}^{k} 2^{t-1} 3n \leq 3n2^k$. ◀

4 Conditional application of phase gate using garden-hose protocols

The following lemma connects the difficulty of removing an unwanted phase gate that is applied conditional on a function f, to the garden-hose complexity of f. This lemma is the

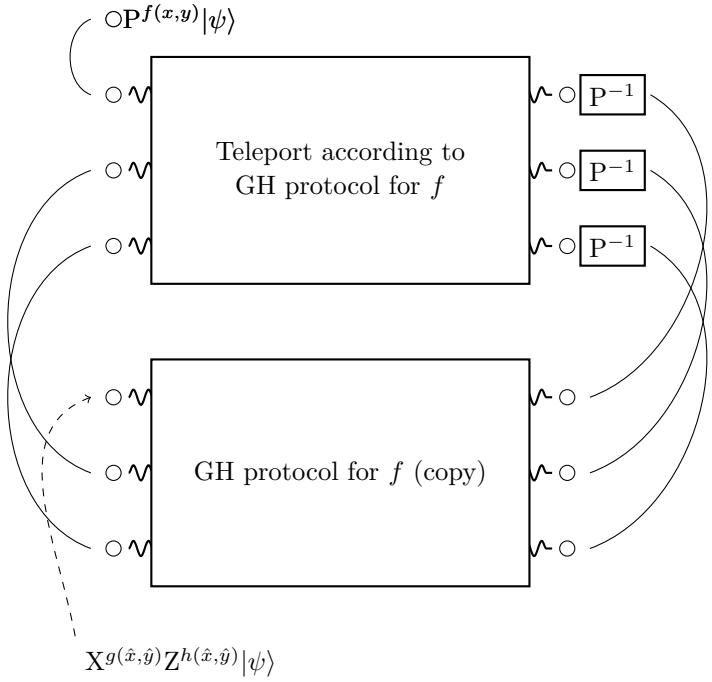

Figure 2 Schematic overview of the quantum protocol to undo the conditionally-present phase gate on $|\psi\rangle$. The solid connections correspond to Bell measurements.

main technical tool which we use to non-locally compute quantum circuits with a dependence on the T-depth.

▶ **Lemma 4.** *Assume Alice has a single qubit with state $\mathrm{P}^{f(x,y)}|\psi\rangle$, for binary strings $x, y \in \{0,1\}^n$, where Alice knows the string x and Bob knows y. Let $GH(f)$ be the garden-hose complexity of the function f. The following two statements hold:*

1. *There exists an instantaneous protocol without any communication which uses $2GH(f)$ pre-shared EPR pairs after which a known qubit of Alice is in the state $\mathrm{X}^{g(\hat{x},\hat{y})}\mathrm{Y}^{h(\hat{x},\hat{y})}|\psi\rangle$. Here $\hat{x}$ depends only on x and the $2GH(f)$ bits that describe the measurement outcomes of Alice, and $\hat{y}$ depends on y and the measurement outcomes of Bob.*
2. *The garden-hose complexities of the functions g and h are at most linear in the complexity of the function f. More precisely, $GH(g) \leq 4GH(f) + 1$ and $GH(h) \leq 11GH(f) + 2$.*

Proof. To prove the first statement we will construct a quantum protocol that uses $2GH(f)$ EPR pairs, which is able to remove the conditional phase gate. The quantum protocol uses the garden-hose protocol for f as a black box.

For the second part of the statement of the lemma, we construct garden-hose protocols which are able to compute the teleportation corrections that were incurred by executing our quantum protocol. By explicitly exhibiting these protocols, we give an upper bound to the garden-hose complexity of the X correction g and the Z correction h.

The quantum protocol is shown as Figure 2. Alice and Bob execute the garden-hose protocol with the state $\mathrm{P}^{f(x,y)}|\psi\rangle$, i.e. they teleport the state back and forth, with the EPR pairs chosen depending on x and y. Afterwards, if $f(x,y) = 0$, the qubit will be at one of the unmeasured EPR halves on Alice's side, and if $f(x,y) = 1$ the qubit will be on Bob's side. The state of the qubit will be $\mathrm{X}^{g'(x',y')}\,\mathrm{Z}^{h'(x',y')}\,\mathrm{P}^{f(x,y)}|\psi\rangle = \mathrm{P}^{f(x,y)}\mathrm{X}^{g'(x',y')}\,\mathrm{Z}^{h'(x',y') \oplus f(x,y)g'(x',y')}\,|\psi\rangle$, for some functions g' and h'.

On each qubit on Bob's side, corresponding with an 'open pipe' in the garden-hose model, Bob applies P^{-1}, so that the state of the qubit is now equal to $\mathrm{X}^{g'(x',y')}\,\mathrm{Z}^{h'(x',y')\oplus f(x,y)g'(x',y')}\,|\psi\rangle$. The exact location of our qubit depends on the protocol, and is unknown to both players. Here x' and y' are the measurement outcomes of Alice and Bob in this first half of the protocol.

To return the qubit to a known position without an extra communication step, we employ a trick that uses the reversibility of the garden-hose model. Alice and Bob repeat the exact same garden-hose strategy, except they leave the start open, and connect the open ends between the original and the copy. Alice performs a Bell measurement between the first open qubit in the original, and the first open qubit in the copy, etc. Bob does the same, after he applied the P gates. Afterwards, the qubit will be present in the start location, 'water tap' in garden-hose terminology, of the copied game, since it has followed the exact same path backwards. The final state of the qubit now is $\mathrm{X}^{g(\hat{x},\hat{y})}\,\mathrm{Z}^{h(\hat{x},\hat{y})}\,|\psi\rangle$, for some functions g and h and $\hat{x}$ and $\hat{y}$ the measurement outcomes of Alice and Bob respectively. The total entanglement consumption is $2GH(f)$.

Every measurement corresponds to a connection of two pipes in the garden-hose model, therefore each player performs at most $GH(f)$ teleportation measurements, of which the outcomes can be described by $2GH(f)$ bits.

Label the EPR pairs with numbers from $\{1, 2, \ldots, 2GH(f)\}$, and use the label 0 for the register holding the starting qubit $|\psi\rangle$. Let $\mathcal{A}$ be a list of disjoint pairs of the indices of the EPR pairs that Alice uses for teleportation in this protocol, and let $a_x, a_z \in \{0,1\}^{|\mathcal{A}|}$ be the bit strings that respectively hold the X and Z outcomes of the corresponding Bell measurements. Similarly, let $\mathcal{B}$ be a list of the indices of the EPR pairs that Bob uses, and let $b_x, b_z \in \{0,1\}^{|\mathcal{B}|}$ be the bit strings that hold the measured X and Z corrections.

To show the second part of the statement, we will construct a garden-hose protocol which tracks the newly-incurred Pauli corrections from teleporting the qubit back-and-forth, by following the qubit through the path defined by $\mathcal{A}$ and $\mathcal{B}$.

We will first construct the protocol for the final X-correction, a function we denoted by g. The protocol is also schematically shown as Figure 3. Note that to compute the X correction the conditional presence of the phase gate is not important: independent of whether $f(x, y)$ equals 1 or 0, we only need to track the X teleportation corrections that the qubit incurred by being teleported back-and-forth by Alice and Bob. An efficient garden-hose protocol for g is given by the following.

Use two pipes for each EPR pair in the protocol, $2GH(f)$ pairs of 2 pipes each. Label the top pipe of some pair i by I_i, and the bottom pipe by X_i. We will iterate over all elements of $\mathcal{A}$, i.e. all performed Bell measurements by Alice. Consider some element of $\mathcal{A}$, say the k-th pair $\mathcal{A}_k$ which consists of $\{i, j\}$. If the corresponding correction $b_{x,k}$ equals 0, we connect the pipe labeled I_i with the pipe labeled I_j and the pipe labeled X_i with the pipe labeled X_j. Otherwise, if $b_{x,k}$ equals 1, we connect them crosswise, so we connect I_i with X_j and X_i with I_j. Finally, the place where the qubit ends up after the protocol is unique (and is the only unmeasured qubit out of all $2GH(f)$ EPR pairs). For the set of open pipes corresponding to that EPR pair, say number i^*, we use one extra pipe to which we connect X_{i^*}, so that the water ends up at Bob's side for the 1-output. This garden-hose protocol computes the X correction on the qubit, and uses $4GH(f) + 1$ pipes in total, therefore $GH(g) \leq 4GH(f) + 1$.

For the Z-correction we can build a garden-hose protocol using the same idea, but there is one complication we have to take care of. At the start of the protocol, there might be an unwanted phase gate present on the state. If some teleportation is performed before this phase gate is corrected, say by Alice with outcomes a_x, a_z, then the effective correction can be

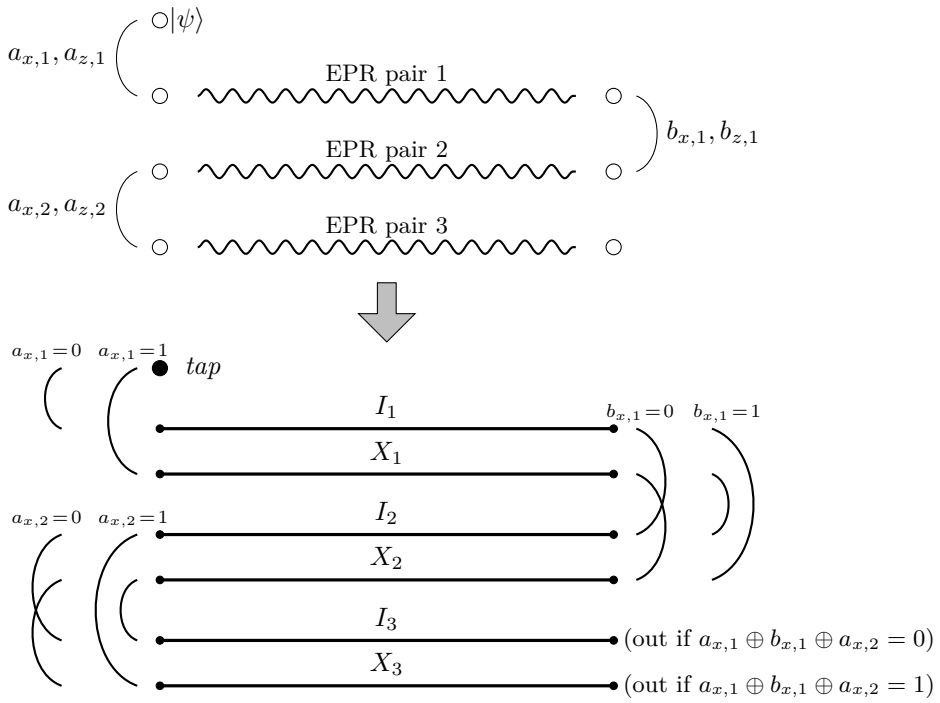

Figure 3 Example garden-hose protocol to compute the Pauli X incurred by Alice and Bob teleporting a qubit back-and-forth. When a teleportation requires a Pauli X correction, the corresponding pipes are connected crosswise, and otherwise they are connected in parallel.

written as $\mathrm{X}^{a_x}\mathrm{Z}^{a_z}\mathrm{P} = \mathrm{P}\mathrm{X}^{a_x}\mathrm{Z}^{a_x \oplus a_z}$. That is, for the part of the protocol that the unwanted phase gate is present, a Bell measurement gives a Z-correction whenever the *exclusive or* of the X- and Z-outcomes is 1, instead of just when the Z-outcome is 1. We will therefore use the garden-hose protocol that computes whether $f(x, y) = 1$, that is, compute whether the phase gate is present, and then execute a slightly different garden-hose protocol for each case.

See Figure 4 for an overview of the different parts of this garden-hose protocol for the Z-correction h. Using Lemma 1 we can transform the garden-hose protocol for f into a garden-hose protocol for f with unique 0 and 1 outputs at Alice's side, of size $3GH(f)$.[8] For the 0 output, that is if there was no unwanted phase gate present, we can track the Z corrections in exactly the same way as we did for the X corrections, for a subprotocol of size $4GH(f) + 1$. For the 1 output there was in fact a phase gate present, for the teleportations that happened in the protocol before the P^{-1} corrections. For that part of the protocol, we execute the correction-tracking protocol using the XOR of the X- and Z-measurement outcomes. For all teleportations after the phase correction, we again track the correction using just the Z-outcomes, since there is no phase gate present anymore. This part of the garden-hose protocol also uses $4GH(f) + 1$ pipes, for a total of $11GH(f) + 2$. ◀

[8] If the unique 0 output has to be at Alice's side, and the unique 1 output at Bob's side, the construction uses $3GH(f) + 1$ pipes. It is an easy exercise to show that the construction of Lemma 1 needs one pipe less if Alice wants to have both the designated 0 output and the 1 output.

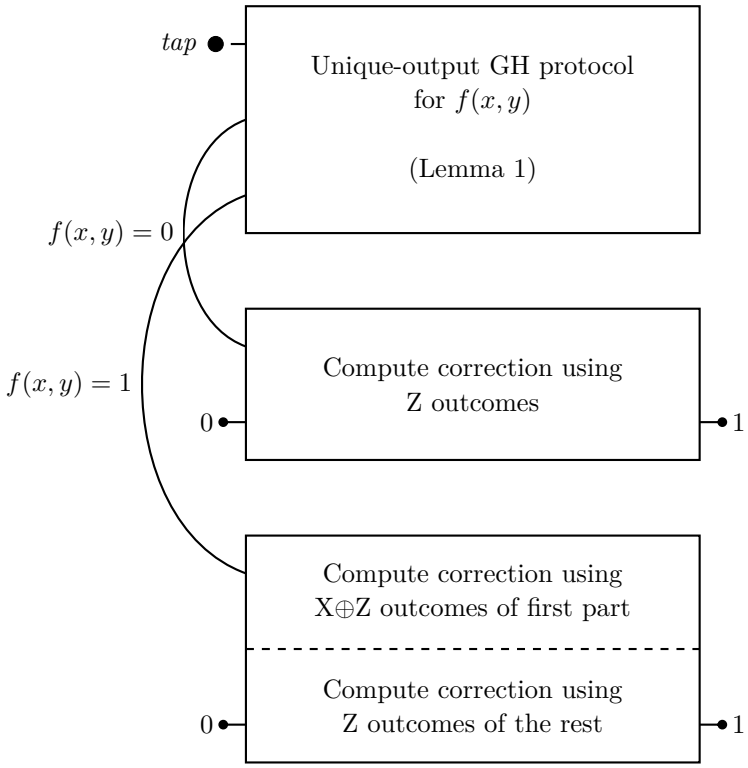

Figure 4 Sketch of garden-hose protocol for the Z correction. The bottom two boxes use the construction which was used for the X-correction; in the top case using the Z-outcomes for all measurements, in the bottom case using the parity of the X- and Z-outcomes for those teleportations that happened before the unwanted phase gate was removed.

5 Low T-depth quantum circuits

▶ **Theorem 5.** *Let C be an n-qubit quantum circuit with gates out of the Clifford+T gate set, where C has T-depth d. Then there exists a protocol for two-party instantaneous non-local computation of C, where each party receives $n/2$ qubits, which uses a pre-shared entangled state of $O((68n)^d)$ EPR pairs. That is,* $\mathrm{INQC}(C) \leq O((68n)^d)$.

Proof. As in the proof of Theorem 3, we write the input state $|\psi\rangle$, and write the correct quantum state after step t of the circuit as $|\psi_t\rangle$. At a step t, the circuit alternates between a *layer* of T gates[9] and a subcircuit consisting of only Clifford gates, C_t.

The high-level idea of this protocol is as follows. During steps 1 to t, Alice will hold the entire uncorrected state and performs a layer of the circuit: she performs a layer of T gates and then a Clifford subcircuit. The Pauli corrections at each step are a function of earlier teleportation outcomes of both Alice and Bob. These functions determine for each qubit whether that qubit now has obtained an unwanted extra P gate when Alice performs the layer of T gates. The players then, for each qubit, correct this extra gate using Lemma 4 – removing the unwanted phase gate from the qubit in a way that both players still know its location.

[9] We will assume that for each layer of T gates *all* wires have a T gate. This is only done to avoid introducing extra notation needed when instead the gates are only applied to a subset – the protocol easily generalizes to the more common general situation.

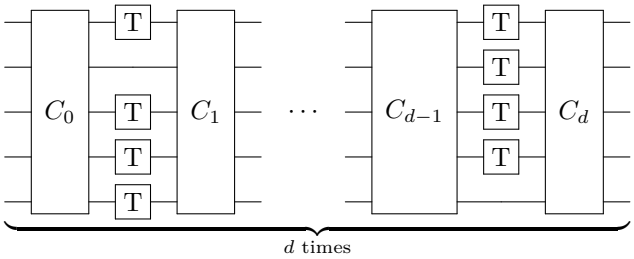

Figure 5 An example circuit with T-depth d. The C_i gates represent subcircuits consisting only of operations from the Clifford group $\mathcal{C}$. A layer does not necessarily have a T gate on all wires.

At each step we express the corrections as functions of earlier measurements and consider their garden-hose complexity, which is important when using Lemma 4. The Clifford subcircuit takes the correction functions to the XOR of several earlier functions. We can bound the growth in garden-hose complexity by taking XORs using Lemma 2. Taken together, the garden-hose complexity grows with a factor of at most a constant times n each step.

We will use $f^t_{x,i}$ to denote the function that describes the presence of an X correction on qubit i, at step t of the protocol. Similarly, $f^t_{z,i}$ is the function that describes the Z correction on qubit i at step t. Both will always be functions of outcomes of earlier teleportation measurements of Alice and Bob. For any t, let m_t be the maximum garden-hose complexity over all the key functions at step t.

Step 0. Bob teleports his qubits, the qubits labeled $n/2$ up to n, to Alice, obtaining the measurement outcomes $b^0_{x,1},\dots,b^0_{x,n/2}$ and $b^0_{z,1},\dots,b^0_{z,n/2}$. On these uncorrected qubits, Alice executes the Clifford subcircuit C_0.

Then, since Bob also knows how C_0 transforms the keys, the functions describing the Pauli corrections can all either be described by a single bit of information which is locally computable by Bob, or are constant and therefore known by both players. Let $f^0_{x,i}$ and $f^0_{z,i}$ be the resulting key function for any qubit i. The garden-hose complexity of all these key functions is constant: $GH(f^0_{x,i}) \leq 3$ and $GH(f^0_{z,i}) \leq 3$, and therefore also for the maximum garden-hose complexity we have $m_0 \leq 3$.

Step $t = 1,\dots,d$. At the start of the step, the X and Z corrections on any wire i are given by $f^{t-1}_{x,i}$ and $f^{t-1}_{z,i}$ respectively.

Alice applies the T gates on all wires. Any wire i now has an unwanted P if and only if $f^t_{x,i}$ equals 1.

Alice and Bob apply the construction of Lemma 4, which removes this unwanted phase gate. Let g^t_i be the function describing the extra X correction incurred by this protocol, so that the new X correction can be written as $f^t_{x,i} \oplus g^t_i$. Let h^t_i be the function describing the Z correction, so that the total Z correction is $f^t_{z,i} \oplus h^t_i$. The entanglement cost of this protocol is given by $2GH(f^t_{x,i})$ and the garden-hose complexities of the new functions are at most $GH(g^t_i) \leq 4GH(f^t_{x,i}) + 1$ and $GH(h^t_i) \leq 11GH(f^t_{x,i}) + 2$.

Alice now executes the Clifford subcircuit C_t. The circuit C_t determines how the current Pauli corrections, i.e. the key functions, transform. For a specification of the possible transformations, see Section 2.2. These new keys are formed by taking the exclusive OR of some subset of keys that were present in the previous step[10].

[10] This is slightly more general than necessary, since not all possible key transformations of this form are actually possible – only those transformations generated by the possibilities in Section 2.2 can occur.

Consider the worst case key for our construction: a key which is given by the XOR of all keys that were present when the Clifford subcircuit was executed. Applying Lemma 2, the worst-case key function of the form $\bigoplus_{i=1}^{n} f_{x,i}^{t-1} \oplus g_i^t \oplus f_{z,i}^{t-1} \oplus h_i^t$ has garden-hose complexity at most

$$\begin{aligned} m_t &\leq 4\left(\sum_{i=1}^{n} GH(f_{x,i}^{t-1}) + GH(g_i^t) + GH(f_{z,i}^{t-1}) + GH(h_i^t)\right) + 1 \\ &\leq 4\left(\sum_{i=1}^{n} GH(f_{x,i}^{t-1}) + 4GH(f_{x,i}^{t-1}) + 1 + GH(f_{z,i}^{t-1}) + 11GH(f_{x,i}^{t-1}) + 2\right) + 1 \\ &\leq 4\left(\sum_{i=1}^{n} m_{t-1} + 4m_{t-1} + 1 + m_{t-1} + 11m_{t-1} + 2\right) + 1 \\ &= 68nm_{t-1} + 12n + 1\,. \end{aligned} \tag{3}$$

Step $d+1$, final step. Alice teleports the last $n/2$ qubits back to Bob. Alice and Bob exchange all results of teleportation measurements and locally perform the needed corrections, using both players' measurement outcomes.

At every step t, the protocol uses at most $2nm_{t-1}$ EPR pairs for the protocol which corrects the phase gate. Using that $m_0 \leq 3$, we can write the upper bound of Equation 3 as the closed form $m_t \leq c_1(68n)^t + c_2$, with $c_1 = \frac{216n-2}{68n-1} \approx \frac{54}{17}$ and $c_2 = 3 - \frac{216n-2}{68n-1} \approx -\frac{3}{17}$. The total entanglement use therefore is bounded by $\sum_{t=1}^{d} 2nm_{t-1} \leq O(\,(68n)^d\,)$. ◀

6 The Interleaved Product protocol

Chakraborty and Leverrier [14] recently proposed a scheme for quantum position verification based on the interleaved multiplication of unitaries, the *Interleaved Product protocol*, denoted by $G_{\mathrm{IP}}(n, t, \eta_{\mathrm{err}}, \eta_{\mathrm{loss}})$. The parameter n concerns the number of qubits that are involved in the protocol in parallel, while t scales with the amount of classical information that the protocol uses. Their paper analyzed several different attacks on this scheme, which all required exponential entanglement in the parameter t. In this section, as an application of the proof strategy of Theorem 5, we present an attack on the Interleaved Product protocol which requires entanglement polynomial in t.

The original protocol is described in terms of the actions of hypothetical honest parties and also involves checking of timings at spatial locations. For simplicity, we instead only describe a two-player game, for players Alice and Bob, such that a high probability of winning this game suffices to break the scheme. Let x be a string $x \in_R \{0,1\}^n$, and let U be a random (single-qubit) unitary operation, i.e. a random element of $U(2)$. Alice receives t unitaries $(u_i)_{i=1}^t$, and Bob receives t unitaries $(v_i)_{i=1}^t$ such that $U = \prod_{i=1}^t u_i v_i$. Alice receives the state $U^{\otimes n}|x\rangle$. The players are allowed one round of simultaneous communication. To break the protocol $G_{\mathrm{IP}}(n, t, \eta_{\mathrm{err}}, \eta_{\mathrm{loss}})$, after the round of simultaneous communication the players need to output an identical string $y \in \{\emptyset, 0, 1\}^n$ such that the number of bits where y is different from x is at most $\eta_{\mathrm{err}} n$ and the number of empty results $\emptyset$ is at most $\eta_{\mathrm{loss}} n$. We will consider attacks on the strongest version of the protocol, where we take $\eta_{\mathrm{loss}} = 0$.

▶ **Theorem 6.** *There exists an attack on $G_{\mathrm{IP}}(n, t, \eta_{\mathrm{err}}, \eta_{\mathrm{loss}} = 0)$ that requires $p(t/\eta_{\mathrm{err}})$ EPR pairs per qubit of the protocol, for some polynomial p, and succeeds with high probability.*

The detailed attack is included as Appendix D.

7 Discussion

We combined ideas from the garden-hose model with techniques from quantum cryptography to find a class of quantum circuits for which instantaneous non-local computation is efficient. These constructions can be used as attacks on protocols for quantum position-verification, and could also be translated back into the settings related to physics (most notable the relation between the constraints of relativity theory and quantum measurements) and distributed computing.

The resource usage of instantaneous non-local quantum computation quantifies the non-locality present in a bi- or multi-partite quantum operation, and there is still room for new upper and lower bounds. Any such bounds will result in new insights, both in terms of position-based quantum cryptography, but also in the other mentioned settings.

Some possible approaches for continuing this line of research are as follows:

- Computing the Pauli corrections happens without error in our current construction. Perhaps introducing randomness and a small probability of error – or the usage of entanglement as given in the *quantum garden-hose model* of [13, Section 2.5] – could make this scheme more efficient.
- Future research might be able to extend this type of construction to a wider gate set or model of computation. One could think for example of a Clifford+cyclotomic gate set [20], match-gate computation [27], or measurement-based quantum computation [6, 9].
- We presented an attack on the Interleaved Product protocol which required entanglement polynomial in t. Since the exponent of this polynomial was quite large, the scheme could still be secure under realistic assumptions. Since the parameter t concerns the *classical* information that the verifiers send, requiring attackers to manipulate an amount of entanglement which scales linearly with the classical information would already make a scheme unpractical to break in practice – let alone a quadratic or cubic dependence.
- The combination of the garden-hose model with the tool set of blind quantum computation is potentially powerful in other settings. For example, following up on Broadbent and Jeffery who published constructions for quantum homomorphic encryption for circuits of low T-gate complexity [10], Dulek, Speelman, and Schaffner [17] developed a scheme for quantum homomorphic encryption, based on this combination as presented in (a preprint of) this work.

Acknowledgments. The author is supported by the EU projects SIQS and QALGO, and thanks Anne Broadbent, Harry Buhrman, Yfke Dulek and Christian Schaffner for useful discussions.

A Definition of INQC

An *instantaneous non-local quantum protocol that uses k qubits of entanglement* is a protocol of the following form.

Alice and Bob start with a fixed, chosen $2k$-qubit state $\eta_{A_eB_e} \in \mathbb{C}^{2^k} \otimes \mathbb{C}^{2^k}$, the entanglement. (Our protocols all use the special case where this state is a tensor product of k EPR pairs.) The players receive an input state $\rho \in \mathcal{S}(A_{in} \otimes B_{in})$, where $\mathcal{S}(A)$ is used for the set of density matrices on some Hilbert space A. Let A_m, A_s, B_m, A_s denote arbitrary-sized quantum registers. Alice applies some quantum operation, i.e. completely positive trace-preserving map, $\mathcal{A}_\infty : \mathcal{S}(A_{in} \otimes A_e) \to \mathcal{S}(A_m \otimes A_s)$ and Bob applies the quantum operation $\mathcal{B}_\infty : \mathcal{S}(B_{in} \otimes B_e) \to \mathcal{S}(B_m \otimes B_s)$. Alice sends the register A_s to Bob, while simultaneously Bob sends B_s to Alice.

Afterwards Alice applies the quantum operation $\mathcal{A}_\in : \mathcal{S}(A_m \otimes B_s) \to \mathcal{S}(A_{out})$ on her memory and the state she received from Bob, and outputs the result. Likewise Bob applies the operation $\mathcal{B}_\in : \mathcal{S}(B_m \otimes A_s) \to \mathcal{S}(B_{out})$ on the part of the quantum state he kept and outputs the result of this operation.

▶ **Definition 7.** Let $\Phi : \mathcal{S}(A_{in} \otimes B_{in}) \to \mathcal{S}(A_{out} \otimes B_{out})$ be a bipartite quantum operation, i.e. a completely positive trace-preserving map, for some input registers A_{in}, B_{in} and output registers A_{out}, B_{out}.

We say that $\mathrm{INQC}_\varepsilon(\Phi)$ is the smallest number k such that there exists an instantaneous non-local quantum protocol that uses k qubits of entanglement, with induced channel $\Psi : \mathcal{S}(A_{in} \otimes B_{in}) \to \mathcal{S}(A_{out} \otimes B_{out})$, so that $\|\Phi - \Psi\|_\diamond \leq \varepsilon$.

For any unitary U, we write $\mathrm{INQC}_\varepsilon(U)$ as a shorthand for $\mathrm{INQC}_\varepsilon(\Phi_U)$, where Φ_U is the induced quantum operation defined by $\rho_{AB} \to U\rho_{AB}U^\dagger$. In this chapter, we assume for simplicity that Alice's and Bob's input and output registers all consist of n qubits.

These definitions are mostly compatible with those given in [5], but differ in two ways – both are unimportant for our results in this chapter, but might be relevant for follow-up results, especially when proving lower bounds. Firstly, we made the choice for generality to allow the players to communicate using qubits, instead of just classical messages. As long as the number of communicated qubits is not too large, quantum communication could potentially be replaced by classical communication using teleportation, at the cost of extra entanglement – the counted resource. Secondly, we make the choice to explicitly separate the shared entangled state from the local memory in notation – Beigi and König split the state in a measured and unmeasured part, but do not introduce notation for (free) extra local memory in addition to the shared entangled state.

Whether these choices are reasonable or not will also depend on the exact application. Since we mostly think about applications to position-based quantum cryptography, giving the players, i.e. 'attackers', as much power as possible seems the most natural.

B The Clifford hierarchy

The Clifford hierarchy, also called the Gottesman–Chuang hierarchy, generalizes the definition of the Clifford group of Equation 1 in the following way [24]. Define $\mathcal{C}_1 = \mathcal{P}$, the first level of the hierarchy, as the Pauli group. Recursively define the k-th level as

$$\mathcal{C}_k = \{U \in U(2^n) \mid \forall \sigma \in \mathcal{P} : U\sigma U^\dagger \in \mathcal{C}_{k-1}\}\,.$$

Then $\mathcal{C}_2$ is the Clifford group and the next levels consist of increasingly more quantum operations – although for $k \geq 3$ the set $\mathcal{C}_k$ is no longer a group [41].

The method behind the protocol of Theorem 3 immediately translates to the related setting of the Clifford hierarchy. Since the dependence on n is exponential, Proposition 8 will only be a qualitative improvement over Beigi and König's port-based teleportation construction when both n and the level k are small.

The results of Chakraborty and Leverrier [14] contain a complete proof of Proposition 8, proven independently and made available earlier than (the preprint of) the current paper. We still include a proof of the statement as an illustrative application of the proof technique of Section 3.

▶ **Proposition 8.** *Let U be an n-qubit operation in the k-th level of the Clifford hierarchy, where Alice receives $n/2$ qubits and Bob receives $n/2$ qubits, then* $\mathrm{INQC}(U) \leq O(n4^{nk})$.

Proof Sketch. First Bob teleports his qubits to Alice, with n outcomes for X and Z. Alice applies U to the uncorrected state, so that now the state equals $U\mathrm{X}^{b_x}\mathrm{Z}^{b_z}|\psi\rangle = V_{b_x,b_z}U|\psi\rangle$, where V_{b_x,b_z} is an operator in the $(k-1)$-th level of the Clifford hierarchy. Exactly which operator depends on Bob's measurement outcomes b_x, b_z.

Alice teleports the entire state to Bob, with outcomes a_x, a_z, and Bob applies the inverse $V^\dagger_{b_x,b_z}$, so that the state is

$$V^\dagger_{b_x,b_z}\mathrm{X}^{a_x}\mathrm{Z}^{a_z}V_{b_x,b_z}U|\psi\rangle = W_{a_x,a_z,b_x,b_z}U|\psi\rangle\,,$$

with W_{a_x,a_z,b_x,b_z} in the $(k-2)$-th level of the Clifford hierarchy. For every possible value of b_x, b_z, the players share a set of n EPR pairs. Bob teleports the state using the set labeled with his measurement outcome b_x, b_z, obtaining teleportation corrections $\hat{b}_x, \hat{b}_z$.

For every set the players repeat this protocol recursively, in the following way. For any set, Alice repeats the protocol as if it were the set used by Bob. At the correct set, Alice effectively knows the values b_x, b_z from the label, and a_x, a_z she knows as own measurement outcomes. The state present is $\mathrm{X}^{\hat{b}_x}\mathrm{Z}^{\hat{b}_z}W_{a_x,a_z,b_x,b_z}U|\psi\rangle$. When Alice applies $W^\dagger_{a_x,a_z,b_x,b_z}$, the state is given by $F_{a_x,a_z,b_x,b_z,\hat{b}_x,\hat{b}_z}U|\psi\rangle$, with F in the $(k-3)$-th level of the Clifford hierarchy. Of this state, effectively only $\hat{b}_x, \hat{b}_z$ is unknown to Alice. Alice teleports this state to Bob using the EPR pairs labeled with a_x, a_z, and the recursive step is complete.

The players continue these steps until the first level of the hierarchy is reached – formed by Pauli operators – after which they can exchange the outcomes of their measurements to undo these and obtain $U|\psi\rangle$.

After t steps, Every teleportation step after the first uses a set of n EPR pairs, picked out of 4^n possibilities corresponding to the Pauli correction of the n qubits teleported in the previous step.

Summing over all rounds gives a total entanglement use of $n\sum_{t=1}^{k} 4^{nt} = O(n4^{nk})$. ◀

C Proof of Lemma 2: Garden-hose protocols for XOR of functions

To prove: Let $(f_1, f_2, \ldots, f_k)$ be functions, where each function f_i has garden-hose complexity $GH(f_i)$. Let $c \in \{0,1\}$ be an arbitrary bit that is 0 or 1. Then,

$$GH\left(c \oplus \bigoplus_{i=1}^{k} f_i\right) \leq 4\sum_{i=1}^{k} GH(f_i) + 1\,.$$

Proof Sketch. This statement was proven by Klauck and Podder [29, Theorem 18] in a more general form, using the following two steps: First, any garden-hose protocol can be turned into a single-output garden-hose protocol, repeated in this paper as Lemma 1, such that the new complexity is at most three times the old complexity. Then, these single-output garden-hose protocols can be used as nodes in a permutation branching program. Our current case is simply an instantiation of that proof for the particular case of the exclusive OR, together with the observation that we can combine both steps into one for this particular case.

For all functions f_i we build a gadget with two input pipes and two output pipes, such that if the water flows in at input pipe labeled $b \in \{0,1\}$, it flows out at the pipe labeled $f_i \oplus b$. See Figure 6 for an overview. We use four copies of the garden-hose protocol for f_i.

The open 0 output pipes of the protocol for f_i in copy 0-IN$_i$ are connected to the open 0 output pipes in copy 0-OUT$_i$. The designated source pipe of the original protocol for f_i in

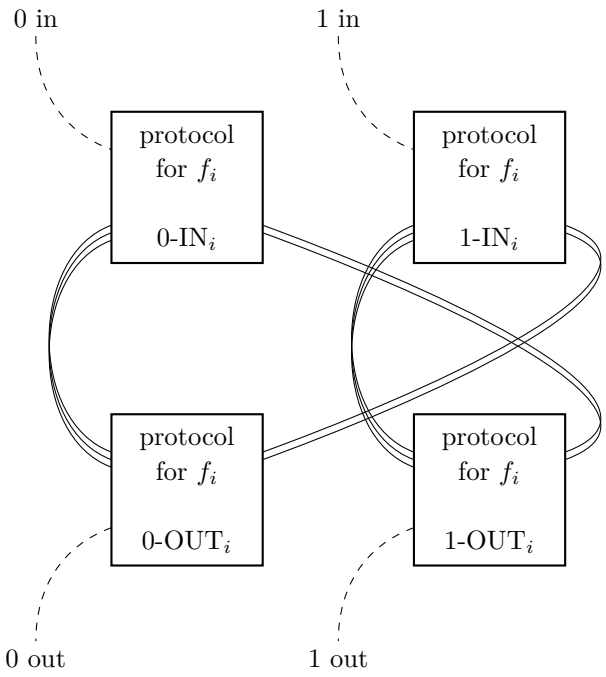

Figure 6 XOR gadget for any function f_i, total complexity $4GH(f_i)$.

copy 0-OUT$_i$ is then guaranteed to be the output.[11] We similarly connect the 1 outputs of 0-IN$_i$ to the 1 outputs of 1-OUT$_i$. This construction, i.e. before adding the 1-IN copy, is exactly the method used to create a single-output protocol. We connect the open 0 pipes of 1-IN$_i$ to the open 0 pipes of 1-OUT$_i$ and the open 1 pipes of the open 1 pipes of 1-IN$_i$ to the open 1 pipes of 0-OUT$_i$.

The gadget then works as claimed by direct inspection. Since all four copies are wired exactly the same, the path of the water through the 'OUT' copy is the reverse of the path it followed through the 'IN' copy, and therefore the water will exit correctly – at the pipe which was the source of the original protocol. ◀

D Proof of Theorem 6: attack on the Interleaved Product scheme

It was shown in [13] that polynomial garden-hose complexity is equivalent to log-space computation – up to a local preprocessing of the inputs. Instead of directly presenting garden-hose protocols, for the current construction it will be easier to argue about space-bounded algorithms and then using this equivalence as a black-box translation.

▶ **Theorem 9** (Theorem 2.12 of [13]). *If $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}$ is log-space computable, then $GH(f)$ is polynomial in n.*

Our attack will involve the computation of the unitary $U = \prod_{i=1}^{t} u_i v_i$ in the garden-hose protocol. This is a simple function, but so far we have only defined the garden-hose model for functions with a binary output. Therefore we define an extension of the garden-hose

[11] This same trick is used in the proof of Lemma 1 in [29, Lemma 11] and in our proof of Lemma 4.

model to functions with a larger output range, where instead of letting the water exit at Alice's or Bob's side, we aim to let the water exit at correctly *labeled pipe*. A short proof of the following proposition is given after the proof of the main theorem.

▶ **Proposition 10.** *Let $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^k$ be a function, such that f is log-space computable and k is at most $O(\log k)$. Then there exists a garden-hose protocol which uses a polynomial number of pipes, and such that for any input x, y the water exists at Alice's side, at a pipe labeled by the output of $f(x, y)$.*

We will also need a decomposition of arbitrary unitary operations into the Clifford+T gate set. The Solovay–Kitaev theorem is a classic result which shows that any single-qubit quantum gate can be approximated up to precision ε using $O(\log^c(1/\varepsilon))$ gates from a finite gate set, where c is approximately equal to 2. See for example [33] for an exposition of the proof. Our constructions use a very particular gate set and we are only concerned with the number of T gates instead of the total number of gates. A recent result by Selinger strengthens the Solovay–Kitaev theorem for this specific case [36][12].

▶ **Theorem 11** (Selinger 2015)**.** *Any single-qubit unitary can be approximated, up to any given error threshold $\epsilon > 0$, by a product of Clifford+T operators with T-count $11 + 12\log(1/\epsilon)$.*

With these auxiliary results in place, we can present our attack on the Interleaved Product protocol.

Proof of Theorem 6. We will describe the actions taken for any single qubit $U|b\rangle$, with $b \in \{0,1\}$, such that the probability of error is at most ε. The protocol will be attacked by performing these actions on each qubit, n times in parallel. Our construction can be divided in the following four steps. For operators A, B, let $\|A\|$ denote the operator norm, and we use $\|A - B\|$ as an associated distance measure.

1. Construct a (polynomial-sized) garden-hose protocol, with a number of pipes s, where the qubit is routed to a pipe labeled with a unitary $\tilde{U}$ which is ε_1-close to the total product U.
2. Decompose the unitaries of all labels in terms of the Clifford+T gate set, using Theorem 11. In particular, we have a Clifford+T circuit C with T-count $k = O(\log \varepsilon_2)$ such that C is ε_2-close to $\tilde{U}$, and therefore C is at most ε-close to U, where $\varepsilon = \varepsilon_1 + \varepsilon_2$.
3. After executing the garden-hose protocol as a series of teleportations, the state at pipe $\tilde{U}$ can be approximated by $X^{f_x} Z^{f_z} C|\psi\rangle$, with f_x and f_z functions of the connections Alice and Bob made in step 1 and their measurement outcomes. By the construction of Figure 3, described in the proof of Lemma 4, the garden-hose complexities $GH(f_x)$ and $GH(f_z)$ are at most linear in s.
   We can now alternate between applying a single gate of the circuit $C^\dagger$ and using Lemma 4, k times in total, to obtain a state which only has Pauli corrections left.
4. After Alice measures this final state, she can broadcast the outcome to Bob. Alice and Bob also broadcast their inputs and measurement outcomes, which together determine whether to flip the outcome of Alice's final measurement.

As the first step, we present a log-space computation solving the following problem (equivalent to the input of the protocol, with simplified notation): The input is given by t two-by-two unitary matrices, $u_1, \ldots, u_t$, and we output a matrix $\tilde{U}$ such that $\|\tilde{U} - u_t \ldots u_2 u_1\| \leq$

[12] When the single-qubit unitary is a z-rotation, an even stronger version of the theorem is available [34].

ε_1, where $\tilde{U}$ is encoded using $O(\log t + \log 1/\varepsilon_1)$ bits. We can then use a simple extension of Theorem 9 to transform this computation to a garden-hose protocol.

Store the current intermediate outcome of the product in the memory of our computation, using $2\ell + 2$ bits for each entry of the two-by-two matrix, $\ell + 1$ for the real and imaginary part each. Let M_r denote the memory of our log-space computation after r steps, obtained by computing the product $u_r M_{r-1}$ with rounding. Since the rounded matrix entry has a difference of at most $2^{-\ell}$ with the unrounded entry, we can write the precision loss at each step as $M_r = u_r M_{r-1} + \Delta_r$, where Δ_r is some matrix with all entries absolute value at most $2^{-\ell}$. Note that $\|\Delta_r\| \leq 2^{-\ell+1}$.

The total error incurred by the repeated rounding can now be upper bounded by

$$\begin{aligned}\|M_t - u_t \dots u_2 u_1\| &\leq \|u_t M_{t-1} + \Delta_t - u_t \dots u_2 u_1\| \\ &\leq \|\Delta_t\| + \|u_t(M_{t-1} - u_{t-1} \dots u_2 u_1)\| \\ &\leq 2^{-\ell+1} + \|M_{t-1} - u_{t-1} \dots u_2 u_1\| \\ &\leq t 2^{-\ell+1}\end{aligned}$$

Here we use that $\|AB\| \leq \|A\|\|B\|$ together with the unitarity of all u_i. The final step is by iteratively applying the earlier steps t times. If we choose $\ell = \log t + \log 1/\varepsilon_1 + 1$ and note that the final output $\tilde{U}$ is given by M_t, we obtain the bound.

By application of Proposition 10 we can convert this log-space computation to a garden-hose protocol, using s pipes, where s is polynomial in ε_1 and t. We then teleport the qubit back-and-forth using Bell measurements given by this garden-hose protocol.

As second step, we approximate the unitaries that label each output pipe of the garden-hose protocol of the previous step. In particular, consider the pipe labeled $\tilde{U}$, and say we approximate $\tilde{U}$ using a Clifford+T circuit C. By Theorem 11, we can write C using $k = 11 + 12\log(1/\varepsilon_2)$ T gates, such that $\|\tilde{U} - C\| \leq \varepsilon_2$. Therefore, defining $\varepsilon = \varepsilon_1 + \varepsilon_2$, we have $\|U - C\| \leq \varepsilon$.

We will perform the next steps for all unmeasured qubits (corresponding to open pipes in the garden-hose model) in parallel. After the simultaneous round of communication, Alice and Bob are then able to pick the correct qubit and ignore the others.

Consider the state of the qubit after the teleportations chosen by the garden-hose protocol. For some functions f_x, f_z, with inputs Alice's and Bob's measurement outcomes, the qubit has state $\mathrm{X}^{f_x}\mathrm{Z}^{f_z}U|b\rangle$. From now on, we will assume this state is exactly equal to $\mathrm{X}^{f_x}\mathrm{Z}^{f_z}C|b\rangle$ – since U is ε-close to C in the operator norm, this assumption adds error probability at most 2ε to the final measurement outcome[13].

Write the inverse of this circuit as alternation between gates from the Clifford group and T gates, $C^\dagger = C_k \mathrm{T} C_{k-1} \mathrm{T} \dots C_1 \mathrm{T} C_0$. We will remove C from the qubit by applying these gates, one by one, by repeated application of Lemma 4. As convenient shorthand, define the state of the qubit after applying the first r layers of $C^\dagger$, i.e. up to and including C_r, of $C^\dagger$ as

$$|\psi_r\rangle = \mathrm{T}^\dagger C_{r+1}^\dagger \mathrm{T}^\dagger C_{r+2} \dots \mathrm{T}^\dagger C_k^\dagger |b\rangle\,.$$

In particular, we have $C_r \mathrm{T}|\psi_{r-1}\rangle = |\psi_r\rangle$.

By exactly the same construction used in the proof of Lemma 4, shown in Figure 3, we observe that the garden-hose complexities of the functions f_x and f_z is at most $2s + 1$. That is, the protocol uses 2 pipes for all of the s EPR pairs, and connects them in parallel if the

[13] See for instance [33, Box 4.1] for a computation of this added error.

corresponding X- or Z-correction is 0, or crosswise if the corresponding X- or Z-correction is 1.

We will use divide f_x^r and f_z^r as the functions describing the X and Z corrections at the end of the step r. Define $m_r = \max\{GH(f_x^i), GH(f_z^i)\}$ to be the maximum garden-hose complexity out the of functions describing the X and Z corrections after step r. After Alice executes the Clifford gate C_0, the new key functions f_x^0 and f_z^0 can be written as (the NOT of) an XOR of subsets of the previous keys, e.g., one of the keys could be $f_x \oplus f_z$. By Lemma 2, we then have that our starting complexities $GH(f_x^0)$ and $GH(f_z^0)$ are at most linear in s.

Now, for any layer $r = 1, 2, \ldots, k$: Our qubit starts in the state $\mathrm{X}^{f_x^{r-1}}\mathrm{Z}^{f_z^{r-1}}|\psi_{r-1}\rangle$, for some functions f_x^{r-1}, f_z^{r-1} that each have garden-hose complexity at most m_{r-1}. After Alice performs a T gate, the qubit is in the state

$$\mathrm{T}\mathrm{X}^{f_x^{r-1}}\mathrm{Z}^{f_z^{r-1}}|\psi_{r-1}\rangle = \mathrm{P}^{f_x^{r-1}}\mathrm{X}^{f_x^{r-1}}\mathrm{Z}^{f_z^{r-1}}\mathrm{T}|\psi_{r-1}\rangle\,.$$

Now, we apply Lemma 4, costing $2GH(f_x^{r-1})$ EPR pairs, so that Alice has the state

$$\mathrm{X}^{f_x^{r-1}\oplus g_r}\mathrm{Z}^{f_z^{r-1}\oplus h_r}\mathrm{T}|\psi_{r-1}\rangle\,,$$

for some functions g_r and h_r that depend on the measurement results by Alice and Bob. We have that $GH(g_r) \leq 4GH(f_x^{r-1}) + 1$ and $GH(g_r) \leq 11GH(f_x^{r-1}) + 2$.

Now Alice applies the Clifford group gate C_r, so that the state becomes

$$C_r\mathrm{X}^{f_x^{r-1}\oplus g_r}\mathrm{Z}^{f_z^{r-1}\oplus h_r}\mathrm{T}|\psi_{r-1}\rangle = \mathrm{X}^{f_x^r}\mathrm{Z}^{f_z^r}|\psi_r\rangle\,.$$

The functions f_x^r and f_z^r can be expressed as XOR of the functions f_x^{r-1}, f_y^{r-1}, g_r, h_r. These functions have garden-hose complexity respectively at most m_{r-1}, m_{r-1}, $4m_{r-1}+1$ and $11m_{r-1}+2$. By application of Lemma 2, the exclusive OR of these functions therefore at most has garden-hose complexity $m_r \leq 4(m_{r-1}+m_{r-1}+4m_{r-1}+1+11m_{r-1}+2)+1 = 68m_{r-1}+13$.

Finally, after application of the gates in $\mathcal{C}^\dagger$, Alice has a qubit in a state which is ε-close to $\mathrm{X}^{f_x^r}\mathrm{Z}^{f_z^r}|b\rangle$. Measurement in the computational basis will produce outcome $b \oplus f_x^r$ with high probability. Besides this final measurement, Alice and Bob both broadcast all teleportation measurement outcomes in their step of simultaneous communication. From these outcomes they can each locally compute f_x^r and so derive the bit b from the outcome, which equals $b \oplus f_x^r$, breaking the protocol.

Our total entanglement usage is s for the first step, and then for each of the at most s output pipes, Alice performs the rest of the protocol. For the part of the protocol that undoes the unitary U, we use at most $2\sum_{r=0}^{k-1} m_r$ EPR pairs (for each of the at most s output pipes of the first part). We have $m_0 \leq O(s)$ and $m_r \leq m_0 \cdot 2^{O(k)}$. Since s is polynomial in t and ε_1 and $k = O(\log \varepsilon_2)$, the total protocol uses entanglement polynomial in t and ε. ◀

Our attack replaces the exponential dependence on t of the attacks presented in [14] by a polynomial dependence. For the case of $\eta_{\mathrm{err}} = 0$, we would need an error per qubit of around $\frac{\varepsilon}{n}$ to achieve total error at most ε. In that case, the entanglement required still grows as a polynomial, now with a super-linear dependence of both parameters n and t.

Only the first step of our attack, i.e. the garden-hose protocol which computes a unitary from the inputs of the players, is specific to the interleaved product protocol. This attack can therefore be seen as a blueprint for attacks on a larger class of protocols: any protocol of this same form, where the unitary operation chosen depends on a log-space computable function with classical inputs, can be attacked with entanglement which scales as a polynomial in the size of the classical inputs.

Proof of Proposition 10. We can split up the computation $f : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^k$ into k functions that each compute a bit, $f_1, \ldots, f_k$. Since f is a log-space computation, each of these functions is also a log-space computation and therefore has a polynomial-size garden-hose protocol by Theorem 9. Using Lemma 1, we can with linear overhead transform each of these protocol into a unique-output protocol, so that the water flows out at a unique pipe when the function is 0 and another unique pipe when the function is 1. Let p be a polynomial so that the single-output garden-hose protocol of each function f_i uses pipes at most $p(n)$.

First use the protocol for f_1, with output pipes labeled 0 and 1. Now each of these output pipes we feed into their own copy of f_2. The 0 output of the first copy we label 00 and its 1 output 10. Similarly, we label the 0 output of the second copy 01 and the 1 output we label 11. By recursively continuing this construction, we build a garden-hose protocol for the function f which uses s pipes, where s is at most

$$s \leq \sum_{i=1}^{k} 2^{i-1} p(n) \leq 2^k p(n) \,.$$

Since we have taken $k = O(\log n)$, this construction uses a number of pipes polynomial in n. ◀

References

1 Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004.

2 Matthew Amy, Dmitri Maslov, and Michele Mosca. Polynomial-time T-depth optimization of Clifford+T circuits via matroid partitioning. *Computer-Aided Design of Integrated Circuits and Systems, IEEE Transactions on*, 33(10):1476–1489, Oct 2014. doi:10.1109/TCAD.2014.2341953.

3 Matthew Amy, Dmitri Maslov, Michele Mosca, and Martin Roetteler. A meet-in-the-middle algorithm for fast synthesis of depth-optimal quantum circuits. *Trans. Comp.-Aided Des. Integ. Cir. Sys.*, 32(6):818–830, June 2013. doi:10.1109/TCAD.2013.2244643.

4 Pablo Arrighi and Louis Salvail. Blind quantum computation. *International Journal of Quantum Information*, 4(05):883–898, 2006.

5 Salman Beigi and Robert König. Simplified instantaneous non-local quantum computation with applications to position-based cryptography. *New Journal of Physics*, 13(9):093036, 2011.

6 HJ Briegel, DE Browne, W Dür, R Raussendorf, and M Van den Nest. Measurement-based quantum computation. *Nature Physics*, 5(1):19–26, 2009.

7 Anne Broadbent. Delegating private quantum computations. *Canadian Journal of Physics*, 93(9):941–946, 2015. doi:10.1139/cjp-2015-0030.

8 Anne Broadbent. Popescu–Rohrlich correlations imply efficient instantaneous nonlocal quantum computation. *arXiv preprint arXiv:1512.04930*, 2015.

9 Anne Broadbent, Joseph Fitzsimons, and Elham Kashefi. Universal blind quantum computation. In *Foundations of Computer Science, 2009. FOCS'09. 50th Annual IEEE Symposium on*, pages 517–526. IEEE, 2009.

10 Anne Broadbent and Stacey Jeffery. Quantum homomorphic encryption for circuits of low T-gate complexity. In Rosario Gennaro and Matthew Robshaw, editors, *Advances in Cryptology – CRYPTO 2015*, volume 9216 of *Lecture Notes in Computer Science*, pages 609–629. Springer Berlin Heidelberg, 2015. doi:10.1007/978-3-662-48000-7_30.

11 H. Buhrman, R. Cleve, M. Laurent, N. Linden, A. Schrijver, and F. Unger. New limits on fault-tolerant quantum computation. In *Foundations of Computer Science, 2006. FOCS'06. 47th Annual IEEE Symposium on*, pages 411–419, Oct 2006. doi:10.1109/FOCS.2006.50.

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

13 Harry Buhrman, Serge Fehr, Christian Schaffner, and Florian Speelman. The garden-hose model. In *Proceedings of the 4th Conference on Innovations in Theoretical Computer Science*, ITCS'13, pages 145–158, New York, NY, USA, 2013. ACM. doi:10.1145/2422436.2422455.

14 Kaushik Chakraborty and Anthony Leverrier. Practical position-based quantum cryptography. *Phys. Rev. A*, 92:052304, Nov 2015. doi:10.1103/PhysRevA.92.052304.

15 Andrew M Childs. Secure assisted quantum computation. *Quantum Information & Computation*, 5(6):456–466, 2005.

16 S R Clark, A J Connor, D Jaksch, and S Popescu. Entanglement consumption of instantaneous nonlocal quantum measurements. *New Journal of Physics*, 12(8):083034, 2010. URL: http://stacks.iop.org/1367-2630/12/i=8/a=083034.

17 Yfke Dulek, Christian Schaffner, and Florian Speelman. Quantum homomorphic encryption for polynomial-sized circuits. *arXiv preprint arXiv:1603.09717*, 2016.

18 Frédéric Dupuis, Jesper Buus Nielsen, and Louis Salvail. Secure two-party quantum evaluation of unitaries against specious adversaries. In *CRYPTO*, pages 685–706, September 2010. arXiv:1009.2096, doi:10.1007/978-3-642-14623-7_37.

19 KAG Fisher, A Broadbent, LK Shalm, Z Yan, J Lavoie, R Prevedel, T Jennewein, and KJ Resch. Quantum computing on encrypted data. *Nature communications*, 5, 2014.

20 Simon Forest, David Gosset, Vadym Kliuchnikov, and David McKinnon. Exact synthesis of single-qubit unitaries over Clifford-cyclotomic gate sets. *Journal of Mathematical Physics*, 56(8):–, 2015. doi:10.1063/1.4927100.

21 Brett Giles and Peter Selinger. Exact synthesis of multiqubit Clifford+T circuits. *Physical Review A*, 87(3):032332, 2013.

22 Daniel Gottesman. The Heisenberg representation of quantum computers. In *Group theoretical methods in physics. Proceedings, 22nd International Colloquium, Group22, ICGTMP'98, Hobart, Australia, July 13-17, 1998*, 1998. arXiv:quant-ph/9807006.

23 Daniel Gottesman. Theory of fault-tolerant quantum computation. *Phys. Rev. A*, 57:127–137, Jan 1998. doi:10.1103/PhysRevA.57.127.

24 Daniel Gottesman and Isaac L. Chuang. Quantum Teleportation is a Universal Computational Primitive. *Nature*, 402:390–393, August 1999. arXiv:9908010, doi:10.1038/46503.

25 Satoshi Ishizaka and Tohya Hiroshima. Asymptotic teleportation scheme as a universal programmable quantum processor. *Phys. Rev. Lett.*, 101(24):240501, Dec 2008. doi:10.1103/PhysRevLett.101.240501.

26 Satoshi Ishizaka and Tohya Hiroshima. Quantum teleportation scheme by selecting one of multiple output ports. *Phys. Rev. A*, 79(4):042306, Apr 2009. doi:10.1103/PhysRevA.79.042306.

27 Richard Jozsa, Barbara Kraus, Akimasa Miyake, and John Watrous. Matchgate and space-bounded quantum computations are equivalent. In *Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences*, page rspa20090433. The Royal Society, 2009.

28 Adrian Kent, William J. Munro, and Timothy P. Spiller. Quantum tagging: Authenticating location via quantum information and relativistic signaling constraints. *Phys. Rev. A*, 84:012326, Jul 2011. doi:10.1103/PhysRevA.84.012326.

29 Hartmut Klauck and Supartha Podder. New bounds for the garden-hose model. In Venkatesh Raman and S. P. Suresh, editors, *34th International Conference on Foundation of Software Technology and Theoretical Computer Science (FSTTCS 2014)*, volume 29 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 481–492, Dagstuhl, Germany, 2014. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. doi:10.4230/LIPIcs.FSTTCS.2014.481.

30 Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Fast and efficient exact synthesis of single-qubit unitaries generated by Clifford and T gates. *Quantum Info. Comput.*, 13(7-8):607–630, July 2013.

31 Hoi-Kwan Lau and Hoi-Kwong Lo. Insecurity of position-based quantum-cryptography protocols against entanglement attacks. *Phys. Rev. A*, 83(1):012322, Jan 2011. doi:10.1103/PhysRevA.83.012322.

32 Gabriele Nebe, Eric M. Rains, and Neil J. A. Sloane. The invariants of the Clifford groups. *Designs, Codes and Cryptography*, 24(1):99–122, 2001. doi:10.1023/A:1011233615437.

33 Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge university press, 2000.

34 Neil J Ross and Peter Selinger. Optimal ancilla-free Clifford+T approximation of z-rotations. *arXiv preprint arXiv:1403.2975*, 2014.

35 Peter Selinger. Quantum circuits of T-depth one. *Physical Review A*, 87(4):042302, 2013.

36 Peter Selinger. Efficient Clifford+T approximation of single-qubit operators. *Quantum Information & Computation*, 15(1-2):159–180, January 2015.

37 Florian Speelman. Position-based quantum cryptography and the garden-hose game. Master's thesis, University of Amsterdam, 2011.

38 Lev Vaidman. Instantaneous measurement of nonlocal variables. *Phys. Rev. Lett.*, 90(1):010402, Jan 2003. doi:10.1103/PhysRevLett.90.010402.

39 Li Yu. Fast controlled unitary protocols using group or quasigroup structures. *arXiv preprint arXiv:1112.0307*, 2011.

40 Li Yu, Robert B Griffiths, and Scott M Cohen. Fast protocols for local implementation of bipartite nonlocal unitaries. *Physical Review A*, 85(1):012304, 2012.

41 Bei Zeng, Xie Chen, and Isaac L Chuang. Semi-Clifford operations, structure of C_k hierarchy, and gate complexity for fault-tolerant quantum computation. *Physical Review A*, 77(4):042313, 2008.