# Studies in Models of Quantum Proof Systems

Attila Pereszlényi

National University of Singapore

2014

# Studies in Models of Quantum Proof Systems

Attila Pereszlényi
(M.Sc., BME)

A thesis submitted for the degree of
Doctor of Philosophy

Centre for Quantum Technologies
National University of Singapore

2014

## Declaration

I hereby declare that this thesis is my original work and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis.
This thesis has also not been submitted for any degree in any university previously.

Attila Pereszlényi
26th September, 2014

# Acknowledgements

First and foremost, I would like to thank my adviser Rahul Jain for giving me the opportunity to work with him and for his support and guidance. I am thankful for the freedom I had to pursue my own interests.

I would like to thank Sándor Imre for introducing me to quantum computing and for guiding me in my undergraduate projects.

I am very grateful to my previous supervisor Katalin Friedl for her guidance and for teaching me a lot about computer science. Her door was always open for friendly discussions, independently of them being academic or non-academic.

I would like to thank the PI's of our group, Hartmut Klauck, Troy Lee, and Miklos Santha, for their friendly and helpful attitude whenever I approached them with questions. I am very grateful to Miklos for his immediate help every time I faced some problem.

Life in the office would have been very different without the warm and friendly atmosphere created by post-docs and fellow students and I feel lucky that I was part of it. Because of them, doing PhD was actually fun. They also gave me invaluable help and support over the past years. Without going into specifics, I would like to express my warmest thanks to Anurag Anshu, Itai Arad, Thomas Decker, Vamsi Krishna Devabathini, Tanvirul Islam, Raghav Kulkarni, Matthew McKague, Priyanka Mukhopadhyay, Supartha Podder, Ved Prakash, Youming Qiao, Bill Rosgen, Jamie Sikora, Aarthi Sundaram, Sarvagya Upadhyay, Antonios Varvitsiotis, and Penghui Yao.

I have benefited greatly from the conversations I had with the visitors of CQT. A very partial list of them includes Peter Høyer, Iordanis Kerenidis, Anupam Prakash, Seung Woo Shin, Mario Szegedy, Thomas Vidick, and Shengyu Zhang.

I would also like to thank the administrative and IT staff of CQT for their excellent support.

Last but not least, I am very grateful to my family and to my girlfriend for

their constant love, support, and encouragement.

# Contents

**Acknowledgements** **v**

**Summary** **ix**

**Publications** **xiii**

**1 Introduction** **1**
- 1.1 Quantum Proof Systems . . . 4
  - 1.1.1 Perfect Completeness for QMA . . . 5
  - 1.1.2 Short Messages . . . 6
  - 1.1.3 Small Gap Merlin-Arthur Proof Systems . . . 7
- 1.2 Entangled Games . . . 11

**2 Preliminaries** **15**
- 2.1 Quantum Information . . . 15
  - 2.1.1 The SWAP Test . . . 22
  - 2.1.2 Choi-Jamiołkowski Representations and Post-Selection . . . 23
- 2.2 Some Complexity Classes . . . 24
- 2.3 Information Theory . . . 31

**3 Results on Quantum Merlin-Arthur Proof Systems** **35**
- 3.1 Eliminating Short Messages . . . 35
  - 3.1.1 The Idea Behind the Proof of Theorem 1.1.3 . . . 35
  - 3.1.2 The Detailed Proof . . . 36
  - 3.1.3 An Open Problem . . . 40
- 3.2 Perfect Completeness with Shared EPR Pairs . . . 40
  - 3.2.1 Some Preliminaries . . . 41
  - 3.2.2 Modified Post-Selection . . . 42
  - 3.2.3 The Idea Behind the Proof . . . 43
  - 3.2.4 The Detailed Proof . . . 46

3.3 Multi-Prover QMA with Small Gap . . . . . 57
3.3.1 QMA$[k]$ with Small Gap Equals NEXP . . . . . 58
3.3.2 BellQMA$[n^{\varepsilon}]$ with Small Gap Equals NEXP . . . . . 60
3.3.3 Conclusions and Open Problems . . . . . 64

**4 Parallel Repetition of Entangled Games** **69**
4.1 The Ideas Behind the Proof . . . . . 69
4.2 Simulating Measurements with Unitaries . . . . . 71
4.3 Proof of the Parallel Repetition Theorem . . . . . 75

**A Deferred Proofs about Small-Gap QMA** **83**
A.1 Proof of Completeness and Soundness for Lemma 3.3.3 . . . . . 83
A.1.1 Proof of Completeness . . . . . 83
A.1.2 Proof of Soundness . . . . . 84
A.2 Proof of Completeness and Soundness for Lemma 3.3.8 . . . . . 89
A.2.1 Proof of Completeness . . . . . 89
A.2.2 Proof of Soundness . . . . . 90

**Bibliography** **93**

# Summary

In this thesis, we study several problems related to quantum proof systems. The simplest quantum proof system is captured by the complexity class QMA, which stands for quantum Merlin-Arthur. Here, the prover is called Merlin and Arthur is the verifier. In QMA, a polynomial-time bounded quantum verifier has to solve a decision problem with the help of a quantum state given to him as a proof. Interestingly, it is not known whether the class retains its expressive power if we force it to have *perfect completeness*. Perfect completeness means that the verifier can only make an error in case of a no instance of the problem. Currently, the strongest result towards settling this question is by Kobayashi, Le Gall, and Nishimura [KLGN13]. They showed that any QMA protocol can be converted to a one-sided error protocol, where Arthur and Merlin initially share a constant number of EPR pairs and then Merlin sends his proof to Arthur.

- Our contribution is a conceptually simpler and more direct proof of the result of Kobayashi et al. Our protocol is similar but somewhat simpler than the original. The main contribution is a simpler and more direct analysis of the soundness property that uses well-known results in quantum information such as the quantum de Finetti theorem and properties of the trace distance and the fidelity.

Quantum interactive proof systems extend the class QMA by allowing the prover and the verifier to interact with each other. The corresponding class, QIP, is well understood and, in particular, has the same expressive power as PSPACE [JJUW11]. However, there are also several variants of QIP that are not that well understood. For example, researchers studied cases when some of the messages are short, meaning at most logarithmic in the input length [BSW11]. Our contribution to this area is the following.

- We answer one of the open problems posed by Beigi, Shor, and Watrous [BSW11]. We consider quantum interactive proof systems where, in the

beginning, the verifier and the prover send messages to each other, with the combined length of all messages being at most logarithmic (in the input length); and at the end, the prover sends a polynomial-length message to the verifier. We show that this class has the same expressive power as QMA.

An interesting consequence of the continuous nature of the quantum proofs is that it allows for arbitrary acceptance probabilities. Contrary to this, in any classical proof system the acceptance probabilities must be separated by a gap that is at least single-exponentially big. Ito, Kobayashi, and Watrous [IKW12] studied quantum classes where the gap between the completeness and soundness parameter is very small. Very small means that the gap is only lower bounded with a function that is exponentially or double-exponentially small or even smaller. Their main result is that quantum interactive proofs with double-exponentially small gap are exactly characterized by EXP, i.e., deterministic exponential time. We study multiple-proof QMA proof systems in the above setting. In multi-prover QMA, the verifier gets more than one proofs and these proofs are guaranteed to be unentangled. Our contributions are the following.

- We observe that the protocol of Blier and Tapp [BT12] scales up which implies that, in the case when the gap is exponentially or double-exponentially small, the proof system has the same expressive power as non-deterministic exponential time (NEXP). Since single-proof QMA proof systems, with the same bound on the gap, have expressive power at most exponential time (EXP), we get a separation between single and multi-prover proof systems in the 'small-gap setting' under the assumption that $\mathsf{EXP} \neq \mathsf{NEXP}$. This implies, among others, the nonexistence of certain operators called disentanglers (defined by Aaronson et al. [ABD$^+$09]) with good approximation parameters.

- We also show that the above multi-prover proof system retains its expressive power of NEXP, if we restrict the verifier to be able to perform only Bell-measurements, i.e., restricting to a BellQMA verifier. In the usual setting, when the gap is bounded by at least an inverse-polynomial function of the input length, BellQMA with polynomially-many provers is equal to single-prover QMA [BH13], but in the small-gap setting, it has the full power of multi-prover QMA. To show this, we use the protocol of Chen and Drucker [CD10] with a similar but simpler analysis. The only caveat here is that we need at least super-constant number of proofs to achieve

the desired complexity-theoretic equivalence, while in the previous setting two proofs were enough.

Non-local games can be viewed as two-prover one-round interactive proof systems where the verifier's predicate is given explicitly. In the terminology of games, the provers are called players and the verifier is called referee. The game is played as follows. Before the game starts, the players can agree on a joint strategy and can share an arbitrary entangled state. Then the referee randomly selects questions for them according to some known distribution. The players are separated during the game and are not allowed to communicate. In particular, they don't know each other's questions. After receiving the questions, they generate their answers by measuring their part of the shared entangled state. Upon receiving the answers, the referee evaluates his predicate which decides whether the players won or lost the game. The value of the game is the supremum of the achievable winning probability by such a strategy. One of the fundamental problems arising in this model is the *parallel repetition* question, which concerns with the behavior of multiple instances of the game played simultaneously. Roughly speaking, a parallel repetition theorem states that the winning probability goes down exponentially with the number of repetitions. It is known to hold in the classical case [Raz98]. This result had deep consequences in the theory of inapproximability. Similarly to the classical case, the study of the parallel repetition question in the entangled setting may have potential applications in quantum complexity theory. Although the question is still open for the general case, it was shown to hold for several classes of entangled games [CSUU08, KRT10, DSV14, CS14a]. Our contribution to this area is the following result.

- We show a parallel repetition theorem for the entangled value $\omega^*(G)$ of any two-player one-round game $G$, where the questions to the players are drawn from a product distribution. We show that for the $k$-fold repetition $G^k$ of the game $G$ (which represents the game $G$ played simultaneously $k$ times independently)

$$\omega^*\left(G^k\right) = \left(1 - (1 - \omega^*(G))^3\right)^{\Omega\left(\frac{k}{\log(|\mathscr{A}|\cdot|\mathscr{B}|)}\right)}$$

  where $\mathscr{A}$ and $\mathscr{B}$ represent the sets from which the answers of the players are drawn.

# Publications

The research during my PhD studies has resulted in the following publications. This thesis contains the materials of Refs. [1, 2, 4, 5]. Reference [3] is excluded because it's on a different topic.


[1] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A parallel repetition theorem for entangled two-player one-round games under product distributions. In *Proceedings of the 29$^{th}$ Annual IEEE Conference on Computational Complexity*, CCC '14, pages 209–216, June 2014, arXiv:1311.6309.

[2] Attila Pereszlényi. One-sided error QMA with shared EPR pairs—A simpler proof. June 2013, arXiv:1306.5406. Contributed talk at AQIS '13. To appear in *Theoretical Computer Science*.

[3] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A direct product theorem for the two-party bounded-round public-coin communication complexity. In *Proceedings of the 53$^{rd}$ Annual IEEE Symposium on Foundations of Computer Science*, FOCS '12, pages 167–176, 2012, arXiv:1201.1666.

[4] Attila Pereszlényi. On quantum interactive proofs with short messages. *Chicago Journal of Theoretical Computer Science*, 2012(9):1–10, December 2012, arXiv:1109.0964.

[5] Attila Pereszlényi. Multi-prover quantum Merlin-Arthur proof systems with small gap. May 2012, arXiv:1205.2761.

# 1 Introduction

Proof systems are central concepts in computational complexity. In their simplest form, they consist of a verifier who is a polynomial time Turing machine and a proof, a bit string, that is given to the verifier. Solving a decision problem formally means that we are given an input $x$ and we want to decide if it belongs to a language $L$. In the above proof system, the verifier gets the input together with the proof, which depends on the input, and he has to compute a binary answer which determines whether he accepts or rejects. "Accept" means that the verifier thinks that $x \in L$ while "reject" means that he thinks that $x \notin L$. There are two conditions that such a proof system must satisfy. Valid statements must be provable while invalid statements shouldn't fool the verifier. More formally, we say that if $x \in L$ then there must exist a proof with which the verifier accepts and if $x \notin L$ then he must reject all proofs. Problems solvable this way correspond to the complexity class NP [AB09]. NP is at the heart of complexity theory and has a very rich literature [GJ79].

The complexity class MA was defined by Babai [Bab85] as the natural probabilistic extension of the class NP. MA stands for Merlin-Arthur where the prover who produces the proof is referred to as Merlin and the verifier is called Arthur. Babai gave these names from an old legend where Arthur was a king of medieval England and Merlin was his magician. In MA, Merlin gives a polynomial length proof to Arthur, the same way as in NP, but now Arthur is allowed to run a polynomial time *randomized* computation. We can further generalize the above model by adding *interaction* to it, i.e., the prover and the verifier can exchange a polynomial number of messages before the verifier makes his

decision. This way we get the class IP [GMR89] where IP stands for interactive proofs.[1] We do not put any computational restriction on the prover so he is able to compute any function. The verifiers of the above proof systems are allowed to make some small error in their decision, but they must satisfy two conditions, analogously to the conditions in NP.

- If $x \in L$ then the verifier has to accept a valid proof with high probability. The probability that the verifier rejects such proof is called the *completeness* error.
- If $x \notin L$ then no matter what proof the verifier receives, he must reject with high probability. The maximum probability that the verifier accepts an invalid proof is called the *soundness* error.

We can generalize interactive proofs even further by adding more provers. In multi-prover interactive proof systems the verifier can communicate with many provers. The corresponding class is denoted by MIP. In MIP, the provers can agree on a strategy before the protocol starts but they are separated during the protocol and not allowed to talk to each other.

One of the first questions one may ask about the above proof systems is whether it is possible to get rid of one or both types of error. It is easy to see that forcing the soundness error to zero collapses MIP (and also IP and MA) to NP [AB09]. So we can't eliminate the soundness error completely, but it is known that we can make it to be at most an inverse-exponential function of the input length, without reducing the expressive power of MA, IP, or MIP. On the other hand, it was shown by Zachos and Fürer [ZF87] that having *perfect completeness*, also called *one-sided error*, doesn't change the power of MA. More formally, it holds that $\mathsf{MA} = \mathsf{MA}_1$, where $\mathsf{MA}_1$ is the class with perfect completeness. The class IP can also be made to have one-sided error, which follows, for example, from the characterization of IP being equal to PSPACE, the class of problems decidable in polynomial space [LFKN92, Sha92, She92]. We also know that MIP is equal to NEXP, the class of problems decidable in non-deterministic exponential time [BFL91]. MIP also has the power of NEXP if we restrict the number of provers to two, only allow one-sided, exponentially small error, and the interaction can only be one question to each prover and one answer from each prover [FL92]. For more information on these classes see e.g., the book of Arora and Barak [AB09].

[1] Babai also defined an interactive version of MA, that can be thought of as a 'public-coin' version of IP. Later Goldwasser and Sipser [GS86] showed that this class has the same expressive power as IP.

Another way of viewing two-prover, one-round MIP proof systems is by *non-local games*, where the acceptance predicate of the verifier is given explicitly. In the terminology of games, we call the provers "players" and the verifier "referee". If there are two players, it is customary to call them Alice and Bob. Formally, a *two-player one-round game* $G$ is specified by finite sets $\mathscr{X}$, $\mathscr{Y}$, $\mathscr{A}$, and $\mathscr{B}$, a distribution $\mu$ over $\mathscr{X} \times \mathscr{Y}$, and a predicate $V : \mathscr{X} \times \mathscr{Y} \times \mathscr{A} \times \mathscr{B} \to \{0,1\}$. It is played as follows. The referee selects questions $(x,y) \in \mathscr{X} \times \mathscr{Y}$ randomly, according to the distribution $\mu$. He sends $x$ to Alice and $y$ to Bob. As in the case of MIP, Alice and Bob are separated and not allowed to communicate. In particular, they don't know each other's questions. After receiving the questions, Alice and Bob separately choose answers $a \in \mathscr{A}$ and $b \in \mathscr{B}$ and they send them back to the referee. The referee then evaluates the predicate $V(x,y,a,b)$ and if it evaluates to 1, we say that the referee accepts or that the players win the game. If $V$ evaluates to 0 then we say that the referee rejects or the players lose the game. Note that, in this setting there is no input of which the verifier has to decide membership. Rather, the acceptance predicate is fixed and given explicitly. Here we are interested about the maximum probability with which the players can win the game. This quantity is defined as the value of the game and denoted by $\omega(G)$. More concretely, $\omega(G)$ denotes the maximum winning probability, averaged over the distribution $\mu$, where the maximum is taken over all deterministic strategies of the players.

These games played an important and pivotal role in the study of the rich theory of inapproximability, leading to the development of *Probabilistically Checkable Proofs* [ALM+98, AS98, Din07] and the famous *Unique Games Conjecture* [Kho02]. One of the most fundamental problems regarding this model is the so called *parallel repetition* question, which concerns the behavior of multiple copies of the game played in parallel. For a game $G = (\mathscr{X}, \mathscr{Y}, \mathscr{A}, \mathscr{B}, \mu, V)$, its $k$-fold product is given by $G^k = \left(\mathscr{X}^k, \mathscr{Y}^k, \mathscr{A}^k, \mathscr{B}^k, \mu^k, V^k\right)$ where $\mu^k$ denotes $k$ independent copies of $\mu$ and $V^k(x,y,a,b) = 1$ if and only if $V(x_i,y_i,a_i,b_i) = 1$ for all $i \in \{1,2,\ldots,k\}$. Simply put, Alice and Bob play $k$ copies of game $G$ in parallel and they win if and only if they win in all the copies. By playing each copy independently, it is easy to see that $\omega\left(G^k\right) \geq \omega(G)^k$ for any game $G$. The equality of the two quantities, for all games, was conjectured by Ben-Or, Goldwasser, Kilian, and Wigderson [BOGKW88] but the conjecture was shown to be false by Fortnow [For89].

However, one could still expect that $\omega\left(G^k\right)$ goes down exponentially in $k$. This is referred to as the parallel repetition, also known as the direct product, question. It was shown to be true in a seminal paper by Raz [Raz98] who

showed that

$$\omega\left(G^k\right) = \left(1 - (1 - \omega(G))^c\right)^{\Omega\left(\frac{k}{\log(|\mathscr{A}||\mathscr{B}|)}\right)}$$

where $c$ is a universal constant. This result, along with the the PCP theorem had deep consequences for the theory of inapproximability. A series of works later exhibited improved results for general and specific games [Holo9, Rao11, Raz11, BRR$^+$09, RR12].

## 1.1 Quantum Proof Systems

Quantum Merlin-Arthur proof systems, and the class QMA, were introduced by Knill [Kni96], Kitaev [KSV02], and also by Watrous [Wat00] as a natural extension of MA and NP to the quantum computational setting. In QMA, the proof of Merlin is a quantum state on polynomially many qubits. When Arthur receives the proof he performs a polynomial-time quantum computation. After the computation, he measures his dedicated output qubit, say in the standard basis, and the output of the measurement will be his decision to accept or reject. Similarly to QMA, quantum interactive proof systems, and the class QIP, were introduced by Watrous [Wat03] as a quantum analogue of IP. In QIP, the prover and the verifier can exchange quantum messages, the verifier is still a polynomial-time quantum computation, while the prover is only limited by the laws of quantum mechanics. These classes have also been well studied and now it's known that the power of quantum interactive proof systems is the same as the classical ones, i.e., QIP = IP = PSPACE [JJUW11]. Furthermore, quantum interactive proof systems still have the same expressive power if we restrict the number of messages to three and have exponentially small one-sided error [KW00].

The class QMA is not as well understood as QIP, but we do have a reasonable amount of knowledge about it. We know from the early results that it can be made to have exponentially small two-sided error [KSV02, AN02, MW05]. It also has natural complete problems, such as the '$k$-local Hamiltonian' problem [KSV02, AN02], for $k \geq 2$ [KKR06], which can be thought of as a quantum analogue of $k$-SAT. With respect to the relation of QMA to classical complexity classes, we know that MA $\subseteq$ QMA $\subseteq$ PP [MW05].[2]

[2] A slightly stronger bound of QMA $\subseteq$ $\mathsf{A_0PP}$ was shown by Vyalyi [Vya03].

### 1.1.1 Perfect Completeness for QMA

Interestingly, we don't know if $\mathsf{QMA} \stackrel{?}{=} \mathsf{QMA}_1$, i.e., whether QMA can be made to have perfect completeness. It is a long-standing open problem which was already mentioned in an early survey by Aharonov and Naveh [AN02]. Besides its inherent importance, giving a positive answer to it would immediately imply that the $\mathsf{QMA}_1$-complete problems are also complete for QMA. Most notable of these is the 'Quantum $k$-SAT' problem of Bravyi [Bra06], for $k \geq 3$ [GN13], which is considered as a more natural quantum generalization of $k$-SAT than the $k$-local Hamiltonian problem.[3] Unfortunately, all previous techniques used to show one-sided error properties of quantum interactive proof systems require adding extra messages to the protocol [KW00, KKMV08, KLGN13], so they can't be used directly in QMA. Aaronson [Aar09] gave an evidence that shows that proving $\mathsf{QMA} = \mathsf{QMA}_1$ may be difficult. He proved that there exists a quantum oracle relative to which $\mathsf{QMA} \neq \mathsf{QMA}_1$. Another difficulty with QMA, compared to MA, is that in a QMA proof system the acceptance probability can be an arbitrary irrational number. However, if certain assumptions are made about the maximum acceptance probability then QMA can be made to have one-sided error [NWZ09]. Recently, Jordan, Kobayashi, Nagaj, and Nishimura [JKNN12] showed that if Merlin's proof is classical (in which case the class is denoted by QCMA) then perfect completeness is achievable, i.e., it holds that $\mathsf{QCMA} = \mathsf{QCMA}_1$. Later we will observe that, as a side-product of one of our theorems, perfect completeness is also achievable in another, less common, variant of QMA. See Section 1.1.3 for details on this. The most recent and strongest result towards proving the original QMA versus $\mathsf{QMA}_1$ question is by Kobayashi, Le Gall, and Nishimura [KLGN13]. They showed that we can convert a QMA proof system to have one-sided error if we allow the prover and the verifier of the resulting $\mathsf{QMA}_1$ protocol to share a constant number of EPR pairs before the prover sends the proof to the verifier. The corresponding class is denoted by $\mathsf{QMA}_1^{\text{const-EPR}}$. With this notation, their result can be formalized as the following theorem.

**Theorem 1.1.1** ([KLGN13])**.** $\mathsf{QMA} \subseteq \mathsf{QMA}_1^{\text{const-EPR}}$.

Since sharing an EPR pair can be done by the verifier preparing it and sending half of it to the prover, the above result implies that QMA is contained in the class of languages provable by one-sided error, two-message quantum interactive proof systems ($\mathsf{QMA} \subseteq \mathsf{QIP}_1(2)$). This is a nontrivial upper bound.

[3] For a list of QMA- and $\mathsf{QMA}_1$-complete problems, see e.g., [Boo12].

Moreover, the result of Beigi, Shor, and Watrous [BSW11], which is described in the next section and formally stated in Theorem 2.2.14, implies that equality in Theorem 1.1.1 holds, resulting in the following characterization of QMA.

**Corollary 1.1.2** ([KLGN13]). $\mathsf{QMA} = \mathsf{QMA}_1^{\text{const-EPR}} = \mathsf{QMA}^{\text{const-EPR}}$.

Contribution

Our contribution is a conceptually simpler and more direct proof of Theorem 1.1.1, compared to the original one by Kobayashi et al. [KLGN13]. The algorithm of our verifier is also simpler, but the main difference is in the proof of its soundness. We believe that our proof helps to understand the result better and we think that it may be simplified further. The detailed description of our new proof is presented in Section 3.2.

### 1.1.2 Short Messages

Several variants of QIP and QMA have been studied in the literature. We also studied the case where some or all of the messages are short, meaning at most logarithmic in the input length. These cases are usually not interesting in the classical setting since a logarithmic-length message can be eliminated by the verifier by enumerating all possibilities. This is not true in the quantum case. Indeed, a variant of QMA that uses two unentangled, logarithmic-length proofs contains NP [BT12], hence is not believed to be equal to BQP. On the other hand, if in QMA there is only one logarithmic-length proof then it has the same expressive power as BQP [MW05].

Beigi, Shor, and Watrous [BSW11] proved that in other variants of quantum interactive proof systems short messages can also be eliminated without changing the power of the proof system. Besides other results, they showed that in the setting where the verifier sends a short message to the prover and the prover responds with an ordinary, polynomial-length message, the short message can be discarded and so the class has the same power as QMA. They have raised the question if this is also true if we replace the short question of the verifier with a 'short interaction', i.e., considering quantum interactive proof systems where, in the beginning, the verifier and the prover send messages to each other with the combined length of all messages being at most logarithmic and at the end the prover sends a polynomial-length message to the verifier.

Contribution

We show that the above class has the same power as QMA, or in other words, the short interaction can be discarded. This is formalized by the following theorem.

**Theorem 1.1.3.** *Let $c, s : \mathbb{N} \to (0,1)$ be polynomial-time computable functions such that $c(n) - s(n) \in 1/\mathsf{poly}(n)$. Then*

$$\mathsf{QIP}_{\mathrm{short}}(O(\log n), c, s) = \mathsf{QMA}.$$

Here $\mathsf{QIP}_{\mathrm{short}}(O(\log n), c, s)$ is the class described above, with completeness-soundness gap being separated by some inverse-polynomial function of the input length. For a rigorous description of the class see Definition 2.2.12. The detailed description of the proof of Theorem 1.1.3, with the underlying ideas, are presented in Section 3.1.

### 1.1.3 Small Gap Merlin-Arthur Proof Systems

Several other variants of QIP and QMA have also been studied. Ito, Kobayashi, and Watrous [IKW12] studied quantum classes where the gap between the completeness and soundness parameter is very small. Very small means that the gap is only lower bounded with a function that is exponentially or double-exponentially small or even smaller. The main result of Ito et al. [IKW12] is that quantum interactive proofs with double-exponentially small gap are exactly characterized by EXP, i.e., deterministic exponential time. This increase of power of QIP from PSPACE to EXP is a purely quantum behavior. In any classical proof system, the verifier uses at most a polynomial amount of random bits so the acceptance probabilities must be separated by a gap that is at least single-exponentially big. Moreover, classical proof systems with single-exponentially small gaps are still characterized by PSPACE. In the quantum setting, arbitrary small gaps are possible due to the continuous nature of quantum proofs. The result of Ito et al. [IKW12] shows that it also has the possibility to strengthen the power of the proof system. We studied variants of quantum Merlin-Arthur proof systems that only have such weak bounds on the gap.

Probably the most interesting generalization of QMA is by Kobayashi, Matsumoto, and Yamakami [KMY03] who defined the class $\mathsf{QMA}[k]$. In this setting there are $k$ provers who send $k$ quantum proofs to the verifier, and these proofs are guaranteed to be unentangled. In the classical setting this generalization is not interesting since we can just concatenate the $k$ proofs and treat them as one

proof. However, in the quantum case a single prover can entangle the $k$ proofs and no method is known to detect such cheating behavior.

Obviously the most important question is whether more provers make the class more powerful or not. In a later version of their paper, Kobayashi et al. [KMY03] (and independently Aaronson et al. [ABD+09]) showed that $\mathsf{QMA}[2] = \mathsf{QMA}[k]$ for all polynomially-bounded $k$ if and only if $\mathsf{QMA}[2]$ can be amplified to exponentially small error. Later Harrow and Montanaro [HM13] showed that the above equality indeed holds. The question now is whether QMA is equal to $\mathsf{QMA}[2]$, or in other words, does unentanglement actually help? There are signs that show that the above two classes are probably not equal. For example, Liu, Christandl, and Verstraete [LCV07] found a problem that has a $\mathsf{QMA}[2]$ proof system, but not known to belong to QMA. Blier and Tapp [BT12] showed that all problems in NP have a $\mathsf{QMA}[2]$ proof system where the length of both proofs are logarithmic in the input length. On the other hand, if QMA has one logarithmic-length proof then it has the same expressive power as BQP [MW05]. Since BQP is not believed to contain NP, $\mathsf{QMA}[2]$ with logarithmic length proofs is probably more powerful than QMA with a logarithmic proof. The above proof system had some inverse-polynomial gap, and this gap was later improved by several papers [Bei10, CF13, GNN12]. However, in all of these improvements the gap is still an inverse-polynomial function of the input length.[4] There is another evidence by Aaronson et al. [ABD+09] who found a $\mathsf{QMA}\left[\widetilde{O}\left(\sqrt{n}\right)\right]$ proof system for 3SAT with constant gap and where each proof consist of $O(\log n)$ qubits. Again, it seems unlikely that 3SAT has a proof system with one $\widetilde{O}(\sqrt{n})$-length proof.

We study multiple-proof QMA proof systems in the setting where the completeness-soundness gap is exponentially small or even smaller. We examine two variants of these proof systems as described below.

### $\mathsf{QMA}[k]$ with Small Gap

Contribution

The first variant we look at is the small-gap version of $\mathsf{QMA}[k]$ mentioned above. We observe that this class is exactly characterized by NEXP if the number of proofs are between 2 and polynomial and the completeness-soundness gap is exponentially or double-exponentially small. The

[4] It is not believed that the gap in this setting can be improved to a constant because it would imply that $\mathsf{QMA}[2] = \mathsf{NEXP}$. [ABD+09]

power of the proof system is still NEXP if we require it to have one-sided error. More precisely, we show the following theorem.

**Theorem 1.1.4.** *For all $\varepsilon > 0$, it holds that*

$$\mathsf{NEXP} = \mathsf{QMA}\left(2, 1, 1 - 2^{-n^{\varepsilon}}\right) = \bigcup_{\substack{0<s<c\leq 1,\\ c-s\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{QMA}(\mathsf{poly}, c, s)$$

*where $c(n)$ and $s(n)$ can be calculated in time at most exponential in $n$ on a classical computer.*

This result is discussed in details in Section 3.3.1.

In the notation above, the first parameter of QMA denotes the number of unentangled proofs the verifier receives, where each proof is at most polynomial in length. The second parameter is the completeness and the third is the soundness parameter. For a precise definition of the above notation see Definition 2.2.9. Note that, in Theorem 1.1.4 the NEXP upper bound is trivial, as it follows from exactly the same argument that shows the NEXP upper bound to the normal-gap $\mathsf{QMA}[2]$. Interestingly, there is no other upper bound known for $\mathsf{QMA}[2]$ and it is a big open question to strengthen this bound [ABD+09, AIM14]. The surprising phenomenon is that if we relax the bound on the gap, then the expressive power of the class jumps all the way up to the trivial upper bound. Note that an EXP upper bound for the small-gap, single-prover QMA is easily seen, so we have a separation between QMA and $\mathsf{QMA}[k]$ in the small-gap setting.[5]

The nontrivial part of the proof is proving the NEXP lower bound. For this, we use the protocol of Blier and Tapp [BT12] on a NEXP-complete language which we call Succinct3Col, the succinct version of graph 3-coloring. The detailed proof of Theorem 1.1.4 is presented in Section 3.3.1.

### $\mathsf{BellQMA}[k]$ with Small Gap

The class $\mathsf{BellQMA}[k]$ was defined by Aaronson et al. [ABD+09], Brandão [Bra08], and Chen and Drucker [CD10]. The above definitions are not exactly the same but the subtle difference doesn't matter in any of the above papers nor does it in this thesis. The exact definition of the class we use can be found in Section 2.2. Roughly speaking, the difference between $\mathsf{QMA}[k]$ and

[5] For more discussion about this and other consequences see Section 3.3.3.

$\mathsf{BellQMA}[k]$ is that in the latter the verifier has to measure each proof separately and non-adaptively, then based on the outcomes has to make his decision. Aaronson et al. [ABD+09] asked the question whether $\mathsf{BellQMA}[k]$ has the same power as $\mathsf{QMA}[k]$ and if there is a $\mathsf{BellQMA}$ protocol for 3SAT with similar parameters as theirs. Regarding the first question, Brandão [Brao8] showed that $\mathsf{BellQMA}[O(1)] = \mathsf{QMA}$ and later Gharibian, Sikora, and Upadhyay [GSU13] showed that $\mathsf{BellQMA}[k] = \mathsf{QMA}$ for any polynomial $k$ if we also have the promise that the possible number of outcomes of the verifier's measurements are also polynomial. Superseding both these results, Brandão and Harrow [BH13] settled this question by proving that $\mathsf{BellQMA}[k] = \mathsf{QMA}$, for all polynomially bounded $k$. A positive answer to the second question of Aaronson et al. was given by Chen and Drucker [CD10].

Here we study the small-gap version of $\mathsf{BellQMA}[k]$, where again small means exponentially or double-exponentially small. One can observe that Brandão's proof of

$$\mathsf{BellQMA}[O(1)] = \mathsf{QMA}$$

doesn't go through if the gap is so small.[6] We don't know the power of $\mathsf{BellQMA}[k]$ with constant $k$ in the small-gap setting.

Contribution

However, we show that if $k \geq n^{\varepsilon}$, for any $\varepsilon > 0$, then $\mathsf{BellQMA}[k]$ has the same power as $\mathsf{QMA}[k]$, i.e., it also equals to $\mathsf{NEXP}$. This is expressed by the following theorem.

**Theorem 1.1.5.** *For any $\varepsilon, \delta > 0$, it holds that*

$$\mathsf{NEXP} = \mathsf{BellQMA}(n^{\varepsilon}, c, s) = \bigcup_{\substack{0<s'<c'\leq 1,\\ c'-s'\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{BellQMA}\left(\mathsf{poly}, c', s'\right)$$

*for some $c$ and $s$ with $c(n) - s(n) \geq 2^{-n^{\delta}}$ and where $c'(n)$ and $s'(n)$ can be calculated in time at most exponential in $n$ on a classical computer.*

This result is discussed in Section 3.3.2.

In the above, the $\mathsf{NEXP}$ upper bound is again trivial so the only thing we need to do is give a $\mathsf{BellQMA}$ protocol for $\mathsf{NEXP}$. Just as in the case of

[6] Brandão, Christandl, and Yard [BCY11a, BCY11b] showed that a variant of multi-prover $\mathsf{QMA}$, with constant many proofs and where we require the verifier to measure the proofs with a one-way LOCC measurement, still has the same expressive power as single-prover $\mathsf{QMA}$. This proof also breaks down if the gap is small.

Theorem 1.1.4, we will use the same language (SUCCINCT3COL) and give a proof system for that. For this we will use the protocol of Chen and Drucker [CD10]. The details of the proof are presented in Section 3.3.2.

This shows an interesting phenomenon with respect to BellQMA. In the normal-gap setting, $\mathsf{BellQMA}[k] = \mathsf{QMA}[1]$ for all polynomially bounded $k$, whereas in the small-gap setting $\mathsf{BellQMA}[k] = \mathsf{QMA}[k]$! The power of $\mathsf{BellQMA}[k]$ with small gap and constant $k$ is still an open problem.

## 1.2 Entangled Games

Multi-prover interactive proofs and the class MIP was generalized to the quantum setting by Kobayashi and Matsumoto [KM03]. They defined the class QMIP, where a polynomial-time bounded quantum verifier exchanges quantum messages with polynomially many provers. The most interesting difference between QMIP and its classical counterpart is that in QMIP the provers can share an arbitrary entangled state. Indeed, if we disallow the provers to share entanglement then quantum messages won't help and the class will have the same power as $\mathsf{MIP} = \mathsf{NEXP}$ [KM03]. Given the crucial role of entanglement, researchers have studied multi-prover interactive proof systems where the verifier and all the messages are classical but the provers are allowed to share entanglement [CHTW04]. This scenario is captured by the complexity class $\mathsf{MIP}^*$ which turned out to have the exact same power as QMIP [RUV13]. The power of $\mathsf{MIP}^*$ is not well understood. We don't know any upper bound on it so currently we can't even rule out the possibility that it contains uncomputable languages. Until recently, it was also not clear that $\mathsf{MIP}^*$ is at least as powerful as MIP since provers with entanglement may potentially have more power to fool the verifier. This, however, was settled recently by the breakthrough result of Ito and Vidick [IV12] who showed that $\mathsf{MIP} \subseteq \mathsf{MIP}^*$.

One-round $\mathsf{MIP}^*$ proof systems can also be viewed as non-local games, the same way as we did earlier with MIP. The only difference now is that we allow Alice and Bob to share a quantum state before the games starts. The questions and answers in the game remain classical. After receiving the questions, Alice and Bob can generate their answers by making quantum measurements on their shared entangled state. The entangled value of game $G$ is denoted by $\omega^*(G)$. The study of entangled games is deeply related to the foundations of quantum mechanics and the understanding of quantum entanglement. These games have been used to give a novel interpretation to Bell inequalities, one of the most famous and useful methods for differentiating classical and quantum the-

ories [CHSH69]. Recently, these games were also studied from cryptographic motivations, such as in Refs. [HR10, TFKW13, MPA11]. Analogously to the classical case, the study of the parallel repetition question in this setting may potentially have applications in quantum complexity theory.

The parallel repetition conjecture was shown to hold for several sub-classes of entangled games. Cleve, Slofstra, Unger, and Upadhyay [CSUU08] showed that perfect parallel repetition holds for XOR games, which means that for these games $\omega^*\left(G^k\right) = \omega^*(G)^k$. This follows from a characterization of these games using semidefinite programming. In XOR games, the answers of the players are single bits and the referee only uses the XOR of these bits in his predicate. Later, Kempe, Regev, and Toner [KRT10] used semidefinite programming to approximate the value of the more general class of unique games and as a consequence they showed a parallel repetition theorem for these games. In unique games, for each pair of questions there is some permutation and the verifier accepts if and only if the answer of the first player is mapped to the answer of the second player with this permutation. Before Raz's result for classical games [Raz98], Feige and Kilian [FK00] showed that the classical value decreases polynomially with the number of repetitions for projection games. They used a modified parallel repetition procedure in which a fraction of the repetitions were made of "confuse rounds". In projection games, if we fix the questions for the players and the answer of the first player then there is at most one possible answer for the second player with which the referee accepts. Projection games are hence more general than unique games and include most of the interesting games. Kempe and Vidick [KV11] extended the framework of Feige and Kilian [FK00] to entangled games and got polynomial decay for projection games. They also used a modified parallel repetition procedure. However, if the questions in the game were drawn independently then no modification was required so the polynomial decay applied to the standard parallel repetition. They also showed polynomial decay to almost all games by further modifying the repetition procedure. Recently, Dinur and Steurer [DS14] introduced an analytical framework to show parallel repetition with exponential decay for the classical value of projection games. This framework was extended to the entangled case by Dinur, Steurer, and Vidick [DSV14] to establish parallel repetition for the entangled value. In a recent work, Chailloux and Scarpa [CS14a] showed exponential decay in $\omega^*\left(G^k\right)$ using information theoretic arguments. Their result is closely related to ours so we present their theorem below.

**Theorem 1.2.1** ([CS14a]). *For any game $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \mu, V)$, where $\mu$ is the*

*uniform distribution on $\mathscr{X} \times \mathscr{Y}$, it holds that*

$$\omega^*\left(G^k\right) = \left(1 - (1 - \omega^*(G))^2\right)^{\Omega\left(\frac{k}{\log(|\mathscr{A}||\mathscr{B}||\mathscr{X}||\mathscr{Y}|)}\right)}.$$

*As a corollary, for a general distribution $\mu$,*

$$\omega^*\left(G^k\right) = \left(1 - (1 - \omega^*(G))^2\right)^{\Omega\left(\frac{k}{Q^4 \log(Q) \log(|\mathscr{A}||\mathscr{B}|)}\right)}$$

*where*

$$Q = \max\left\{\left\lceil \frac{1}{\min\left\{\sqrt{\mu(x,y)} \,:\, x,y \text{ for which } \mu(x,y) \neq 0\right\}}\right\rceil, |\mathscr{X}| \cdot |\mathscr{Y}|\right\}.$$

Here $\omega^*\left(G^k\right)$ depends also on $|\mathscr{X}| \cdot |\mathscr{Y}|$ and not just on $|\mathscr{A}| \cdot |\mathscr{B}|$. Also, the value of $Q$ can be very large, depending on the distribution $\mu$.

**Contribution**

We consider the case when the questions to the players are drawn independently or, in other words, the distribution $\mu$ is product across $\mathscr{X} \times \mathscr{Y}$. Formally, there are distributions $\mu_X$ on $\mathscr{X}$ and $\mu_Y$ on $\mathscr{Y}$ such that for all $(x,y) \in \mathscr{X} \times \mathscr{Y}$ it holds that $\mu(x,y) = \mu_X(x) \cdot \mu_Y(y)$. Our result is formalized by the following theorem.

**Theorem 1.2.2.** *For any game $G = (\mathscr{X}, \mathscr{Y}, \mathscr{A}, \mathscr{B}, \mu, V)$, where $\mu$ is a product distribution on $\mathscr{X} \times \mathscr{Y}$, it holds that*

$$\omega^*\left(G^k\right) = \left(1 - (1 - \omega^*(G))^3\right)^{\Omega\left(\frac{k}{\log(|\mathscr{A}| \cdot |\mathscr{B}|)}\right)}.$$

The proof of Theorem 1.2.2, with the underlying ideas, are presented in Chapter 4.

Note that, the uniform distribution is a product distribution and our result has no dependence on the size of $\mathscr{X}$ and $\mathscr{Y}$. Hence, our result implies and strengthens the result of Chailloux and Scarpa [CS14a], up to the exponent of $1 - \omega^*(G)$.[7]

[7] Recent works in Refs. [CS14b, CWY14] have superseded both our result and the result of [CS14a] in terms of the dependence on the parameters.

# 2
# Preliminaries

The purpose of this chapter is to present the notations and background information (definitions, theorems) required to understand the results of this thesis. We start with some general notations. In this document, we denote the imaginary unit by $\imath$ instead of $i$, which we use as an index in summations, for example. We denote the set of positive functions of $n$ that are upper bounded by some polynomial in $n$ by $\mathsf{poly}(n)$. If the argument is clear, we omit it and just write $\mathsf{poly}$. For a positive integer $n \in \mathbb{Z}^+$, we sometimes use $[n]$ to represent the set $\{1, 2, \ldots, n\}$. Generally, we use Ralph Smith's script font to denote finite sets, such as $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, etc. For set $\mathscr{X}$ and $k \in \mathbb{Z}^+$, $\mathscr{X}^k$ denotes the set $\mathscr{X} \times \cdots \times \mathscr{X}$, the cross product of $\mathscr{X}$, $k$ times.

## 2.1 Quantum Information

In this thesis, we only deal with complex Euclidean spaces that are finite dimensional, which we will also simply call as Hilbert spaces. We generally try to follow the notations used in [Wato8b]. An $N$-dimensional Hilbert space is denoted by $\mathbb{C}^N$. We use Hermann Zapf's Euler script symbols to denote complex Euclidean spaces, such as $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$, etc. For vectors in Hilbert spaces, we use lowercase Greek letters and Dirac's bra–ket notation. For example, we denote a vector in a Hilbert space $\mathcal{H}$ by $|\varphi\rangle \in \mathcal{H}$. A qubit is an object that has associated Hilbert space $\mathbb{C}^2$. In our terminology, quantum registers are collections of qubits which we denote by uppercase sans serif letters, such as $\mathsf{A}$, $\mathsf{B}$, $\mathsf{C}$, etc. When we talk about a quantum register $\mathsf{H}$ of size $k$, we mean the

object made up of $k$ qubits. It has associated Hilbert space $\mathcal{H} = \mathbb{C}^{2^k}$. We always assume that some standard basis of $\mathcal{H} = \mathbb{C}^{2^k}$ have been fixed and we index those basis vectors by bit strings of length $k$. So the standard basis of $\mathcal{H}$ is denoted by $\left\{|s\rangle \,:\, s \in \{0,1\}^k\right\}$. We denote the all zero string by $\bar{0} \stackrel{\text{def}}{=} 00\ldots0$.

We denote the space of all linear mappings from $\mathcal{H}$ to itself by $\mathrm{L}(\mathcal{H})$. Linear operators are usually denoted by uppercase bold letters, such as **A**, **B**, **C**, etc. For operators **A** and **B**, $\mathbf{A} \otimes \mathbf{B}$ denotes the tensor product, also known as the Kronecker product, of **A** and **B**. The adjoint of $\mathbf{A} \in \mathrm{L}(\mathcal{H})$ is denoted by $\mathbf{A}^*$ and the adjoint of $|\varphi\rangle \in \mathcal{H}$ is denoted by $\langle\varphi| \stackrel{\text{def}}{=} (|\varphi\rangle)^*$. We denote the identity operator on some Hilbert space $\mathcal{H}$ by $\mathbb{1}_{\mathcal{H}}$ and we sometimes omit the subscript if it is clear from the context. We also use some well-known unitary operators (also called quantum gates), such as the controlled-NOT (**CNOT**) gate, the Hadamard gate (**H**), the $\pi/8$ gate (**T**), and the Pauli operators (**X**, **Z**, **Y**). The definition of these operators can be found in any standard quantum textbook, for example in [NCoo].

Density operators, also called as quantum states, are denoted by lowercase Greek letters, such as $\rho$, $\sigma$, $\tau$, etc. The set of all density operators on $\mathcal{H}$ is denoted by $\mathrm{D}(\mathcal{H})$. Formally,

$$\mathrm{D}(\mathcal{H}) \stackrel{\text{def}}{=} \{\rho \,:\, \rho \in \mathrm{L}(\mathcal{H})\,,\ \rho \geq 0,\ \mathrm{Tr}(\rho) = 1\}$$

where $\rho \geq 0$ means that $\rho$ is positive-semidefinite and the trace function is defined as

$$\mathrm{Tr}(\mathbf{A}) \stackrel{\text{def}}{=} \sum_i \langle i|\, \mathbf{A}\, |i\rangle$$

for $\mathbf{A} \in \mathrm{L}(\mathcal{H})$ and where the summation is taken over an arbitrary orthonormal basis of $\mathcal{H}$. A density operator, or quantum state, $\rho \in \mathrm{D}(\mathcal{H})$ is called pure if $\rho = |\varphi\rangle\langle\varphi|$ for some $|\varphi\rangle \in \mathcal{H}$. We will use the following quantum states often so it is convenient to introduce notations for them. Let

$$|\phi^+\rangle \stackrel{\text{def}}{=} \frac{|0\rangle + |1\rangle}{\sqrt{2}}, \qquad |\phi^-\rangle \stackrel{\text{def}}{=} \frac{|0\rangle - |1\rangle}{\sqrt{2}}, \qquad |\phi^+\rangle\,, |\phi^-\rangle \in \mathbb{C}^2.$$

Note that $|\phi^+\rangle$ and $|\phi^-\rangle$ can be obtained by applying **H** on $|0\rangle$ and $|1\rangle$. They are often denoted simply by $|+\rangle$ and $|-\rangle$ in the literature. We will also use the Bell basis.

**Definition 2.1.1.** The following states form a basis of $\mathbb{C}^4$ and are called the *Bell*

*basis*.

$$\left|\Phi^+\right\rangle \stackrel{\text{def}}{=} \frac{|00\rangle + |11\rangle}{\sqrt{2}}, \qquad \left|\Phi^-\right\rangle \stackrel{\text{def}}{=} \frac{|00\rangle - |11\rangle}{\sqrt{2}},$$
$$\left|\Psi^+\right\rangle \stackrel{\text{def}}{=} \frac{|01\rangle + |10\rangle}{\sqrt{2}}, \qquad \left|\Psi^-\right\rangle \stackrel{\text{def}}{=} \frac{|01\rangle - |10\rangle}{\sqrt{2}}.$$

We will often call the state $\left|\Phi^+\right\rangle$ as the *EPR pair*.[1] The *Euclidean norm*, or *length*, of a vector $|\varphi\rangle \in \mathcal{H}$ is defined as

$$\|\varphi\| \stackrel{\text{def}}{=} \sqrt{\langle\varphi|\varphi\rangle}.$$

For operators we will need two different norms.

**Definition 2.1.2.** The *trace norm* of $\mathbf{A} \in \mathrm{L}(\mathcal{H})$ is defined by

$$\|\mathbf{A}\|_{\mathrm{Tr}} \stackrel{\text{def}}{=} \mathrm{Tr}\left(\sqrt{\mathbf{A}^*\mathbf{A}}\right)$$

and the *operator norm* of $\mathbf{A}$ is

$$\|\mathbf{A}\|_{\infty} \stackrel{\text{def}}{=} \max\left\{\|\mathbf{A}\,|\varphi\rangle\| \,:\, |\varphi\rangle \in \mathcal{H},\ \|\varphi\| = 1\right\}.$$

The following inequality is a special case of the Hölder inequality for Schatten norms, which is a generalization of the Cauchy-Schwarz inequality for operators. For more information, see e.g., [Wato8b].

**Lemma 2.1.3.** *For any Hilbert space $\mathcal{H}$ and operators $\mathbf{A}, \mathbf{B} \in \mathrm{L}(\mathcal{H})$, it holds that*

$$|\mathrm{Tr}(\mathbf{B}^*\mathbf{A})| \leq \|\mathbf{A}\|_{\mathrm{Tr}} \cdot \|\mathbf{B}\|_{\infty}.$$

A quantum channel or super-operator ($\Phi$) is a completely positive and trace-preserving linear map of the form $\Phi : \mathrm{L}(\mathcal{Q}) \to \mathrm{L}(\mathcal{R})$. The set of all such channels is denoted by $\mathrm{C}(\mathcal{Q}, \mathcal{R})$. The trace norm of a super-operator $\Phi \in \mathrm{C}(\mathcal{Q}, \mathcal{R})$ is defined as

$$\|\Phi\|_{\mathrm{Tr}} \stackrel{\text{def}}{=} \max\left\{\|\Phi(\mathbf{X})\|_{\mathrm{Tr}} \,:\, \mathbf{X} \in \mathrm{L}(\mathcal{Q}),\ \|\mathbf{X}\|_{\mathrm{Tr}} \leq 1\right\}$$

and the *diamond norm* of $\Phi$ is

$$\|\Phi\|_{\diamond} \stackrel{\text{def}}{=} \left\|\Phi \otimes \mathbb{1}_{\mathrm{L}(\mathcal{Q})}\right\|_{\mathrm{Tr}}$$

[1] The EPR pair was named after Albert Einstein, Boris Podolsky, and Nathan Rosen. In their famous paper [EPR35], they wanted to demonstrate that quantum mechanics is incomplete by showing one of its consequences that looked unreasonable at that time.

where $\mathbb{1}_{\mathrm{L}(\mathcal{Q})}$ is the identity super-operator on $\mathrm{L}(\mathcal{Q})$. More on these norms can be found in [Wato8b].

The following definition is used to quantify the distance between operators.

**Definition 2.1.4.** The *trace distance* between operators $\mathbf{A}, \mathbf{B} \in \mathrm{L}(\mathcal{H})$ is defined as

$$d(\mathbf{A}, \mathbf{B}) \stackrel{\text{def}}{=} \frac{\|\mathbf{A} - \mathbf{B}\|_{\mathrm{Tr}}}{2}.$$

If the operators represent pure quantum states, i.e., $\mathbf{A} = |\varphi\rangle\langle\varphi|$ and $\mathbf{B} = |\psi\rangle\langle\psi|$, for some $|\varphi\rangle, |\psi\rangle \in \mathcal{H}$, for which $\|\varphi\| = \|\psi\| = 1$, then the trace distance can be more conveniently written as

$$d(|\varphi\rangle, |\psi\rangle) = \sqrt{1 - |\langle\varphi|\psi\rangle|^2}. \tag{2.1}$$

Let $\rho \in \mathrm{D}(\mathcal{H})$ be a density operator and $\mathcal{X}$ be a Hilbert space. We say that $|\varphi\rangle \in \mathcal{X} \otimes \mathcal{H}$ is a *purification* of $\rho$ if $\mathrm{Tr}_{\mathcal{X}}(|\varphi\rangle\langle\varphi|) = \rho$. The *partial trace* is defined as

$$\mathrm{Tr}_{\mathcal{X}}(\mathbf{A}) \stackrel{\text{def}}{=} \sum_{i=0}^{\dim(\mathcal{X})-1} (\langle i| \otimes \mathbb{1}_{\mathcal{H}})\, \mathbf{A}\, (|i\rangle \otimes \mathbb{1}_{\mathcal{H}})$$

for any $\mathbf{A} \in \mathrm{L}(\mathcal{X} \otimes \mathcal{H})$. Sometimes we want to emphasize that some state $\rho$ is a state of some registers $(\mathsf{A}, \mathsf{B})$. We then denote the state by $\rho^{\mathsf{AB}}$, i.e., by putting the registers in the superscript. Following this notation, the state $\rho^{\mathsf{B}}$ denotes the marginal of $\rho^{\mathsf{AB}}$ on $\mathsf{B}$. Formally, $\rho^{\mathsf{B}} \stackrel{\text{def}}{=} \mathrm{Tr}_{\mathcal{A}}(\rho^{\mathsf{AB}})$.

**Theorem 2.1.5** (Unitary equivalence of purifications)**.** *Let $\rho \in \mathrm{D}(\mathcal{H})$ and $\mathcal{X}$ be a Hilbert space. If $|\varphi\rangle \in \mathcal{X} \otimes \mathcal{H}$ and $|\psi\rangle \in \mathcal{X} \otimes \mathcal{H}$ are both purifications of $\rho$ then there exists a unitary operator $\mathbf{U} \in \mathrm{L}(\mathcal{X})$ such that*

$$(\mathbf{U} \otimes \mathbb{1}_{\mathcal{H}})\, |\varphi\rangle = |\psi\rangle\,.$$

Another way of quantifying the similarity between density operators is by the fidelity defined below.

**Definition 2.1.6.** The *fidelity* between $\rho, \sigma \in \mathrm{D}(\mathcal{H})$ is defined as

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

If $\rho = |\varphi\rangle\langle\varphi|$ then the fidelity can be more conveniently written as

$$\mathrm{F}(|\varphi\rangle\langle\varphi|, \sigma) = \sqrt{\langle\varphi|\, \sigma\, |\varphi\rangle}. \tag{2.2}$$

The following alternate characterization of the fidelity will be useful later.

**Theorem 2.1.7** (Uhlmann's Theorem, see e.g., [Wato8b] for a proof)**.** *Let* $\rho, \sigma \in \mathrm{D}(\mathcal{H})$ *and* $\mathcal{X}$ *be a Hilbert space such that* $\dim(\mathcal{X}) \geq \dim(\mathcal{H})$*. Let* $|\varphi\rangle \in \mathcal{X} \otimes \mathcal{H}$ *be any purification of* $\rho$*, i.e.,* $\mathrm{Tr}_{\mathcal{X}}(|\varphi\rangle\langle\varphi|) = \rho$*. Then*

$$\mathrm{F}(\rho, \sigma) = \max\left\{|\langle\varphi|\psi\rangle| \,:\, |\psi\rangle \in \mathcal{X} \otimes \mathcal{H},\ \mathrm{Tr}_{\mathcal{X}}(|\psi\rangle\langle\psi|) = \sigma\right\}.$$

We now list some properties of the trace distance.

**Lemma 2.1.8** (triangle inequality)**.** *For any* $\mathbf{A}, \mathbf{B}, \mathbf{C} \in \mathrm{L}(\mathcal{H})$*, it holds that*

$$d(\mathbf{A}, \mathbf{B}) \leq d(\mathbf{A}, \mathbf{C}) + d(\mathbf{C}, \mathbf{B}).$$

The following theorem states that super-operators can't increase the trace distance.

**Theorem 2.1.9** (Theorem 9.2 from [NCoo])**.** *Let* $\Phi \in \mathrm{C}(\mathcal{H}, \mathcal{K})$ *be a quantum super-operator and let* $\rho, \sigma \in \mathrm{D}(\mathcal{H})$*. Then*

$$d(\Phi(\rho), \Phi(\sigma)) \leq d(\rho, \sigma).$$

**Lemma 2.1.10.** *Let* $\mathbf{A}, \mathbf{B} \in \mathrm{L}(\mathcal{H})$*. If* $0 \leq \mathbf{B}$ *and* $\mathrm{Tr}(\mathbf{B}) \leq \varepsilon$*, for some* $0 \leq \varepsilon$*, then*

$$d(\mathbf{A} + \mathbf{B}, \mathbf{A}) \leq \frac{\varepsilon}{2}.$$

*Proof.* From the definition of the trace norm and the trace distance, together with the fact that $\sqrt{\mathbf{B}^*\mathbf{B}} = \mathbf{B}$, we get that

$$\begin{aligned} d(\mathbf{A} + \mathbf{B}, \mathbf{A}) &= \frac{\|\mathbf{A} + \mathbf{B} - \mathbf{A}\|_{\mathrm{Tr}}}{2} \\ &= \frac{\|\mathbf{B}\|_{\mathrm{Tr}}}{2} \\ &= \frac{\mathrm{Tr}(\mathbf{B})}{2} \\ &\leq \frac{\varepsilon}{2}. \end{aligned}$$

❑

**Lemma 2.1.11.** *Let* $\rho, \sigma \in \mathrm{D}(\mathcal{H})$ *and* $0 \leq \varepsilon < 1$*. It holds that*

$$d((1-\varepsilon)\rho + \varepsilon\sigma, \rho) \leq \varepsilon.$$

*Proof.* Using the triangle inequality (Lemma 2.1.8) and Lemma 2.1.10, we get

that

$$\begin{aligned} d((1-\varepsilon)\rho + \varepsilon\sigma, \rho) &\leq d((1-\varepsilon)\rho + \varepsilon\sigma, (1-\varepsilon)\rho) + d(\rho, (1-\varepsilon)\rho) \\ &\leq \frac{\varepsilon}{2} + \frac{\|\rho - (1-\varepsilon)\rho\|_{\mathrm{Tr}}}{2} \\ &= \frac{\varepsilon}{2} + \frac{\mathrm{Tr}(\varepsilon\rho)}{2} \\ &= \varepsilon. \end{aligned}$$

❑

The following lemma will be used to quantify how much a projective measurement changes a state. It is a variant of Winter's gentle measurement lemma [Win99].

**Lemma 2.1.12** (Lemma 4 from [JN12])**.** *Let $\rho \in \mathrm{D}(\mathcal{H})$ be a density operator and $\Pi \in \mathrm{L}(\mathcal{H})$ be a projector such that $\mathrm{Tr}(\rho\Pi) < 1$. Then*

$$1 - \mathrm{Tr}(\rho\Pi) \leq \mathrm{F}\left(\rho, \frac{(\mathbb{1} - \Pi)\,\rho\,(\mathbb{1} - \Pi)}{\mathrm{Tr}(\rho\,(\mathbb{1} - \Pi))}\right)^2.$$

The following theorem gives a relation between the trace distance and the fidelity.

**Theorem 2.1.13** (Fuchs-van de Graaf Inequalities, see e.g., [Wato8b] for a proof)**.** *For any $\rho, \sigma \in \mathrm{D}(\mathcal{H})$, it holds that*

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

The following argument has appeared before, for example in [BSW11]. We present it here as a separate lemma and include its proof for convenience.

**Lemma 2.1.14.** *Let $0 \leq \varepsilon \leq 1$, $\rho \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$, and $\sigma \in \mathrm{D}(\mathcal{B})$. If*

$$d(\mathrm{Tr}_{\mathcal{A}}(\rho), \sigma) \leq \varepsilon$$

*then there exists a $\tau \in \mathrm{D}(\mathcal{A} \otimes \mathcal{B})$ for which*

$$\mathrm{Tr}_{\mathcal{A}}(\tau) = \sigma \qquad \text{and} \qquad d(\rho, \tau) \leq \sqrt{2\varepsilon}.$$

*Proof.* Let us take an auxiliary Hilbert space $\mathcal{X} \cong \mathcal{A} \otimes \mathcal{B}$ and let $|\varphi\rangle \in \mathcal{X} \otimes \mathcal{A} \otimes \mathcal{B}$

be a purification of $\rho$, i.e., $\mathrm{Tr}_{\mathcal{X}}(|\varphi\rangle\langle\varphi|) = \rho$. We have that

$$\begin{aligned}
1 - \varepsilon &\leq 1 - d(\mathrm{Tr}_{\mathcal{A}}(\rho), \sigma) \\
&\leq \mathrm{F}(\mathrm{Tr}_{\mathcal{A}}(\rho), \sigma) && (2.3) \\
&= \max\{|\langle\varphi|\psi\rangle| : |\psi\rangle \in \mathcal{X} \otimes \mathcal{A} \otimes \mathcal{B},\ \mathrm{Tr}_{\mathcal{X}\otimes\mathcal{A}}(|\psi\rangle\langle\psi|) = \sigma\} && (2.4)
\end{aligned}$$

where Eq. (2.3) follows from Theorem 2.1.13 and Eq. (2.4) follows from Theorem 2.1.7. This means that there exists a $|\psi\rangle \in \mathcal{X} \otimes \mathcal{A} \otimes \mathcal{B}$, such that $1 - \varepsilon \leq |\langle\varphi|\psi\rangle|$ and $\mathrm{Tr}_{\mathcal{X}\otimes\mathcal{A}}(|\psi\rangle\langle\psi|) = \sigma$. Let

$$\tau \stackrel{\mathrm{def}}{=} \mathrm{Tr}_{\mathcal{X}}(|\psi\rangle\langle\psi|).$$

We only need to bound the distance between $\rho$ and $\tau$.

$$\begin{aligned}
d(\rho, \tau) &\leq d(|\varphi\rangle, |\psi\rangle) && (2.5) \\
&= \sqrt{1 - |\langle\varphi|\psi\rangle|^2} && (2.6) \\
&\leq \sqrt{1 - (1-\varepsilon)^2} \\
&\leq \sqrt{2\varepsilon},
\end{aligned}$$

where Eq. (2.5) follows from Theorem 2.1.9 and Eq. (2.6) follows from Eq. (2.1). ❑

The following theorem is used to eliminate the entanglement between registers.

**Theorem 2.1.15** (quantum de Finetti theorem [CKMR07]; this form is from [Wato8b])**.** *Let $\mathsf{X}_1, \ldots, \mathsf{X}_n$ be identical quantum registers, each having associated space $\mathbb{C}^d$, and let $\rho \in \mathrm{D}(\mathbb{C}^{d^n})$ be the state of these registers. Suppose that $\rho$ is invariant under the permutation of the registers. Then for any choice of $k \in \{2, 3, \ldots, n-1\}$ there exists a number $m \in \mathbb{Z}^+$, a probability distribution $\{p_i : i \in \{1, 2, \ldots, m\}\}$, and a collection of density operators $\{\xi_i : i \in \{1, 2, \ldots, m\}\} \subset \mathrm{D}(\mathbb{C}^d)$ such that*

$$\left\| \rho^{\mathsf{X}_1 \cdots \mathsf{X}_k} - \sum_{i=1}^{m} p_i \xi_i^{\otimes k} \right\|_{\mathrm{Tr}} < \frac{4d^2k}{n}.$$

The following claims describe how efficiently we can simulate an arbitrary unitary operator with a quantum circuit made up of some fixed set of gates.

**Theorem 2.1.16** ([NCoo], Chapter 4.5.2)**.** *An arbitrary unitary operator on $\ell$ qubits can be implemented using a circuit containing $O(\ell^2 4^\ell)$ single qubit and* **CNOT** *gates.*

The next theorem follows from the Solovay-Kitaev theorem [Kit97, NC00, DN06].

**Theorem 2.1.17.** *For any unitary operator* $\mathbf{U}$ *on one qubit and* $\varepsilon > 0$*, there exists a circuit* $C_{\mathbf{U},\varepsilon}$ *such that* $C_{\mathbf{U},\varepsilon}$ *is made up of* $O\left(\log^4(1/\varepsilon)\right)$ *gates from the set* $\{\mathbf{H}, \mathbf{T}\}$ *and*

$$\|\Phi_{\mathbf{U}} - C_{\mathbf{U},\varepsilon}\|_\diamond \leq \varepsilon$$

*where* $\Phi_{\mathbf{U}}(\rho) \stackrel{\text{def}}{=} \mathbf{U}\rho\mathbf{U}^* \in \mathrm{C}\left(\mathbb{C}^2, \mathbb{C}^2\right)$.

The following is corollary to Theorems 2.1.16 and 2.1.17.

**Corollary 2.1.18.** *For any unitary operator* $\mathbf{U}$ *on* $\ell$ *qubits and* $\varepsilon > 0$*, there exists a circuit* $C_{\mathbf{U},\varepsilon}$ *such that* $C_{\mathbf{U},\varepsilon}$ *is made up of* $O\left(5^\ell \cdot \log^4\left(5^\ell/\varepsilon\right)\right)$ *gates from the set* $\{\mathbf{H}, \mathbf{T}, \mathbf{CNOT}\}$ *and*

$$\|\Phi_{\mathbf{U}} - C_{\mathbf{U},\varepsilon}\|_\diamond \leq \varepsilon$$

*where* $\Phi_{\mathbf{U}}(\rho) = \mathbf{U}\rho\mathbf{U}^* \in \mathrm{C}\left(\mathbb{C}^{2^\ell}, \mathbb{C}^{2^\ell}\right)$.

**Corollary 2.1.19.** *Let* $\Phi_{\mathbf{U}}$ *and* $C_{\mathbf{U},\varepsilon}$ *be given by Corollary 2.1.18 and let* $\mathcal{H}$ *be an arbitrary finite dimensional complex Euclidean space. From the properties of the diamond norm, it follows that for all* $\rho \in \mathrm{D}\left(\mathbb{C}^{2^\ell} \otimes \mathcal{H}\right)$*,*

$$\left\| \left(\Phi_{\mathbf{U}} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{H})}\right)(\rho) - \left(C_{\mathbf{U},\varepsilon} \otimes \mathbb{1}_{\mathrm{L}(\mathcal{H})}\right)(\rho) \right\|_{\mathrm{Tr}} \leq \varepsilon.$$

The following lemma states how well we can perform state tomography on an unknown quantum state.

**Lemma 2.1.20** (Lemma 1 of [BSW11])**.** *Let* $\rho \in \mathrm{D}\left(\mathbb{C}^{2^q}\right)$ *be a state on* $q = O(\log n)$ *qubits. For any* $\varepsilon \in 1/\mathsf{poly}(n)$*, choose* $N$ *such that* $N \geq 2^{10q}/\varepsilon^3$ *and* $N \in \mathsf{poly}(n)$*. If* $\rho^{\otimes N}$ *is given to a* $\mathsf{poly}(n)$*-time quantum machine then it can perform quantum state tomography and get a classical description* $\xi \in \mathrm{L}\left(\mathbb{C}^{2^q}\right)$ *of* $\rho$ *which, with probability at least* $1 - \varepsilon$*, satisfies*

$$\|\rho - \xi\|_{\mathrm{Tr}} < \varepsilon.$$

### 2.1.1 The SWAP Test

The *SWAP Test* [BBD+97, BCWdW01] is a well-known method for testing if two pure states are the same or far from each other. The test is described in Algorithm 1. Note that, to perform the test, we need two Hadamard gates and $O(\log(\dim(\mathcal{H})))$-number of **CNOT** gates, besides the measurement of qubit $\mathsf{C}$.

The following theorem establishes the success probability of Algorithm 1 when the input state is separable.

**Algorithm 1** SWAP Test

**INPUT:** registers A and B, each having associated Hilbert space $\mathcal{H}$
**OUTPUT:** success or failure

1: Create a qubit C and initialize its state to $|0\rangle$.
2: Apply **H** on C.
3: Perform a controlled-**SWAP** operation between A and B with the control qubit being C.
4: Apply **H** on C.
5: Measure C in the standard basis.
6: **IF** the output is 0 **THEN**
7: **RETURN** success
8: **ELSE**
9: **RETURN** failure
10: **END IF**

**Theorem 2.1.21** ([BCWdW01, KMY03]). *When the SWAP Test is applied to $\rho \otimes \sigma$, where $\rho, \sigma \in \mathrm{D}(\mathcal{H})$, it succeeds with probability*

$$\frac{1 + \mathrm{Tr}(\rho\sigma)}{2}.$$

If the states are pure, i.e., $\rho = |\varphi\rangle\langle\varphi|$ and $\sigma = |\psi\rangle\langle\psi|$, then the success probably of the SWAP Test can be more conveniently written as

$$\frac{1 + |\langle\varphi|\psi\rangle|^2}{2}. \tag{2.7}$$

From Eq. (2.7), we can see that if the input states are the same pure states then the SWAP Test succeeds with probability 1 and if the states are orthogonal then the success probability is $1/2$.

### 2.1.2 Choi-Jamiołkowski Representations and Post-Selection

Let $\Phi \in \mathrm{C}\left(\mathbb{C}^{2^k}, \mathbb{C}^{2^\ell}\right)$ be a quantum super-operator. The normalized Choi-Jamiołkowski representation of $\Phi$ is defined as

$$\rho_\Phi \stackrel{\text{def}}{=} \frac{1}{2^k} \sum_{x,y \in \{0,1\}^k} \Phi(|x\rangle\langle y|) \otimes |x\rangle\langle y| \in \mathrm{D}\left(\mathbb{C}^{2^\ell} \otimes \mathbb{C}^{2^k}\right).$$

The state $\rho_\Phi$ can be created by applying $\Phi$ to one-half of $k$ EPR pairs. Let's introduce registers L, K, and X with associated Hilbert spaces $\mathcal{L} \cong \mathbb{C}^{2^\ell}$, $\mathcal{K} \cong \mathbb{C}^{2^k}$, and $\mathcal{X} \cong \mathbb{C}^{2^k}$. If we are given $\rho_\Phi$ in (L, K) and an arbitrary $\sigma \in \mathrm{D}(\mathcal{X})$ in X then there exists a simple procedure which produces $\Phi(\sigma)$ with probability $1/4^k$.

The procedure is described in Algorithm 2.

Note that Algorithm 2 is basically teleportation, where we want to teleport the state of $\mathsf{X}$ to register $\mathsf{L}$. If we only get outputs $|\Phi^+\rangle$ then no correction is needed in the teleportation. As mentioned above, we can say that $\rho_\Phi$ was prepared by applying $\Phi$ to one-half of $k$ EPR pairs. Since Algorithm 2 doesn't touch $\mathsf{L}$, in case of success the final state of $\mathsf{L}$ is the same as if the application of $\Phi$ happened after the execution of Algorithm 2, in which case the state produced is $\Phi(\sigma)$. If any of the output happens to be $|\Psi^+\rangle$, $|\Phi^-\rangle$, or $|\Psi^-\rangle$ then there is a Pauli-$\mathbf{X}$, $\mathbf{Z}$, or $\mathbf{Y}$ error in the teleportation that we can't correct, so we declare failure. This idea of simulating a quantum operator with Choi-Jamiołkowski representations has appeared before in the context of quantum interactive proof and quantum Merlin-Arthur proof systems, such as in Refs. [BSW11, KLGN13]. We summarize the above discussion in the following lemma.

**Lemma 2.1.22.** *Suppose that the inputs to Algorithm 2 are $\rho_\Phi$ in $(\mathsf{L},\mathsf{K})$, for some $\Phi \in \mathrm{C}(\mathcal{K},\mathcal{L})$, and an arbitrary $\sigma$ in $\mathsf{X}$. Then the algorithm will succeed with probability $4^{-k}$ and in that case it will output $\Phi(\sigma)$ in $\mathsf{L}$.*

If $\Phi$ is unitary, i.e., $\Phi(\sigma) = \mathbf{U}^*\sigma\mathbf{U}$, for some unitary operator $\mathbf{U}$, then $\rho_\Phi$ is pure, in which case we use the notation $|J(\mathbf{U})\rangle$, where $|J(\mathbf{U})\rangle\langle J(\mathbf{U})| = \rho_\Phi$. For more information see Section 2.1 of [BSW11].

## 2.2 Some Complexity Classes

We assume the reader is familiar with computational complexity and basic complexity classes like P, NP, PSPACE, IP, EXP, NEXP, etc. A good textbook on complexity theory is the one by Arora and Barak [AB09]. We also assume some familiarity with quantum computational complexity but we will define the rele-

**Algorithm 2** Post-Selection

**INPUT:** registers $\mathsf{L}$, $\mathsf{K}$, and $\mathsf{X}$ *{$(\mathsf{L},\mathsf{K})$ are supposed to contain the state $\rho_\Phi$.}*
**OUTPUT:** success and $\mathsf{L}$, or failure

1: Perform a measurement in the Bell basis on each qubit of $\mathsf{K}$ and its corresponding qubit in $\mathsf{X}$.
2: **IF** all the outputs are $|\Phi^+\rangle$ **THEN**
3: **RETURN** success and $\mathsf{L}$
4: **ELSE**
5: **RETURN** failure
6: **END IF**

vant quantum complexity classes in the rest of this section. A good overview of quantum computational complexity is the survey by Watrous [Wato8a]. At the end of this section, Table 2.1 summarizes all the complexity classes mentioned in this thesis.

Before we define quantum complexity classes, let us briefly describe what we mean by polynomial-time quantum algorithms or quantum verifiers. Quantum verifiers are polynomial-time uniformly generated quantum circuits consisting of some universal set of gates. There are many different universal sets and we assume that one of them has been chosen beforehand. One example of a universal set is $\{\mathbf{CNOT}, \mathbf{H}, \mathbf{T}\}$. Usually it doesn't matter which set we choose when we define quantum verifiers and classes like BQP or QMA, because it is known that each universal set can approximate any other set with exponential precision. However, later we will have quantum proof systems with one-sided error and exponentially or double-exponentially small gap, in which case the gate set may matter. This is because simulating one set of gates with another may introduce an exponentially small error. In this thesis, we only assume that the verifier can perform or perfectly simulate the **CNOT** and the **H** gate with his universal set, besides being able to perform any polynomial-time classical computation. Note that, with **CNOT** and **H**, one can perform all Pauli operators. The above assumption is enough for all of our results, so we won't bother about the gate set in the rest of the thesis.

**Definition 2.2.1** ([Watoo, ANo2])**.** For functions $c, s : \mathbb{Z}^+ \to (0, 1]$, a language $L$ is in $\mathsf{QMA}(c, s)$ if there exists a quantum verifier $V$ with the following properties. For all $n \in \mathbb{Z}^+$ and inputs $x \in \{0,1\}^n$, the circuit of $V$ on input $x$, denoted by $\mathbf{V}_x$, is a polynomial-time uniformly generated quantum circuit acting on two polynomial-size registers $\mathsf{P}$ and $\mathsf{A}$. One output qubit of $\mathbf{V}_x$ is designated as the acceptance qubit. We say that $\mathbf{V}_x$ on input $|\varphi\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}}$ accepts if the acceptance qubit of $\mathbf{V}_x\left(|\varphi\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}}\right)$ is projected to $|1\rangle$ and we say that $\mathbf{V}_x$ rejects if it's projected to $|0\rangle$. $\mathbf{V}_x$ must satisfy the following properties.

**Completeness.** If $x \in L$ then there exists a quantum state $|\varphi\rangle \in \mathcal{P}$ such that the acceptance probability of $\mathbf{V}_x$, on input $|\varphi\rangle \otimes |\bar{0}\rangle_{\mathcal{A}}$, is at least $c(n)$.

**Soundness.** If $x \notin L$ then for all states $|\varphi\rangle \in \mathcal{P}$, $\mathbf{V}_x$ accepts with probability at most $s(n)$, given $|\varphi\rangle \otimes |\bar{0}\rangle_{\mathcal{A}}$ as its input.

Note that $\mathsf{P}$ is the register in which the verifier receives his proof and $\mathsf{A}$ is his private register, which is, without loss of generality, always initialized to $|\bar{0}\rangle$. Without causing confusion, we will denote both the circuit of the verifier and the unitary operator it represents by $\mathbf{V}_x$.

**Definition 2.2.2.** The class QMA is defined as $\mathsf{QMA} \stackrel{\text{def}}{=} \mathsf{QMA}\left(\frac{2}{3}, \frac{1}{3}\right)$ and the class $\mathsf{QMA}_1$ is defined as $\mathsf{QMA}_1 \stackrel{\text{def}}{=} \mathsf{QMA}\left(1, \frac{1}{2}\right)$.

It is easy to see that the choice of constants in the above definition are arbitrary. This is formalized by the following theorem for the case of QMA.

**Theorem 2.2.3** ([KSV02, AN02, MW05]). *Let $c \in (0,1)$ be a constant and $p(n)$ be a positive polynomial in $n$. It holds that*

$$\mathsf{QMA} = \mathsf{QMA}\left(c, c - \frac{1}{p(n)}\right) = \mathsf{QMA}\left(1 - 2^{-p(n)}, 2^{-p(n)}\right).$$

We now define QIP, the generalization of QMA to multiple messages. For the purpose of this thesis, an informal definition will suffice. For the detailed, rigorous definition see Refs. [Wat03, KW00, Wat08a].

**Definition 2.2.4** (informal)**.** For functions $c, s : \mathbb{Z}^+ \to (0,1]$, a language $L$ is in $\mathsf{QIP}(c,s)$ if the following holds. For all $n \in \mathbb{Z}^+$ and inputs $x \in \{0,1\}^n$, there exists a polynomial $p(n)$ which denotes the number of message exchanges between the prover and the verifier. It is always assumed that the prover sends the last message. So, the first message is sent by the prover or the verifier depending on whether $p(n)$ is even or odd. The provers and verifiers are now a sequence of quantum operations. Without loss of generality, we assume that they are unitary operations. In any round, the verifier applies his operation to his private register and a dedicated message register and sends the message register to the prover. The prover then applies his operation to his private register and the message register and sends back the message register. This procedure repeats for as many times as $p(n)$ dictates. After the interaction ends, the verifier performs his last operation and measures his output qubit which determines acceptance or rejection. We don't put any computational restrictions on the prover so, in particular, his private space can be arbitrary large and his operations can be arbitrary unitaries. Similarly to the case of QMA, all the operators of the verifier must be polynomial-time uniformly generated circuits. So, the message and his private registers are also polynomial-sized. The conditions for the completeness and soundness errors are the usual:

- If $x \in L$ then there must exist a prover that makes the verifier accept with probability at least $c(n)$.
- If $x \notin L$ then, no matter what prover the verifier is interacting with, his acceptance probability must be at most $s(n)$.

**Definition 2.2.5.** The classes $\mathsf{QIP}$ and $\mathsf{QIP}_1$ are defined analogously to $\mathsf{QMA}$ and $\mathsf{QMA}_1$. Formally, $\mathsf{QIP} \stackrel{\text{def}}{=} \mathsf{QIP}\left(\frac{2}{3}, \frac{1}{3}\right)$ and $\mathsf{QIP}_1 \stackrel{\text{def}}{=} \mathsf{QIP}\left(1, \frac{1}{2}\right)$.

**Definition 2.2.6.** The class $\mathsf{QMA}^{\text{const-EPR}}(c, s)$ is defined the same way as $\mathsf{QMA}(c, s)$ in Definition 2.2.1, except that before the prover sends the proof to the verifier, they can share a constant number of EPR pairs (the two-qubit state $|\Phi^+\rangle$).

**Definition 2.2.7.** The class $\mathsf{QMA}_1^{\text{const-EPR}}$ is defined as

$$\mathsf{QMA}_1^{\text{const-EPR}} \stackrel{\text{def}}{=} \mathsf{QMA}^{\text{const-EPR}}\left(1, \frac{1}{2}\right).$$

Similarly as before, the choice of $1/2$ is arbitrary. This holds because a $\mathsf{QMA}_1^{\text{const-EPR}}$ proof system is a special case of a two-message $\mathsf{QIP}_1$ proof system and perfect parallel repetition holds even for three-message $\mathsf{QIP}_1$ [KW00]. Since changing the soundness from one constant to another requires only a constant number of repetitions, we have the following lemma.

**Lemma 2.2.8.** *For any constant $s \in (0, 1)$, it holds that*

$$\mathsf{QMA}_1^{\text{const-EPR}} = \mathsf{QMA}^{\text{const-EPR}}(1, s).$$

The following definition generalizes Definition 2.2.1 to multiple provers.

**Definition 2.2.9** ([KMY03, ABD+09])**.** For functions $k : \mathbb{Z}^+ \to \mathbb{Z}^+$ and $c, s : \mathbb{Z}^+ \to (0, 1]$, a language $L$ is in $\mathsf{QMA}(k, c, s)$ if there exists a quantum verifier $V$ such that, for all $n \in \mathbb{Z}^+$ and inputs $x \in \{0,1\}^n$, the circuit of $V$ on input $x$ is a polynomial-time uniformly generated quantum circuit and the following holds.

**Completeness:** If $x \in L$ then there exist quantum states $|\phi_1\rangle, \ldots, \left|\phi_{k(n)}\right\rangle$ where, for all $i$, $|\phi_i\rangle$ is a state on $\mathsf{poly}(n)$ qubits and the acceptance probability of $V$ on inputs $x$ and $|\phi_1\rangle \otimes \cdots \otimes \left|\phi_{k(n)}\right\rangle$ is at least $c(n)$.

**Soundness:** If $x \notin L$ then for all states $|\phi_1\rangle, \ldots, \left|\phi_{k(n)}\right\rangle$ where, for all $i$, $|\phi_i\rangle$ is a state on at most $\mathsf{poly}(n)$ qubits, $V$ accepts with probability at most $s$, given $x$ and $|\phi_1\rangle \otimes \cdots \otimes \left|\phi_{k(n)}\right\rangle$ as the inputs.

*Remark* 2.2.10. If we just give one parameter to $\mathsf{QMA}$, then it indicates the number of proofs. So the notation $\mathsf{QMA}[k]$ is defined as $\mathsf{QMA}[k] \stackrel{\text{def}}{=} \mathsf{QMA}\left(k, \frac{2}{3}, \frac{1}{3}\right)$. With this notation the class $\mathsf{QMA}$ is defined as $\mathsf{QMA} \stackrel{\text{def}}{=} \mathsf{QMA}[1]$.

**Definition 2.2.11** ([Brao8, ABD$^+$09])**.** The class $\mathsf{BellQMA}(k, c, s)$ is defined almost the same way as $\mathsf{QMA}(k, c, s)$ in Definition 2.2.9, except that the verifier $V$ is not an arbitrary polynomial-time quantum computation. The restriction we put on the verifier is the following. The verifier, upon seeing $x$, performs a classical randomized polynomial-time computation and produces circuits for measurements $M_1, \ldots, M_{k(n)}$, where each $M_i$ is a POVM. Then, for all $i$, he measures $|\phi_i\rangle$ with $M_i$ and obtains outcome $m_i$. After all measurements were performed, he runs a classical computation on inputs $m_1, \ldots, m_{k(n)}$ and decides whether to accept or reject.

Note that in the above definition the verifier has to measure each proofs separately. Moreover, none of the circuits of the measurements can depend on the outcome of any previous measurement. Chen and Drucker [CD10] defined $\mathsf{BellQMA}$ in a slightly different way by allowing the verifier to do quantum computations before and after the measurements. Our result also holds if we take their definition. The reason we chose the above definition is because we will prove a lower bound for our $\mathsf{BellQMA}$ class so with the more restricted definition our result is slightly stronger.

Now we define quantum interactive proof systems where in the beginning there is a $O(\log n)$-long interaction which is followed by a $\mathsf{poly}(n)$-length message from the prover. Note that in this setting we can assume, without loss of generality, that all messages, except the last one, consist of a single qubit and the total number of rounds is at most $O(\log n)$. This is because we can add dummy qubits that are interspersed with the qubits sent by the other party. We define the class according to this observation.

**Definition 2.2.12.** Let the class $\mathsf{QIP}_{\mathrm{short}}(m, c, s)$ be the set of languages for which there exists a quantum interactive proof system with the following properties. The completeness parameter is $c$ and the soundness is $s$. The proof system consists of $m$ rounds and each round is a question-answer pair. All questions and answers are one qubits except for the last answer which is $\mathsf{poly}(n)$ qubits, where $n$ is the length of the input. See Fig. 2.1 for an example with $m = 3$.

A similar class, $\mathsf{QIP}([\mathsf{log}, \mathsf{poly}], c, s)$ was defined in [BSW11] to be the class of problems for which there exists a one round quantum interactive proof system, with completeness and soundness parameters $c$ and $s$. Additionally, the verifier's question has length $O(\log n)$ and the prover's answer is $\mathsf{poly}(n)$ qubits.

*Remark* 2.2.13. The following inclusion is trivially true between the above

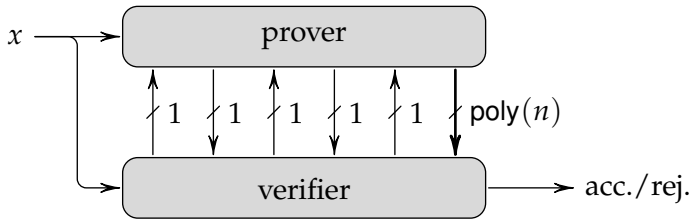


Figure 2.1: The interaction in the proof system of Definition 2.2.12 in case $m = 3$.

classes.

$$\mathsf{QMA}(c,s) \subseteq \mathsf{QIP}([\log, \mathsf{poly}], c, s) \subseteq \mathsf{QIP}_{\text{short}}(O(\log n), c, s)$$

for all values of $c$ and $s$.

In Ref. [BSW11], it was proven that in their setting the question from the verifier is unnecessary. This is formulated by the following theorem.

**Theorem 2.2.14** ([BSW11])**.** *Let $c, s : \mathbb{N} \to (0,1)$ be polynomial-time computable functions such that $c(n) - s(n) \in 1/\mathsf{poly}(n)$. Then $\mathsf{QIP}([\log, \mathsf{poly}], c, s) = \mathsf{QMA}$.*

Later we will prove that the seemingly stronger class of Definition 2.2.12 also has the same power as QMA if $m = O(\log n)$.

We now formally define the succinct version of the graph 3-coloring problem. It will be used to prove the lower bounds on the small gap multi-prover QMA classes.

**Definition 2.2.15** ([GW83])**.** Let $G(V, E)$ be an undirected graph where $V = \{0, 1, \ldots, m-1\}$ and $m \leq 2^{\tilde{n}}$ for some $\tilde{n} \in \mathbb{N}$. We define $C_G$ to be a *small circuit representation* of $G$ if the following conditions hold.

- $C_G$ is a circuit containing AND, OR, and NOT gates.
- $C_G$ has two inputs of $\tilde{n}$ bits each.
- $C_G$ has $\mathsf{poly}(\tilde{n})$ gates.
- The output of $C_G$ is given by $C_G(u,v) = \begin{cases} 00 & \text{if } u \notin V \text{ or } v \notin V \text{ or } u \geq v, \\ 10 & \text{if } u < v \text{ and } (u,v) \notin E, \\ 11 & \text{if } u < v \text{ and } (u,v) \in E. \end{cases}$

**Definition 2.2.16.** Let the decision problem Succinct3Col be the set of small circuit representations of graphs that are 3-colorable.

**Theorem 2.2.17** ([PY86])**.** Succinct3Col *is* NEXP*-complete.*

| Complexity classes | |
|---|---|
| Class | Description |
| NP | The class of problems decidable by a *nondeterministic polynomial-time* Turing machine. |
| MA | The class of problems for which a positive instance can be verified in probabilistic polynomial time with the help of a proof. Same as IP with one message. |
| PP | The class of problems decidable in *probabilistic polynomial time* without any guarantee on the gap. |
| PSPACE | The class of problems solvable by a Turing machine in *polynomial space*. |
| EXP | The class of problems solvable in deterministic *exponential time*. |
| NEXP | The class of problems decidable in *nondeterministic exponential time*. |
| IP | The class of problems for which a positive instance can be verified by an *interactive proof*. Equals PSPACE. |
| MIP | The class of problems for which a positive instance can be verified by a *multi-prover* interactive proof. Equals NEXP. |
| BQP | The class of problems solvable in *polynomial* time by a *quantum* computer with *bounded* error. |
| QMA | The class of problems such that a positive instance can be verified by a one-message quantum interactive proof. See Definitions 2.2.1 and 2.2.2. |
| QCMA | The same as QMA but the proof must be a classical string. |
| $\mathsf{QMA}[k]$ | The same as QMA, except that the verifier is given $k$ polynomial-size quantum proofs, which are guaranteed to be *unentangled*. See Definition 2.2.9 and Remark 2.2.10. |
| BellQMA | The same as $\mathsf{QMA}[k]$ but the verifier has to measure each proofs separately. See Definition 2.2.11. |
| QIP | The class of problems such that a positive instance can be verified by a *quantum interactive proof*. See Definition 2.2.4. |
| $\mathsf{QIP}(2)$ | QIP with two messages. |
| QMIP | The quantum generalization of MIP and the multi-prover generalization of QIP. |
| $\mathsf{MIP}^*$ | QMIP with a classical verifier. Alternatively, MIP with shared entanglement between the provers. Equals QMIP. |
| $\mathsf{QMA}^{\text{const-EPR}}$ | The same as $\mathsf{QIP}(2)$ but the first message consists only of a constant number of halves of EPR pairs. See Definitions 2.2.6 and 2.2.7. |
| $\mathsf{QIP}([\mathsf{log},\mathsf{poly}],.,.)$ | The same as $\mathsf{QIP}(2)$ but the length of the first message is logarithmic. |
| $\mathsf{QIP}_{\text{short}}$ | The same as QIP but the combined length of all but the last messages is logarithmic. See Definition 2.2.12. |

Table 2.1: Summary of complexity classes mentioned in this thesis. Some of the above classes have one-sided error variants which are denoted by the subscript 1 and are not included in this list.

## 2.3 Information Theory

Let $\mu$ be a probability distribution on some finite set $\mathscr{X}$ and let $\mu(x)$ represent the probability of $x \in \mathscr{X}$ according to $\mu$. Let $X$ be a random variable distributed according to $\mu$, i.e., $\Pr[X = x] = \mu(x)$. We often use the same symbol to represent the random variable and its distribution when it is clear from the context. The expectation value of a function $f$ on $\mathscr{X}$ is defined as

$$\mathop{\mathbb{E}}_{x \leftarrow X}[f(x)] \stackrel{\text{def}}{=} \sum_{x \in \mathscr{X}} \mu(x) \cdot f(x)$$

where $x \leftarrow X$ means that $x$ is drawn from the distribution of $X$. A classical distribution $\mu$ can be viewed as a quantum state with diagonal entries $\mu(x)$ and non-diagonal entries 0. A good textbook on classical information theory is the one by Cover and Thomas [CT06] and a good text on quantum information is the lecture notes of Watrous [Wat08b].

**Definition 2.3.1.** The *entropy* of a quantum state $\rho \in \mathrm{D}(\mathcal{X})$ is defined as

$$\mathrm{S}(\rho) \stackrel{\text{def}}{=} -\mathrm{Tr}(\rho \log \rho)\,.$$

We also use the notation $\mathrm{S}(\mathsf{X})_\rho$ to represent $\mathrm{S}(\rho)$.

**Definition 2.3.2.** The *relative entropy* between quantum states $\rho$ and $\sigma$ is defined as

$$\mathrm{S}(\rho\|\sigma) \stackrel{\text{def}}{=} \mathrm{Tr}(\rho \log \rho) - \mathrm{Tr}(\rho \log \sigma)$$

and the *relative min-entropy* between them is defined as

$$\mathrm{S}_\infty(\rho\|\sigma) \stackrel{\text{def}}{=} \min\left\{\lambda \,:\, \rho \leq 2^\lambda \sigma\right\}. \tag{2.8}$$

Since the logarithm is operator-monotone, $\mathrm{S}(\rho\|\sigma) \leq \mathrm{S}_\infty(\rho\|\sigma)$.

**Definition 2.3.3.** Let $\rho^{\mathsf{XY}}$ be the state of registers $(\mathsf{X}, \mathsf{Y})$. The *mutual information* between registers $\mathsf{X}$ and $\mathsf{Y}$ is defined as

$$\mathrm{I}(\mathsf{X} : \mathsf{Y})_\rho \stackrel{\text{def}}{=} \mathrm{S}(\mathsf{X})_\rho + \mathrm{S}(\mathsf{Y})_\rho - \mathrm{S}(\mathsf{XY})_\rho\,.$$

It is easy to see that

$$\mathrm{I}(\mathsf{X} : \mathsf{Y})_\rho = \mathrm{S}\big(\rho^{\mathsf{XY}} \big\| \rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}}\big)\,. \tag{2.9}$$

If $\mathsf{X}$ is a classical register, i.e., $\rho$ can be written in the form $\rho = \sum_{x \in \mathscr{X}} \mu(x) \cdot$

$|x\rangle\langle x| \otimes \sigma_x$ for some finite set $\mathscr{X}$ and where $\mu$ is a probability distribution on $\mathscr{X}$, then

$$\mathrm{I}(\mathsf{X} : \mathsf{Y})_\rho = \mathrm{S}(\mathsf{Y})_\rho - \mathrm{S}(\mathsf{Y}|\mathsf{X})_\rho$$

where the *conditional entropy* is defined as

$$\mathrm{S}(\mathsf{Y}|\mathsf{X})_\rho \stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(\sigma_x)] .$$

**Definition 2.3.4.** Let $\rho^{\mathsf{XYZ}}$ be a quantum state where $\mathsf{Y}$ is a classical register. The mutual information between $\mathsf{X}$ and $\mathsf{Z}$, conditioned on $\mathsf{Y}$, is defined as

$$\begin{aligned} \mathrm{I}(\mathsf{X} : \mathsf{Z}|\mathsf{Y})_\rho &\stackrel{\text{def}}{=} \mathop{\mathbb{E}}_{y \leftarrow \mathsf{Y}} \Big[ \mathrm{I}(\mathsf{X} : \mathsf{Z}|\mathsf{Y} = y)_\rho \Big] \\ &= \mathrm{S}(\mathsf{X}|\mathsf{Y})_\rho + \mathrm{S}(\mathsf{Z}|\mathsf{Y})_\rho - \mathrm{S}(\mathsf{XZ}|\mathsf{Y})_\rho . \end{aligned}$$

The following lemma follows easily from the definitions.

**Lemma 2.3.5** (*Chain rule* for the mutual information)**.** *Let $\rho^{\mathsf{XYZ}}$ be a quantum state where $\mathsf{Y}$ is a classical register. It holds that*

$$\mathrm{I}(\mathsf{X} : \mathsf{YZ})_\rho = \mathrm{I}(\mathsf{X} : \mathsf{Y})_\rho + \mathrm{I}(\mathsf{X} : \mathsf{Z}|\mathsf{Y})_\rho .$$

**Theorem 2.3.6** (Joint convexity of the relative entropy)**.** *Let $\rho_1$, $\rho_2$, $\sigma_1$, and $\sigma_2$ be quantum states and let $p \in [0, 1]$. Then*

$$\mathrm{S}(p\rho_1 + (1-p)\,\rho_2 \| p\sigma_1 + (1-p)\,\sigma_2) \leq p \cdot \mathrm{S}(\rho_1\|\sigma_1) + (1-p) \cdot \mathrm{S}(\rho_2\|\sigma_2) .$$

**Theorem 2.3.7** (*Chain rule* for the relative entropy)**.** *Let $\rho_1, \rho_2 \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be classical-quantum states with the following form,*

$$\begin{aligned} \rho_1 &= \sum_{x \in \mathscr{X}} \mu_1(x) \cdot |x\rangle\langle x| \otimes \sigma_x \\ \rho_2 &= \sum_{x \in \mathscr{X}} \mu_2(x) \cdot |x\rangle\langle x| \otimes \xi_x. \end{aligned}$$

*Then it holds that*

$$\mathrm{S}(\rho_1\|\rho_2) = \mathrm{S}(\mu_1\|\mu_2) + \mathop{\mathbb{E}}_{x \leftarrow \mu_1} [\mathrm{S}(\sigma_x\|\xi_x)] .$$

**Corollary 2.3.8.** *Let $\rho \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$ be a classical-quantum state with the following*

*form,*

$$\rho = \sum_{x \in \mathscr{X}} \mu(x) \cdot |x\rangle\langle x| \otimes \sigma_x$$

*and let* $\sigma \stackrel{\text{def}}{=} \mathrm{Tr}_{\mathcal{X}}(\rho)$*. It holds that*

$$\mathrm{I}(\mathsf{X} : \mathsf{Y})_\rho = \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(\sigma_x \| \sigma)] \,.$$

*Proof.* The corollary follows from the chain rule for the relative entropy.

$$\begin{aligned} \mathrm{I}(\mathsf{X} : \mathsf{Y})_\rho &= \mathrm{S}\left(\rho \middle\| \sum_{x \in \mathscr{X}} \mu(x) \cdot |x\rangle\langle x| \otimes \sigma\right) \\ &= \mathrm{S}(\mu \| \mu) + \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(\sigma_x \| \sigma)] \\ &= \mathop{\mathbb{E}}_{x \leftarrow \mu} [\mathrm{S}(\sigma_x \| \sigma)] \end{aligned}$$

where the first equality follows from Eq. (2.9) and the second equality follows from Theorem 2.3.7. ❑

**Lemma 2.3.9.** *For states* $\rho^{\mathsf{XY}} \in \mathrm{D}(\mathcal{X} \otimes \mathcal{Y})$*,* $\sigma^{\mathsf{X}} \in \mathrm{D}(\mathcal{X})$*, and* $\tau^{\mathsf{Y}} \in \mathrm{D}(\mathcal{Y})$*, it holds that*

$$\mathrm{S}\big(\rho^{\mathsf{XY}} \big\| \sigma^{\mathsf{X}} \otimes \tau^{\mathsf{Y}}\big) \geq \mathrm{S}\big(\rho^{\mathsf{XY}} \big\| \rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}}\big) \,.$$

*Proof.* The lemma follows from the following calculation.

$$\begin{aligned} \mathrm{S}\big(\rho^{\mathsf{XY}} \big\| \sigma^{\mathsf{X}} \otimes \tau^{\mathsf{Y}}\big) - \mathrm{S}\big(\rho^{\mathsf{XY}} \big\| \rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}}\big) &= \mathrm{Tr}\big(\rho^{\mathsf{XY}} \log\big(\rho^{\mathsf{X}} \otimes \rho^{\mathsf{Y}}\big)\big) - \mathrm{Tr}\big(\rho^{\mathsf{XY}} \log\big(\sigma^{\mathsf{X}} \otimes \tau^{\mathsf{Y}}\big)\big) \\ &= \mathrm{Tr}\big(\rho^{\mathsf{X}} \log \rho^{\mathsf{X}}\big) + \mathrm{Tr}\big(\rho^{\mathsf{Y}} \log \rho^{\mathsf{Y}}\big) \\ &\qquad - \mathrm{Tr}\big(\rho^{\mathsf{X}} \log \sigma^{\mathsf{X}}\big) - \mathrm{Tr}\big(\rho^{\mathsf{Y}} \log \tau^{\mathsf{Y}}\big) \\ &= \mathrm{S}\big(\rho^{\mathsf{X}} \big\| \sigma^{\mathsf{X}}\big) + \mathrm{S}\big(\rho^{\mathsf{Y}} \big\| \tau^{\mathsf{Y}}\big) \\ &\geq 0 \end{aligned}$$

where the second equality follows from the formula

$$\log(\mathbf{A} \otimes \mathbf{B}) = \log(\mathbf{A}) \otimes \mathbb{1} + \mathbb{1} \otimes \log(\mathbf{B})$$

and the last inequality holds because the relative entropy is always non-negative. ❑

**Theorem 2.3.10** (Pinsker's inequality, [Wat08b, JRS03])**.** *For quantum states* $\rho$ *and*

*$\sigma$, it holds that*

$$\|\rho - \sigma\|_1 \leq \sqrt{S(\rho\|\sigma)} \qquad \textit{and} \qquad 1 - F(\rho,\sigma) \leq S(\rho\|\sigma)\,.$$

The relative entropy is non-increasing when subsystems are considered. This is formalized by the following theorem.

**Theorem 2.3.11.** *Let $\rho^{XY}, \sigma^{XY} \in D(\mathcal{X} \otimes \mathcal{Y})$ be density operators. It holds that*

$$S\left(\rho^{XY}\middle\|\sigma^{XY}\right) \geq S\left(\rho^{X}\middle\|\sigma^{X}\right).$$

**Lemma 2.3.12.** *Let $|\psi\rangle \in \mathcal{A} \otimes \mathcal{B}$ be a bipartite pure state and let $\rho \stackrel{\text{def}}{=} \operatorname{Tr}_{\mathcal{A}}(|\psi\rangle\langle\psi|)$. Let $\mathbf{M} \in L(\mathcal{A})$ be a POVM element, i.e., $0 \leq \mathbf{M} \leq \mathbb{1}_{\mathcal{A}}$. Let $q$ be the probability of getting $\mathbf{M}$ when measuring $|\psi\rangle$. Formally, let $q \stackrel{\text{def}}{=} \langle\psi|\left(\mathbf{M} \otimes \mathbb{1}_{\mathcal{B}}\right)|\psi\rangle$. Let $\sigma$ be the resulting state on $\mathcal{B}$ after the measurement, i.e.,*

$$\sigma \stackrel{\text{def}}{=} \operatorname{Tr}_{\mathcal{A}}\left(\frac{\left(\sqrt{\mathbf{M}} \otimes \mathbb{1}_{\mathcal{B}}\right)|\psi\rangle\langle\psi|\left(\sqrt{\mathbf{M}} \otimes \mathbb{1}_{\mathcal{B}}\right)}{q}\right).$$

*Then it holds that $S_\infty(\sigma\|\rho) \leq -\log q$.*

*Proof.* It is easy to see that $\rho = q\sigma + (1-q)\,\xi$, where

$$\xi \stackrel{\text{def}}{=} \operatorname{Tr}_{\mathcal{A}}\left(\frac{\left(\sqrt{\mathbb{1}_{\mathcal{A}} - \mathbf{M}} \otimes \mathbb{1}_{\mathcal{B}}\right)|\psi\rangle\langle\psi|\left(\sqrt{\mathbb{1}_{\mathcal{A}} - \mathbf{M}} \otimes \mathbb{1}_{\mathcal{B}}\right)}{1-q}\right).$$

From this and Eq. (2.8), it follows that $S_\infty(\sigma\|\rho) \leq -\log q$. ❑

# 3

# Results on Quantum Merlin-Arthur Proof Systems

This chapter discusses our results on quantum Merlin-Arthur proof systems. The three main results are divided into three sections. The first result is about eliminating a logarithmic-length interaction.

## 3.1 Eliminating Short Messages

In this section we show that in an interactive proof system where, in the beginning, there is a logarithmic-length interaction and at the end there is a polynomial-length message from the prover, the logarithmic-length interaction can be eliminated, so the proof system has the same power as QMA. This was posed as an open problem by Beigi, Shor, and Watrous in Ref. [BSW11]. The result itself is formulated by Theorem 1.1.3 which we restate below for convenience. The material of this section has appeared in Ref. [Per12b].

**Theorem 1.1.3.** *Let $c, s : \mathbb{N} \to (0,1)$ be polynomial-time computable functions such that $c(n) - s(n) \in 1/\mathsf{poly}(n)$. Then*

$$\mathsf{QIP}_{\mathrm{short}}(O(\log n), c, s) = \mathsf{QMA}.$$

### 3.1.1 The Idea Behind the Proof of Theorem 1.1.3

We observe that it's sufficient for the QIP prover to have only $O(\log n)$ qubits in his private work register in all but the last round without changing the

acceptance probability. So the prover's unitaries in these rounds can be approximated by polynomial-size quantum circuits. The prover in the QMA proof system gives the classical descriptions of these circuits to the verifier who approximately produces the state of the whole system appearing in the beginning of the last round of the QIP protocol. This system is composed of the prover's private space, the question to the prover and the verifier's private space. While simulating the last round, we don't care about the prover's private space, so we treat his operation as a quantum channel whose input is the private space of the prover and the question from the verifier, and whose output is the answer to the verifier. Since the input is on $O(\log n)$-many qubits, to perform the action of this channel, we can use the same method as was used in Section 3 of Ref. [BSW11]. For this step, the QMA prover sends many copies of the normalized Choi-Jamiołkowski representation of the channel, with which the verifier can simulate the channel using the post-selection procedure described in Algorithm 2.

### 3.1.2 The Detailed Proof

Before we prove Theorem 1.1.3, we give a lemma that will be the key to handle the short interaction.

**Lemma 3.1.1.** *Let us have a* $\mathsf{QIP}_{\text{short}}(m+1, c, s)$ *proof system.*[1] *Without loss of generality (i.e., without changing completeness c and soundness s) we can assume that during the first m rounds the prover only uses* $2m$ *qubits in his private register, in both the honest and the dishonest case. Moreover, the actions of the prover in each of these rounds are unitary transformations.*

The above lemma is a special case of Lemma 11 of [KM03] (when there is only one prover) and also appears in the proof of Theorem 6 of [GW07]. The intuitive reason why it holds is the following. Before the verifier and the prover interact, the state of the whole system (i.e., the verifier's and the prover's private spaces) has Schmidt number one. With each qubit sent, the Schmidt number of this system increases at most by a factor of two. At the end of the $m$-th round the Schmidt number of the system is at most $2^{2m}$. This means that we can find a purification of the verifier's state, in each step, which has at most $2m$ qubits at the prover's side. For each round we find two purifications; first when the prover receives the question and second after the prover generates the answer. From the unitary equivalence of purifications (Theorem 2.1.5),

[1] See Definition 2.2.12 for the definition of $\mathsf{QIP}_{\text{short}}$.

there exist unitary transformations on the prover's side that transform between these purifications. Corollary 2.1.18 will be used to put an upper bound on the number of gates we need to simulate these unitaries.

We are now ready to prove the main theorem of the section.

*Proof of Theorem 1.1.3.* As mentioned in Remark 2.2.13, the inclusion $\mathsf{QMA} \subseteq \mathsf{QIP}_{\text{short}}(O(\log n), c, s)$ is trivial, so we only need to prove that

$$\mathsf{QIP}_{\text{short}}(O(\log n), c, s) \subseteq \mathsf{QMA}.$$

Let $L \in \mathsf{QIP}_{\text{short}}(m+1, c, s)$, where $m = O(\log n)$, and let $V$ be the corresponding verifier. We will construct a verifier $W$ for the $\mathsf{QMA}$ proof system. Because of Lemma 3.1.1, we can assume that any prover strategy in the first $m$ rounds are unitary operators on $2m$ qubits, say $\mathbf{U}_1, \ldots, \mathbf{U}_m$. The constructed $W$ expects to get, as part of the proof, the classical descriptions of circuits $C_{\mathbf{U}_1, 3^{-n}}, \ldots, C_{\mathbf{U}_m, 3^{-n}}$, i.e., the circuits that approximate the prover's operators with precision $1/3^n$. According to Corollary 2.1.18, the length of this proof is $O\left(m \cdot 5^{2m} \cdot \log^4\left(5^{2m} \cdot 3^n\right)\right) \in \mathsf{poly}(n)$. $W$ uses this classical proof to simulate the first $m$ rounds of the proof system and to produce the state of the whole system at the end of the $m$-th round. This means the prover's and verifier's private spaces and the answer to the verifier from the $m$-th round. We denote this state by $|\psi\rangle$. Using Corollary 2.1.19 and the fact that each circuit approximates the corresponding unitary with precision $1/3^n$, we get that after applying $O(\log n)$-many of them, it is true that

$$\| |\psi\rangle\langle\psi| - |\phi\rangle\langle\phi| \|_{\text{Tr}} \leq \frac{m}{3^n} \leq \frac{1}{2^n}$$

for sufficiently large $n$, and where $|\phi\rangle$ is the state of the whole system after the $m$-th round in the case when the unitaries $\mathbf{U}_1, \ldots, \mathbf{U}_m$ were applied instead of the circuits.

We are left with specifying how $W$ simulates the prover in the last, $(m+1)$-th round. We use exactly the same method as was used in the proof of Theorem 2.2.14 in [BSW11]. Our proof closely follows that proof as well. Since we are in the last round, we don't have to keep track of the prover's private space, so we can just describe its strategy as a quantum channel that transforms the private space of the prover with the question from the verifier to the answer to the verifier. Let's call this channel $\Phi \in \mathrm{C}(\mathcal{S}, \mathcal{R})$ from now on, where $\mathcal{S}$ is the joint space associated to the prover's private space and the question, and $\mathcal{R}$ is the space associated to the answer. The input space $\mathcal{S}$ is on

$q \stackrel{\text{def}}{=} 2m+1 = O(\log n)$ qubits and the output space $\mathcal{R}$ is on $\mathsf{poly}(n)$ qubits. $W$ expects to get $\rho_\Phi^{\otimes(N+k)}$ as the quantum part of its proof, where $\rho_\Phi \in \mathrm{D}(\mathcal{R}\otimes\mathcal{S})$ is the normalized Choi-Jamiołkowski representation of $\Phi$, for $N$ and $k$ to be specified later. Let's divide up the quantum certificate given to $W$ into registers $\mathsf{R}_1, \mathsf{S}_1, \mathsf{R}_2, \mathsf{S}_2, \ldots, \mathsf{R}_{N+k}, \mathsf{S}_{N+k}$, where the space of each $\mathsf{R}_i$ is $\mathcal{R}$ and the space of each $\mathsf{S}_i$ is $\mathcal{S}$. $W$ expects each $(\mathsf{R}_i, \mathsf{S}_i)$ to contain a copy of $\rho_\Phi$. To simulate the last round of the interactive proof system, $W$ does the following.

1. Randomly permute the pairs $(\mathsf{R}_1, \mathsf{S}_1), \ldots, (\mathsf{R}_{N+k}, \mathsf{S}_{N+k})$ according to a uniformly chosen permutation and discard all but the first $(N+1)$ pairs.

2. Perform quantum state tomography on the registers $(\mathsf{S}_2, \ldots, \mathsf{S}_{N+1})$ and reject if the resulting approximation is not within trace-distance $\delta/2$ of the completely mixed state $(1/2^q)\,\mathbb{1}$, for $\delta$ to be specified later.

3. Simulate the channel specified by $(\mathsf{R}_1, \mathsf{S}_1)$ by post-selection. Reject if post-selection fails, otherwise simulate the last operation of $V$ and accept if and only if $V$ accepts.

Let $g(n) \in \mathsf{poly}(n)$ be such that $c(n) - s(n) \geq 1/g(n)$. We now set the parameters.

$$\varepsilon \stackrel{\text{def}}{=} \frac{1}{g \cdot 4^{q+1}}, \qquad \delta \stackrel{\text{def}}{=} \frac{\varepsilon^2}{4}, \qquad N \stackrel{\text{def}}{=} \left\lceil \frac{2^{10q}}{(\delta/2)^3} \right\rceil, \qquad k \stackrel{\text{def}}{=} \left\lceil \frac{(N+1)\cdot 4^{2q+1}}{\varepsilon} \right\rceil.$$

Note that $1/\varepsilon, 1/\delta, N, k \in \mathsf{poly}(n)$.

**Completeness.** Suppose there exists a prover $P$ that causes $V$ to accept with probability $\geq c$. Let the certificate to $W$ be the classical descriptions of circuits $C_{\mathbf{U}_1, 3^{-n}}, \ldots, C_{\mathbf{U}_m, 3^{-n}}$, together with the state $\rho_\Phi^{\otimes(N+k)}$, where each $(\mathsf{R}_i, \mathsf{S}_i)$ contains a copy of $\rho_\Phi$, for $i \in \{1, 2, \ldots, N+k\}$. After simulating the first $m$ rounds, $W$ produces $|\psi\rangle$ which is $\leq 1/2^n$ far from the correct $|\phi\rangle$ in the trace distance, just as described above. Note that in the simulation of the last round, step 1 doesn't change the state of registers $(\mathsf{R}_1, \mathsf{S}_1), \ldots, (\mathsf{R}_{N+1}, \mathsf{S}_{N+1})$. According to Lemma 2.1.20, $W$ rejects in step 2 with probability $\leq \delta/2$. In step 3, post-selection succeeds with probability $1/4^q$. If $W$ was using $|\phi\rangle$ instead of $|\psi\rangle$ the probability of acceptance would be at least

$$\left(1 - \frac{\delta}{2}\right) \frac{c}{4^q}.$$

So using $|\psi\rangle$, the probability that $W$ accepts is at least

$$\left(1 - \frac{\delta}{2}\right) \frac{c}{4^q} - \frac{1}{2^n} \geq \frac{c}{4^q} - \varepsilon - \frac{1}{2^n}.$$

**Soundness.** Suppose that all $P$ cause $V$ to accept with probability $\leq s$. Note that, without loss of generality, any classical proof specifies some set of unitaries that corresponds to a valid prover strategy. Hence, it is still true that after $W$ simulates the first $m$ rounds using the given circuits, he ends up with a state $|\psi\rangle$ that is at most $1/2^n$ far from a state $|\phi\rangle$, where $|\phi\rangle$ can be produced by some $P$ interacting with $V$.

Now consider the situation that the state of $(\mathsf{S}_1, \ldots, \mathsf{S}_{N+1})$ before step 2 has the form

$$\sigma^{\otimes(N+1)} \tag{3.1}$$

for some $\sigma \in \mathrm{D}(\mathcal{S})$. (The classical part of the proof has been used up and discarded before step 1.) We consider two cases:

- Suppose that $\|\sigma - (1/2^q)\,\mathbb{1}\|_{\mathrm{Tr}} < \delta$. Let the state of $(\mathsf{R}_1, \mathsf{S}_1)$ before step 3 be $\xi \in \mathrm{D}(\mathcal{R} \otimes \mathcal{S})$, so we have $\mathrm{Tr}_{\mathcal{R}}(\xi) = \sigma$. From Lemma 2.1.14, there exists a state $\tau \in \mathrm{D}(\mathcal{R} \otimes \mathcal{S})$ such that $\mathrm{Tr}_{\mathcal{R}}(\tau) = (1/2^q)\,\mathbb{1}$ and $d(\tau, \xi) \leq \varepsilon$. Given this $\tau$, the post-selection in step 3 succeeds with probability $1/4^q$, so the acceptance in step 3 occurs with probability at most $s/4^q + 1/2^n$. Given $\xi$ instead of $\tau$, $W$ will accept with probability at most

$$\frac{s}{4^q} + \frac{1}{2^n} + \varepsilon.$$

- If $\|\sigma - (1/2^q)\,\mathbb{1}\|_{\mathrm{Tr}} \geq \delta$ then, in step 2, $W$ will accept with probability $\leq \delta/2$. This follows from Lemma 2.1.20.

Since $\delta/2 \leq s/4^q + 1/2^n + \varepsilon$ then in both cases acceptance occurs with probability $\leq s/4^q + 1/2^n + \varepsilon$.

Now suppose that the state of $(\mathsf{S}_1, \ldots, \mathsf{S}_{N+1})$ before step 2 has the form

$$\sum_i p_i \sigma_i^{\otimes(N+1)} \tag{3.2}$$

for some probability vector $p$ and some set $\{\sigma_i\} \subset \mathrm{D}(\mathcal{S})$. Since the state in Eq. (3.2) is a convex combination of states of the form given by Eq. (3.1), acceptance will occur with probability $\leq s/4^q + 1/2^n + \varepsilon$. In the real scenario,

by Theorem 2.1.15, it is true that the state of $(\mathsf{S}_1, \ldots, \mathsf{S}_{N+1})$ after step 1 will be $\varepsilon$ close to a state of the form given by Eq. (3.2), in the trace distance. So the probability of acceptance of $W$ will be $\leq s/4^q + 2\varepsilon + 1/2^n$. Since

$$\frac{c}{4^q} - \varepsilon - \frac{1}{2^n} - \left(\frac{s}{4^q} + 2\varepsilon + \frac{1}{2^n}\right) \geq \frac{1}{h(n)}$$

for some $h(n) \in \mathsf{poly}(n)$, it holds that $L \in \mathsf{QMA}$. ❑

### 3.1.3 An Open Problem

We mention an open problem that we think is interesting and that is related to the above result. Let us consider interactive proof systems which are similar to the ones studied in this section but the polynomial-length message is at the beginning of the interaction, not at the end. More precisely, the interaction starts with a polynomial-length message from the prover and then continues with a conversation between the prover and the verifier, where the combined length of all messages is at most logarithmic. What is the power of this class?

Note that the power of this class doesn't change if we allow a logarithmic-length interaction both before and after the polynomial-length message. The reason is that in this case we can start the interaction with the prover sending the long message along with the private space of the verifier. Then the verifier flips a coin and decides to continue the protocol forwards or backwards. He accepts if he ends up in the accepting state or initial state, respectively. This idea has appeared, for example, in [KKMV08].

Also note that this proof system is 'somewhere in between' BQP and QIP. If there is no long message from the prover (i.e., the length of the whole interaction is at most logarithmic), then the proof system has the same power as BQP [BSW11]. On the other hand, if there are two polynomial-length messages from the prover then the proof system has the full power of QIP [KW00].

## 3.2 Perfect Completeness with Shared EPR Pairs

In this section, we give a new, simpler proof of one of the results of Kobayashi, Le Gall, and Nishimura [KLGN13], which shows that any QMA protocol can be converted to a one-sided error protocol, in which Arthur and Merlin initially share a constant number of EPR pairs and then Merlin sends his proof to Arthur. We restate the corresponding theorem here from the introduction.

**Theorem 1.1.1.** $\mathsf{QMA} \subseteq \mathsf{QMA}_1^{\text{const-EPR}}$.

As mentioned before, our protocol is similar but somewhat simpler than the original. Our main contribution is a simpler and more direct analysis of the soundness property that uses well-known results in quantum information such as properties of the trace distance and the fidelity, and the quantum de Finetti theorem. This section is based on Ref. [Per13].

### 3.2.1 Some Preliminaries

A key to our one-sided error algorithm will be the following operator which will be used to reduce the acceptance probability of a QMA verifier to $1/2$. Let $q \in [0,1]$ and

$$\mathbf{W}_q \stackrel{\text{def}}{=} \begin{bmatrix} \sqrt{1-q} & -\imath\sqrt{q} \\ -\imath\sqrt{q} & \sqrt{1-q} \end{bmatrix}.$$

Note that $\mathbf{W}_q$ corresponds to a rotation about the $\hat{x}$ axis in the Bloch sphere and it is very similar to the corresponding operator in [KLGN13]. The following lemma will be the basic building block to prove perfect completeness, similarly to [KLGN13].

**Lemma 3.2.1.** *Let $\Delta, \Pi \in \mathrm{L}(\mathcal{H})$ be projectors. Suppose that one of the eigenvalues of $\Delta\Pi\Delta$ is $1/2$ with corresponding eigenstate $|\omega\rangle$. Then*

$$\Delta\left(\mathbb{1} - 2\Pi\right)\Delta\,|\omega\rangle = 0.$$

*Proof.* Using the fact that $\Delta\,|\omega\rangle = |\omega\rangle$, we get

$$\begin{aligned} \Delta\left(\mathbb{1} - 2\Pi\right)\Delta\,|\omega\rangle &= \left(\Delta - 2\Delta\Pi\Delta\right)|\omega\rangle \\ &= |\omega\rangle - 2\left(\frac{1}{2}\,|\omega\rangle\right) \\ &= 0. \end{aligned}$$

❑

In [KLGN13], the procedure defined by applying $\Delta\left(\mathbb{1} - 2\Pi\right)\Delta$ is called 'Reflection Procedure'. The procedure is very similar to the quantum rewinding technique of Watrous [Wat09], which has been used before to achieve perfect completeness for quantum multi-prover interactive proofs [KKMV08]. Also note that the idea behind the quantum rewinding technique dates back to the strong gap amplification for QMA [MW05].

It should be mentioned here that Lemma 3.2.1 will only be used in the honest case, while in the dishonest case we will argue about the rejection

probability directly. This is why we can have a much simpler lemma compared to the description of the Reflection Procedure in [KLGN13].

### 3.2.2 Modified Post-Selection

Suppose we have an EPR pair ($|\Phi^+\rangle$) in registers $(\mathsf{S}, \mathsf{S}')$ and let $q \in [0,1]$. As we described in Section 2.1.2, the normalized Choi-Jamiołkowski representation of $\mathbf{W}_q$, denoted by $\left|J(\mathbf{W}_q)\right\rangle$, can be generated by applying $\mathbf{W}_q$ on register $\mathsf{S}$. By simple calculation, we get that

$$\begin{aligned}\left|J(\mathbf{W}_q)\right\rangle &= (\mathbf{W}_q \otimes \mathbb{1})\left|\Phi^+\right\rangle = \sqrt{1-q}\left|\Phi^+\right\rangle - \imath\sqrt{q}\left|\Psi^+\right\rangle \\ \left|J\left(\mathbf{W}_q^*\right)\right\rangle &= \left(\mathbf{W}_q^* \otimes \mathbb{1}\right)\left|\Phi^+\right\rangle = \sqrt{1-q}\left|\Phi^+\right\rangle + \imath\sqrt{q}\left|\Psi^+\right\rangle.\end{aligned}$$

Later, in Algorithm 4, we will be given two copies of $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$ and we will have to create the state $\mathbf{W}_q\left|0\right\rangle$ with the help of the first copy. Using the second copy, we will need to apply $\mathbf{W}_q^*$ on an arbitrary input state. The way this can be done is as follows. Suppose now that we are given $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$ and we want to create $\mathbf{W}_q\left|0\right\rangle$. This can easily be done by applying the following unitary

$$\mathbf{Q} \stackrel{\text{def}}{=} \left|00\right\rangle\left\langle\Phi^+\right| - \left|10\right\rangle\left\langle\Psi^+\right| + \left|01\right\rangle\left\langle\Phi^-\right| - \left|11\right\rangle\left\langle\Psi^-\right| \tag{3.3}$$

because $\mathbf{Q}\left|J\left(\mathbf{W}_q^*\right)\right\rangle = \left(\mathbf{W}_q\left|0\right\rangle\right) \otimes \left|0\right\rangle$. Note that $\mathbf{Q}$ can be implemented with **CNOT** and Hadamard gates so our assumption on the gate set we made in Section 2.2 is still valid. Now assume that we want to apply $\mathbf{W}_q^*$ on an arbitrary state $|\varphi\rangle$, with the help of $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$. As we discussed in Section 2.1.2, this can be accomplished with probability $1/4$ using post-selection (Algorithm 2). Here we use a slightly modified version that succeeds with probability $1/2$. This modified procedure is described in Algorithm 3. In Algorithm 3, we want to teleport the state of $\mathsf{X}$ (let's say it's $|\varphi\rangle$) to register $\mathsf{S}$. If we get output $|\Phi^+\rangle$ then no correction is needed in the teleportation. Since $\mathbf{W}_q^*$ was applied to $\mathsf{S}$ before, we get $\mathbf{W}_q^*\left|\varphi\right\rangle$ in $\mathsf{S}$. If the output is $|\Psi^+\rangle$ then there is a 'Pauli-$\mathbf{X}$ error' in the teleportation so we get $\mathbf{W}_q^*\mathbf{X}\left|\varphi\right\rangle$, which we can correct since $\mathbf{W}_q^*$ and $\mathbf{X}$ commute. In case of the other two outputs ($|\Phi^-\rangle$ and $|\Psi^-\rangle$), there is a $\mathbf{Z}$ or a $\mathbf{Y}$ error that we can't correct, so we declare failure. We state a lemma below, analogous to Lemma 2.1.22, that we will use in the honest case. In the dishonest case, we will argue about the success probability and the output of Algorithm 3 in the analysis of Algorithm 4.

**Algorithm 3** Modified Post-Selection

**INPUT:** single qubit registers $\mathsf{S}$, $\mathsf{S}'$, $\mathsf{X}$ *{$(\mathsf{S},\mathsf{S}')$ are supposed to contain the state $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$.}*

**OUTPUT:** success and $\mathsf{S}$, or failure

1: Perform a measurement in the Bell basis on $(\mathsf{S}',\mathsf{X})$.
2: **IF** the output is $|\Phi^+\rangle$ **THEN**
3: **RETURN** success and $\mathsf{S}$
4: **ELSE IF** the output is $|\Psi^+\rangle$ **THEN**
5: Apply $\mathbf{X}$ on $\mathsf{S}$.
6: **RETURN** success and $\mathsf{S}$
7: **ELSE**
8: **RETURN** failure
9: **END IF**

**Lemma 3.2.2.** *Suppose that the inputs to Algorithm 3 are $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$ in $(\mathsf{S},\mathsf{S}')$, for some $q \in [0,1]$, and an arbitrary $|\varphi\rangle$ in $\mathsf{X}$. Then the algorithm will succeed with probability $1/2$ and in that case it will output $\mathbf{W}_q^* |\varphi\rangle$ in $\mathsf{S}$.*

### 3.2.3 The Idea Behind the Proof

Before we give the detailed proof of Theorem 1.1.1, let us briefly describe the intuition behind our proof. We also point out the similarities and the differences between our proof and the proof in [KLGN13].

The basic idea to achieve perfect completeness is very similar to the idea in Ref. [KLGN13]. For any input $x$, let us define

$$\mathbf{M}_x \stackrel{\text{def}}{=} \left(\mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}}\right) \mathbf{V}_x^* \Pi_{\text{acc}} \mathbf{V}_x \left(\mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}}\right)$$

where $\mathbf{V}_x$ is the circuit of the verifier corresponding to input $x$, as defined in Definition 2.2.1, and $\Pi_{\text{acc}}$ is the projector that corresponds to projecting the acceptance qubit of $\mathbf{V}_x$ to $|1\rangle$. Note that $0 \leq \mathbf{M}_x \leq \mathbb{1}_{\mathcal{P}\otimes\mathcal{A}}$. As was observed in [MW05], the maximum acceptance probability of $\mathbf{V}_x$ is $\|\mathbf{M}_x\|_\infty$, or in other words, the maximum eigenvalue of $\mathbf{M}_x$. We will use Lemma 3.2.1 to construct a test that succeeds with probability 1 in case $x \in L$. In order to achieve this, we need that for all $x \in L$, $\|\mathbf{M}_x\|_\infty = 1/2$. Unfortunately, this is not true in general. Instead, we have that if $x \in L$ then $\|\mathbf{M}_x\|_\infty \geq 1/2$. Our first objective is to modify $\mathbf{M}_x$ such that its maximum eigenvalue is exactly $1/2$. We do this by using an auxiliary qubit (stored in register $\mathsf{S}$) and defining

$$\mathbf{M}'_x \stackrel{\text{def}}{=} \mathbf{M}_x \otimes \left(|0\rangle\langle 0|_{\mathsf{S}} \mathbf{W}_q^* |1\rangle\langle 1|_{\mathsf{S}} \mathbf{W}_q |0\rangle\langle 0|_{\mathsf{S}}\right)$$

$$= \left(\mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}\otimes\mathcal{S}}\right) \left(\mathbf{V}_x \otimes \mathbf{W}_q\right)^* \left(\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}}\right) \left(\mathbf{V}_x \otimes \mathbf{W}_q\right) \left(\mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}\otimes\mathcal{S}}\right)$$

where $q \stackrel{\text{def}}{=} 1/\left(2 \cdot \|\mathbf{M}_x\|_\infty\right) \in [1/2, 1]$. It is now easy to see that $\|\mathbf{M}'_x\|_\infty = 1/2$ and we can also write $\mathbf{M}'_x$ as $\mathbf{M}'_x = \Delta\Pi\Delta$, for

$$\Delta \stackrel{\text{def}}{=} \mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}\otimes\mathcal{S}} \qquad \text{and} \qquad \Pi \stackrel{\text{def}}{=} \left(\mathbf{V}_x \otimes \mathbf{W}_q\right)^* \left(\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}}\right) \left(\mathbf{V}_x \otimes \mathbf{W}_q\right).$$

Now, we can use Lemma 3.2.1 and obtain the following test. Let the principal eigenvector of $\mathbf{M}'_x$ (that corresponds to eigenvalue $1/2$) be denoted by $|\omega\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}\otimes\mathcal{S}}$. The test receives this eigenstate as the input, applies the unitary operator $\mathbb{1} - 2\Pi$, and performs a measurement defined by operators $\{\Delta, \mathbb{1} - \Delta\}$. If the state is projected to $\Delta$ the test rejects and otherwise it accepts. Lemma 3.2.1 guarantees that we never project to $\Delta$.

However, a polynomial-time verifier may not be able to perform this test, because it is possible that $\mathbf{W}_q$ can't be expressed by a polynomial-size quantum circuit and the verifier may not even know the exact value of $q$. To overcome this difficulty, the verifier expects the prover to give several copies of the normalized Choi-Jamiołkowski representations of $\mathbf{W}_q^*$, besides $|\omega\rangle_{\mathcal{P}}$. As explained in Section 3.2.2, these can be used to perform $\mathbf{W}_q$ and $\mathbf{W}_q^*$, by using unitary $\mathbf{Q}$ to do $\mathbf{W}_q$, and Algorithm 3 to do $\mathbf{W}_q^*$. Note that Algorithm 3 may fail, in which case we have to accept in order to maintain perfect completeness. This is the main idea to prove perfect completeness, and it is basically the same as in [KLGN13].

The harder part is to prove the soundness and this is where our proof differs from the one in [KLGN13]. Let us first give a high-level overview of the soundness proof of Kobayashi et al. [KLGN13]. The main idea in their proof is to perform a sequence of tests (i.e., quantum algorithms with measurements at the end) which together ensure that the registers that are supposed to contain the Choi-Jamiołkowski representations of the desired operator actually contain the Choi-Jamiołkowski representations of *some* operator. Then they show that doing the so-called 'Reflection Simulation Test', the one just described above, with these states in the registers, will cause rejection with some constant probability. The tests they use to ensure that the states are close to Choi-Jamiołkowski representations are the 'Distillation Procedure' (which is used to remove the entanglement between the register of the original proof and the registers of the Choi-Jamiołkowski representations), the 'Space Restriction Test' (which tests that the states are in a certain subspace), and the SWAP Test. In their analysis they also use the de Finetti theorem. We don't describe these tests here, as

the interested reader can find them in [KLGN13]. We just list them in order to compare them to the tools we use.

Our main idea behind the soundness proof is conceptually different. We don't argue that the states are close to Choi-Jamiołkowski representations, but we analyze our version of the Reflection Simulation Test directly. As we described this test above, there are two measurements in it. The first measurement is in Algorithm 3 and the second is given by $\{\Delta, \mathbb{1} - \Delta\}$. So, roughly speaking, we have to prove two things. First, we have to show that Algorithm 3 can't always fail, as otherwise we would end up always accepting without reaching the end of the procedure. This will be formalized later in Lemma 3.2.5. In order to prove Lemma 3.2.5, we only need two assumptions. The first assumption is that the state being measured in Algorithm 3 is separable, which is guaranteed by the de Finetti theorem (Theorem 2.1.15). The second assumption is that the state of some registers is close to being completely mixed, which is obviously true because these registers hold parts of EPR pairs.

The second part of the soundness proof is to show that conditioned on Algorithm 3 being successful, we get a state that projects to $\Delta$ with constant probability. To prove this, we first argue that the private register of the verifier (register $\mathsf{A}$) projects to $|\bar{0}\rangle\langle\bar{0}|$. This follows from simple properties of the trace distance. We then show that the state of register $\mathsf{S}$ projects to $|0\rangle\langle 0|$. To prove this, we use the SWAP Test on the registers that are supposed to contain the Choi-Jamiołkowski representations. This ensures that the state of these registers are close to the same pure state. This property is formalized in Lemma 3.2.6. We also use a simplified version of the Space Restriction Test, which is not really a test but an application of a super-operator on the above mentioned registers. This super-operator will be defined later in Eq. (3.4). We can think of it as performing a projective measurement that corresponds to the Space Restriction Test and forgetting the outcome. Using the above tools, it will follow by direct calculation that the state of $\mathsf{S}$ projects to $|0\rangle\langle 0|$.

Note that we don't use the Distillation Procedure of [KLGN13] and we use a simpler form of the Space Restriction Test. Besides that, it's worth mentioning that the tools we use can be grouped into two sets based on whether we use them in the analysis of the first or the second measurement. For the analysis of the first measurement, we need that some state is close to being maximally mixed, while in the analysis of the second, we use the SWAP Test and the above mentioned super-operator. One exception is the de Finetti theorem, as we need that the states are separable in both parts. This property of the proof may be useful for simplifying it further, because for example, to omit the SWAP Test,

one would only need to re-prove that the state of $\mathsf{S}$ projects to $|0\rangle\langle 0|$ in the last measurement.

### 3.2.4 The Detailed Proof

This section presents the detailed proof of Theorem 1.1.1. Let $L \in \mathsf{QMA}$ and $V$ be the corresponding verifier. Let $x$ be an input to language $L$ and let us denote its length by $n$. We denote the circuit of $V$ on input $x$ (and also the unitary transformation it represents) by $\mathbf{V}_x$. Let the private register of $\mathbf{V}_x$ be denoted by $\mathsf{A}$ and the register in which the proof is received by $\mathsf{P}$. As in the previous section, let $\Pi_{\text{acc}} \in \mathrm{L}(\mathcal{P} \otimes \mathcal{A})$ be the projector that corresponds to projecting the acceptance qubit of $\mathbf{V}_x$ to $|1\rangle$. By Theorem 2.2.3, we assume that the completeness of $V$ is at least $1/2$ and his soundness is at most $4^{-n}$. Let $N \stackrel{\text{def}}{=} 2^{127}$. We construct a verifier $W$ which recognizes the same language $L$ with completeness 1, constant soundness, and with the additional property that $W$ possesses $N$ halves of EPR pairs in registers $\mathsf{S}'_1, \ldots, \mathsf{S}'_N$ before the protocol begins. The other halves of the EPR pairs are held by the prover. $W$ gets his proof in registers $\mathsf{P}, \mathsf{S}_1, \ldots, \mathsf{S}_N$, where the $\mathsf{S}_i$'s are single qubit registers, which had contained the other halves of the EPR pairs before the prover performed some transformation on them. $W$ expects to get the original proof of $V$ in $\mathsf{P}$ and the state of each $(\mathsf{S}_i, \mathsf{S}'_i)$ is supposed to be $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$, for some $q \in [0,1]$. In the description of $W$ we will use the following notations. Let $\mathcal{W}^+$ be the subspace of $\mathbb{C}^4$ spanned by $|\Phi^+\rangle$ and $|\Psi^+\rangle$, and $\mathcal{W}^-$ be the subspace spanned by $|\Phi^-\rangle$ and $|\Psi^-\rangle$. Let

$$\Pi^+ \stackrel{\text{def}}{=} |\Phi^+\rangle\langle\Phi^+| + |\Psi^+\rangle\langle\Psi^+| \qquad \text{and} \qquad \Pi^- \stackrel{\text{def}}{=} |\Phi^-\rangle\langle\Phi^-| + |\Psi^-\rangle\langle\Psi^-|,$$

i.e., the projections to subspaces $\mathcal{W}^+$ and $\mathcal{W}^-$. Let $\Psi \in \mathrm{C}\left(\mathbb{C}^4, \mathbb{C}^4\right)$ be a quantum super-operator defined as

$$\Psi(\mathbf{A}) \stackrel{\text{def}}{=} \Pi^+ \mathbf{A} \Pi^+ + \Pi^- \mathbf{A} \Pi^-. \tag{3.4}$$

$\mathbf{Q}$ still denotes the operator defined by Eq. (3.3). With these notations, the procedure of $W$ is described in Algorithm 4.

Note that Algorithm 4 runs in polynomial time and besides performing the circuit $\mathbf{V}_x$ and its inverse, it only uses $\mathbf{H}$, $\mathbf{CNOT}$, $\mathbf{Q}$, Pauli gates, and classical logical gates. (This justifies our assumption we made about the gate set in Section 2.2.) We have to prove completeness and soundness in order to prove Theorem 1.1.1. Lemma 3.2.3 proves that in the honest case $W$ always accepts,

**Algorithm 4** Description of verifier $W$ in the proof of Theorem 1.1.1.

**INPUT:** description of a circuit $\mathbf{V}_x$, polynomial-size register $\mathsf{P}$ compatible with $\mathbf{V}_x$, and single qubit registers $\mathsf{S}_1, \ldots, \mathsf{S}_N, \mathsf{S}'_1, \ldots, \mathsf{S}'_N$, where the state of $(\mathsf{S}'_1, \ldots, \mathsf{S}'_N)$ is guaranteed to be $\mathbb{1}/2^N$ *{For all $i$, $(\mathsf{S}_i, \mathsf{S}'_i)$ are supposed to contain $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$.}*

**OUTPUT:** accept or reject

Permute registers $(\mathsf{S}_1, \mathsf{S}'_1), \ldots, (\mathsf{S}_N, \mathsf{S}'_N)$ uniformly at random and discard all but $(\mathsf{S}_1, \mathsf{S}'_1)$ and $(\mathsf{S}_2, \mathsf{S}'_2)$.
Apply $\Psi$ on both $(\mathsf{S}_1, \mathsf{S}'_1)$ and $(\mathsf{S}_2, \mathsf{S}'_2)$.
Choose $b \in_R \{0, 1\}$ uniformly at random.
**IF** $b = 0$ **THEN**
    Apply $\mathbf{Q}$ on $(\mathsf{S}_1, \mathsf{S}'_1)$. *{This creates $\mathbf{W}_q |0\rangle$ in $\mathsf{S}_1$. $\mathsf{S}'_1$ can be discarded.}*
    Create register $\mathsf{A}$, compatible with $\mathbf{V}_x$, and initialize its state to $|\bar{0}\rangle$.
    Apply $\mathbf{V}_x$ on $(\mathsf{P}, \mathsf{A})$.
    Apply a phase-flip if both the acceptance qubit and register $\mathsf{S}_1$ are 1. *{This is done by applying the unitary $\mathbb{1}_{\mathcal{P}\otimes\mathcal{A}\otimes\mathcal{S}_1} - 2\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}_1}$ on $(\mathsf{P}, \mathsf{A}, \mathsf{S}_1)$.}*
    Apply $\mathbf{V}_x^*$ on $(\mathsf{P}, \mathsf{A})$.
    Execute Algorithm 3 with input $(\mathsf{S}_2, \mathsf{S}'_2, \mathsf{S}_1)$.
    **IF** Algorithm 3 fails **THEN**
        **RETURN** accept
    **END IF**
    Measure $(\mathsf{A}, \mathsf{S}_2)$ in the standard basis.
    **IF** the output of the measurement is $\bar{0}$ **THEN**
        **RETURN** reject
    **ELSE**
        **RETURN** accept
    **END IF**
**ELSE**
    Apply the SWAP Test on $(\mathsf{S}_1, \mathsf{S}'_1)$ and $(\mathsf{S}_2, \mathsf{S}'_2)$.
    **IF** the SWAP Test succeeds **THEN**
        **RETURN** accept
    **ELSE**
        **RETURN** reject
    **END IF**
**END IF**

while Lemma 3.2.4 proves that in the dishonest case $W$ rejects with probability at least $2^{-62}$. This shows that $L \in \mathsf{QMA}^{\text{const-EPR}}\left(1, 1-2^{-62}\right)$. By Lemma 2.2.8, $\mathsf{QMA}^{\text{const-EPR}}\left(1, 1-2^{-62}\right) = \mathsf{QMA}_1^{\text{const-EPR}}$ so Theorem 1.1.1 follows.

**Lemma 3.2.3** (Completeness)**.** *If $x \in L$ then the prover can prepare registers $\mathsf{P}$ and $\mathsf{S}_1, \ldots, \mathsf{S}_N$ in such a way that verifier $W$ of Algorithm 4 accepts with probability 1.*

*Proof.* Let $p_x \in [1/2, 1]$ be the maximum probability with which $V$ can be made to accept $x$, where the maximum is taken over all states in $\mathsf{P}$. Let

$$q \stackrel{\text{def}}{=} \frac{1}{2p_x}$$

and note that $q \in [1/2, 1]$. The honest Merlin prepares $|\omega_x\rangle$ in $\mathsf{P}$, where $|\omega_x\rangle$ is the original witness of $V$ that makes it accept with probability exactly $p_x$. Furthermore, for all $i \in \{1, 2, \ldots, N\}$, Merlin applies $\mathbf{W}_q^*$ to $\mathsf{S}_i$. This creates $\left|J\left(\mathbf{W}_q^*\right)\right\rangle$ in all $(\mathsf{S}_i, \mathsf{S}_i')$. Then Merlin sends registers $\mathsf{P}, \mathsf{S}_1, \ldots, \mathsf{S}_N$ to $W$.

Note that steps 1 and 2 of Algorithm 4 don't change the state because

$$\left|J\left(\mathbf{W}_q^*\right)\right\rangle = \sqrt{1-q}\left|\Phi^+\right\rangle + \imath\sqrt{q}\left|\Psi^+\right\rangle \in \mathcal{W}^+.$$

If, in step 3, $b$ is chosen to be 1 then the SWAP Test in step 21 succeeds with certainty, by Theorem 2.1.21. So, from now on, suppose that $b$ is chosen to be 0, in which case we continue to step 5. From the arguments of Section 3.2.2, we have that the state of $\mathsf{S}_1$ after step 5 is $\mathbf{W}_q|0\rangle$. So the state of $(\mathsf{P}, \mathsf{A}, \mathsf{S}_1)$ before entering step 10 is

$$\left(\mathbf{V}_x^* \otimes \mathbb{1}_{\mathcal{S}_1}\right)\left(\mathbb{1} - 2\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}_1}\right)\left(\mathbf{V}_x \otimes \mathbf{W}_q\right)\left(|\omega_x\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{S}_1}\right).$$

We assume that Algorithm 3 in step 10 succeeds, as otherwise we accept. In this case, by Lemma 3.2.2, the state of $(\mathsf{P}, \mathsf{A}, \mathsf{S}_2)$ after step 10 will be

$$\left(\mathbf{V}_x^* \otimes \mathbf{W}_q^*\right)\left(\mathbb{1} - 2\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}_2}\right)\left(\mathbf{V}_x \otimes \mathbf{W}_q\right)\left(|\omega_x\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}} \otimes |0\rangle_{\mathcal{S}_2}\right).$$

Let

$$\Delta \stackrel{\text{def}}{=} \mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}} \otimes |0\rangle\langle 0|_{\mathcal{S}_2} \qquad \text{and} \qquad \Pi \stackrel{\text{def}}{=} (\mathbf{V}_x^* \Pi_{\text{acc}} \mathbf{V}_x) \otimes \left(\mathbf{W}_q^* |1\rangle\langle 1|_{\mathcal{S}_2} \mathbf{W}_q\right).$$

Note that the maximum eigenvalue of operator $\Delta\Pi\Delta$ is $1/2$, with corresponding eigenstate $|\omega_x\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}\otimes\mathcal{S}_2}$. From Lemma 3.2.1,

$$0 = \Delta\left(\mathbb{1} - 2\Pi\right)\Delta\left(|\omega_x\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}\otimes\mathcal{S}_2}\right)$$

$$= \left(\mathbb{1}_{\mathcal{P}} \otimes |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}\otimes\mathcal{S}_2}\right)\left(\mathbf{V}_x^* \otimes \mathbf{W}_q^*\right)\left(\mathbb{1} - 2\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}_2}\right)$$
$$\cdot \left(\mathbf{V}_x \otimes \mathbf{W}_q\right)\left(|\omega_x\rangle_{\mathcal{P}} \otimes |\bar{0}\rangle_{\mathcal{A}\otimes\mathcal{S}_2}\right).$$

It means that the measurement of step 14 will never output $\bar{0}$. This finishes the proof of the lemma. ❑

**Lemma 3.2.4** (Soundness)**.** *Let $x \notin L$ and $n$ sufficiently large. Suppose that the input to Algorithm 4 is such that the reduced state on $(\mathsf{S}'_1, \ldots, \mathsf{S}'_N)$ is $\mathbb{1}/2^N$. Then Algorithm 4 rejects with probability at least $2^{-62}$.*

*Proof.* Let's denote the state of $(\mathsf{P}, \mathsf{S}_1, \mathsf{S}'_1, \mathsf{S}_2, \mathsf{S}'_2)$, after step 1, by $\rho_1$. Theorem 2.1.15 implies that

$$d\left(\text{Tr}_{\mathcal{P}}(\rho_1), \sum_{i=1}^{m} p_i \xi_i \otimes \xi_i\right) \leq \frac{16}{N}.$$

Let's denote the state of the same registers, after step 2, by $\rho_2$. It can be checked by direct calculation that

$$\text{Tr}_{\mathcal{P}\otimes\mathcal{S}_1\otimes\mathcal{S}_2}(\rho_2) = \frac{\mathbb{1}_{\mathcal{S}'_1\otimes\mathcal{S}'_2}}{4}. \tag{3.5}$$

From Theorem 2.1.9, it holds that

$$d\left(\text{Tr}_{\mathcal{P}}(\rho_2), \sum_{i=1}^{m} p_i \sigma_i \otimes \sigma_i\right) \leq \frac{16}{N}$$

where $\sigma_i \stackrel{\text{def}}{=} \Psi(\xi_i)$. By Lemma 2.1.14, there exists a $\rho'_2$ such that

$$\text{Tr}_{\mathcal{P}}\left(\rho'_2\right) = \sum_{i=1}^{m} p_i \sigma_i \otimes \sigma_i$$

and

$$d\left(\rho_2, \rho'_2\right) \leq \sqrt{\frac{32}{N}}. \tag{3.6}$$

Let us suppose, from now on, that before entering step 3 the state of the system is $\rho'_2$. This will result in a bias of at most $\sqrt{32/N}$ in the trace distance in the rest of the states that we calculate. Throughout the rest of the proof, we will assume that the SWAP Test on input $\text{Tr}_{\mathcal{P}}(\rho'_2)$ rejects with probability at most $\varepsilon \stackrel{\text{def}}{=} 2 \cdot 2^{-62} + \sqrt{32/N} = 2^{-60}$, as otherwise we are done with the proof. With this in mind, the rest of the proof will only deal with the case when $b$ is chosen

to be 0 in step 3. In this case we continue to step 5. With these assumptions, the state of the system after step 5 is

$$\rho_5 \stackrel{\text{def}}{=} \left(\mathbf{Q} \otimes \mathbb{1}_{\mathcal{P}\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right) \rho'_2 \left(\mathbf{Q}^* \otimes \mathbb{1}_{\mathcal{P}\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right).$$

Let's denote the state of the whole system after step 7 by

$$\rho_7 \stackrel{\text{def}}{=} \left(\mathbf{V}_x \otimes \mathbb{1}_{\mathcal{S}_1\otimes\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right) \left(|\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}} \otimes \rho_5\right) \left(\mathbf{V}_x^* \otimes \mathbb{1}_{\mathcal{S}_1\otimes\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right).$$

Since the acceptance probability of $\mathbf{V}_x$ is at most $4^{-n}$, we have that

$$\mathrm{Tr}\left(\rho_7 \widetilde{\Pi}_{\text{acc}}\right) \leq \frac{1}{4^n}$$

where $\widetilde{\Pi}_{\text{acc}} \stackrel{\text{def}}{=} \Pi_{\text{acc}} \otimes \mathbb{1}_{\mathcal{S}_1\otimes\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}$. Let $\rho'_7$ be the projection of $\rho_7$ to the rejection subspace, i.e.,

$$\rho'_7 \stackrel{\text{def}}{=} \frac{\left(\mathbb{1} - \widetilde{\Pi}_{\text{acc}}\right) \rho_7 \left(\mathbb{1} - \widetilde{\Pi}_{\text{acc}}\right)}{\mathrm{Tr}\left(\rho_7 \left(\mathbb{1} - \widetilde{\Pi}_{\text{acc}}\right)\right)}.$$

From Lemma 2.1.12 and Theorem 2.1.13, we have that

$$1 - \frac{1}{4^n} \leq \mathrm{F}\left(\rho_7, \rho'_7\right)^2 \leq 1 - d\left(\rho_7, \rho'_7\right)^2$$

from which it follows that

$$d\left(\rho_7, \rho'_7\right) \leq \frac{1}{2^n}.$$

Now suppose that before entering step 8 the state of the system is $\rho'_7$ instead of $\rho_7$. This will result in an additional bias of at most $2^{-n}$ in the trace distance in the rest of the states that we calculate. Since $\rho'_7$ lies in the rejection subspace,

$$\left(\left(\mathbb{1} - 2\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}_1}\right) \otimes \mathbb{1}_{\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right) \rho'_7 \left(\left(\mathbb{1} - 2\Pi_{\text{acc}} \otimes |1\rangle\langle 1|_{\mathcal{S}_1}\right) \otimes \mathbb{1}_{\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right) = \rho'_7$$

which means that step 8 doesn't change the state. So the state of the system before entering step 9 is $\rho'_7$. Let us change the state again, at this time from $\rho'_7$ back to $\rho_7$. This will result in another bias of at most $2^{-n}$. If the state of the system is $\rho_7$ before entering step 9 then the state after step 9 will be

$$\left(\mathbf{V}_x^* \otimes \mathbb{1}_{\mathcal{S}_1\otimes\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right) \rho_7 \left(\mathbf{V}_x \otimes \mathbb{1}_{\mathcal{S}_1\otimes\mathcal{S}'_1\otimes\mathcal{S}_2\otimes\mathcal{S}'_2}\right) = |\bar{0}\rangle\langle\bar{0}|_{\mathcal{A}} \otimes \rho_5.$$

From Lemma 3.2.6, together with the assumption we made about the success probability of the SWAP Test, we get that there exists a set of states

$\{|\varphi_i\rangle : |\varphi_i\rangle \in \mathcal{W}^+ \text{ or } |\varphi_i\rangle \in \mathcal{W}^-\}$ such that

$$d\left(\mathrm{Tr}_{\mathcal{P}}\left(\rho_2'\right), \sum_{i=1}^{m} p_i \left(|\varphi_i\rangle\langle\varphi_i|\right)^{\otimes 2}\right) \leq 6\sqrt{\varepsilon}. \tag{3.7}$$

This implies that

$$d(\mathrm{Tr}_{\mathcal{P}}(\rho_5), \rho_9) \leq 6\sqrt{\varepsilon}$$

where

$$\rho_9 \stackrel{\text{def}}{=} \sum_{i=1}^{m} p_i \left(\mathbf{Q}\,|\varphi_i\rangle\langle\varphi_i|\,\mathbf{Q}^*\right) \otimes |\varphi_i\rangle\langle\varphi_i| \in \mathrm{D}\left(\mathcal{S}_1 \otimes \mathcal{S}_1' \otimes \mathcal{S}_2 \otimes \mathcal{S}_2'\right).$$

Now let us change the state of $(\mathsf{S}_1, \mathsf{S}_1', \mathsf{S}_2, \mathsf{S}_2')$ from $\mathrm{Tr}_{\mathcal{P}}(\rho_5)$ to $\rho_9$. This will result in another bias of at most $6\sqrt{\varepsilon}$. (Note that $\mathsf{P}$ is not touched by the algorithm after step 9, so we don't keep track of its state.) From Eqs. (3.5) to (3.7), it follows that

$$d\left(\mathrm{Tr}_{\mathcal{S}_1 \otimes \mathcal{S}_2}\left(\sum_{i=1}^{m} p_i \left(|\varphi_i\rangle\langle\varphi_i|\right)^{\otimes 2}\right), \frac{\mathbb{1}_{\mathcal{S}_1' \otimes \mathcal{S}_2'}}{4}\right) \leq \sqrt{\frac{32}{N}} + 6\sqrt{\varepsilon} < \frac{1}{16}.$$

So $\rho_9$ satisfies the requirements of Lemma 3.2.5 below. This means that Algorithm 3 in step 10 succeeds with probability at least $2^{-25}$, in which case we continue to step 14.

We now argue that, conditioned on Algorithm 3 being successful, the measurement in step 14 outputs $\bar{0}$ with certainty. This will finish the proof. Note that Algorithm 3 can't change the state of $\mathsf{A}$ as it was independent of $(\mathsf{S}_2, \mathsf{S}_2', \mathsf{S}_1)$ before executing Algorithm 3. So before entering step 14, the state of $\mathsf{A}$ is still $|\bar{0}\rangle$. Now we argue that after successfully executing Algorithm 3, the state of $\mathsf{S}_2$ will be $|0\rangle$. Let us take some $|\varphi\rangle \in \mathcal{S}_1 \otimes \mathcal{S}_1'$ that belongs to either $\mathcal{W}^+$ or $\mathcal{W}^-$. Here we only argue about the case when $|\varphi\rangle \in \mathcal{W}^+$ as the other case can be proven by exactly the same way. We can write $|\varphi\rangle$ as

$$|\varphi\rangle = a\left|\Phi^+\right\rangle + b\left|\Psi^+\right\rangle, \qquad a, b \in \mathbb{C}, \qquad |a|^2 + |b|^2 = 1.$$

It is easy to see that after applying $\mathbf{Q}$ to $|\varphi\rangle$, the resulting state on $\mathsf{S}_1$ will be $a\,|0\rangle - b\,|1\rangle$. Suppose that the state of $(\mathsf{S}_2, \mathsf{S}_2')$ is $|\varphi\rangle$ and the state of $\mathsf{S}_1$ is $a\,|0\rangle - b\,|1\rangle$. It can be shown by direct calculation that

$$\left(|1\rangle\langle 1|_{\mathcal{S}_2} \otimes \left|\Phi^+\right\rangle\left\langle\Phi^+\right|_{\mathcal{S}_2' \otimes \mathcal{S}_1}\right)|\varphi\rangle \otimes (a\,|0\rangle - b\,|1\rangle) = 0.$$

This means that if Algorithm 3 is executed with the above input and the measurement in the algorithm results in $|\Phi^+\rangle$, then the state of $\mathsf{S}_2$ will be $|0\rangle$. Similarly to the above, it can also be shown that

$$\left(|0\rangle\langle 0|_{\mathcal{S}_2} \otimes \left|\Psi^+\right\rangle\left\langle\Psi^+\right|_{\mathcal{S}'_2\otimes\mathcal{S}_1}\right)|\varphi\rangle \otimes (a\,|0\rangle - b\,|1\rangle) = 0.$$

This means that if the measurement in Algorithm 3 results in $|\Psi^+\rangle$ then the state of $\mathsf{S}_2$ will be $|1\rangle$. In this case, Algorithm 3 applies $\mathbf{X}$ on $\mathsf{S}_2$ so the state of this register, after the algorithm, will be $|0\rangle$. Since $\rho_9$ is a convex combination of states of the above form, we got that if the state of $(\mathsf{S}_1, \mathsf{S}'_1, \mathsf{S}_2, \mathsf{S}'_2)$ is $\rho_9$, before entering step 10, then Algorithm 3 succeeds with probability at least $2^{-25}$ and, conditioned on success, Algorithm 4 rejects in step 16 with certainty.

However, we did modify the state during our analysis four times, so we have to account for the bias they caused, which is at most

$$\frac{1}{2^{n-1}} + \sqrt{\frac{32}{N}} + 6\sqrt{\varepsilon}.$$

So the real rejection probability, with the original input, is at least

$$\frac{1}{2^{25}} - \left(\frac{1}{2^{n-1}} + \sqrt{\frac{32}{N}} + 6\sqrt{\varepsilon}\right) > \frac{1}{2^{26}} - \frac{1}{2^{n-1}} \geq \frac{1}{2^{27}}$$

where the last inequality is true for $n \geq 28$. ❑

**Lemma 3.2.5.** *Suppose that before entering step 10 of Algorithm 4, the state of* $(\mathsf{S}_1, \mathsf{S}'_1, \mathsf{S}_2, \mathsf{S}'_2)$ *is*

$$\rho \stackrel{\text{def}}{=} \sum_{i=1}^{m} p_i \left(\mathbf{Q}\sigma_i\mathbf{Q}^*\right) \otimes \sigma_i$$

*for some* $m \in \mathbb{Z}^+$, *probability distribution* $\{p_i : i = 1, \ldots, m\}$, *and states* $\sigma_i \in \mathrm{D}(\mathcal{S}_2 \otimes \mathcal{S}'_2) \cong \mathrm{D}(\mathcal{S}_1 \otimes \mathcal{S}'_1)$. *Further assume that*

$$d\left(\mathrm{Tr}_{\mathcal{S}_1\otimes\mathcal{S}_2}\left(\sum_{i=1}^{m} p_i\sigma_i \otimes \sigma_i\right), \frac{\mathbb{1}_{\mathcal{S}'_1\otimes\mathcal{S}'_2}}{4}\right) \leq \frac{1}{16}. \tag{3.8}$$

*Then Algorithm 3, in step 10, will succeed with probability at least* $2^{-25}$.

The idea behind the proof of Lemma 3.2.5 is very simple. We show that if the measurement in Algorithm 3 fails with high probability on a state of the form $\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \otimes \zeta$, where $\zeta \in \mathrm{D}(\mathcal{S}_1)$ is an arbitrary state, then $\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i)$ must

be close to either $|\phi^+\rangle$ or $|\phi^-\rangle$. But then the convex combination of the states $\mathrm{Tr}_{\mathcal{S}_1}(\sigma_i) \otimes \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i)$ won't be close to the maximally mixed state.

*Proof of Lemma 3.2.5.* Let us group the states in ensemble $\rho$ with respect to their reduced state on $\mathcal{S}_2'$ being close to $|\phi^+\rangle$, or to $|\phi^-\rangle$, or being far from both. Formally, let $\varepsilon_1 \stackrel{\text{def}}{=} 2^{-11}$ and

$$\begin{aligned} A^+ &\stackrel{\text{def}}{=} \left\{ i \,:\, 1 \le i \le m,\ d\left(\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i), |\phi^+\rangle\langle\phi^+|\right) \le \varepsilon_1 \right\} \\ A^- &\stackrel{\text{def}}{=} \left\{ i \,:\, 1 \le i \le m,\ d\left(\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i), |\phi^-\rangle\langle\phi^-|\right) \le \varepsilon_1 \right\} \\ B &\stackrel{\text{def}}{=} \{1, 2, \ldots, m\} \setminus \left(A^+ \cup A^-\right). \end{aligned}$$

Since $d(|\phi^+\rangle, |\phi^-\rangle) = 1$ and $\varepsilon_1 < 1/2$, from the triangle inequality we have that $A^+ \cap A^- = \emptyset$.

We first show that if the probability of $B$ is at least $\varepsilon_2 \stackrel{\text{def}}{=} 1/8$ then we are done. So assume for now that $\varepsilon_2 \le \sum_{i \in B} p_i$. For all $i \in B$ we have that

$$\begin{aligned} \sqrt{\langle\phi^+| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) |\phi^+\rangle} &= \mathrm{F}\left(\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i), |\phi^+\rangle\langle\phi^+|\right) && (3.9) \\ &\le \sqrt{1 - d(\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i), |\phi^+\rangle\langle\phi^+|)^2} && (3.10) \\ &< \sqrt{1 - \varepsilon_1^2} && (3.11) \end{aligned}$$

where Eq. (3.9) follows from Eq. (2.2), Eq. (3.10) follows from Theorem 2.1.13, and Eq. (3.11) is from the definition of $B$. The above implies that

$$\left\langle\phi^+\right| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \left|\phi^+\right\rangle < 1 - \varepsilon_1^2 \qquad \text{and similarly} \qquad \left\langle\phi^-\right| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \left|\phi^-\right\rangle < 1 - \varepsilon_1^2.$$

From the above and using the fact that

$$\left\langle\phi^+\right| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \left|\phi^+\right\rangle + \left\langle\phi^-\right| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \left|\phi^-\right\rangle = \mathrm{Tr}(\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i)) = 1$$

we get that

$$\varepsilon_1^2 < \left\langle\phi^+\right| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \left|\phi^+\right\rangle \qquad \text{and} \qquad \varepsilon_1^2 < \left\langle\phi^-\right| \mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \left|\phi^-\right\rangle.$$

Let us take an arbitrary state

$$|\psi\rangle \stackrel{\text{def}}{=} a \left|\phi^+\right\rangle + b \left|\phi^-\right\rangle \in \mathcal{S}_1, \qquad a, b \in \mathbb{C}, \qquad |a|^2 + |b|^2 = 1.$$

If the state of $(\mathsf{S}_2', \mathsf{S}_1)$, in the input to Algorithm 3, is $\mathrm{Tr}_{\mathcal{S}_2}(\sigma_i) \otimes |\psi\rangle\langle\psi|$ then the

algorithm will succeed with probability

$$\begin{aligned}\operatorname{Tr}\left(\left(\operatorname{Tr}_{\mathcal{S}_2}(\sigma_i)\otimes|\psi\rangle\langle\psi|\right)\Pi^+\right) &= |a|^2\cdot\langle\phi^+|\operatorname{Tr}_{\mathcal{S}_2}(\sigma_i)|\phi^+\rangle + |b|^2\cdot\langle\phi^-|\operatorname{Tr}_{\mathcal{S}_2}(\sigma_i)|\phi^-\rangle\\ &> \varepsilon_1^2\left(|a|^2+|b|^2\right)\\ &= \varepsilon_1^2\end{aligned}$$

where the first equality follows from direct calculation using

$$\begin{aligned}|\Phi^+\rangle &= \frac{|\phi^+\rangle\otimes|\phi^+\rangle + |\phi^-\rangle\otimes|\phi^-\rangle}{\sqrt{2}}\\ |\Psi^+\rangle &= \frac{|\phi^+\rangle\otimes|\phi^+\rangle - |\phi^-\rangle\otimes|\phi^-\rangle}{\sqrt{2}}.\end{aligned}$$

This implies that if the state of $(\mathsf{S}_2', \mathsf{S}_1)$ is $\operatorname{Tr}_{\mathcal{S}_2}(\sigma_i)\otimes\zeta$, for any $\zeta\in\mathrm{D}(\mathcal{S}_1)$, then the probability that Algorithm 3 succeeds is at least $\varepsilon_1^2$. We got that if $\varepsilon_2 \le \sum_{i\in B} p_i$ then Algorithm 3 succeeds with probability at least $\varepsilon_1^2\varepsilon_2 = 2^{-25}$, in which case we are done.

So, from now on, assume that $\sum_{i\in B} p_i < \varepsilon_2$. We will show that this assumption leads to a contradiction, which will finish the proof. Lemma 2.1.14 implies that

$$\begin{aligned}\forall i\in A^+,\ \exists\tau_i\in\mathrm{D}(\mathcal{S}_2)\ :\ d\left(\sigma_i, \tau_i\otimes|\phi^+\rangle\langle\phi^+|\right) &\le \sqrt{2\varepsilon_1}\\ \forall i\in A^-,\ \exists\tau_i\in\mathrm{D}(\mathcal{S}_2)\ :\ d\left(\sigma_i, \tau_i\otimes|\phi^-\rangle\langle\phi^-|\right) &\le \sqrt{2\varepsilon_1}.\end{aligned}$$

We now replace $\sigma_i$ with $\tau_i\otimes|\phi^+\rangle\langle\phi^+|$ or $\tau_i\otimes|\phi^-\rangle\langle\phi^-|$ in $\rho$. Formally, let us define

$$\begin{aligned}\mu_B &\stackrel{\text{def}}{=} \sum_{i\in B} p_i\left(\mathbf{Q}\sigma_i\mathbf{Q}^*\right)\otimes\sigma_i\\ \rho' &\stackrel{\text{def}}{=} \sum_{i\in A^+} p_i\left(\mathbf{Q}\left(\tau_i\otimes|\phi^+\rangle\langle\phi^+|\right)\mathbf{Q}^*\right)\otimes\tau_i\otimes|\phi^+\rangle\langle\phi^+|\\ &\quad + \sum_{i\in A^-} p_i\left(\mathbf{Q}\left(\tau_i\otimes|\phi^-\rangle\langle\phi^-|\right)\mathbf{Q}^*\right)\otimes\tau_i\otimes|\phi^-\rangle\langle\phi^-|\\ &\quad + \mu_B\end{aligned}$$

where $\operatorname{Tr}(\mu_B) < \varepsilon_2$. Note that $d(\rho,\rho') < 2\sqrt{2\varepsilon_1}$ which, together with Eq. (3.8), implies that

$$d\left(\xi, \frac{\mathbb{1}_{\mathcal{S}_1'\otimes\mathcal{S}_2'}}{4}\right) \le 2\sqrt{2\varepsilon_1} + \frac{1}{16} = \frac{1}{8} \tag{3.12}$$

where

$$\xi \stackrel{\text{def}}{=} \mathrm{Tr}_{\mathsf{S}_1\otimes\mathsf{S}_2}\left(\left(\mathbf{Q}^*\otimes \mathbb{1}_{\mathsf{S}_2\otimes\mathsf{S}_2'}\right)\rho'\left(\mathbf{Q}\otimes\mathbb{1}_{\mathsf{S}_2\otimes\mathsf{S}_2'}\right)\right).$$

On the other hand, we have that

$$\xi = p_+\left(|\phi^+\rangle\langle\phi^+|\right)^{\otimes 2} + p_-\left(|\phi^-\rangle\langle\phi^-|\right)^{\otimes 2} + \nu_B$$

for some $\nu_B$, where we used the shorthand $p_+ \stackrel{\text{def}}{=} \sum_{i\in A^+} p_i$ and $p_- \stackrel{\text{def}}{=} \sum_{i\in A^-} p_i$. Note that $\mathrm{Tr}(\nu_B) < \varepsilon_2$, so Lemma 2.1.10 implies that

$$d\left(\xi, p_+\left(|\phi^+\rangle\langle\phi^+|\right)^{\otimes 2} + p_-\left(|\phi^-\rangle\langle\phi^-|\right)^{\otimes 2}\right) \leq \frac{\varepsilon_2}{2}. \tag{3.13}$$

The following calculation will lead us to a contradiction.

$$\begin{aligned}
\frac{1}{4} &\leq \frac{1}{2}\left(\left|\frac{1}{4}-p_+\right| + \left|\frac{1}{4}-p_-\right| + \frac{1}{2}\right)\\
&= \frac{1}{2}\left\|\frac{\mathbb{1}_{\mathsf{S}_1'\otimes\mathsf{S}_2'}}{4} - \left(p_+\left(|\phi^+\rangle\langle\phi^+|\right)^{\otimes 2} + p_-\left(|\phi^-\rangle\langle\phi^-|\right)^{\otimes 2}\right)\right\|_{\mathrm{Tr}} &(3.14)\\
&= d\left(\frac{\mathbb{1}_{\mathsf{S}_1'\otimes\mathsf{S}_2'}}{4}, p_+\left(|\phi^+\rangle\langle\phi^+|\right)^{\otimes 2} + p_-\left(|\phi^-\rangle\langle\phi^-|\right)^{\otimes 2}\right)\\
&\leq d\left(\xi, \frac{\mathbb{1}_{\mathsf{S}_1'\otimes\mathsf{S}_2'}}{4}\right) + d\left(\xi, p_+\left(|\phi^+\rangle\langle\phi^+|\right)^{\otimes 2} + p_-\left(|\phi^-\rangle\langle\phi^-|\right)^{\otimes 2}\right) &(3.15)\\
&\leq d\left(\xi, \frac{\mathbb{1}_{\mathsf{S}_1'\otimes\mathsf{S}_2'}}{4}\right) + \frac{\varepsilon_2}{2} &(3.16)
\end{aligned}$$

where Eq. (3.14) is because the eigenvalues of

$$\frac{\mathbb{1}_{\mathsf{S}_1'\otimes\mathsf{S}_2'}}{4} - \left(p_+\left(|\phi^+\rangle\langle\phi^+|\right)^{\otimes 2} + p_-\left(|\phi^-\rangle\langle\phi^-|\right)^{\otimes 2}\right)$$

are $\frac{1}{4}-p_+$, $\frac{1}{4}-p_-$, and $\frac{1}{4}$ with multiplicity 2. Equation (3.15) follows from the triangle inequality and at Eq. (3.16) we used Eq. (3.13). Equation (3.16) implies that

$$d\left(\xi, \frac{\mathbb{1}_{\mathsf{S}_1'\otimes\mathsf{S}_2'}}{4}\right) \geq \frac{1}{4} - \frac{\varepsilon_2}{2} = \frac{3}{16}$$

which contradicts to Eq. (3.12). So we conclude that it must be that $\varepsilon_2 \leq \sum_{i\in B} p_i$, in which case Algorithm 3 succeeds with the desired probability, as argued above. ❑

The following lemma is similar to Proposition 24 of [KLGN13].

**Lemma 3.2.6.** *Let* $\mathsf{S}_1$, $\mathsf{S}_1'$, $\mathsf{S}_2$, *and* $\mathsf{S}_2'$ *be single-qubit registers and let the state of*

$(\mathsf{S}_1, \mathsf{S}_1', \mathsf{S}_2, \mathsf{S}_2')$ *be*

$$\rho \stackrel{\text{def}}{=} \sum_{i=1}^{m} p_i \sigma_i \otimes \sigma_i$$

*where* $m \in \mathbb{Z}^+$, $\{p_i : i = 1, \ldots, m\}$ *is a probability distribution, and* $\sigma_i = \Psi(\xi_i)$, *for some* $\xi_i \in \mathrm{D}(\mathcal{S}_1 \otimes \mathcal{S}_1') \cong \mathrm{D}(\mathcal{S}_2 \otimes \mathcal{S}_2')$. *Let* $0 \leq \varepsilon < 1$. *If the SWAP Test, applied between* $(\mathsf{S}_1, \mathsf{S}_1')$ *and* $(\mathsf{S}_2, \mathsf{S}_2')$, *succeeds with probability at least* $1 - \varepsilon$ *then there exist a set of states*

$$\left\{ |\varphi_i\rangle : 1 \leq i \leq m,\ |\varphi_i\rangle \in \mathcal{W}^+ \text{ or } |\varphi_i\rangle \in \mathcal{W}^- \right\}$$

*such that*

$$d\left( \rho, \sum_{i=1}^{m} p_i \, |\varphi_i\rangle\langle\varphi_i| \otimes |\varphi_i\rangle\langle\varphi_i| \right) \leq 6\sqrt{\varepsilon}.$$

*Proof.* On input $\sigma_i \otimes \sigma_i$ the SWAP Test succeeds with probability $\left(1 + \mathrm{Tr}\left(\sigma_i^2\right)\right)/2$, by Theorem 2.1.21. So with input $\rho$ the SWAP Test succeeds with probability

$$\sum_{i=1}^{m} p_i \frac{1 + \mathrm{Tr}\left(\sigma_i^2\right)}{2} \geq 1 - \varepsilon.$$

If $\varepsilon = 0$ it implies that all $\sigma_i$'s are pure and the statement of the lemma follows. So, from now on, assume that $0 < \varepsilon$. Then the above inequality intuitively means that for most of the $i$'s, $\mathrm{Tr}\left(\sigma_i^2\right)$ must be close to 1. Formally, let

$$\begin{aligned} B &\stackrel{\text{def}}{=} \left\{ i : 1 \leq i \leq m,\ \mathrm{Tr}\left(\sigma_i^2\right) \leq 1 - 2\sqrt{\varepsilon} \right\} \\ A &\stackrel{\text{def}}{=} \{1, 2, \ldots, m\} \setminus B. \end{aligned}$$

Suppose towards contradiction that $2\sqrt{\varepsilon} \leq \sum_{i \in B} p_i$. Then the probability that the SWAP Test fails is

$$\begin{aligned} \sum_{i=1}^{m} p_i \frac{1 - \mathrm{Tr}\left(\sigma_i^2\right)}{2} &\geq \sum_{i \in B} p_i \frac{1 - \mathrm{Tr}\left(\sigma_i^2\right)}{2} \\ &\geq \sum_{i \in B} p_i \frac{1 - \left(1 - 2\sqrt{\varepsilon}\right)}{2} \\ &\geq \sqrt{\varepsilon} \cdot \sum_{i \in B} p_i \\ &\geq 2\varepsilon \end{aligned}$$

which is a contradiction. This implies that $\sum_{i \in B} p_i < 2\sqrt{\varepsilon}$. For all $i \in A$, let $\lambda_i$ be the maximum eigenvalue of $\sigma_i$ and $|\varphi_i\rangle$ be the corresponding eigenstate.

Note that either $|\varphi_i\rangle \in \mathcal{W}^+$ or $|\varphi_i\rangle \in \mathcal{W}^-$. From the definition of $A$, we have that

$$1 - 2\sqrt{\varepsilon} < \mathrm{Tr}\left(\sigma_i^2\right) \leq \|\sigma_i\|_{\mathrm{Tr}} \cdot \|\sigma_i\|_\infty = \|\sigma_i\|_\infty = \lambda_i$$

where the second inequality follows from Lemma 2.1.3. The above calculation, together with Lemma 2.1.11, imply that

$$\forall i \in A \,:\, d(\sigma_i, |\varphi_i\rangle\langle\varphi_i|) \leq 2\sqrt{\varepsilon}. \tag{3.17}$$

For all $i \in B$, let $|\varphi_i\rangle$ be an arbitrary state from $\mathcal{W}^+$ or $\mathcal{W}^-$. We can now bound the required trace distance.

$$\begin{aligned} d\left(\rho, \sum_{i=1}^m p_i \left(|\varphi_i\rangle\langle\varphi_i|\right)^{\otimes 2}\right) &\leq d\left(\sum_{i=1}^m p_i \sigma_i^{\otimes 2}, \sum_{i\in A} p_i \sigma_i^{\otimes 2}\right) \\ &\quad + d\left(\sum_{i\in A} p_i \sigma_i^{\otimes 2}, \sum_{i\in A} p_i \left(|\varphi_i\rangle\langle\varphi_i|\right)^{\otimes 2}\right) \\ &\quad + d\left(\sum_{i\in A} p_i \left(|\varphi_i\rangle\langle\varphi_i|\right)^{\otimes 2}, \sum_{i=1}^m p_i \left(|\varphi_i\rangle\langle\varphi_i|\right)^{\otimes 2}\right) \quad (3.18) \\ &\leq 6\sqrt{\varepsilon} \quad (3.19) \end{aligned}$$

where Eq. (3.18) follows from the triangle inequality and at Eq. (3.19) we used Lemma 2.1.10 twice and Eq. (3.17). ❑

## 3.3 Multi-Prover QMA with Small Gap

In this section, we study multiple-proof quantum Merlin-Arthur proof systems in the setting where the completeness-soundness gap is small. Small means that we only lower bound the gap with an inverse-exponential function of the input length, or with an even smaller function. In Section 3.3.1, we observe that the protocol of Blier and Tapp [BT12] scales up which implies that, in this case, the proof system has the same expressive power as non-deterministic exponential time. Since single-proof QMA proof systems, with the same bound on the gap, have expressive power at most exponential time, we get a separation between single and multi-prover proof systems in the 'small-gap setting' under the assumption that EXP $\neq$ NEXP. This implies, among others, the nonexistence of certain operators called disentanglers (defined by Aaronson et al. [ABD$^+$09]) with good approximation parameters. These conclusions are discussed in

Section 3.3.3.

In Section 3.3.2, we show that, in the above setting, the proof system has the same expressive power if we restrict the verifier to be able to perform only Bell-measurements, i.e., using a BellQMA verifier. In the usual setting, when the gap is bounded by at least an inverse-polynomial function of the input length, BellQMA with polynomially-many provers collapses to single-prover QMA [BH13], but here, in the small-gap setting, it has the full power of multi-prover QMA. To show this, we use the protocol of Chen and Drucker [CD10] with a similar but simpler analysis. The only caveat here is that we need at least some super-constant number of proofs to achieve the power of NEXP while in the previous setting two proofs were enough. Section 3.3 is based on Ref. [Per12a].

Before we begin, let us introduce a simple notation that will be used throughout the section.

**Definition 3.3.1.** We define the state $|u_m\rangle$ as the uniform superposition of the standard basis states. That is,

$$|u_m\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{m}} \sum_{i=0}^{m-1} |i\rangle \,.$$

We also define the projective measurement that projects onto this state, or more formally the measurement $\{\mathbf{P}_0, \mathbf{P}_1\}$ where $\mathbf{P}_0 \stackrel{\text{def}}{=} |u_m\rangle\langle u_m|$ and $\mathbf{P}_1 \stackrel{\text{def}}{=} \mathbb{1} - |u_m\rangle\langle u_m|$. We say that $\mathbf{P}_0$ (and $\mathbf{P}_1$) corresponds to outcome 0 (and 1).

Note that the above measurement can be performed using $\lceil \log m \rceil$ Hadamard gates and single-qubit measurements.

### 3.3.1 QMA$[k]$ with Small Gap Equals NEXP

This section proves Theorem 1.1.4, i.e., we show that QMA$[k]$ equals NEXP if $k$ is at least 2 and at most poly$(n)$ and the completeness-soundness gap is bounded away by an inverse-exponential or doubly exponential function of $n$. The theorem is restated here for convenience.

**Theorem 1.1.4.** *For all $\varepsilon > 0$, it holds that*

$$\mathsf{NEXP} = \mathsf{QMA}\left(2, 1, 1 - 2^{-n^{\varepsilon}}\right) = \bigcup_{\substack{0<s<c\leq 1,\\ c-s\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{QMA}(\mathsf{poly}, c, s)$$

*where $c(n)$ and $s(n)$ can be calculated in time at most exponential in $n$ on a classical computer.*

The proof of this theorem is divided into Lemma 3.3.2 and Theorem 3.3.4, according to the two directions of the containment. This proof is essentially a scaled version of the proof of Blier and Tapp [BT12].

**Lemma 3.3.2.** $\bigcup_{\substack{0<s<c\leq 1,\\ c-s\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{QMA}(\mathsf{poly}, c, s) \subseteq \mathsf{NEXP}$, *where $c(n)$ and $s(n)$ can be calculated in time at most exponential in $n$ on a classical computer.*

*Proof sketch.* Let $L \in \mathsf{QMA}(\mathsf{poly}, c, s)$ with some $c$ and $s$ satisfying the conditions in the lemma. The proofs in the QMA proof system are polynomially-many quantum states on polynomially-many qubits, which are vectors in the complex euclidean space with exponential dimension. These vectors can be described up to an exponential number of bits of accuracy by a classical proof of exponential length. Given this proof to an exponential-time classical computer, it can calculate the acceptance probability of the QMA verifier to an exponential number of bits of accuracy and it can decide whether this probability is more than $c$ or less than $s$. This means that $L \in \mathsf{NEXP}$. ❑

The following lemma is the key for the other direction of the containment.

**Lemma 3.3.3.** Succinct3Col $\in \mathsf{QMA}\left(2, 1, 1 - \Omega\left(4^{-\tilde{n}}\right)\right)$, *where $\tilde{n}$ is the length of both of the inputs of the circuit representing the graph.*

Before we prove this lemma let us see how it is used to prove Theorem 1.1.4. The other direction of the containment is formulated by the following theorem.

**Theorem 3.3.4.** *For all $\varepsilon > 0$, it holds that* $\mathsf{NEXP} \subseteq \mathsf{QMA}\left(2, 1, 1 - 2^{-n^{\varepsilon}}\right)$.

*Proof.* By Lemma 3.3.3, Succinct3Col $\in \mathsf{QMA}\left(2, 1, 1 - K \cdot 4^{-\tilde{n}}\right)$, for some constant $K$. By Definition 2.2.15, there exists $\kappa \in \mathbb{Z}^+$ such that $n \leq \tilde{n}^{\kappa}$, where $n$ is the size of the input. Let's pick an arbitrary constant $\kappa' \in \mathbb{Z}^+$ for which $\kappa' \geq \max\left\{\kappa, \frac{2}{\varepsilon}\right\}$. Now suppose that we modify the language Succinct3Col by inflating the input circuits of size $n \leq \tilde{n}^{\kappa} \leq \tilde{n}^{\kappa'}$ to be $\geq \tilde{n}^{\kappa'}$. This can be done by adding dummy gates to the circuit, for example adding an even number of NOT gates to the first input wire. For this modified language we have that the input size $n$ is at least $\tilde{n}^{\kappa'}$. The soundness parameter is upper bounded by

$$1 - K \cdot 4^{-\tilde{n}} \leq 1 - K \cdot 4^{-n^{1/\kappa'}} \leq 1 - 2^{-n^{2/\kappa'}} \leq 1 - 2^{-n^{\varepsilon}}$$

for sufficiently large $n$. This, together with the fact that the modified language is still NEXP-complete, implies the statement of the theorem. ❑

*Proof of Theorem 1.1.4.* The theorem immediately follows from Lemma 3.3.2, Theorem 3.3.4, and the trivial observation that for constant $\varepsilon$,

$$\mathsf{QMA}\left(2,1,1-2^{-n^{\varepsilon}}\right) \subseteq \bigcup_{\substack{0<s<c\leq 1,\\ c-s\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{QMA}(\mathsf{poly},c,s)\,. \qquad \text{❑}$$

We now prove the key lemma.

*Proof of Lemma 3.3.3.* We give a QMA verifier $V$ for the language Succinct3Col. We use the protocol of Blier and Tapp [BT12], with essentially the same analysis as theirs. Let the input to Succinct3Col be denoted by $C_G$ and its length by $n$. With our notation, $C_G$ has two inputs of $\tilde{n}$ bits each. Let $V$ get his two unentangled proofs in registers $\mathsf{R}_1$ and $\mathsf{R}_2$. Both $\mathsf{R}_i$'s have two parts, $\mathsf{R}_i = \mathsf{N}_i\mathsf{C}_i$, where $\mathsf{N}_i$ is the 'node' part and $\mathsf{C}_i$ is the 'color' part. $\mathsf{N}_1$ and $\mathsf{N}_2$ have associated Hilbert space $\mathbb{C}^{2^{\tilde{n}}}$ while $\mathsf{C}_1$ and $\mathsf{C}_2$ have associated space $\mathbb{C}^3$. The procedure $V$ performs is described in Algorithm 5.

Note that $V$ runs in $\mathsf{poly}(n)$-time because the SWAP Test, evaluating the circuit $C_G$, and performing the measurement of Definition 3.3.1, for $m = 2^{\tilde{n}}$, can all be performed in polynomial time. We are left to prove completeness and soundness. The completeness is formulated by Lemma 3.3.5 and the soundness is by Lemma 3.3.6. ❑

**Lemma 3.3.5** (Completeness)**.** *If $C_G \in$ Succinct3Col then there exist a pair of proofs with which $V$ will accept with probability* 1*.*

**Lemma 3.3.6** (Soundness)**.** *If $C_G \notin$ Succinct3Col then verifier $V$ described by Algorithm 5 will reject with probability at least $\frac{1}{3\cdot 10^{10}\cdot 4^{\tilde{n}}}$.*

The proofs of Lemmas 3.3.5 and 3.3.6 are presented in Appendix A.1 on page 83.

### 3.3.2 $\mathsf{BellQMA}[n^{\varepsilon}]$ with Small Gap Equals NEXP

In this section we prove Theorem 1.1.5, i.e., we show that multi-prover QMA with exponentially small gap still equals to NEXP if we restrict the verifier to only be able to perform Bell-measurements. However, we will need at least $n^{\varepsilon}$ proofs. The theorem is restated below.

**Theorem 1.1.5.** *For any $\varepsilon, \delta > 0$, it holds that*

$$\mathsf{NEXP} = \mathsf{BellQMA}(n^{\varepsilon},c,s) = \bigcup_{\substack{0<s'<c'\leq 1,\\ c'-s'\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{BellQMA}\left(\mathsf{poly},c',s'\right)$$

**Algorithm 5** Description of verifier $V$ in the proof of Lemma 3.3.3.

**INPUT:** classical circuit $C_G$, quantum registers $\mathsf{R}_1 = \mathsf{N}_1\mathsf{C}_1$ and $\mathsf{R}_2 = \mathsf{N}_2\mathsf{C}_2$, where the state of $\mathsf{R}_1$ and $\mathsf{R}_2$ is separable

**OUTPUT:** accept or reject

1: With probability 1/3 do the **Equality Test** (line 2), the **Consistency Test** (line 8), or the **Uniformity Test** (line 16).
2: **Equality Test.** Perform the SWAP Test on $\mathsf{R}_1$ and $\mathsf{R}_2$.
3: **IF** the SWAP Test fails **THEN**
4: **RETURN** reject *{The two registers are not equal.}*
5: **ELSE**
6: **RETURN** accept
7: **END IF**
8: **Consistency Test.** Measure $\mathsf{N}_1$, $\mathsf{C}_1$, $\mathsf{N}_2$, and $\mathsf{C}_2$ in the computational basis and denote the outcomes by $v_1$, $c_1$, $v_2$, and $c_2$.
9: **IF** $(v_1 = v_2)$ **AND** $(c_1 \neq c_2)$ **THEN**
10: **RETURN** reject *{The same vertex has two colors.}*
11: **ELSE IF** $(\mathrm{C}(v_1, v_2) = 11)$ *{Assume that $v_1 < v_2$ otherwise swap them.}* **AND** $(c_1 = c_2)$ **THEN**
12: **RETURN** reject *{Adjacent vertices have same color.}*
13: **ELSE**
14: **RETURN** accept
15: **END IF**
16: **Uniformity Test.** Measure $\mathsf{N}_1$ and $\mathsf{C}_1$ separately according to the measurement of Definition 3.3.1.
17: **IF** (the outcome on $\mathsf{C}_1$ is 0) **AND** (the outcome on $\mathsf{N}_1$ is 1) **THEN**
18: **RETURN** reject *{Not all nodes are present.}*
19: **ELSE**
20: **RETURN** accept
21: **END IF**

*for some c and s with $c(n) - s(n) \geq 2^{-n^\delta}$ and where $c'(n)$ and $s'(n)$ can be calculated in time at most exponential in n on a classical computer.*

We essentially use the algorithm of Chen and Drucker [CD10] on the succinct version of graph 3-coloring (Succinct3Col). We also use one of their lemmas but our proof will be simpler than theirs because we don't aim for constant gap. We don't use the PCP theorem either. Note that, in the previous section, we already argued about the NEXP upper bound on the QMA classes. The same argument applies here too. It's also easy to see that restricting the verifier can only make the power of the proof system weaker. So the only statement left to prove, in order to prove Theorem 1.1.5, is the following.

**Theorem 3.3.7.** *For any $\varepsilon, \delta > 0$, it holds that*

$$\mathsf{NEXP} \subseteq \mathsf{BellQMA}(n^\varepsilon, c, s)$$

*for some c and s with $c(n) - s(n) \geq 2^{-n^\delta}$.*

Just as in the previous section, in order to show the above theorem it is enough to prove the following lemma. The argument is the same as in the previous section so we omit it from here.

**Lemma 3.3.8.** Succinct3Col $\in \mathsf{BellQMA}(\Omega(\tilde{n}), c, s)$*, for some c and s with $c(n) - s(n) = \Omega\left(4^{-\tilde{n}}\right)$ and where n is the size of the input circuit and $\tilde{n}$ is the length of both of the inputs of the circuit.*

*Proof.* We construct a BellQMA verifier $V$ for Succinct3Col. Just as in the previous section, let the input to Succinct3Col be denoted by $C_G$ and its length by $n$. Verifier $V$ will receive $k$ quantum proofs in registers $\mathsf{N}_1, \mathsf{C}_1, \ldots, \mathsf{N}_k, \mathsf{C}_k$ where, for each $i \in \{1, 2, \ldots, k\}$, the state of $\mathsf{N}_i\mathsf{C}_i$ is separable from the rest of the registers. We will set $k$, the number of provers, to be some function in $\Omega(\tilde{n})$ later. Registers $\mathsf{N}_i$ have associated space $\mathbb{C}^{2^{\tilde{n}}}$ and registers $\mathsf{C}_i$ have associated space $\mathbb{C}^3$ similarly as in the previous section. The behavior of $V$ is described in Algorithm 6.

Note that Algorithm 6 runs in polynomial time. Furthermore, for both the Consistency and the Uniformity Test, the algorithm starts with measuring all the quantum registers according to a fixed measurement. So $V$ is a proper BellQMA verifier. Lemma 3.3.9 below shows that the completeness of the protocol is $c > 1 - 2^{-\frac{k}{40}}$, while Lemma 3.3.10 shows that the soundness is $s < 1 - 12000^{-1} \cdot 2^{-2\tilde{n}}$. If $k \geq 120\tilde{n}$ then $c - s = \Omega\left(2^{-2\tilde{n}}\right)$ so the lemma follows. ❑

**Algorithm 6** Description of verifier $V$ in the proof of Lemma 3.3.8.

**INPUT:** classical circuit $C_G$, quantum registers $\mathsf{N}_1, \mathsf{C}_1, \ldots, \mathsf{N}_k, \mathsf{C}_k$ where, $\forall i \in \{1, 2, \ldots, k\}$, the state of $\mathsf{N}_i\mathsf{C}_i$ is separable from the rest of the registers

**OUTPUT:** accept or reject

1: With probability $\frac{1}{2}$ do the **Consistency Test** (line 2) or the **Uniformity Test** (line 14).
2: **Consistency Test.**
3: **FOR ALL** $i \in \{1, 2, \ldots, k\}$ **DO**
4: Measure $\mathsf{N}_i$ and $\mathsf{C}_i$ in the computational basis and get $v_i$ and $c_i$.
5: **END FOR**
6: **FOR ALL** $1 \leq i < j \leq k$ **DO**
7: **IF** $(v_i = v_j)$ **AND** $(c_i \neq c_j)$ **THEN**
8: **RETURN** reject *{The same vertex has two colors.}*
9: **ELSE IF** $(C_G(v_i, v_j) = 11)$ **AND** $(c_i = c_j)$ **THEN**
10: **RETURN** reject *{Adjacent vertices have same color.}*
11: **END IF**
12: **END FOR**
13: **RETURN** accept
14: **Uniformity Test.**
15: **FOR ALL** $i \in \{1, 2, \ldots, k\}$ **DO**
16: Measure $\mathsf{C}_i$ with the measurement of Definition 3.3.1 and denote the outcome by $x_i$.
17: Measure $\mathsf{N}_i$ with the measurement of Definition 3.3.1 and denote the outcome by $y_i$.
18: **END FOR**
19: Let $\mathscr{Z} \stackrel{\text{def}}{=} \{i : x_i = 0\}$.
20: **IF** $|\mathscr{Z}| < k/6$ **THEN**
21: **RETURN** reject
22: **END IF**
23: **FOR ALL** $i \in \mathscr{Z}$ **DO**
24: **IF** $y_i = 1$ **THEN**
25: **RETURN** reject *{Not all nodes are present.}*
26: **END IF**
27: **END FOR**
28: **RETURN** accept

**Lemma 3.3.9** (Completeness). *If $C_G \in$ Succinct3Col then there exist quantum states on registers $\mathsf{N}_1, \mathsf{C}_1, \ldots, \mathsf{N}_k, \mathsf{C}_k$, such that if they are input to $V$, defined by Algorithm 6, then $V$ will accept with probability at least $1 - 2^{-\frac{k}{40}}$.*

**Lemma 3.3.10** (Soundness). *If $C_G \notin$ Succinct3Col then $V$ of Algorithm 6 will reject with probability at least $12000^{-1} \cdot 4^{-\tilde{n}}$.*

The proofs of Lemmas 3.3.9 and 3.3.10 are deferred to Appendix A.2 on page 89.

### 3.3.3 Conclusions and Open Problems

In this section we discuss some of the consequences of the previous results, i.e., the consequences of Theorems 1.1.4 and 1.1.5. We also raise some related open problems.

#### Tightness of the Soundness Analyses

One can observe that both the $\mathsf{QMA}[2]$ verifier of Algorithm 5 and the $\mathsf{BellQMA}[k]$ verifier of Algorithm 6 have soundness parameter $1 - \Omega\left(4^{-\tilde{n}}\right)$ and gap $\Omega\left(4^{-\tilde{n}}\right)$. (As shown by Lemma 3.3.6 and Lemma 3.3.10.) Note that this bound is tight up to a constant factor in case of Algorithm 5 and tight up to some low-order terms in case of Algorithm 6. The reason for this is the same as what was observed in one of the remark in [CF13].

The argument is briefly the following. Suppose that $C_G \notin$ Succinct3Col and that $G$ is such, that there exist a coloring such that only one pair of nodes are colored inconsistently. If the prover gives states of the form defined by Eq. (A.1) on page 83, but using this coloring, then the verifier won't notice this in either of the Uniformity Tests nor in the Equality Test. The only place where the verifier can catch the prover is in the Consistency Test when he checks the colors of the nodes according to the constraints posed by the graph $G$. The prover gets caught if the verifier gets the inconsistently colored nodes in the measurement outputs. This happens with probability $O\left(2^{-2\tilde{n}}\right)$ in case of Algorithm 5 and $O\left(\tilde{n}^2 2^{-2\tilde{n}}\right)$ in case of Algorithm 6. This means that it is possible to fool the verifier of Algorithm 5 with probability $1 - O\left(4^{-\tilde{n}}\right)$ and to fool the verifier of Algorithm 6 with probability $1 - \widetilde{O}\left(4^{-\tilde{n}}\right)$.

#### Separation Between $\mathsf{QMA}$ and $\mathsf{QMA}[2]$ in the Small-Gap Setting

As we said before, it is a big open problem whether $\mathsf{QMA}$ is equal to $\mathsf{QMA}[2]$ and we mentioned some evidences that suggest us that they are not equal.

Here we show that under plausible complexity-theoretic assumptions $\mathsf{QMA}[2]$ is *strictly* more powerful than $\mathsf{QMA}$ in the low-gap setting. We elaborate on this in the following.

Theorem 1.1.4 shows that $\mathsf{QMA}[2]$ with exponentially or double-exponentially small gap is exactly characterized by $\mathsf{NEXP}$. So it is natural to ask, what is the power of $\mathsf{QMA}$ with the same gap, or what upper bounds can we give for it? In a related paper, Ito et al. [IKW12] showed that quantum interactive proof systems (or the class $\mathsf{QIP}$) with double-exponentially small gap are exactly characterized by $\mathsf{EXP}$. Since $\mathsf{QIP}$ contains $\mathsf{QMA}$, with the same gap, we have a separation between $\mathsf{QMA}$ and $\mathsf{QMA}[2]$ in the setting where the gap is exponentially or double-exponentially small, unless $\mathsf{EXP} = \mathsf{NEXP}$. The result of Ito et al. is quite involved but if we are only interested in the upper bound on $\mathsf{QMA}$ then we can give a very simple argument for it, which we state here in the following lemma.

**Lemma 3.3.11.** $\bigcup_{\substack{0<s<c\leq 1,\\ c-s\geq 2^{-2^{\mathsf{poly}}}}} \mathsf{QMA}(1,c,s) \subseteq \mathsf{EXP}$, *where $c(n)$ and $s(n)$ can be calculated in time at most exponential in $n$ on a classical computer.*

*Proof sketch.* Let $x$ be an input to a problem in $\mathsf{QMA}(1,c,s)$ with $c$ and $s$ having the given property. The action of the verifier can be described by a binary-valued measurement $\{\mathbf{P}_0^x, \mathbf{P}_1^x\}$ on the proof state, where $\mathbf{P}_1^x$ corresponds to acceptance and $\mathbf{P}_0^x$ corresponds to rejecting. Note that the maximum acceptance probability of the verifier is equal to the spectral norm of $\mathbf{P}_1^x$. (Or, in other words, the biggest eigenvalue of $\mathbf{P}_1^x$.) Since the proof is on polynomially-many qubits, the dimension of $\mathbf{P}_1^x$ is exponential. An $\mathsf{EXP}$-machine, knowing $x$, can approximate $\mathbf{P}_1^x$ with up to an exponential amount of digits of accuracy. This is because the verifier is a uniform quantum circuit of polynomial size. Now the $\mathsf{EXP}$-machine can approximate the spectral norm of $\mathbf{P}_1^x$ up to an exponential amount of bits of accuracy. ❑

### One-Sided Error Case

Note that the $\mathsf{NEXP}$ characterization of the small-gap $\mathsf{QMA}[2]$ proof system still holds if we restrict the proof system to have one-sided error. It is not known whether $\mathsf{QMA}$ can be made to have one-sided error, so we can investigate the relation between these classes as well. Interestingly, it turns out that we can state an even stronger separation in this case. This is due to a result by Ito et al. [IKW12].

**Theorem 3.3.12** (Theorem 6 of [IKW12]). $\mathsf{QMA}(1,1,<1) \subseteq \mathsf{PSPACE}$, *where the notation* $< 1$ *in the third parameter means that we only require the soundness to be strictly less than* 1.

This means that, in the one-sided error case, $\mathsf{QMA}[2]$ with exponentially small gap is *strictly* more powerful then $\mathsf{QMA}$ with even unbounded gap, unless $\mathsf{PSPACE} = \mathsf{NEXP}$!

### Nonexistence of Disentanglers

The above discussions have an interesting consequence to the existence question of certain operators called disentanglers. They were defined by Aaronson et al. [ABD+09] as the following.

**Definition 3.3.13** (Definition 40 of [ABD+09]). A super-operator $\Phi \in \mathrm{C}\left(\mathbb{C}^N, \mathbb{C}^M \otimes \mathbb{C}^M\right)$ is an $(\varepsilon, \delta)$-disentangler if

- $\Phi(\rho)$ is $\varepsilon$-close to a separable state for every $\rho \in \mathrm{D}\left(\mathbb{C}^N\right)$ and
- for every separable state $\sigma \in \mathrm{D}\left(\mathbb{C}^M \otimes \mathbb{C}^M\right)$, there exists a $\rho \in \mathrm{D}\left(\mathbb{C}^N\right)$ such that $\Phi(\rho)$ is $\delta$-close to $\sigma$.

Note that if there exists a $\left(\frac{1}{\mathsf{poly}(\log M)}, \frac{1}{\mathsf{poly}(\log M)}\right)$-disentangler with $\log N = \mathsf{poly}(\log M)$ and if that disentangler can be implemented in $\mathsf{poly}(\log M)$-time, then $\mathsf{QMA} = \mathsf{QMA}[2]$. So it is not believed that such a disentangler exists. Towards proving this, Aaronson et al. showed that no $(0,0)$-disentangler exists for any finite $N$ and $M$. The discussion in the previous section implies that there exists no disentangler with approximation error inverse of the square of the dimension. More precisely, we get the following corollary.

**Corollary 3.3.14.** *There exists a function* $f(M) = \Omega\left(M^{-2}\right)$, *such that there exists no* $\mathsf{poly}(\log M)$*-time implementable* $(f(M), f(M))$*-disentangler with* $\log N = \mathsf{poly}(\log M)$, *unless* $\mathsf{EXP} = \mathsf{NEXP}$.

*Proof.* Suppose that there exists such a disentangler $\Phi$, for $f(M) = \kappa_1 \cdot M^{-2}$, for some constant $\kappa_1$ to be specified later. Lemma 3.3.3 implies that

$$\textsc{Succinct3Col} \in \mathsf{QMA}\left(2, 1, 1 - \kappa_2 \cdot 4^{-\tilde{n}}\right),$$

for some constant $\kappa_2$ and where the dimension of both proof states are $3 \cdot 2^{\tilde{n}}$. Let $V$ be the corresponding verifier. We show that $\textsc{Succinct3Col} \in \mathsf{QMA}(1, c, s)$ with $c - s = \Omega\left(4^{-\tilde{n}}\right)$ by constructing a verifier $W$ that uses only one proof.

By Lemma 3.3.11, it holds that $\mathsf{QMA}(1,c,s) \subseteq \mathsf{EXP}$ and since Succinct3Col is NEXP-complete, we get that $\mathsf{EXP} = \mathsf{NEXP}$.

We are left to define verifier $W$. $W$ fist applies $\Phi$ on its quantum proof then simulates $V$ on the output of $\Phi$ and outputs whatever $V$ outputs. Note that $\Phi$ is polynomial-time implementable and the size of the proof of $V$ is also polynomial. To see completeness for $W$, note that there exist a state $|\psi\rangle \otimes |\psi\rangle$ with which $V$ accepts with probability 1. From Definition 3.3.13 there exist a $\rho$ such that $\Phi(\rho)$ is $f\left(3 \cdot 2^{\tilde{n}}\right)$-close to $|\psi\rangle \otimes |\psi\rangle$. Since $f\left(3 \cdot 2^{\tilde{n}}\right) = \frac{\kappa_1}{9} \cdot 4^{-\tilde{n}}$, the probability of acceptance of $W$ is at least $1 - \frac{\kappa_1}{9} \cdot 4^{-\tilde{n}}$. Similarly, for the soundness of $W$, we have that for all separable states, $V$ accepts with probability at most $1 - \kappa_2 \cdot 4^{-\tilde{n}}$. Again from Definition 3.3.13, for all $\rho$, $\Phi(\rho)$ is $f\left(3 \cdot 2^{\tilde{n}}\right)$-close to a separable state. So the probability of acceptance of $W$ is at most $1 - \kappa_2 \cdot 4^{-\tilde{n}} + \frac{\kappa_1}{9} \cdot 4^{-\tilde{n}}$. If $\kappa_1$ is sufficiently small then $c - s = \Omega\left(4^{-\tilde{n}}\right)$ so the corollary follows. ❑

**Notes on BellQMA Proof Systems**

An interesting consequence of Theorems 1.1.4 and 1.1.5 is that, in the small-gap setting, $\mathsf{BellQMA}[k]$ proof systems have the same power as $\mathsf{QMA}[k]$ proof systems if $k \geq n^{\epsilon}$. As we mentioned before, in the normal-gap setting $\mathsf{BellQMA}[k] = \mathsf{QMA}$ for any $k \in \mathsf{poly}(n)$ [BH13]. This means that if a verifier is restricted to Bell-measurements then he gains a lot of extra power if we decrease the bound on the gap. The proof of $\mathsf{BellQMA}[O(1)] = \mathsf{QMA}$, by Brandão [Bra08], doesn't generalize to the small gap setting (and nor does the proof of $\mathsf{BellQMA}[\mathsf{poly}(n)] = \mathsf{QMA}$, of course). So the power of $\mathsf{BellQMA}[O(1)]$ in the small gap setting is still an open problem. There are two possibilities. Either $\mathsf{BellQMA}[2] = \mathsf{NEXP}$, which would supersede both Theorems 1.1.4 and 1.1.5, or $\mathsf{BellQMA}[2] \subset \mathsf{NEXP}$, which would imply that somewhere between constant and $n^{\epsilon}$ number of provers the power of $\mathsf{BellQMA}[k]$ increases. We leave the study of this class for future work.

**Error and Proof Reduction**

As a side-product of our results, in both the $\mathsf{BellQMA}[k]$ and $\mathsf{QMA}[k]$ proof systems we can amplify the error from double-exponentially small gap to single-exponentially small gap. Also, in the case of $\mathsf{QMA}[k]$, we can make the proof system to have *one-sided error* which, up to our knowledge, has only been shown to hold for some restricted versions of single-prover QMA [NWZ09, JKNN12]. Additionally, the number of proofs in $\mathsf{QMA}[k]$ can be reduced to two, but this also follows from [HM13] where an essentially different argument was

used.

**More Open Problems**

Here we list some more open problems that we think may be interesting to work on.

- What is the power of $\mathsf{QMA}[k]$ and $\mathsf{BellQMA}[k]$ with unbounded gap? Can we at least show some upper bounds?
- What is the power of $\mathsf{QMA}$ if we allow only one qubit as its proof but we allow double-exponentially small or unbounded gap? Is it the same as $\mathsf{PQP}$, i.e., $\mathsf{BQP}$ with unbounded gap? Note that the known proofs that show that $\mathsf{QMA}$ with normal gap and a logarithmic length proof equals $\mathsf{BQP}$ [MW05, BSW11] break down if the gap is so small.

# 4

# Parallel Repetition of Entangled Games

In this chapter we prove Theorem 1.2.2, our parallel repetition theorem for entangled games.

**Theorem 1.2.2.** *For any game $G = (\mathscr{X}, \mathscr{Y}, \mathscr{A}, \mathscr{B}, \mu, V)$, where $\mu$ is a product distribution on $\mathscr{X} \times \mathscr{Y}$, it holds that*

$$\omega^*\left(G^k\right) = \left(1 - (1 - \omega^*(G))^3\right)^{\Omega\left(\frac{k}{\log(|\mathscr{A}|\cdot|\mathscr{B}|)}\right)}.$$

This chapter is based on Ref. [JPY14]. We begin with explaining the high level ideas.

## 4.1 The Ideas Behind the Proof

The arguments we use are information theoretic and are broadly on similar lines as that of Raz [Raz98] and Holenstein [Holo9] for classical games. The additional quantum ingredients we need, to deal with entangled games, are inspired by the work of Jain, Radhakrishnan, and Sen [JRS08], where quantum information theoretic arguments were used to achieve message compression in quantum communication protocols.

Given the $k$-fold game $G^k$, let us condition on success on a set $\mathscr{C} \subseteq [k]$ of coordinates. That is, we condition on the event that the players win all the games whose coordinates are in $\mathscr{C}$. If the overall success in coordinates in $\mathscr{C}$ is already as small as we want, then we are done. Otherwise, we exhibit another coordinate $j \notin \mathscr{C}$ such that the success in the $j$-th coordinate, even when we

condition on success in coordinates inside $\mathscr{C}$, is bounded away from 1. Here we assume that $\omega^*(G)$ is bounded away from 1. We add this $j$ to $\mathscr{C}$ and repeat the argument. This way the overall success keeps going down and becomes exponentially small in $k$ after we have identified $\Omega(k)$ such coordinates. To argue that the probability with which the players win the game in the $j$-th coordinate, conditioned on success in $\mathscr{C}$, is bounded away from 1, we show that close to this success probability can be achieved for a single instance of $G$. That is, given inputs $(x', y')$ drawn from $\mu$, for a single instance of $G$, Alice and Bob can embed $(x', y')$ to the $j$-th coordinate of $G^k$, conditioned on success in $\mathscr{C}$, and generate the rest of the state with good approximation. This state consists of the questions and answers to all the $k$ coordinates as well as the shared entangled state. So, if the probability of success in the $j$-th coordinate, conditioned on success in $\mathscr{C}$, is very close to 1 then there is a strategy for $G$ with probability of success strictly larger than $\omega^*(G)$. This is a contradiction to the definition of $\omega^*(G)$.

We now describe how to embed the input $(x', y')$ and generate the state of the whole system. Suppose that the global state in $G^k$, conditioned on success in $\mathscr{C}$, is of the form

$$\sigma^{\mathsf{XYAB}} = \sum_{x \in \mathscr{X}^k, y \in \mathscr{Y}^k} \tilde{\mu}(x,y) \, |xy\rangle\langle xy|^{\mathsf{XY}} \otimes |\phi_{xy}\rangle\langle\phi_{xy}|^{\mathsf{AB}}$$

where $\tilde{\mu}$ is a distribution, potentially different from $\mu^k$ because of the conditioning on success. Registers $\mathsf{X}$ and $\mathsf{Y}$ contain the questions to Alice and Bob, while $\mathsf{A}$ and $\mathsf{B}$ contain the answers and the shared entangled state. In $\sigma^{\mathsf{XYAB}}$, we also fix the questions and answers in $\mathscr{C}$ to specific values but, to keep the notation simple, we don't explicitly denote it. In protocol $\mathcal{P}$, for a single instance of $G$, we let Alice and Bob start with the shared pure state

$$|\varphi\rangle = \sum_{x \in \mathscr{X}^k, y \in \mathscr{Y}^k} \sqrt{\tilde{\mu}(x,y)} \, |xxyy\rangle^{\tilde{\mathsf{X}}\mathsf{X}\tilde{\mathsf{Y}}\mathsf{Y}} \otimes |\phi_{xy}\rangle^{\mathsf{AB}}.$$

Note that, $|\varphi\rangle$ is a purification of $\sigma^{\mathsf{XYAB}}$ where registers $\tilde{\mathsf{X}}$ and $\tilde{\mathsf{Y}}$ are identical to $\mathsf{X}$ and $\mathsf{Y}$. The reason we want the global state, including the questions, answers, and the shared state, to be pure is because we can then use properties of purifications such as the unitary equivalence of purifications and Uhlmann's theorem. If we trace out $\tilde{\mathsf{X}}$ and $\tilde{\mathsf{Y}}$ then the state of $\mathsf{X}$ and $\mathsf{Y}$ becomes classical. So, we will view these registers as being classical. This will be important in the following argument.

Using the chain rule for mutual information, we are able to argue that both $\mathrm{I}(\mathsf{X}_j : \mathsf{Y}\tilde{\mathsf{Y}}\mathsf{B})$ and $\mathrm{I}(\mathsf{Y}_j : \mathsf{X}\tilde{\mathsf{X}}\mathsf{A})$ are close to 0 in $|\varphi\rangle$. This, obviously, is only possible when the distribution $\mu$ is product. In addition, the distribution of questions in the $j$-th coordinate, in $|\varphi\rangle$, remains close to $\mu$ in the $L_1$-distance. Suppose that in protocol $\mathcal{P}$, when Alice and Bob get questions $x'$ and $y'$ they measure registers $\mathsf{X}_j$ and $\mathsf{Y}_j$ in $|\varphi\rangle$. Let's denote the outcomes they get by $x'_j$ and $y'_j$ and let $\left|\varphi_{x'_j y'_j}\right\rangle$ be the resulting state. If, by luck, it happens that $(x', y') = \left(x'_j, y'_j\right)$ then they can further measure the answer registers in $\left|\varphi_{x'_j y'_j}\right\rangle$ and send back the answers to the referee. However, the problem is that the probability that $(x', y') = \left(x'_j, y'_j\right)$ can be very small. So the question is: Is there a way they can generate the post-measurement state $\left|\varphi_{x'_j y'_j}\right\rangle$, at least approximately, without measurements? We describe next how this can be achieved.

Let $\left|\varphi_{x'_j}\right\rangle$ be the resulting state after we measure register $\mathsf{X}_j$ in $|\varphi\rangle$ and get outcome $x'_j$. The fact that $\mathrm{I}(\mathsf{X}_j : \mathsf{Y}\tilde{\mathsf{Y}}\mathsf{B})$ is close to 0 implies that Bob's side of $\left|\varphi_{x'_j}\right\rangle$ is almost independent of $x'_j$. By the unitary equivalence of purifications and Uhlmann's theorem, there is a unitary transformation $\mathbf{U}_{x'_j}$ that Alice can apply to take the state $|\varphi\rangle$ quite close to the state $\left|\varphi_{x'_j}\right\rangle$. Similarly, let us define $\left|\varphi_{y'_j}\right\rangle$ and again $\mathrm{I}(\mathsf{Y}_j : \mathsf{X}\tilde{\mathsf{X}}\mathsf{A})$ being close to 0 implies that Alice's side of $\left|\varphi_{y'_j}\right\rangle$ is mostly independent of $y'_j$. Again, by Uhlmann's theorem, there is a unitary transformation $\mathbf{U}_{y'_j}$ that Bob can apply to take the state $|\varphi\rangle$ close to the state $\left|\varphi_{y'_j}\right\rangle$. Interestingly, as was argued in [JRS08], when Alice and Bob simultaneously apply $\mathbf{U}_{x'_j}$ and $\mathbf{U}_{y'_j}$, they take $|\varphi\rangle$ close to the state $\left|\varphi_{x'_j y'_j}\right\rangle$! This again requires the distribution of questions to be independent across Alice and Bob.

## 4.2 Simulating Measurements with Unitaries

Before we proceed to the detailed proof, we state two lemmas here that show how we create the post-measurement states using unitaries. In this chapter we slightly abuse notations and use $\psi$ to represent the density matrix $|\psi\rangle\langle\psi|$ associated with $|\psi\rangle$. The following lemma states that when the concerned mutual information is small, then a measurement on Alice's side can be simulated by a unitary operation on her side.

**Lemma 4.2.1.** *Let $\mu$ be a probability distribution on $\mathscr{X}$. Let*

$$|\varphi\rangle \stackrel{\text{def}}{=} \sum_{x\in\mathscr{X}} \sqrt{\mu(x)}\,|xx\rangle^{\tilde{\mathsf{X}}\mathsf{X}} \otimes |\psi_x\rangle^{\mathsf{AB}}$$

*be a joint pure state of Alice and Bob, where registers $(\tilde{\mathsf{X}}, \mathsf{X}, \mathsf{A})$ are with Alice and register $\mathsf{B}$ is with Bob. Let $\mathrm{I}(\mathsf{X}:\mathsf{B})_\varphi \leq \varepsilon$ and $|\varphi_x\rangle \stackrel{\text{def}}{=} |xx\rangle \otimes |\psi_x\rangle$. There exist unitary operators $\{\mathbf{U}_x\}_{x\in\mathscr{X}}$ acting on $(\tilde{\mathsf{X}}, \mathsf{X}, \mathsf{A})$ such that*

$$\mathop{\mathbb{E}}_{x\leftarrow\mu}[\|\,|\varphi_x\rangle\langle\varphi_x| - (\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}})\,|\varphi\rangle\langle\varphi|\,(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}})\|_1] \leq 4\sqrt{\varepsilon}.$$

*Proof.* Let us denote the reduced state of Bob in $|\varphi_x\rangle$ and $|\varphi\rangle$ by

$$\rho_x \stackrel{\text{def}}{=} \mathrm{Tr}_{\mathcal{A}}(|\psi_x\rangle\langle\psi_x|) \qquad \text{and} \qquad \rho \stackrel{\text{def}}{=} \mathrm{Tr}_{\tilde{\mathcal{X}}\otimes\mathcal{X}\otimes\mathcal{A}}(|\varphi\rangle\langle\varphi|)\,.$$

From the condition on the mutual information, we get that

$$\begin{aligned}
\varepsilon &\geq \mathrm{I}(\mathsf{X}:\mathsf{B})_\varphi \\
&= \mathop{\mathbb{E}}_{x\leftarrow\mu}[\mathrm{S}(\rho_x\|\rho)] \\
&\geq 1 - \mathop{\mathbb{E}}_{x\leftarrow\mu}[\mathrm{F}(\rho_x,\rho)]
\end{aligned}$$

where the first equality follows from Corollary 2.3.8 and the last inequality follows from Theorem 2.3.10. By Theorems 2.1.5 and 2.1.7, there exists a unitary $\mathbf{U}_x$ for each $x \in \mathscr{X}$ such that

$$|\langle\varphi_x|\,(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}})\,|\varphi\rangle| = \mathrm{F}(\rho_x,\rho)\,.$$

The lemma follows from the following calculation.

$$\begin{aligned}
&\mathop{\mathbb{E}}_{x\leftarrow\mu}[\|\,|\varphi_x\rangle\langle\varphi_x| - (\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}})\,|\varphi\rangle\langle\varphi|\,(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}})\|_1] \\
&\quad= 2\mathop{\mathbb{E}}_{x\leftarrow\mu}\left[\sqrt{1 - |\langle\varphi_x|\,(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}})\,|\varphi\rangle|^2}\right] && (4.1)\\
&\quad\leq 2\sqrt{1 - \mathop{\mathbb{E}}_{x\leftarrow\mu}[|\langle\varphi_x|\,(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}})\,|\varphi\rangle|]^2} && (4.2)\\
&\quad= 2\sqrt{1 - \mathop{\mathbb{E}}_{x\leftarrow\mu}[\mathrm{F}(\rho_x,\rho)]^2} \\
&\quad\leq 4\sqrt{\varepsilon}
\end{aligned}$$

where Eq. (4.1) follows from Eq. (2.1) and at Eq. (4.2) we used the concavity of the function $\sqrt{1-\alpha^2}$. ❑

The following is an extension of the above lemma that states that when both the mutual informations are small then simultaneous measurements on Alice's and Bob's side can be simulated by unitary operations on their respective sides. It is a special case of a more general result in Ref. [JRS08].

**Lemma 4.2.2** ([JRS08]). *Let $\mu$ be a probability distribution on $\mathscr{X} \times \mathscr{Y}$. Let $\mu_X$ and $\mu_Y$ be the marginals of $\mu$ on $\mathscr{X}$ and $\mathscr{Y}$. Let*

$$|\varphi\rangle \stackrel{\text{def}}{=} \sum_{x \in \mathscr{X}, y \in \mathscr{Y}} \sqrt{\mu(x,y)}\, |xxyy\rangle^{\tilde{\mathsf{X}}\mathsf{X}\tilde{\mathsf{Y}}\mathsf{Y}} \otimes |\psi_{x,y}\rangle^{\mathsf{AB}}$$

*be a joint pure state of Alice and Bob, where registers $(\tilde{\mathsf{X}}, \mathsf{X}, \mathsf{A})$ belong to Alice and registers $(\tilde{\mathsf{Y}}, \mathsf{Y}, \mathsf{B})$ belong to Bob. Let*

$$\mathrm{I}(\mathsf{X} : \mathsf{B}\mathsf{Y}\tilde{\mathsf{Y}})_{\varphi} \leq \varepsilon \qquad \text{and} \qquad \mathrm{I}(\mathsf{Y} : \mathsf{A}\mathsf{X}\tilde{\mathsf{X}})_{\varphi} \leq \varepsilon.$$

*Let $|\varphi_{x,y}\rangle \stackrel{\text{def}}{=} |xxyy\rangle \otimes |\psi_{x,y}\rangle$. There exist unitary operators $\{\mathbf{U}_x\}_{x \in \mathscr{X}}$ on $(\tilde{\mathsf{X}}, \mathsf{X}, \mathsf{A})$ and $\{\mathbf{V}_y\}_{y \in \mathscr{Y}}$ on $(\tilde{\mathsf{Y}}, \mathsf{Y}, \mathsf{B})$ such that*

$$\begin{aligned} &\mathop{\mathbb{E}}_{(x,y) \leftarrow \mu} \left[ \left\| |\varphi_{x,y}\rangle\langle\varphi_{x,y}| - \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right) \right\|_1 \right] \\ &\qquad \leq 8\sqrt{\varepsilon} + 2 \left\| \mu - \mu_X \otimes \mu_Y \right\|_1 . \end{aligned}$$

To keep our presentation self-contained, and because the proof in [JRS08] is complicated as it proves a more general result, we give a proof for the above lemma below. Before we do that, we state a small lemma here that is easy to verify and that will be used in the proof.

**Lemma 4.2.3.** *Let $0 < \varepsilon, \varepsilon' < 1$, $0 < c$, $\mu$ and $\mu'$ be probability distributions on a set $\mathscr{X}$, and $f : \mathscr{X} \to [0, c]$ be a function. If $\mathbb{E}_{x \leftarrow \mu}[f(x)] \leq \varepsilon$ and $\|\mu - \mu'\|_1 \leq \varepsilon'$ then $\mathbb{E}_{x \leftarrow \mu'}[f(x)] \leq \varepsilon + c\varepsilon'$.*

*Proof of Lemma 4.2.2.* Let $|\varphi_x\rangle$ be the state obtained when we measure register $\mathsf{X}$ in $|\varphi\rangle$ and obtain $x$. Similarly, let $|\varphi_y\rangle$ be the state obtained when we measure register $\mathsf{Y}$ in $|\varphi\rangle$ and obtain $y$. By Lemma 4.2.1, there exist unitary operators $\{\mathbf{U}_x\}_{x \in \mathscr{X}}$ and $\{\mathbf{V}_y\}_{y \in \mathscr{Y}}$ such that

$$\begin{aligned} \mathop{\mathbb{E}}_{x \leftarrow \mu_X} \left[ \left\| |\varphi_x\rangle\langle\varphi_x| - \left(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}}\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}}\right) \right\|_1 \right] &\leq 4\sqrt{\varepsilon} \\ \mathop{\mathbb{E}}_{y \leftarrow \mu_Y} \left[ \left\| |\varphi_y\rangle\langle\varphi_y| - \left(\mathbb{1}_{\mathcal{A}} \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbb{1}_{\mathcal{A}} \otimes \mathbf{V}_y^*\right) \right\|_1 \right] &\leq 4\sqrt{\varepsilon}. \end{aligned}$$

Using the above, we get that

$$
\begin{aligned}
&\Big\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} \left[ |xy\rangle\langle xy| \otimes |\varphi_{x,y}\rangle\langle\varphi_{x,y}| \right] \\
&\qquad - \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu_X\otimes\mu_Y} \left[ |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right) \right] \Big\|_1 \\
&\leq \Big\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} \left[ |xy\rangle\langle xy| \otimes |\varphi_{x,y}\rangle\langle\varphi_{x,y}| \right] \\
&\qquad - \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu_X\otimes\mu_Y} \left[ |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}}\right) |\varphi_y\rangle\langle\varphi_y| \left(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}}\right) \right] \Big\|_1 \\
&\qquad + \Big\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu_X\otimes\mu_Y} \left[ |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}}\right) |\varphi_y\rangle\langle\varphi_y| \left(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}}\right) \right. \\
&\qquad\qquad \left. - |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right) \right] \Big\|_1 \qquad (4.3) \\
&\leq \Big\| \mathop{\mathbb{E}}_{x\leftarrow\mu_X} \left[ |x\rangle\langle x| \otimes |\varphi_x\rangle\langle\varphi_x| - |x\rangle\langle x| \otimes \left(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}}\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}}\right) \right] \Big\|_1 \\
&\qquad + \Big\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu_X\otimes\mu_Y} \left[ |xy\rangle\langle xy| \otimes |\varphi_y\rangle\langle\varphi_y| \right. \\
&\qquad\qquad \left. - |xy\rangle\langle xy| \otimes \left(\mathbb{1}_{\mathcal{A}} \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbb{1}_{\mathcal{A}} \otimes \mathbf{V}_y^*\right) \right] \Big\|_1 \qquad (4.4) \\
&= \mathop{\mathbb{E}}_{x\leftarrow\mu_X} \left[ \| |\varphi_x\rangle\langle\varphi_x| - \left(\mathbf{U}_x \otimes \mathbb{1}_{\mathcal{B}}\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbb{1}_{\mathcal{B}}\right) \|_1 \right] \\
&\qquad + \mathop{\mathbb{E}}_{y\leftarrow\mu_Y} \left[ \Big\| |\varphi_y\rangle\langle\varphi_y| - \left(\mathbb{1}_{\mathcal{A}} \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbb{1}_{\mathcal{A}} \otimes \mathbf{V}_y^*\right) \Big\|_1 \right] \\
&\leq 8\sqrt{\varepsilon} \qquad (4.5)
\end{aligned}
$$

where Eq. (4.3) follows from the triangle inequality, the second term in Eq. (4.4) is because $\mathbf{U}_x$ doesn't change the trace distance, and the first term in Eq. (4.4) follows from Theorem 2.1.9 with the super-operator that corresponds to measuring $Y$ in the standard basis and storing the outcome in a new register. The lemma follows from the following calculation.

$$
\begin{aligned}
&\mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} \left[ \Big\| |\varphi_{x,y}\rangle\langle\varphi_{x,y}| - \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right) \Big\|_1 \right] \\
&\quad = \Big\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} \left[ |xy\rangle\langle xy| \otimes |\varphi_{x,y}\rangle\langle\varphi_{x,y}| \right. \\
&\qquad \left. - |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right) \right] \Big\|_1 \\
&\quad \leq \Big\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} \left[ |xy\rangle\langle xy| \otimes |\varphi_{x,y}\rangle\langle\varphi_{x,y}| \right] \\
&\qquad - \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu_X\otimes\mu_Y} \left[ |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right) \right] \Big\|_1
\end{aligned}
$$

$$+ \left\| \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu_X\otimes\mu_Y} \left[ |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right)\right] \right.$$
$$\left. - \mathop{\mathbb{E}}_{(x,y)\leftarrow\mu} \left[ |xy\rangle\langle xy| \otimes \left(\mathbf{U}_x \otimes \mathbf{V}_y\right) |\varphi\rangle\langle\varphi| \left(\mathbf{U}_x^* \otimes \mathbf{V}_y^*\right)\right] \right\|_1$$
$$\leq 8\sqrt{\varepsilon} + 2\left\|\mu - \mu_X \otimes \mu_Y\right\|_1$$

where the first inequality follows from the triangle inequality and at the last inequality we used Eq. (4.5) and Lemma 4.2.3. ❑

## 4.3 Proof of the Parallel Repetition Theorem

Let a game $G = (\mathscr{X}, \mathscr{Y}, \mathscr{A}, \mathscr{B}, \mu, V)$ be given. From now on, we assume that the distribution $\mu = \mu_X \otimes \mu_Y$ is product across $\mathscr{X}$ and $\mathscr{Y}$. Now, let's consider the game $G^k$. Let $x = (x_1, \ldots, x_k) \in \mathscr{X}^k$, $y = (y_1, \ldots, y_k) \in \mathscr{Y}^k$, $a = (a_1, \ldots, a_k) \in \mathscr{A}^k$, and $b = (b_1, \ldots, b_k) \in \mathscr{B}^k$. To make notations short, we denote $\mu(x,y) = \prod_i \mu(x_i, y_i)$ and $V(x,y,a,b) = \prod_i V(x_i, y_i, a_i, b_i)$, whenever it is clear from the context. Without loss of generality, we assume that, before the game starts, Alice and Bob share a pure state on the registers $(\mathsf{A}, \mathsf{E}'_A, \mathsf{B}, \mathsf{E}'_B)$, where $(\mathsf{A}, \mathsf{E}'_A)$ belong to Alice and $(\mathsf{B}, \mathsf{E}'_B)$ belong to Bob. Registers $\mathsf{A}$ and $\mathsf{B}$ will be used to store the answers for Alice and Bob, respectively, while $\mathsf{E}'_A$ and $\mathsf{E}'_B$ are some extra registers that hold a possibly entangled state. After getting the questions, Alice and Bob perform unitary operations independently and then they measure registers $\mathsf{A}$ and $\mathsf{B}$ in the standard basis. The outcomes of the measurements are sent to the referee. Let $\mathscr{C} \subseteq [k]$ and let $\overline{\mathscr{C}}$ represent its *complement* in $[k]$. Let $x_{\mathscr{C}}$ represent the substring of $x$ corresponding to the indices in $\mathscr{C}$. (Similarly, we will use $y_{\mathscr{C}}$, $a_{\mathscr{C}}$, and $b_{\mathscr{C}}$.) Let's define

$$|\theta\rangle \stackrel{\text{def}}{=} \sum_{x\in\mathscr{X}^k, y\in\mathscr{Y}^k} \sqrt{\mu(x,y)}\, |xxyy\rangle^{\tilde{\mathsf{X}}\mathsf{X}\tilde{\mathsf{Y}}\mathsf{Y}}$$
$$\otimes \sum_{a_{\mathscr{C}}\in\mathscr{A}^{|\mathscr{C}|}, b_{\mathscr{C}}\in\mathscr{B}^{|\mathscr{C}|}} |a_{\mathscr{C}} b_{\mathscr{C}}\rangle^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}} \otimes \left|\gamma_{x,y,a_{\mathscr{C}},b_{\mathscr{C}}}\right\rangle^{\mathsf{E}_A\mathsf{E}_B}$$

where $\mathsf{E}_A \stackrel{\text{def}}{=} \mathsf{E}'_A\mathsf{A}_{\overline{\mathscr{C}}}$, $\mathsf{E}_B \stackrel{\text{def}}{=} \mathsf{E}'_B\mathsf{B}_{\overline{\mathscr{C}}}$, and $\sum_{a_{\mathscr{C}},b_{\mathscr{C}}} |a_{\mathscr{C}} b_{\mathscr{C}}\rangle \otimes \left|\gamma_{x,y,a_{\mathscr{C}},b_{\mathscr{C}}}\right\rangle$ is the shared state after Alice and Bob performed their unitary operations corresponding to questions $x$ and $y$. (Note that $\left|\gamma_{x,y,a_{\mathscr{C}},b_{\mathscr{C}}}\right\rangle$ is unnormalized.) Consider the state

$$|\varphi\rangle \stackrel{\text{def}}{=} \frac{1}{\sqrt{q}} \sum_{x,y} \sqrt{\mu(x,y)}\, |xxyy\rangle^{\tilde{\mathsf{X}}\mathsf{X}\tilde{\mathsf{Y}}\mathsf{Y}}$$

$$\otimes \sum_{a_{\mathscr{C}},b_{\mathscr{C}}\,:\,V(x_{\mathscr{C}},y_{\mathscr{C}},a_{\mathscr{C}},b_{\mathscr{C}})=1} \left|a_{\mathscr{C}}b_{\mathscr{C}}\right\rangle^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}} \otimes \left|\gamma_{x,y,a_{\mathscr{C}},b_{\mathscr{C}}}\right\rangle^{\mathsf{E}_A\mathsf{E}_B}$$

where $q$ is the probability that, in $G^k$, all the instances of $G$ in coordinates $\mathscr{C}$ succeed. The following lemma states that if we take the state $\theta$ and condition it on having success in coordinates $\mathscr{C}$, i.e., we go from $\theta$ to $\varphi$, and additionally we fix the answers in these coordinates then the resulting state will be close to the original. This closeness is measured in the relative entropy.

**Lemma 4.3.1.** *For states $\theta$ and $\varphi$ defined above, it holds that*

$$\mathop{\mathbb{E}}_{\substack{x_{\mathscr{C}}y_{\mathscr{C}}a_{\mathscr{C}}b_{\mathscr{C}}\\ \leftarrow \varphi^{\mathsf{X}_{\mathscr{C}}\mathsf{Y}_{\mathscr{C}}\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}}} \left[\mathrm{S}\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{x_{\mathscr{C}}y_{\mathscr{C}}a_{\mathscr{C}}b_{\mathscr{C}}} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{x_{\mathscr{C}}y_{\mathscr{C}}}\right)\right] \leq -\log q + |\mathscr{C}|\cdot\log(|\mathscr{A}|\cdot|\mathscr{B}|)\,.$$

*Proof.* Let $p(a_{\mathscr{C}},b_{\mathscr{C}})$ be the probability of obtaining $(a_{\mathscr{C}},b_{\mathscr{C}})$ when measuring registers $(\mathsf{A}_{\mathscr{C}},\mathsf{B}_{\mathscr{C}})$ in $|\varphi\rangle$. From Lemma 2.3.12, we get that

$$\begin{aligned}
\mathrm{S}_\infty\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{a_{\mathscr{C}}b_{\mathscr{C}}} \middle\| \varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right) &\leq -\log p(a_{\mathscr{C}},b_{\mathscr{C}})\\
\mathrm{S}_\infty\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right) &\leq -\log q.
\end{aligned}$$

We first get a bound for the relative min-entropy,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{a_{\mathscr{C}}b_{\mathscr{C}}\leftarrow\varphi^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}} \left[\mathrm{S}_\infty\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{a_{\mathscr{C}}b_{\mathscr{C}}} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right)\right]\\
&\quad\leq \mathop{\mathbb{E}}_{a_{\mathscr{C}}b_{\mathscr{C}}\leftarrow\varphi^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}} \left[\mathrm{S}_\infty\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{a_{\mathscr{C}}b_{\mathscr{C}}} \middle\| \varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right)\right.\\
&\qquad\left. + \mathrm{S}_\infty\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right)\right]\\
&\quad\leq \mathop{\mathbb{E}}_{a_{\mathscr{C}}b_{\mathscr{C}}\leftarrow\varphi^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}} \left[-\log p(a_{\mathscr{C}},b_{\mathscr{C}}) - \log q\right]\\
&\quad= -\log q + \mathrm{S}\left(\varphi^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}\right)\\
&\quad\leq -\log q + |\mathscr{C}|\cdot(\log|\mathscr{A}| + \log|\mathscr{B}|)
\end{aligned}$$

where the first inequality follows from the definition of the relative min-entropy. From the above, the required bound for the relative entropy follows easily.

$$\begin{aligned}
&-\log q + |\mathscr{C}|\cdot\log(|\mathscr{A}|\cdot|\mathscr{B}|)\\
&\qquad\geq \mathop{\mathbb{E}}_{a_{\mathscr{C}}b_{\mathscr{C}}\leftarrow\varphi^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}} \left[\mathrm{S}_\infty\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{a_{\mathscr{C}}b_{\mathscr{C}}} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right)\right]\\
&\qquad\geq \mathop{\mathbb{E}}_{a_{\mathscr{C}}b_{\mathscr{C}}\leftarrow\varphi^{\mathsf{A}_{\mathscr{C}}\mathsf{B}_{\mathscr{C}}}} \left[\mathrm{S}\left(\varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}_{a_{\mathscr{C}}b_{\mathscr{C}}} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{XYE}_A\mathsf{E}_B}\right)\right]
\end{aligned}$$

$$\geq \mathop{\mathbb{E}}_{x_{\mathscr{C}} y_{\mathscr{C}} a_{\mathscr{C}} b_{\mathscr{C}} \leftarrow \varphi^{\mathsf{X}_{\mathscr{C}} \mathsf{Y}_{\mathscr{C}} \mathsf{A}_{\mathscr{C}} \mathsf{B}_{\mathscr{C}}}} \left[ \mathrm{S}\left( \varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}} \tilde{\mathsf{Y}}_{\overline{\mathscr{C}}} \mathsf{X}\mathsf{Y}\mathsf{E}_A \mathsf{E}_B}_{x_{\mathscr{C}} y_{\mathscr{C}} a_{\mathscr{C}} b_{\mathscr{C}}} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}} \tilde{\mathsf{Y}}_{\overline{\mathscr{C}}} \mathsf{X}\mathsf{Y}\mathsf{E}_A \mathsf{E}_B}_{x_{\mathscr{C}} y_{\mathscr{C}}} \right) \right]$$

where the last inequality follows from Theorem 2.3.7. ❑

For each $i \in [k]$, let us define a binary random variable $T_i \in \{0,1\}$ which indicates success in the $i$-th repetition. That is, $T_i = V(\mathsf{X}_i, \mathsf{Y}_i, A_i, B_i)$. Our main theorem will follow directly from the following lemma.

**Lemma 4.3.2.** *Let $0.1 > \delta_1, \delta_2, \delta_3 > 0$ such that $\delta_3 = \delta_2 + \delta_1 \cdot \log(|\mathscr{A}| \cdot |\mathscr{B}|)$. Let $k' \stackrel{\text{def}}{=} \lfloor \delta_1 k \rfloor$. For any quantum strategy for the $k$-fold game $G^k$, there exists a set $\{i_1, \ldots, i_{k'}\}$, such that for each $1 \leq r \leq k' - 1$, either*

$$\Pr\left[T^{(r)} = 1\right] \leq 2^{-\delta_2 k}$$

*or*

$$\Pr\left[T_{i_{r+1}} = 1 \middle| T^{(r)} = 1\right] \leq \omega^*(G) + 12\sqrt{10\delta_3}$$

*where $T^{(r)} \stackrel{\text{def}}{=} \prod_{j=1}^{r} T_{i_j}$.*

*Proof.* In the following, we assume that $1 \leq r < k'$. However, the same argument also works when $r = 0$, i.e., for identifying the first coordinate, which we skip for the sake of avoiding repetition. Suppose that we have already identified $r$ coordinates $i_1, \ldots, i_r$ satisfying that

$$\Pr[T_{i_1} = 1] \leq \omega^*(G) + 12\sqrt{10\delta_3}$$

and

$$\Pr\left[T_{i_{j+1}} = 1 \middle| T^{(j)} = 1\right] \leq \omega^*(G) + 12\sqrt{10\delta_3}$$

for $1 \leq j \leq r-1$. If $\Pr\left[T^{(r)} = 1\right] \leq 2^{-\delta_2 k}$ then we are done so, from now on, we assume that $\Pr\left[T^{(r)} = 1\right] > 2^{-\delta_2 k}$. Let $\mathscr{C} \stackrel{\text{def}}{=} \{i_1, \ldots, i_r\}$. To simplify notations, let $\tilde{\mathsf{A}} \stackrel{\text{def}}{=} (\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}, \mathsf{X}, \mathsf{E}_A)$, $\tilde{\mathsf{B}} \stackrel{\text{def}}{=} (\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}, \mathsf{Y}, \mathsf{E}_B)$, and $\mathsf{R}_i \stackrel{\text{def}}{=} \left(\mathsf{X}_{\mathscr{C} \cup [i-1]}, \mathsf{Y}_{\mathscr{C} \cup [i-1]}, \mathsf{A}_{\mathscr{C}}, \mathsf{B}_{\mathscr{C}}\right)$. For coordinate $i$, let $\left|\varphi_{x_{<i} y_{<i}}\right\rangle$ be the pure state that results when we measure registers $(\mathsf{X}_{<i}, \mathsf{Y}_{<i})$ (i.e., registers $(\mathsf{X}_1, \ldots, \mathsf{X}_{i-1}, \mathsf{Y}_1, \ldots, \mathsf{Y}_{i-1})$) in $|\varphi\rangle$ and get outcome $x_{<i} y_{<i}$. We argue now that for a typical coordinate outside $\mathscr{C}$, the distribution of questions is close to $\mu$ in the state $\varphi$. We also prove that, for this coordinate,

the questions and $\mathsf{R}_i$ are almost independent. From Lemma 4.3.1, we get that

$$\begin{aligned}
\delta_3 k &\geq \delta_2 k + |\mathscr{C}| \cdot \log(|\mathscr{A}| \cdot |\mathscr{B}|) \\
&\geq \mathop{\mathbb{E}}_{x_{<\mathscr{C}} y_{<\mathscr{C}} a_{<\mathscr{C}} b_{<\mathscr{C}} \leftarrow \varphi^{\mathsf{X}_{<\mathscr{C}}\mathsf{Y}_{<\mathscr{C}}\mathsf{A}_{<\mathscr{C}}\mathsf{B}_{<\mathscr{C}}}} \left[ \mathrm{S}\left( \varphi^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{X}\mathsf{Y}\mathsf{E}_A\mathsf{E}_B}_{x_{<\mathscr{C}} y_{<\mathscr{C}} a_{<\mathscr{C}} b_{<\mathscr{C}}} \middle\| \theta^{\tilde{\mathsf{X}}_{\overline{\mathscr{C}}}\tilde{\mathsf{Y}}_{\overline{\mathscr{C}}}\mathsf{X}\mathsf{Y}\mathsf{E}_A\mathsf{E}_B}_{x_{<\mathscr{C}} y_{<\mathscr{C}}} \right) \right] && (4.6) \\
&\geq \mathop{\mathbb{E}}_{x_{<\mathscr{C}} y_{<\mathscr{C}} a_{<\mathscr{C}} b_{<\mathscr{C}} \leftarrow \varphi^{\mathsf{X}_{<\mathscr{C}}\mathsf{Y}_{<\mathscr{C}}\mathsf{A}_{<\mathscr{C}}\mathsf{B}_{<\mathscr{C}}}} \left[ \mathrm{S}\left( \varphi^{\mathsf{X}\mathsf{Y}}_{x_{<\mathscr{C}} y_{<\mathscr{C}} a_{<\mathscr{C}} b_{<\mathscr{C}}} \middle\| \theta^{\mathsf{X}\mathsf{Y}}_{x_{<\mathscr{C}} y_{<\mathscr{C}}} \right) \right] && (4.7) \\
&= \sum_{i \notin \mathscr{C}} \mathop{\mathbb{E}}_{r_i \leftarrow \varphi^{\mathsf{R}_i}} \left[ \mathrm{S}\left( \varphi^{\mathsf{X}_i\mathsf{Y}_i}_{r_i} \middle\| \theta^{\mathsf{X}_i\mathsf{Y}_i} \right) \right] && (4.8) \\
&= \sum_{i \notin \mathscr{C}} \mathrm{S}\left( \varphi^{\mathsf{X}_i\mathsf{Y}_i\mathsf{R}_i} \middle\| \varphi^{\mathsf{R}_i} \otimes \theta^{\mathsf{X}_i\mathsf{Y}_i} \right) && (4.9) \\
&\geq \sum_{i \notin \mathscr{C}} \mathrm{S}\left( \varphi^{\mathsf{X}_i\mathsf{Y}_i\mathsf{R}_i} \middle\| \varphi^{\mathsf{R}_i} \otimes \varphi^{\mathsf{X}_i\mathsf{Y}_i} \right) && (4.10) \\
&\geq \sum_{i \notin \mathscr{C}} \mathop{\mathbb{E}}_{x_i y_i \leftarrow \varphi^{\mathsf{X}_i\mathsf{Y}_i}} \left[ \mathrm{S}\left( \varphi^{\mathsf{R}_i}_{x_i y_i} \middle\| \varphi^{\mathsf{R}_i} \right) \right] && (4.11) \\
&\geq \sum_{i \notin \mathscr{C}} \mathop{\mathbb{E}}_{x_i y_i \leftarrow \varphi^{\mathsf{X}_i\mathsf{Y}_i}} \left[ \left\| \varphi^{\mathsf{R}_i}_{x_i y_i} - \varphi^{\mathsf{R}_i} \right\|_1^2 \right] && (4.12) \\
&\geq \sum_{i \notin \mathscr{C}} \left( \mathop{\mathbb{E}}_{x_i y_i \leftarrow \varphi^{\mathsf{X}_i\mathsf{Y}_i}} \left[ \left\| \varphi^{\mathsf{R}_i}_{x_i y_i} - \varphi^{\mathsf{R}_i} \right\|_1 \right] \right)^2 && (4.13)
\end{aligned}$$

where Eq. (4.6) follows from Lemma 4.3.1, Eq. (4.7) follows from Theorem 2.3.11, Eqs. (4.8), (4.9) and (4.11) follow from Theorem 2.3.7, Eq. (4.10) follows from Lemma 2.3.9, Eq. (4.12) follows from Theorem 2.3.10, and Eq. (4.13) follows from the convexity of the function $\alpha^2$. Next, we argue that for a typical coordinate outside $\mathscr{C}$, the information between Alice's questions and Bob's registers is small in $|\varphi\rangle$. Again, from Lemma 4.3.1 and Theorem 2.3.11, we get that

$$\begin{aligned}
\delta_3 k &\geq \mathop{\mathbb{E}}_{x_{<\mathscr{C}} y_{<\mathscr{C}} a_{<\mathscr{C}} b_{<\mathscr{C}} \leftarrow \varphi^{\mathsf{X}_{<\mathscr{C}}\mathsf{Y}_{<\mathscr{C}}\mathsf{A}_{<\mathscr{C}}\mathsf{B}_{<\mathscr{C}}}} \left[ \mathrm{S}\left( \varphi^{\mathsf{X}\tilde{\mathsf{B}}}_{x_{<\mathscr{C}} y_{<\mathscr{C}} a_{<\mathscr{C}} b_{<\mathscr{C}}} \middle\| \theta^{\mathsf{X}\tilde{\mathsf{B}}}_{x_{<\mathscr{C}} y_{<\mathscr{C}}} \right) \right] \\
&\geq \mathrm{I}(\mathsf{X} : \tilde{\mathsf{B}} | \mathsf{R}_1)_\varphi && (4.14) \\
&\geq \sum_{i \notin \mathscr{C}} \mathrm{I}(\mathsf{X}_i : \tilde{\mathsf{B}} | \mathsf{R}_1 \mathsf{X}_{<i})_\varphi && (4.15) \\
&\geq \sum_{i \notin \mathscr{C}} \mathrm{I}(\mathsf{X}_i : \tilde{\mathsf{B}} | \mathsf{R}_i)_\varphi && (4.16)
\end{aligned}$$

where at Eq. (4.14) we used Eq. (2.9) and the fact that $\theta^{\mathsf{X}\tilde{\mathsf{B}}}_{x_{<\mathscr{C}} y_{<\mathscr{C}}} = \theta^{\mathsf{X}}_{x_{<\mathscr{C}} y_{<\mathscr{C}}} \otimes \theta^{\tilde{\mathsf{B}}}_{x_{<\mathscr{C}} y_{<\mathscr{C}}}$. Equations (4.15) and (4.16) follow from the chain rule for the mutual information and at Eq. (4.16) we also used the observation that $\tilde{\mathsf{B}}$ contains register $\mathsf{Y}$.

Similarly to the above, for Bob's questions we have

$$\delta_3 k \geq \sum_{i \notin \mathscr{C}} \mathrm{I}(\mathsf{Y}_i : \tilde{\mathsf{A}}|\mathsf{R}_i)_\varphi \,. \tag{4.17}$$

From Eqs. (4.8), (4.13), (4.16) and (4.17) and using standard application of Markov's inequality, we get that there exists a coordinate $j \notin \mathscr{C}$ such that

$$\mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \mathrm{S}\left( \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} \middle\| \theta^{\mathsf{X}_j\mathsf{Y}_j} \right) \right] \leq \frac{5\delta_3}{1-\delta_1} \leq 10\delta_3 \tag{4.18}$$

$$\mathop{\mathbb{E}}_{x_j y_j \leftarrow \varphi^{\mathsf{X}_j\mathsf{Y}_j}} \left[ \left\| \varphi_{x_j y_j}^{\mathsf{R}_j} - \varphi^{\mathsf{R}_j} \right\|_1 \right] \leq \sqrt{\frac{5\delta_3}{1-\delta_1}} \leq \sqrt{10\delta_3} \tag{4.19}$$

$$\mathrm{I}\big(\mathsf{X}_j : \tilde{\mathsf{B}} \big| \mathsf{R}_j\big)_\varphi \leq \frac{5\delta_3}{1-\delta_1} \leq 10\delta_3 \tag{4.20}$$

$$\mathrm{I}\big(\mathsf{Y}_j : \tilde{\mathsf{A}} \big| \mathsf{R}_j\big)_\varphi \leq \frac{5\delta_3}{1-\delta_1} \leq 10\delta_3. \tag{4.21}$$

Let $\left|\varphi_{r_j}\right\rangle$ be the pure state that we get when we measure register $\mathsf{R}_j$ in $|\varphi\rangle$ and get outcome $r_j$. Suppose that there exists a protocol $\mathcal{P}_0$ for $G^k$ which wins all coordinates in $\mathscr{C}$ with probability greater than $2^{-\delta_2 k}$. Moreover, conditioning on success on all coordinates in $\mathscr{C}$, the probability it wins the game in the $j$-th coordinate is $\omega$.

- Let us construct a new protocol $\mathcal{P}_1$, that starts with the joint state $\varphi^{\mathsf{X}_j\mathsf{Y}_j\mathsf{R}_j\mathsf{E}_A\mathsf{E}_B}$, where $\mathsf{X}_j\mathsf{E}_A$ and $\mathsf{Y}_j\mathsf{E}_B$ are given to Alice and Bob, respectively, and $\mathsf{R}_j$ is shared between them. From our assumption, the probability that Alice and Bob win the game in the $j$-th coordinate is $\omega$.

- Let us consider a new protocol $\mathcal{P}_2$, where Alice and Bob are given questions $(x_j, y_j) \leftarrow \varphi^{\mathsf{X}_j\mathsf{Y}_j}$ and they share $r_j \leftarrow \varphi_{x_j y_j}^{\mathsf{R}_j}$ as public coins. By Lemma 4.2.2, they are able to create a joint state that is close to the starting state of $\mathcal{P}_1$ by sharing $\left|\varphi_{r_j}\right\rangle$ and applying local unitary operations. More concretely, Eqs. (4.20) and (4.21) show the conditions for the mutual informations required by Lemma 4.2.2. From Eq. (4.18), we can get

$$\begin{aligned} 10\delta_3 &\geq \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \mathrm{S}\left( \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} \middle\| \theta^{\mathsf{X}_j\mathsf{Y}_j} \right) \right] \\ &\geq \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \mathrm{S}\left( \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} \middle\| \varphi_{r_j}^{\mathsf{X}_j} \otimes \varphi_{r_j}^{\mathsf{Y}_j} \right) \right] \qquad (4.22) \\ &\geq \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \left\| \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} - \varphi_{r_j}^{\mathsf{X}_j} \otimes \varphi_{r_j}^{\mathsf{Y}_j} \right\|_1^2 \right] \qquad (4.23) \end{aligned}$$

$$\geq \left( \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \left\| \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} - \varphi_{r_j}^{\mathsf{X}_j} \otimes \varphi_{r_j}^{\mathsf{Y}_j} \right\|_1 \right] \right)^2 \tag{4.24}$$

where Eq. (4.22) follows from Lemma 2.3.9, Eq. (4.23) follows from Theorem 2.3.10, and at Eq. (4.24) we used the convexity of the function $\alpha^2$. This implies

$$\mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \left\| \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} - \varphi_{r_j}^{\mathsf{X}_j} \otimes \varphi_{r_j}^{\mathsf{Y}_j} \right\|_1 \right] \leq \sqrt{10\delta_3}.$$

Thus, using the above and Lemma 4.2.2, we conclude that they can win the game with probability at least $\omega - 10\sqrt{10\delta_3}$.

- Let us construct a new protocol $\mathcal{P}_3$, where Alice and Bob are given questions $(x_j, y_j) \leftarrow \varphi^{\mathsf{X}_j\mathsf{Y}_j}$. They share public coins $r_j \leftarrow \varphi^{\mathsf{R}_j}$ and execute the same strategy as in $\mathcal{P}_2$. By Eq. (4.19), the probability that they win the game is at least $\omega - 11\sqrt{10\delta_3}$.

- Let us consider a new protocol $\mathcal{P}_4$, where Alice and Bob are given questions $(x, y) \leftarrow \mu$ and they execute the same strategy as in $\mathcal{P}_3$. Similarly as above, from Eq. (4.18) together with the fact that $\theta^{\mathsf{X}_j\mathsf{Y}_j} = \mu$, we get that

$$\begin{aligned} 10\delta_3 &\geq \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \mathrm{S}\left( \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} \middle\| \mu \right) \right] \\ &\geq \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \left\| \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} - \mu \right\|_1^2 \right] \\ &\geq \left( \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \left\| \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} - \mu \right\|_1 \right] \right)^2 \end{aligned}$$

where again the second inequality follows from Theorem 2.3.10 and at the last inequality we used the convexity of the function $\alpha^2$. From the above, it follows that

$$\sqrt{10\delta_3} \geq \mathop{\mathbb{E}}_{r_j \leftarrow \varphi^{\mathsf{R}_j}} \left[ \left\| \varphi_{r_j}^{\mathsf{X}_j\mathsf{Y}_j} - \mu \right\|_1 \right] \geq \left\| \varphi^{\mathsf{X}_j\mathsf{Y}_j} - \mu \right\|_1 .$$

This means that the probability that Alice and Bob win the game, using protocol $\mathcal{P}_4$, is at least $\omega - 12\sqrt{10\delta_3}$. Note that $\mathcal{P}_4$ is a strategy for game $G$ under distribution $\mu$. This means that $\omega - 12\sqrt{10\delta_3} \leq \omega^*(G)$.

We conclude the lemma. ❑

We can now prove our main result. We restate it here for convenience.

**Theorem 1.2.2.** *Let $\varepsilon > 0$. Given a game $G$ with value $\omega^*(G) \leq 1 - \varepsilon$, it holds that*

$$\begin{aligned}\omega^*\left(G^k\right) &\leq \left(1 - \frac{\varepsilon}{2}\right)^{\frac{\varepsilon^2 k}{12000 \cdot (\log|\mathscr{A}| + \log|\mathscr{B}|)}} \\ &= \left(1 - \varepsilon^3\right)^{\Omega\left(\frac{k}{\log|\mathscr{A}| + \log|\mathscr{B}|}\right)}.\end{aligned}$$

*Proof.* We set

$$\begin{aligned}\delta_1 &= \frac{\varepsilon^2}{12000 \cdot (\log|\mathscr{A}| + \log|\mathscr{B}|)} \\ \delta_2 &= \frac{\varepsilon^2}{12000} \\ \delta_3 &= \frac{\varepsilon^2}{6000}.\end{aligned}$$

Given any strategy for $G^k$, using Lemma 4.3.2, either $\omega^*\left(G^k\right) \leq 2^{-\delta_2 k}$ or there are $\lfloor \delta_1 k \rfloor$ coordinates $\left\{i_1, \ldots, i_{\lfloor \delta_1 k \rfloor}\right\}$ such that the probability that Alice and Bob win the $i_j$-th coordinate, conditioning on success on all the previous coordinates, is at most $1 - \varepsilon/2$. This finishes the proof of the theorem. ❑

# A

# Deferred Proofs about Small-Gap QMA

This appendix contains the proofs that were deferred from Section 3.3. They are fairly straightforward from the proofs of Refs. [BT12, CF13, CD10].

## A.1 Proof of Completeness and Soundness for Lemma 3.3.3

This section proves completeness and soundness for verifier $V$ described by Algorithm 5 on page 61 and so finishes the proof of Lemma 3.3.3. This part closely follows the analysis of Blier and Tapp [BT12] and of Chiesa and Forbes [CF13], albeit with different parameters. The proofs are done through a few lemmas.

### A.1.1 Proof of Completeness

The following lemma, which is essentially the same as Lemma 3.2 of [BT12], proves completeness for $V$.

**Lemma 3.3.5** (Completeness)**.** *If $C_G \in$ Succinct3Col then there exist a pair of proofs with which $V$ will accept with probability* 1*.*

*Proof.* For $i \in \{0, 1, \ldots, m-1\}$ let $c(i) \in \{0, 1, 2\}$ be a valid coloring of the graph $G$, where $m$ is the number of nodes. For $i \in \{m, \ldots, 2^{\tilde{n}} - 1\}$ let $c(i) = 0$. Let the state of both $\mathsf{R}_1$ and $\mathsf{R}_2$ be

$$|\phi\rangle = \frac{1}{\sqrt{2^{\tilde{n}}}} \sum_{i=0}^{2^{\tilde{n}}-1} |i\rangle\,|c(i)\rangle \tag{A.1}$$

where $|i\rangle$ is on the node register (N) and $|c(i)\rangle$ is on the color register (C). From Eq. (2.7), it follows that the Equality Test succeeds with probability 1. Since $c$ is a valid 3-coloring, the Consistency Test succeeds with probability 1 as well. To see the same for the Uniformity Test, let us calculate the resulting state after measuring 0 on $\mathsf{C}_1$ in line 17 of Algorithm 5. Up to some normalization factor, the state is

$$(\mathbb{1} \otimes |u_3\rangle\langle u_3|)\,|\phi\rangle = \frac{1}{3\sqrt{2^{\tilde{n}}}} \left( \sum_{i=0}^{2^{\tilde{n}}-1} |i\rangle \right) \otimes \left( \sum_{k=0}^{2} |k\rangle \right).$$

This means that the state of $\mathsf{N}_1$ is $|u_{2^{\tilde{n}}}\rangle$ so the Uniformity Test always succeeds. ❑

### A.1.2 Proof of Soundness

From now on let us suppose that $C_G \notin$ Succinct3Col and let's denote the state of $\mathsf{R}_1$ by $|\psi\rangle$ and the state of $\mathsf{R}_2$ by $|\varphi\rangle$. These two states can be written in the form

$$|\psi\rangle = \sum_{i=0}^{2^{\tilde{n}}-1} \alpha_i\,|i\rangle \sum_{j=0}^{2} \beta_{i,j}\,|j\rangle \qquad \text{and} \qquad |\varphi\rangle = \sum_{i=0}^{2^{\tilde{n}}-1} \alpha'_i\,|i\rangle \sum_{j=0}^{2} \beta'_{i,j}\,|j\rangle$$

where $\sum_i |\alpha_i|^2 = \sum_i |\alpha'_i|^2 = 1$ and, for all $i$, $\sum_j \left|\beta_{i,j}\right|^2 = \sum_j \left|\beta'_{i,j}\right|^2 = 1$.

The following lemma says that if the Equality Test succeeds with high probability then the distribution of outcomes in the Consistency Test will be similar. This is analogous to Lemma 3.3 of [BT12].

**Lemma A.1.1.** *If the Equality test of Algorithm 5 succeeds with probability at least $1-\varepsilon$, then for all $k$ and $\ell$ it holds that $\left||\alpha_k\beta_{k,\ell}|^2 - \left|\alpha'_k\beta'_{k,\ell}\right|^2\right| \leq \sqrt{8\varepsilon}$.*

*Proof.* Let $p_{i,j} \stackrel{\text{def}}{=} \left|\alpha_i\beta_{i,j}\right|^2$, $q_{i,j} \stackrel{\text{def}}{=} \left|\alpha'_i\beta'_{i,j}\right|^2$ and let us denote the probability vector with elements $p_{i,j}$ by $p$ and similarly for $q$. We have the following.

$$\begin{aligned}
\sqrt{1-|\langle\psi|\varphi\rangle|^2} &= d(|\psi\rangle\langle\psi|, |\varphi\rangle\langle\varphi|) && \text{(A.2)}\\
&\geq \frac{1}{2}\,\|p-q\|_1 && \text{(A.3)}\\
&= \frac{1}{2}\sum_{i,j}\left||\alpha_i\beta_{i,j}|^2 - \left|\alpha'_i\beta'_{i,j}\right|^2\right| \\
&\geq \frac{1}{2}\left||\alpha_k\beta_{k,\ell}|^2 - \left|\alpha'_k\beta'_{k,\ell}\right|^2\right|
\end{aligned}$$

for any $k$ and $\ell$. Equation (A.2) is from Eq. (2.1) and Eq. (A.3) follows from Theorem 2.1.9. Equation (2.7) implies that $\frac{1}{2}\left(1+|\langle\psi|\varphi\rangle|^2\right) \geq 1-\varepsilon$. With the above derivation the claim of the lemma follows. ❑

Similarly to Lemma 3.4 of [BT12] (and to Lemma 6.2 of [CF13]), the next lemma states that vertices with high probability of being observed have a well-defined color.

**Lemma A.1.2.** *Suppose that $|\psi\rangle$ and $|\varphi\rangle$ pass the Equality Test of Algorithm 5 and also line 9 in the Consistency Test with probability at least $1-10^{-10}\cdot 4^{-\tilde{n}}$. Then for all $i$, for which $|\alpha_i|^2 \geq 100^{-1}\cdot 2^{-\tilde{n}}$, there exist one $j$ for which $|\beta_{i,j}|^2 \geq 0.9$.*

*Proof.* Towards contradiction suppose that $\exists i$ such that $|\alpha_i|^2 \geq \frac{1}{100\cdot 2^{\tilde{n}}}$ and $\forall j$ it holds that $|\beta_{i,j}|^2 < \frac{9}{10}$. Then, without loss of generality, we can say that $|\beta_{i,0}|^2 \geq \frac{1}{20}$ and $|\beta_{i,1}|^2 \geq \frac{1}{20}$. Since the probability that the Equality Test succeeds is at least $1-\frac{1}{10^{10}\cdot 4^{\tilde{n}}}$, we can apply Lemma A.1.1 and get that

$$\left||\alpha_i\beta_{i,1}|^2 - |\alpha_i'\beta_{i,1}'|^2\right| \leq \frac{\sqrt{8}}{10^5\cdot 2^{\tilde{n}}}.$$

This implies that

$$\begin{aligned}|\alpha_i'|^2\,|\beta_{i,1}'|^2 &\geq |\alpha_i|^2\,|\beta_{i,1}|^2 - \frac{\sqrt{8}}{10^5\cdot 2^{\tilde{n}}}\\ &\geq \frac{1}{2000\cdot 2^{\tilde{n}}} - \frac{\sqrt{8}}{10^5\cdot 2^{\tilde{n}}}.\end{aligned}$$

The probability that, in line 8, in the Consistency Test we get $v_1 = v_2 = i$, $c_1 = 0$, and $c_2 = 1$ is

$$\begin{aligned}\Pr[v_1 = i \text{ and } c_1 = 0]\cdot\Pr[v_2 = i \text{ and } c_2 = 1] &\geq \frac{1}{2000\cdot 2^{\tilde{n}}}\left(\frac{1}{2000\cdot 2^{\tilde{n}}} - \frac{\sqrt{8}}{10^5\cdot 2^{\tilde{n}}}\right)\\ &> \frac{1}{10^{10}\cdot 4^{\tilde{n}}}.\end{aligned}$$

This contradicts to the assumption that $|\psi\rangle$ and $|\varphi\rangle$ pass line 9 with probability at least $1-10^{-10}\cdot 4^{-\tilde{n}}$. ❑

The next lemma is analogous to Lemma 3.5 of [BT12] and also to Lemma 6.3 of [CF13].

**Lemma A.1.3.** *Suppose that $|\psi\rangle$ and $|\varphi\rangle$ pass the Equality Test of Algorithm 5 and also line 9 in the Consistency Test with probability at least $1-10^{-10}\cdot 4^{-\tilde{n}}$. Then the probability of measuring 0 on $\mathsf{C}_1$ in line 17 in the Uniformity Test is at least 0.05.*

*Proof.* Suppose that we measure $\mathsf{N}_1$ in the standard basis. If the outcome is $i$ then the probability of measuring 0 on $\mathsf{C}_1$ with the measurement of Definition 3.3.1 is

$$\frac{1}{3}\left|\beta_{i,0}+\beta_{i,1}+\beta_{i,2}\right|^2.$$

For all $i$ for which $|\alpha_i|^2 \geq \frac{1}{100\cdot 2^{\tilde{n}}}$ Lemma A.1.2 applies, which means that there exist $k_i$ such that $\left|\beta_{i,k_i}\right|^2 \geq \frac{9}{10}$. Let $\ell_i \stackrel{\text{def}}{=} k_i+1 \mod 3$ and $m_i \stackrel{\text{def}}{=} k_i+2 \mod 3$. Then $\left|\beta_{i,\ell_i}\right|^2+\left|\beta_{i,m_i}\right|^2 < \frac{1}{10}$. We can lower bound the above probability by

$$\begin{aligned}
\frac{1}{3}\left|\beta_{i,k_i}+\beta_{i,\ell_i}+\beta_{i,m_i}\right|^2 &\geq \frac{1}{3}\left|\left|\beta_{i,k_i}\right|-\left|\beta_{i,\ell_i}+\beta_{i,m_i}\right|\right|^2 \\
&\geq \frac{1}{3}\left(\left|\beta_{i,k_i}\right|-\sqrt{2\cdot\left(\left|\beta_{i,\ell_i}\right|^2+\left|\beta_{i,m_i}\right|^2\right)}\right)^2 \\
&\geq \frac{1}{3}\left(\frac{9}{10}-\sqrt{\frac{2}{10}}\right)^2 \\
&> \frac{6}{100}
\end{aligned}$$

where the second inequality follows from the Cauchy-Schwarz inequality. More precisely, we have $\left|\beta_{i,\ell_i}+\beta_{i,m_i}\right|^2 \leq 2\cdot\left(\left|\beta_{i,\ell_i}\right|^2+\left|\beta_{i,m_i}\right|^2\right) < \frac{2}{10} < \frac{9}{10} \leq \left|\beta_{i,k_i}\right|^2$. The probability of measuring 0 on $\mathsf{C}_1$ in line 17 is

$$\begin{aligned}
\sum_{i=0}^{2^{\tilde{n}}-1}|\alpha_i|^2\cdot\frac{1}{3}\cdot\left|\beta_{i,0}+\beta_{i,1}+\beta_{i,2}\right|^2 &\geq \sum_{\substack{i \text{ for which} \\ |\alpha_i|^2\geq\frac{1}{100\cdot 2^{\tilde{n}}}}}|\alpha_i|^2\cdot\frac{1}{3}\cdot\left|\beta_{i,0}+\beta_{i,1}+\beta_{i,2}\right|^2 \\
&> \frac{6}{100}\cdot\sum_{\substack{i \text{ for which} \\ |\alpha_i|^2\geq\frac{1}{100\cdot 2^{\tilde{n}}}}}|\alpha_i|^2 \\
&\geq \frac{6}{100}\cdot\left(1-\frac{2^{\tilde{n}}-1}{100\cdot 2^{\tilde{n}}}\right) \\
&> \frac{5}{100}
\end{aligned}$$

where we used the fact that at most $2^{\tilde{n}}-1$ nodes ($i$'s) can have $|\alpha_i|^2 < \frac{1}{100\cdot 2^{\tilde{n}}}$. ❑

In order to proceed, we need two lemmas, one from [BT12] and one from [CD10]. We present them now together with their proofs.

**Lemma A.1.4** (Lemma 3.6 of [BT12])**.** *For any state $|\xi\rangle = \sum_{i=0}^{m-1}\gamma_i|i\rangle \in \mathbb{C}^m$, if there exists a $k$ such that $|\gamma_k|^2 < \frac{1}{2m}$ then the probability of getting 1 when we measure $|\xi\rangle$ with the measurement of Definition 3.3.1 is at least $\frac{1}{16m^2}$.*

*Proof.* Let $p$ and $q$ be the probability distributions that arise when we measure $|\xi\rangle$ and $|u_m\rangle$ in the computational basis. Or in other words, let $p$ be the probability vector with elements $|\gamma_i|^2$, and $q$ be the vector with all elements equal to $\frac{1}{m}$. Similarly to Lemma A.1.1, we have that

$$\begin{aligned}\sqrt{1-|\langle u_m|\xi\rangle|^2} &= d(|u_m\rangle\langle u_m|, |\xi\rangle\langle\xi|) \\ &\geq \frac{1}{2}\|q-p\|_1 \\ &= \frac{1}{2}\sum_{i=0}^{m-1}\left|\frac{1}{m}-|\gamma_i|^2\right| \\ &\geq \frac{1}{2}\left|\frac{1}{m}-|\gamma_k|^2\right| \\ &> \frac{1}{4m}.\end{aligned}$$

The probability of getting 1 when we measure $|\xi\rangle$ with the measurement of Definition 3.3.1 is $1-|\langle u_m|\xi\rangle|^2$ so the statement of the lemma follows. ❑

The following argument appears in the proof of Lemma 3 of [CD10], which we state here as a separate lemma.

**Lemma A.1.5.** *Suppose that we have a bipartite quantum state $|\psi\rangle \in \mathcal{N}\otimes\mathcal{C}$ with $\mathcal{N}=\mathbb{C}^N$ and $\mathcal{C}=\mathbb{C}^C$. We can write this state as*

$$|\psi\rangle = \sum_{i=0}^{N-1}\alpha_i|i\rangle\sum_{j=0}^{C-1}\beta_{i,j}|j\rangle$$

*where $\sum_i|\alpha_i|^2=1$ and, for all $i$, $\sum_j|\beta_{i,j}|^2=1$. Suppose that the probability of measuring 0 on $\mathcal{C}$ with the measurement of Definition 3.3.1 is $p$ and after the measurement the resulting state on $\mathcal{N}$ is*

$$|\xi\rangle = \sum_{i=0}^{N-1}\gamma_i|i\rangle \in \mathcal{N}$$

*where $\sum_i|\gamma_i|^2=1$. Then, for all $i$, it holds that*

$$|\alpha_i|^2 \geq p\cdot|\gamma_i|^2.$$

*Proof.* Let $q_i$ denote the probability that if we measure the $\mathcal{C}$ part of $|\psi\rangle$ with the measurement of Definition 3.3.1 we get outcome 0, then if we measure the $\mathcal{N}$ part in the standard basis, we get $i$. For all $i$,

$$q_i = p\cdot|\gamma_i|^2.$$

On the other hand,

$$\begin{aligned} q_i &= \langle\psi| \left(|i\rangle\langle i| \otimes |u_C\rangle\langle u_C|\right) |\psi\rangle \\ &= \frac{|\alpha_i|^2}{C} \cdot \left|\sum_{j=0}^{C-1} \beta_{i,j}\right|^2 \\ &\leq \frac{|\alpha_i|^2}{C} \cdot C \cdot \sum_{j=0}^{C-1} \left|\beta_{i,j}\right|^2 \\ &= |\alpha_i|^2 \end{aligned}$$

where the inequality above follows from the Cauchy-Schwarz inequality. The above derivations imply the statement of the lemma. ❑

Analogously to Lemma 3.7 of [BT12] (and to Lemma 6.4 of [CF13]), the following lemma says that if the states pass some of the tests with high probability then it must be that all nodes appear with high enough probability.

**Lemma A.1.6.** *Suppose that $|\psi\rangle$ and $|\varphi\rangle$ pass the Equality Test of Algorithm 5, line 9 in the Consistency Test, and also the Uniformity Test with probability at least $1 - 10^{-10} \cdot 4^{-\tilde{n}}$. Then for all $i$, $|\alpha_i|^2 \geq 100^{-1} \cdot 2^{-\tilde{n}}$.*

*Proof.* Because of Lemma A.1.3 the probability of measuring 0 on $\mathsf{C}_1$ in line 17 of Algorithm 5 is at least $\frac{5}{100}$. Let the state of $\mathsf{N}_1$ be $|\xi\rangle = \sum_{i=0}^{2^{\tilde{n}}-1} \gamma_i |i\rangle$ after we got 0 on $\mathsf{C}_1$. Towards contradiction, suppose that there exists an $i$ such that $|\alpha_i|^2 < \frac{1}{100 \cdot 2^{\tilde{n}}}$. Since we got this measurement result with probability $\geq \frac{5}{100}$, Lemma A.1.5 implies that $|\gamma_i|^2 < \frac{1}{5 \cdot 2^{\tilde{n}}}$. From Lemma A.1.4, the probability of getting 1 when we measure $\mathsf{N}_1$ in line 17 is at least $\frac{1}{16 \cdot 4^{\tilde{n}}}$. So the probability of failing the Uniformity Test is at least $\frac{5}{100} \cdot \frac{1}{16 \cdot 4^{\tilde{n}}} > \frac{1}{10^{10} \cdot 4^{\tilde{n}}}$. This is a contradiction. ❑

The following lemma finishes the proof of soundness for verifier $V$.

**Lemma 3.3.6** (Soundness)**.** *If $C_G \notin$ Succinct3Col then verifier $V$ described by Algorithm 5 will reject with probability at least $\frac{1}{3 \cdot 10^{10} \cdot 4^{\tilde{n}}}$.*

*Proof.* Assume that $|\psi\rangle$ and $|\varphi\rangle$ pass the Equality Test, the Uniformity Test, and line 9 of Algorithm 5 with probability at least $1 - \frac{1}{10^{10} \cdot 4^{\tilde{n}}}$ as otherwise we are done. Let $c(i)$ be equal to the $j$ for which $\left|\beta_{i,j}\right|$ is maximal or, in other words,

$$c(i) \stackrel{\text{def}}{=} \arg\max_j \left|\beta_{i,j}\right| .$$

Because of Lemmas A.1.2 and A.1.6 this maximum is well defined. According to Lemma A.1.6, when measuring $|\psi\rangle$ in line 8, the probability of obtaining $(k, c(k))$, for all $k$, is at least $|\alpha_k|^2 \cdot \frac{9}{10} \geq \frac{1}{100 \cdot 2^{\tilde{n}}} \cdot \frac{9}{10} > \frac{1}{120 \cdot 2^{\tilde{n}}}$. Similarly, from Lemma A.1.1, for all $k$, the probability that we get $(k, c(k))$ when we measure $|\varphi\rangle$ in line 8, is at least $\frac{1}{120 \cdot 2^{\tilde{n}}} - \frac{\sqrt{8}}{10^5 \cdot 2^{\tilde{n}}} > \frac{1}{240 \cdot 2^{\tilde{n}}}$. Since the graph is not 3-colorable $\exists u, v \in V$ such that $(u, v) \in E$ and $c(u) = c(v)$. If, in line 8, we get $(u, c(u))$ and $(v, c(v))$ then the Consistency Test will reject. This happens with probability at least

$$\frac{1}{120 \cdot 2^{\tilde{n}}} \cdot \frac{1}{240 \cdot 2^{\tilde{n}}} > \frac{1}{10^{10} \cdot 4^{\tilde{n}}}.$$

Since the Consistency Test is chosen with probability $\frac{1}{3}$, the statement of the lemma follows. ❑

## A.2 Proof of Completeness and Soundness for Lemma 3.3.8

This section proves completeness and soundness for verifier $V$ described by Algorithm 6 on page 63 and so finishes the proof of Lemma 3.3.8.

### A.2.1 Proof of Completeness

**Lemma 3.3.9** (Completeness). *If $C_G \in$ Succinct3Col then there exist quantum states on registers $\mathsf{N}_1, \mathsf{C}_1, \ldots, \mathsf{N}_k, \mathsf{C}_k$, such that if they are input to $V$, defined by Algorithm 6, then $V$ will accept with probability at least $1 - 2^{-\frac{k}{40}}$.*

*Proof.* For all $i \in \{1, 2, \ldots, k\}$, let the state of $\mathsf{N}_i\mathsf{C}_i$ be $|\phi\rangle$, where $|\phi\rangle$ is defined by Eq. (A.1) on page 83. For exactly the same reason as in the proof of Lemma 3.3.5, the Consistency Test will succeed with probability 1. As for the Uniformity Test, note that for all $i \in \mathscr{Z}$, the measurement of $\mathsf{N}_i$ in line 17 of Algorithm 6 yields 1 with probability 1. The argument for this is also in the proof of Lemma 3.3.5.

This means that given the above input, the only place where Algorithm 6 may reject is at line 21, i.e., when $|\mathscr{Z}| < \frac{k}{6}$. So, in the following we only need to upper bound this probability. We do it similarly to the proof of Lemma 1 in [CD10]. By direct calculation, the probability that $x_i = 0$, in line 16, is

$$\Pr[x_i = 0] = \langle\phi| \left(\mathbb{1} \otimes |u_3\rangle\langle u_3|\right) |\phi\rangle = \frac{1}{3}$$

for all $i$. This means that the expected cardinality of $\mathscr{Z}$ is $\mathbb{E}[|\mathscr{Z}|] = \frac{k}{3}$. Since

the $x_i$'s are independent, we can use the Chernoff bound and get that

$$\Pr\left[|\mathscr{Z}| < \frac{k}{6}\right] < e^{-\frac{k}{48}} < 2^{-\frac{k}{40}}.$$

This finishes the proof of the lemma. ❑

### A.2.2 Proof of Soundness

We are left to prove soundness for $V$. From now on, let's denote the quantum input to Algorithm 6 by $|\varphi_1\rangle \otimes \cdots \otimes |\varphi_k\rangle$. For each $i \in \{1, 2, \ldots, k\}$, we write

$$|\varphi_i\rangle = \sum_{v=0}^{2^{\tilde{n}}-1} \alpha_v^{(i)} |v\rangle \sum_{j=0}^{2} \beta_{v,j}^{(i)} |j\rangle$$

where $|v\rangle$ is a state on $\mathsf{N}_i$, $|j\rangle$ is a state on $\mathsf{C}_i$, $\sum_{v=0}^{2^{\tilde{n}}-1} \left|\alpha_v^{(i)}\right|^2 = 1$ for each $i$, and $\sum_{j=0}^{2} \left|\beta_{v,j}^{(i)}\right|^2 = 1$ for each $i$ and $v$. Similarly to the notation in [CD10] let

$$\mathscr{Z}' \stackrel{\text{def}}{=} \left\{ i \,:\, \Pr[x_i = 0] \geq \frac{1}{12} \right\}.$$

We need a lemma from [CD10] which we will state and use with a bit different parameters. Intuitively, the lemma says that in order to avoid rejection in line 21, we must have a constant fraction of the registers for which the measurement of line 16 yields 0 with at least a constant probability.

**Lemma A.2.1** (Lemma 2 in [CD10])**.** *If $|\mathscr{Z}'| \leq k/6$ and if in line 1 of Algorithm 6 the Uniformity Test is chosen, then the test will reject in line 21 with probability $\Omega(1)$.*

We want all nodes to appear with sufficiently big amplitudes in $|\varphi_i\rangle$, for each $i \in \mathscr{Z}'$. This is formalized by the following lemma.

**Lemma A.2.2.** *Suppose that the Uniformity Test rejects with probability at most $200^{-1} \cdot 4^{-\tilde{n}}$. Then $\forall i \in \mathscr{Z}'$ and $\forall v \in \left\{0, 1, \ldots, 2^{\tilde{n}} - 1\right\}$ it holds that*

$$\left|\alpha_v^{(i)}\right|^2 > \frac{1}{24 \cdot 2^{\tilde{n}}}.$$

*Proof.* Let's pick an $i \in \mathscr{Z}'$ and consider the state $|\varphi_i\rangle$ on register $\mathsf{N}_i\mathsf{C}_i$. Suppose that we measured 0 on $\mathsf{C}_i$ with the measurement of Definition 3.3.1 and denote the resulting state on $\mathsf{N}_i$ by

$$|\xi_i\rangle = \sum_{v=0}^{2^{\tilde{n}}-1} \gamma_v^{(i)} |v\rangle .$$

Since $i \in \mathscr{Z}'$, this outcome happens with probability $\geq \frac{1}{12}$. Assume towards contradiction that $\exists v$ such that $\left|\gamma_v^{(i)}\right|^2 < \frac{1}{2\cdot 2^{\tilde{n}}}$. Using Lemma A.1.4, we get that when we measure $|\xi_i\rangle$ with the measurement of Definition 3.3.1, we get outcome 1 with probability at least $\frac{1}{16\cdot 4^{\tilde{n}}}$. But this means that the Uniformity Test rejects with probability at least $\frac{1}{12\cdot 16\cdot 4^{\tilde{n}}} > \frac{1}{200\cdot 4^{\tilde{n}}}$. This contradicts to the statement of the lemma, so it must be that $\left|\gamma_v^{(i)}\right|^2 \geq \frac{1}{2\cdot 2^{\tilde{n}}}$ for all $v$. Lemma A.1.5 implies that, for all $v$,

$$\left|\alpha_v^{(i)}\right|^2 \geq \frac{1}{12\cdot 2\cdot 2^{\tilde{n}}}.$$ ❑

We are now ready to prove soundness for $V$.

**Lemma 3.3.10** (Soundness)**.** *If $C_G \notin$ Succinct3Col then $V$ of Algorithm 6 will reject with probability at least* $12000^{-1}\cdot 4^{-\tilde{n}}$.

*Proof.* Suppose that the Uniformity Test rejects with probability at most $\frac{1}{200\cdot 4^{\tilde{n}}}$, as otherwise we are done. From Lemma A.2.1, $|\mathscr{Z}'| > \frac{k}{6}$. Since $\frac{k}{6} = \Omega(n)$, we can always take $k \geq 12$ so we have $|\mathscr{Z}'| > 2$. Let's pick two elements $q, r \in \mathscr{Z}'$. We define two colorings $c_1$ and $c_2$ the following way,

$$c_1(v) \stackrel{\text{def}}{=} \arg\max_j \left|\beta_{v,j}^{(q)}\right|$$

and similarly

$$c_2(v) \stackrel{\text{def}}{=} \arg\max_j \left|\beta_{v,j}^{(r)}\right|$$

for all $v \in \{0,1,\ldots,2^{\tilde{n}}-1\}$. If the maximum is not well defined then we just choose an arbitrary $j$ for which $\left|\beta_{v,j}^{(\cdot)}\right|$ is maximal. From Lemma A.2.2, the probability that we get $(v, c_1(v))$ when we measure $|\varphi_q\rangle$ in the standard basis is at least $\left|\alpha_v^{(q)}\right|^2 \cdot \frac{1}{3} > \frac{1}{72\cdot 2^{\tilde{n}}}$, for all $v$, and the same lower bound is true for getting $(v, c_2(v))$ when measuring $|\varphi_r\rangle$. There are two cases.

- Suppose that the two colorings are different, i.e., $\exists v$ such that $c_1(v) \neq c_2(v)$. In this case, in line 4, we get $(v, c_1(v))$ when measuring $\mathsf{N}_q\mathsf{C}_q$ and $(v, c_2(v))$ when measuring $\mathsf{N}_r\mathsf{C}_r$ with probability at least $\left(\frac{1}{72\cdot 2^{\tilde{n}}}\right)^2 > \frac{1}{6000\cdot 4^{\tilde{n}}}$. It means that with at least the above probability the Consistency Test will reject in line 8.

- Suppose that the two colorings are the same, i.e., $\forall v : c_1(v) = c_2(v)$. Since $G$ is not 3-colorable, $\exists v_1, v_2 \in \{0,1,\ldots,2^{\tilde{n}}-1\}$ such that $(v_1, v_2)$ is an edge in $G$ and $c_1(v_1) = c_1(v_2)$, or equivalently $C_G(v_1, v_2) = 11$.

Similarly as above, with probability at least $\frac{1}{6000 \cdot 4^{\bar{n}}}$, we get $(v_1, c_1(v_1))$ when measuring $\mathsf{N}_q\mathsf{C}_q$ and $(v_2, c_1(v_2))$ when measuring $\mathsf{N}_r\mathsf{C}_r$ in line 4 of the algorithm. In this case the Consistency Test will reject with at least the above probability in line 10.

Since in both cases the Consistency Test rejects with probability at least $\frac{1}{6000 \cdot 4^{\bar{n}}}$ and the test is chosen with probability $\frac{1}{2}$, the lemma follows. ❑

# Bibliography


[Aar09] Scott Aaronson. On perfect completeness for QMA. *Quantum Information and Computation*, 9(1):81–89, January 2009, arXiv:0806.0450.

[AB09] Sanjeev Arora and Boaz Barak. *Computational Complexity: A Modern Approach*. Cambridge University Press, New York, NY, USA, 1st edition, 2009.

[ABD+09] Scott Aaronson, Salman Beigi, Andrew Drucker, Bill Fefferman, and Peter Shor. The power of unentanglement. *Theory of Computing*, 5(1):1–42, 2009, arXiv:0804.0802.

[AIM14] Scott Aaronson, Russell Impagliazzo, and Dana Moshkovitz. AM with multiple Merlins. In *Proceedings of the 29th Annual IEEE Conference on Computational Complexity*, CCC '14, pages 44–55, June 2014, arXiv:1401.6848.

[ALM+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. Proof verification and the hardness of approximation problems. *Journal of the ACM*, 45(3):501–555, May 1998.

[AN02] Dorit Aharonov and Tomer Naveh. Quantum NP - a survey. October 2002, arXiv:quant-ph/0210077.

[AS98] Sanjeev Arora and Shmuel Safra. Probabilistic checking of proofs: A new characterization of NP. *Journal of the ACM*, 45(1):70–122, January 1998.

[Bab85] László Babai. Trading group theory for randomness. In *Proceedings of the 17th Annual ACM Symposium on Theory of Computing*, STOC '85, pages 421–429, 1985.

[BBD+97] Adriano Barenco, André Berthiaume, David Deutsch, Artur Ekert, Richard Jozsa, and Chiara Macchiavello. Stabilization of quantum computations by symmetrization. *SIAM Journal on Computing*, 26(5):1541–1557, 1997, arXiv:quant-ph/9604028.

[BCWdW01] Harry Buhrman, Richard Cleve, John Watrous, and Ronald de Wolf. Quantum fingerprinting. *Physical Review Letters*, 87(16):167902, September 2001, arXiv:quant-ph/0102001.

[BCY11a] Fernando G. S. L. Brandão, Matthias Christandl, and Jon Yard. A quasipolynomial-time algorithm for the quantum separability problem. In *Proceedings of the 43rd Annual ACM Symposium on Theory of Computing*, STOC '11, pages 343–352, 2011, arXiv:1011.2751.

[BCY11b] Fernando G.S.L. Brandão, Matthias Christandl, and Jon Yard. Faithful squashed entanglement. *Communications in Mathematical Physics*, 306(3):805–830, 2011, arXiv:1010.1750.

[Bei10] Salman Beigi. NP vs $\mathrm{QMA}_{\log}(2)$. *Quantum Information and Computation*, 10(1):141–151, January 2010, arXiv:0810.5109.

[BFL91] László Babai, Lance Fortnow, and Carsten Lund. Nondeterministic exponential time has two-prover interactive protocols. *Computational Complexity*, 1(1):3–40, 1991.

[BH13] Fernando G.S.L. Brandão and Aram W. Harrow. Quantum de Finetti theorems under local measurements with applications. In *Proceedings of the 45th Annual ACM Symposium on Theory of Computing*, STOC '13, pages 861–870, 2013, arXiv:1210.6367.

[BOGKW88] Michael Ben-Or, Shafi Goldwasser, Joe Kilian, and Avi Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the 20th Annual ACM Symposium on Theory of Computing*, STOC '88, pages 113–131, 1988.

[Boo12] Adam D. Bookatz. QMA-complete problems. December 2012, arXiv:1212.6312.

[Bra06] Sergey Bravyi. Efficient algorithm for a quantum analogue of 2-SAT. February 2006, arXiv:quant-ph/0602108.

[Bra08] Fernando G. S. L. Brandão. *Entanglement Theory and the Quantum Simulation of Many-Body Physics*. PhD thesis, Imperial College London, 2008, arXiv:0810.0026.

[BRR+09] Boaz Barak, Anup Rao, Ran Raz, Ricky Rosen, and Ronen Shaltiel. Strong parallel repetition theorem for free projection games. In Irit Dinur, Klaus Jansen, Joseph Naor, and José Rolim, editors, *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, volume 5687 of *Lecture Notes in Computer Science*, pages 352–365. Springer Berlin / Heidelberg, 2009.

[BSW11] Salman Beigi, Peter Shor, and John Watrous. Quantum interactive proofs with short messages. *Theory of Computing*, 7(1):101–117, 2011, arXiv:1004.0411.

[BT12] Hugue Blier and Alain Tapp. A quantum characterization of NP. *Computational Complexity*, 21(3):499–510, 2012, arXiv:0709.0738.

[CD10] Jing Chen and Andrew Drucker. Short multi-prover quantum proofs for SAT without entangled measurements. November 2010, arXiv:1011.0716.

[CF13] Alessandro Chiesa and Michael A. Forbes. Improved soundness for QMA with multiple provers. *Chicago Journal of Theoretical Computer Science*, 2013(1):1–23, January 2013, arXiv:1108.2098.

[CHSH69] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, October 1969.

[CHTW04] Richard Cleve, Peter Høyer, Ben Toner, and John Watrous. Consequences and limits of nonlocal strategies. In *Proceedings of the 19$^{th}$ Annual IEEE Conference on Computational Complexity*, CCC '04, pages 236–249, June 2004, arXiv:quant-ph/0404076.

[CKMR07] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. One-and-a-half quantum de Finetti theorems. *Communications in Mathematical Physics*, 273(2):473–498, 2007, arXiv:quant-ph/0602130.

[CS14a] André Chailloux and Giannicola Scarpa. Parallel repetition of entangled games with exponential decay via the superposed information cost. In Javier Esparza, Pierre Fraigniaud, Thore Husfeldt, and Elias Koutsoupias, editors, *Automata, Languages, and Programming*, volume 8572 of *Lecture Notes in Computer Science*, pages 296–307. Springer Berlin / Heidelberg, 2014, arXiv:1310.7787.

[CS14b] André Chailloux and Giannicola Scarpa. Parallel repetition of free entangled games: Simplification and improvements. October 2014, arXiv:1410.4397.

[CSUU08] Richard Cleve, William Slofstra, Falk Unger, and Sarvagya Upadhyay. Perfect parallel repetition theorem for quantum xor proof systems. *Computational Complexity*, 17(2):282–299, 2008, arXiv:quant-ph/0608146.

[CT06] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley-Interscience, 2$^{nd}$ edition, 2006.

[CWY14] Kai-Min Chung, Xiaodi Wu, and Henry Yuen. Strong parallel repetition for free entangled games, with any number of players. November 2014, arXiv:1411.1397.

[Din07] Irit Dinur. The PCP theorem by gap amplification. *Journal of the ACM*, 54(3), June 2007.

[DN06] Christopher M. Dawson and Michael A. Nielsen. The Solovay-Kitaev algorithm. *Quantum Information and Computation*, 6(1):81–95, January 2006, arXiv:quant-ph/0505030.

[DS14] Irit Dinur and David Steurer. Analytical approach to parallel repetition. In *Proceedings of the 46$^{th}$ Annual ACM Symposium on Theory of Computing*, STOC '14, pages 624–633, 2014, arXiv:1305.1979.

[DSV14] Irit Dinur, David Steurer, and Thomas Vidick. A parallel repetition theorem for entangled projection games. In *Proceedings of the 29$^{th}$ Annual IEEE Conference on Computational Complexity*, CCC '14, pages 197–208, June 2014, arXiv:1310.4113.

[EPR35] Albert Einstein, Boris Podolsky, and Nathan Rosen. Can quantum-mechanical description of physical reality be considered complete? *Physical Review*, 47:777–780, May 1935.

[FK00] Uriel Feige and Joe Kilian. Two-prover protocols—low error at affordable rates. *SIAM Journal on Computing*, 30(1):324–346, 2000.

[FL92] Uriel Feige and László Lovász. Two-prover one-round proof systems: Their power and their problems (extended abstract). In *Proceedings of the 24$^{th}$ Annual ACM Symposium on Theory of Computing*, STOC '92, pages 733–744, 1992.

[For89] Lance Jeremy Fortnow. *Complexity-Theoretic Aspects of Interactive Proof Systems*. PhD thesis, Massachusetts Institute of Technology, June 1989.

[GJ79] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman, 1$^{st}$ edition, 1979.

[GMR89] Shafi Goldwasser, Silvio Micali, and Charles Rackoff. The knowledge complexity of interactive proof systems. *SIAM Journal on Computing*, 18(1):186–208, 1989.

[GN13] David Gosset and Daniel Nagaj. Quantum 3-SAT is $\mathrm{QMA}_1$-complete. In *Proceedings of the 54$^{th}$ Annual IEEE Symposium on Foundations of Computer Science*, FOCS '13, pages 756–765, October 2013, arXiv:1302.0290.

[GNN12] François Le Gall, Shota Nakagawa, and Harumichi Nishimura. On QMA protocols with two short quantum proofs. *Quantum Information and Computation*, 12(7-8):589–600, July 2012, arXiv:1108.4306.

[GS86] Shafi Goldwasser and Michael Sipser. Private coins versus public coins in interactive proof systems. In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing*, STOC '86, pages 59–68, 1986.

[GSU13] Sevag Gharibian, Jamie Sikora, and Sarvagya Upadhyay. QMA variants with polynomially many provers. *Quantum Information and Computation*, 13(1–2):135–157, January 2013, arXiv:1108.0617.

[GW83] Hana Galperin and Avi Wigderson. Succinct representations of graphs. *Information and Control*, 56(3):183–198, 1983.

[GW07] Gus Gutoski and John Watrous. Toward a general theory of quantum games. In *Proceedings of the 39th Annual ACM Symposium on Theory of Computing*, STOC '07, pages 565–574, 2007, arXiv:quant-ph/0611234.

[HM13] Aram W. Harrow and Ashley Montanaro. Testing product states, quantum Merlin-Arthur games and tensor optimization. *Journal of the ACM*, 60(1):3:1–3:43, February 2013, arXiv:1001.0017.

[Hol09] Thomas Holenstein. Parallel repetition: Simplification and the no-signaling case. *Theory of Computing*, 5(8):141–172, 2009, arXiv:cs/0607139.

[HR10] Esther Hänggi and Renato Renner. Device-independent quantum key distribution with commuting measurements. September 2010, arXiv:1009.1833.

[IKW12] Tsuyoshi Ito, Hirotada Kobayashi, and John Watrous. Quantum interactive proofs with weak error bounds. In *Proceedings of the 3rd Conference on Innovations in Theoretical Computer Science*, ITCS '12, pages 266–275, 2012, arXiv:1012.4427.

[IV12] Tsuyoshi Ito and Thomas Vidick. A multi-prover interactive proof for NEXP sound against entangled provers. In *Proceedings of the 53rd Annual IEEE Symposium on Foundations of Computer Science*, FOCS '12, pages 243–252, October 2012, arXiv:1207.0550.

[JJUW11] Rahul Jain, Zhengfeng Ji, Sarvagya Upadhyay, and John Watrous. QIP = PSPACE. *Journal of the ACM*, 58(6):30:1–30:27, December 2011, arXiv:0907.4737.

[JKNN12] Stephen P. Jordan, Hirotada Kobayashi, Daniel Nagaj, and Harumichi Nishimura. Achieving perfect completeness in classical-witness quantum Merlin-Arthur proof systems. *Quantum Information and Computation*, 12(5–6):461–471, May 2012, arXiv:1111.5306.

[JN12] Rahul Jain and Ashwin Nayak. Short proofs of the quantum substate theorem. *IEEE Transactions on Information Theory*, 58(6):3664–3669, June 2012, arXiv:1103.6067.

[JPY14] Rahul Jain, Attila Pereszlényi, and Penghui Yao. A parallel repetition theorem for entangled two-player one-round games under product distributions. In *Proceedings of the 29$^{th}$ Annual IEEE Conference on Computational Complexity*, CCC '14, pages 209–216, June 2014, arXiv:1311.6309.

[JRS03] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. A lower bound for the bounded round quantum communication complexity of set disjointness. In *Proceedings of the 44$^{th}$ Annual IEEE Symposium on Foundations of Computer Science*, FOCS '03, pages 220–229, October 2003, arXiv:quant-ph/0303138.

[JRS08] Rahul Jain, Jaikumar Radhakrishnan, and Pranab Sen. Optimal direct sum and privacy trade-off results for quantum and classical communication complexity. July 2008, arXiv:0807.1267.

[Kho02] Subhash Khot. On the power of unique 2-prover 1-round games. In *Proceedings of the 34$^{th}$ Annual ACM Symposium on Theory of Computing*, STOC '02, pages 767–775, 2002.

[Kit97] A. Yu Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[KKMV08] Julia Kempe, Hirotada Kobayashi, Keiji Matsumoto, and Thomas Vidick. Using entanglement in quantum multi-prover interactive proofs. In *Proceedings of the 23$^{rd}$ Annual IEEE Conference on Computational Complexity*, CCC '08, pages 211–222, June 2008, arXiv:0711.3715.

[KKR06] Julia Kempe, Alexei Kitaev, and Oded Regev. The complexity of the local hamiltonian problem. *SIAM Journal on Computing*, 35(5):1070–1097, 2006, arXiv:quant-ph/0406180.

[KLGN13] Hirotada Kobayashi, François Le Gall, and Harumichi Nishimura. Stronger methods of making quantum interactive proofs perfectly complete. In *Proceedings of the 4$^{th}$ Conference on Innovations in Theoretical Computer Science*, ITCS '13, pages 329–352, 2013, arXiv:1210.1290.

[KM03] Hirotada Kobayashi and Keiji Matsumoto. Quantum multi-prover interactive proof systems with limited prior entanglement. *Journal of Computer and System Sciences*, 66(3):429–450, 2003, arXiv:cs/0102013.

[KMY03] Hirotada Kobayashi, Keiji Matsumoto, and Tomoyuki Yamakami. Quantum Merlin-Arthur proof systems: Are multiple Merlins more helpful to Arthur? In *Algorithms and Computation*, volume 2906 of *Lecture Notes in Computer Science*, pages 189–198. Springer Berlin / Heidelberg, 2003, arXiv:quant-ph/0306051.

[Kni96] Emanuel Knill. Quantum randomness and nondeterminism. October 1996, arXiv:quant-ph/9610012.

[KRT10] Julia Kempe, Oded Regev, and Ben Toner. Unique games with entangled provers are easy. *SIAM Journal on Computing*, 39(7):3207–3229, 2010, arXiv:0710.0655.

[KSV02] A. Yu. Kitaev, A. H. Shen, and M. N. Vyalyi. *Classical and Quantum Computation*. American Mathematical Society, 2002.

[KV11] Julia Kempe and Thomas Vidick. Parallel repetition of entangled games. In *Proceedings of the 43$^{rd}$ Annual ACM Symposium on Theory of Computing*, STOC '11, pages 353–362, 2011, arXiv:1012.4728.

[KW00] Alexei Kitaev and John Watrous. Parallelization, amplification, and exponential time simulation of quantum interactive proof systems. In *Proceedings of the 32$^{nd}$ Annual ACM Symposium on Theory of Computing*, STOC '00, pages 608–617, 2000.

[LCV07] Yi-Kai Liu, Matthias Christandl, and F. Verstraete. Quantum computational complexity of the $N$-representability problem: QMA complete. *Physical Review Letters*, 98:110503, Mar 2007.

[LFKN92] Carsten Lund, Lance Fortnow, Howard Karloff, and Noam Nisan. Algebraic methods for interactive proof systems. *Journal of the ACM*, 39(4):859–868, October 1992.

[MPA11] Lluís Masanes, Stefano Pironio, and Antonio Acín. Secure device-independent quantum key distribution with causally independent measurement devices. *Nature Communications*, 2(238), March 2011, arXiv:1009.1567.

[MW05] Chris Marriott and John Watrous. Quantum Arthur-Merlin games. *Computational Complexity*, 14(2):122–152, 2005, arXiv:cs/0506068.

[NC00] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2000.

[NWZ09] Daniel Nagaj, Pawel Wocjan, and Yong Zhang. Fast amplification of QMA. *Quantum Information and Computation*, 9(11):1053–1068, November 2009, arXiv:0904.1549.

[Per12a] Attila Pereszlényi. Multi-prover quantum Merlin-Arthur proof systems with small gap. May 2012, arXiv:1205.2761.

[Per12b] Attila Pereszlényi. On quantum interactive proofs with short messages. *Chicago Journal of Theoretical Computer Science*, 2012(9):1–10, December 2012, arXiv:1109.0964.

[Per13] Attila Pereszlényi. One-sided error QMA with shared EPR pairs—A simpler proof. June 2013, arXiv:1306.5406. Contributed talk at AQIS '13. To appear in *Theoretical Computer Science*.

[PY86] Christos H. Papadimitriou and Mihalis Yannakakis. A note on succinct representations of graphs. *Information and Control*, 71(3):181–185, 1986.

[Rao11] Anup Rao. Parallel repetition in projection games and a concentration bound. *SIAM Journal on Computing*, 40(6):1871–1891, 2011.

[Raz98] Ran Raz. A parallel repetition theorem. *SIAM Journal on Computing*, 27(3):763–803, 1998.

[Raz11] Ran Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011.

[RR12] Ran Raz and Ricky Rosen. A strong parallel repetition theorem for projection games on expanders. In *Proceedings of the 27$^{th}$ Annual IEEE Conference on Computational Complexity*, CCC '12, pages 247–257, June 2012.

[RUV13] Ben W. Reichardt, Falk Unger, and Umesh Vazirani. Classical command of quantum systems. *Nature*, 496(7446):456–460, April 2013, arXiv:1209.0448.

[Sha92] Adi Shamir. IP = PSPACE. *Journal of the ACM*, 39(4):869–877, October 1992.

[She92] A. Shen. IP = PSPACE: Simplified proof. *Journal of the ACM*, 39(4):878–880, October 1992.

[TFKW13] Marco Tomamichel, Serge Fehr, Jędrzej Kaniewski, and Stephanie Wehner. A monogamy-of-entanglement game with applications to device-independent quantum cryptography. *New Journal of Physics*, 15(10):103002, 2013, arXiv:1210.4359.

[Vya03] Mikhail N. Vyalyi. QMA = PP implies that PP contains PH. Technical Report 21 (2003), Electronic Colloquium on Computational Complexity, April 2003. TR03-021.

[Wat00] John Watrous. Succinct quantum proofs for properties of finite groups. In *Proceedings of the 41st Annual IEEE Symposium on Foundations of Computer Science*, FOCS '00, pages 537–546, 2000, arXiv:cs/0009002.

[Wat03] John Watrous. PSPACE has constant-round quantum interactive proof systems. *Theoretical Computer Science*, 292(3):575–588, 2003.

[Wat08a] John Watrous. Quantum computational complexity. April 2008, arXiv:0804.3401.

[Wat08b] John Watrous. Theory of quantum information. Lecture notes from Fall 2008, https://cs.uwaterloo.ca/~watrous/quant-info/, 2008.

[Wat09] John Watrous. Zero-knowledge against quantum attacks. *SIAM Journal on Computing*, 39(1):25–58, 2009, arXiv:quant-ph/0511020.

[Win99] Andreas Winter. *Coding Theorems of Quantum Information Theory*. PhD thesis, Universität Bielefeld, 1999, arXiv:quant-ph/9907077.

[ZF87] Stathis Zachos and Martin Fürer. Probabalistic quantifiers vs. distrustful adversaries. In *Proceedings of the 7th Conference on Foundations of Software Technology and Theoretical Computer Science*, FSTTCS '87, volume 287 of *Lecture Notes in Computer Science*, pages 443–455, London, UK, 1987. Springer-Verlag.