

אוניברסיטת תל-אביב

הפקולטה למדעים מדויקים ע"ש ריימונד וברלי סאקלר

בית הספר למדעי המחשב

מחלצי אקראיות משני מקורות הבטוחים כנגד יריבים קוונטיים

חיבור זה מוגש כחלק ממילוי הדרישות

לקבלת התואר "מוסמך למדעים" (M. Sc.)

בבית הספר למדעי המחשב באוניברסיטת תל-אביב

ע"י

רועי כשר

העבודה הוכנה בהדרכתם של

דוקטור יוליה קמפה ופרופסור אמנון תא-שמע

אייר ה'תש"ע

תקציר

אנו פותחים במחקר של מחלצי אקראיות (randomness extractors) ממספר מקורות בלתי תלויים בעולם הקוונטי. מטרתנו היא לחלץ סיביות אקראיות משני מקורות אקראיות-חלשה, החשופים חלקית ליריבים קוונטים.

תוצאתנו המרכזית היא מחלץ אקראיות משני מקורות הבטוח כנד יריבים קוונטיים, עם פרמטרים המתקרבים לאלו בעולם הקלאסי, ואף הדוקים במקרים מסוימים. בנוסף, המלחץ בטוח גם במצב של שזירה קוונטית בין היריבים. המחלץ הוא למעשה מחלץ המכפלה הפנימית הבוליאני של שור-גולדרייך [CG88], ומקבילתו מרובת הסיביות של דודיס ואחרים [DEOR04].

עד עתה, המחקר בתחום התמקד בבנייה של מחלצי אקראיות ממקור בודד הבטוחים כנגד יריבים קוונטים. סביבה של מספר מקורות מציבה אתגרים חדשים, בפרט התמודדות עם שזירה קוונטית שעלולה לשבור את האי-תלות בין המקורות.

TEL-AVIV UNIVERSITY
RAYMOND AND BEVERLY SACKLER
FACULTY OF EXACT SCIENCES
SCHOOL OF COMPUTER SCIENCE

Two-Source Extractors Secure Against Quantum Adversaries

Thesis submitted in partial fulfillment of the requirements for the M.Sc. degree in the
School of Computer Science, Tel-Aviv University

by

Roy Kasher

The research work for this thesis has been carried out at Tel-Aviv University
under the supervision of Dr. Julia Kempe and Prof. Amnon Ta-Shma

May 2010

Acknowledgements

I would like to thank my advisor Dr. Julia Kempe for her patience and expertise, Prof. Amnon Ta-Shma for teaching me extractors and Prof. Ran Canetti for providing a fun environment in which to learn. I am grateful to Prof. Oded Regev, Thomas Vidick and Prof. Ronald de Wolf for their contributions. Lastly, I thank Nir and Omer, my fellow students, for times had.

Abstract

We initiate the study of multi-source extractors in the quantum world. In this setting, our goal is to extract random bits from two independent weak random sources, on which two quantum adversaries store a bounded amount of information. Our main result is a two-source extractor secure against quantum adversaries, with parameters closely matching the classical case and tight in several instances. Moreover, the extractor is secure even if the adversaries share entanglement. The construction is the Chor-Goldreich [CG88] two-source inner product extractor and its multi-bit variant by Dodis et al. [DEOR04]. Previously, research in this area focused on the construction of seeded extractors secure against quantum adversaries; the multi-source setting poses new challenges, among which is the presence of entanglement that could potentially break the independence of the sources.

Contents

1	Introduction and Results	1
1.1	Background	1
1.2	Our Results	3
1.3	Proof Ideas and Tools	5
1.4	Related Work	6
1.5	Discussion and Open Problems	7
1.6	Organization	8
2	Preliminaries and Tools	9
2.1	Quantum Computation	9
2.2	Two-Source Extractors and the DEOR Construction	10
2.3	Classical-Quantum XOR-Lemma	12
3	Extractors Against Quantum Storage	15
3.1	Definition	15
3.2	One Bit Extractor	16
3.2.1	Average Case Lower Bound for Inner Product	16
3.2.2	The Extractor and Tightness Results	18
3.3	Many Bit Extractor	20
4	Extractors Against Quantum Knowledge	23
4.1	Guessing Entropy	23
4.2	Non-entangled Adversaries	24
4.2.1	One-Bit Output	24
4.2.2	Multi-Bit Output	26
4.3	Entangled Adversaries	27
A	Many Bit Extractors Against Quantum Storage from Classical Storage	29
	Bibliography	31

Chapter 1

Introduction and Results

1.1 Background

Randomness extractors are fundamental in many areas of computer science, with numerous applications to derandomization, error-correcting codes, expanders, combinatorics and cryptography, to name just a few. Randomness extractors generate almost uniform randomness from imperfect sources, as they appear either in nature, or in various applications. Typically, the imperfect source is modelled as a distribution over n -bit strings whose *min-entropy* is at least k , i.e., a distribution in which no string occurs with probability greater than 2^{-k} [SV84, CG88, Zuc90]. Such sources are known as *weak sources*. One way to arrive at a weak source is to imagine that an adversary (or some process in nature), when in contact with a uniform source, *stores* $n - k$ bits of information about the string (which are later used to break the security of the extractor, i.e. to distinguish its output from uniform). Then, from the adversary's point of view, the source essentially has min-entropy k .

Ideally, we would like to extract randomness from a weak source. However, it is easy to see that no deterministic function can extract even one bit of randomness from all such sources, even for min-entropies as high as $n - 1$ (see e.g. [SV84]). One main approach to circumvent this problem is to use a short truly random *seed* for extraction from the weak source (*seeded extractors*) (see, e.g., [Sha02]). The other main approach, which is the focus of the current work, is to use several independent weak sources (*seedless extractors*) (e.g. [CG88, Vaz87, DEOR04, Bou05, Raz05] and many more).

With the advent of quantum computation, we must now deal with the possibility of quantum adversaries (or quantum physical processes) interfering with the sources used for randomness extraction. For instance, one could imagine that a quantum adversary now stores $n - k$ *qubits* of information about the string sampled from the source. This scenario of a *bounded storage quantum adversary* arises in several applications, in particular in cryptography.

Some constructions of *seeded* extractors were shown to be secure in the presence of quantum

adversaries: König, Maurer, and Renner [RK05, KMR05, Ren05] proved that the pairwise independent extractor of [ILL89] is also good against quantum adversaries, and with the same parameters. König and Terhal [KT08] showed that any one-bit output extractor is also good against quantum adversaries, with roughly the same parameters. In light of this, it was tempting to conjecture that *any* extractor is also secure against quantum storage. Somewhat surprisingly, Gavinsky et al. [GKK⁺08] gave an example of a seeded extractor that is secure against classical storage but becomes insecure even against very small quantum storage. This example has initiated a series of recent ground-breaking work that examined which seeded extractors stay secure against bounded storage quantum adversaries. Ta-Shma [Ta-09] gave an extractor with a short (polylogarithmic) seed extracting a polynomial fraction of the min-entropy. His result was improved by De and Vidick [DV10] extracting almost all of the min-entropy. Both constructions are based on Trevisan’s extractor [Tre01].

However, the question of whether *seedless* multi-source extractors can remain secure against quantum adversaries has remained wide open. The multi-source scenario corresponds to several independent adversaries, each tampering with one of the sources, and then jointly trying to distinguish the extractor’s output from uniform: One can imagine a malicious entity planting (quantum) storage devices (possibly sharing an entangled state) in remote locations, later collecting the devices and observing their joint state. In the classical setting this just leads to several independent weak sources. In the quantum world, measuring the adversaries’ stored information might break the independence of the sources, thus jeopardizing the performance of the extractor.¹ Moreover, the multi-source setting offers a completely new aspect of the problem: the adversaries could potentially share *entanglement* prior to tampering with the sources. Entanglement between several parties has been known to yield several astonishing effects with no counterpart in the classical world, e.g., non-local correlations [Bel64] and superdense coding [BW92].

We note that the example of Gavinsky et al. can also be viewed as an example in the two-source model; we can imagine that the seed comes from a second source (of full entropy in this case, just like any seeded extractor can be artificially viewed as a two-source extractor). And obviously, in the same way, recent work on quantum secure seeded extractors artificially gives secure two-source extractors, albeit for a limited range of parameters and without allowing for entanglement. However, no one has as of yet explored how more realistic multi-source extractors fare against quantum adversaries, and in particular how entanglement might change the picture. We ask: Are there any good multi-source extractors secure against quantum bounded storage? And does this remain true when considering entanglement?

¹Such an effect appears also in *strong seeded* extractors and has been discussed in more detail in [KT08].

1.2 Our Results

We answer the above questions in the positive. We focus on the inner-product based two-source extractor of Dodis et al. [DEOR04] (DEOR-extractor). Given two independent weak sources X and Y with the same length n and min-entropies k_1 and k_2 satisfying $k_1 + k_2 \gtrsim n$, this extractor gives m close to uniform random bits, where $m \approx \max(k_1, k_2) + k_1 + k_2 - n$. In recent years several two-source extractors with better parameters have been presented; however, the DEOR-construction stands out through its elegance and simplicity and its parameters still fare very well in comparison with recent work (e.g., [Bou05, Raz05]).

A first conceptual step in this paper is to define the model of quantum adversaries and of security in the two-source scenario: Each adversary gets access to an independent weak source X (resp. Y), and is allowed to store a *short* arbitrary quantum state.² In the entangled setting, the two adversaries may share arbitrary prior entanglement, and hence their final joint stored state is the possibly entangled state ρ_{XY} . In the non-entangled case their joint state is of the form $\rho_{XY} = \rho_X \otimes \rho_Y$. In both cases, the security of the extractor is defined with respect to the joint state they store.

Definition 1.1. *[Two-source extractor against (entangled) quantum storage (informal):] A function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ is a (k_1, k_2, ε) extractor against (b_1, b_2) (entangled) quantum storage if for any sources X, Y with min-entropies k_1, k_2 , and any joint stored quantum state ρ_{XY} prepared as above, with X -register of b_1 qubits and Y -register of b_2 qubits, the distribution $E(X, Y)$ is ε -close to uniform even when given access to ρ_{XY} .*

Depending on the type of adversaries, we will say E is secure against *entangled* or *non-entangled* storage. Note again that entanglement between the adversaries is specific to the multi-source scenario and does not arise in the case of seeded extractors.

Having set the framework, we show that the construction of Dodis et al. [DEOR04] is secure, first in the case of non-entangled adversaries.

Theorem 1.2. *The DEOR-construction is a (k_1, k_2, ε) extractor against (b_1, b_2) non-entangled storage with $m = (1 - o(1)) \max(k_1 - \frac{b_1}{2}, k_2 - \frac{b_2}{2}) + \frac{1}{2}(k_1 - b_1 + k_2 - b_2 - n) - 9 \log \varepsilon^{-1} - O(1)$ output bits, provided $k_1 + k_2 - \max(b_1, b_2) > n + \Omega(\log^3(n/\varepsilon))$.*

As we show next the extractor remains secure even in the case of entangled adversaries. Notice the loss of essentially a factor of 2 in the allowed storage; this is related to the fact that superdense coding allows to store n bits using only $n/2$ entangled qubit pairs.

Theorem 1.3. *The DEOR-construction is a (k_1, k_2, ε) extractor against (b_1, b_2) entangled storage with $m = (1 - o(1)) \max(k_1 - b_2, k_2 - b_1) + \frac{1}{2}(k_1 - 2b_1 + k_2 - 2b_2 - n) - 9 \log \varepsilon^{-1} - O(1)$ output bits, provided $k_1 + k_2 - 2 \max(b_1, b_2) > n + \Omega(\log^3(n/\varepsilon))$.*

²In the setting of seeded extractors with one source, this type of adversary was called *quantum encoding* in [Ta-09].

Note that in both cases, when the storage is linear in the source entropy we can output $\Omega(n)$ bits with exponentially small error. To compare to the performance of the DEOR-extractor in the classical case, note that a source with min-entropy k and *classical* storage of size b roughly corresponds to a source of min-entropy $k-b$ (see, e.g., [Ta-09] Lem. 3.1). Using this correspondence, the extractor of [DEOR04] gives $m = \max(k_1, k_2) + k_1 - b_1 + k_2 - b_2 - n - 6 \log \varepsilon^{-1} - O(1)$ output bits against classical storage, whenever $k_1 + k_2 - \max(b_1, b_2) > n + \Omega(\log^3 n + \log n \log \varepsilon^{-1})$ (see Sec. 2.2). Hence the conditions under which we can extract randomness are essentially the same for DEOR and for our Thm. 1.2. The amount of random bits we can extract is somewhat less than in the classical case, even when disregarding storage.

In the non-entangled case, we are able to generalize our result to the stronger notion of guessing entropy adversaries or so called *quantum knowledge*. We show that the DEOR-extractor remains secure even in this case, albeit with slightly weaker parameters.

Theorem 1.4. *The DEOR-construction is a (k_1, k_2, ε) extractor against quantum knowledge with $m = (1 - o(1)) \max(k_1, k_2) + \frac{1}{6}(k_1 + k_2 - n) - 9 \log \varepsilon^{-1} - O(1)$ output bits, provided $k_1 + k_2 > n + \Omega(\log^3(n/\varepsilon))$.*

In this setting, we place a bound on the *guessing entropy* of the source given the adversary's storage (rather than the size of the storage). Informally, a guessing entropy of at least k means that the adversary's probability of correctly guessing the source is at most 2^{-k} (or equivalently, that given the adversary's state, the source has essentially min-entropy at least k). Working with guessing entropy has the advantage that we no longer have to worry about two parameters (min-entropy and storage) instead only working with one parameter (guessing entropy), and that the resulting extractors are stronger (assuming all other parameters are the same), see Chap. 4. In the entangled case, defining the security of extractors is trickier, and we give a couple of impossibility results in Sec.4.3. We note that in the classical world, a guessing entropy of k is more or less equivalent to a source with k min-entropy; in the quantum world, however, things become less trivial. In the case of seeded extractors, this more general model has been successfully introduced and studied in [Ren05, KT08, FS08, DPVR09, TSSR10], where several constructions secure against bounded guessing entropy were shown.³

Strong extractors: The extractor in Thms. 1.2, 1.3 and 1.4 is a so called *weak* extractor, meaning that when trying to break the extractor, no full access to any of the sources is given (which is natural in the multi-source setting). We also obtain several results in the so called *strong* case (see Cor. 3.6, Lem. 3.10, Cor. 4.8 and Lem. 4.9). A *strong* extractor has the additional property that the output remains secure even if the adversaries later gain full access to any one (but obviously not both) of the sources.⁴ See Chap. 3 for details and a discussion of the subtleties in defining a strong extractor

³Renner [Ren05] deals with the notion of *relative min-entropy*, which was shown to be equivalent to guessing entropy [KRS09].

⁴In [DEOR04], this is called a *strong blender*.

in the entangled case, and Chaps. 3, 4 for our results in the strong case.

Tightness: In the one-bit output case, we show that our results are *tight*, both in the entangled and non-entangled setting against storage (see Lem. 3.8).

1.3 Proof Ideas and Tools

To show all of our results, we first focus on the simplest case of one-bit outputs. In this case the DEOR extractor [DEOR04] simply computes the inner product $E(x, y) = x \cdot y \pmod{2}$ of the n -bit strings x and y coming from the two sources. Assume that the two adversaries are allowed quantum storage of b qubits each. Given their stored information they jointly wish to distinguish $E(x, y)$ from uniform, or, in other words, to predict $x \cdot y$. We start by observing that this setting corresponds to the well known simultaneous message passing (SMP) model in communication complexity,⁵ where two parties, Alice and Bob, have access to an input each (which is unknown to the other). They each send a message of length b to a referee, who, upon reception of both messages, is to compute a function $E(x, y)$ of the two inputs. When E is hard to compute, it is a good extractor. Moreover, the entangled adversaries case corresponds to the case of SMP with entanglement between Alice and Bob, a model that has been studied in recent work (see e.g. [GKRdW09, GKdW06]).

Before we proceed, let us remark that there are cases where entanglement is known to add tremendous power to the SMP model. Namely, Gavinsky et al. [GKRdW09] showed an exponential saving in communication in the entangled SMP model, compared to the non-entangled case.⁶ This points to the possibility that some extractors can be secure against a large amount of storage in the non-entangled case, but be insecure against drastically smaller amounts of entangled storage. Our results show that this is not the case for the DEOR extractor, i.e., that this construction is secure against the potentially harmful effects of entanglement.

In the one-bit output DEOR case against bounded storage we can tap into known results on the quantum communication complexity of the inner product problem (IP). Cleve et al. [CvDNT98] and Nayak and Salzman [NS06] have given tight lower bounds in the one-way and two-way communication model, with and without entanglement (which also gives bounds in the SMP model). For instance, in the non-entangled case, to compute IP exactly in the one-way model, n qubits of communication are needed, and in the SMP model, n qubits of communication are needed from Alice and from Bob, just like in the classical case. Note that whereas in the communication setting typically worst case problems are studied, extractors correspond to *average case* (w.r.t. to weak randomness) problems. With some extra work we can adapt the communication lower bounds to

⁵The connection between extractors and communication complexity has been long known, see, e.g., [Vaz87].

⁶This result has been shown for a relation, not a function. It is tempting to conjecture that this result can be turned into an exponential separation for an extractor with entangled vs. non-entangled adversaries. It is, however, not immediate how to turn a worst case relation lower bound into an average case function bound, as needed in the extractor setting, so we leave this problem open.

weak sources and to the average bias which is needed for the extractor result. In fact, the results we obtain hold in the strong case (where later one of the sources is completely exposed), which corresponds to one-way communication complexity.

Tightness of our results comes from matching upper bounds on the one-way and SMP model communication complexity of the inner product. Adapting the work of [CG88] we can obtain tight bounds for any bias ε . Somewhat surprisingly, it seems no one has looked at tight upper bounds for IP in the *entangled SMP model*, where [CvDNT98] give an $n/2$ lower bound for the message length for Alice and Bob. It turns out this bound is tight,⁷ which essentially leads to the factor 2 separation in our results for entangled vs. non-entangled case (see Chap. 3).

In the case of *non-entangled* guessing entropy adversaries, we can show (based on [KT08]) that any classical *one-bit* output two-source extractor remains secure against bounded guessing entropy adversaries, albeit with slightly worse parameters. In the *entangled* adversaries case, extending the one adversary model of seeded extractors to our two adversary scenario is not as trivial. We introduce several possible models and provide some impossibility results. In particular, we show that inner product is insecure in two of the models. See Chap. 4 for details.

To show our results for the case of multi-bit extractors, we use the nice properties of the DEOR construction (and its precursors [Vaz87, DO03]). The extractor outputs bits of the form $Ax \cdot y$. Vazirani’s XOR-Lemma allows to reduce the multi-bit to the one-bit case by relating the distance from uniform of the multi-bit extractor to the sum of biases of XOR’s of subsets of its bits. Each such XOR, in turn, is just a (linearly transformed) inner product, for which we already know how to bound the bias. Our main technical challenge is to adapt the XOR lemma to the case of *quantum* side-information (see Chap. 2). This way we obtain first results for multi-bit extractors, which even hold in the case of strong extractors. Following [DEOR04], we further improve the parameters in the *weak* extractor setting by combining our strong two-source extractor with a good seeded extractor (in our case with the construction of [DPVR09]) to extract even more bits. See Secs. 3.3 and 4.2 for details.

1.4 Related Work

We are the first to consider two-source extractors in the quantum world, especially against entanglement. As mentioned, previous work on seeded extractors against quantum adversaries [RK05, KMR05, Ren05, KT08, Ta-09, DV10, DPVR09, BT10] gives rise to trivial two-source extractors where one of the sources is not touched by the adversaries. However, the only previous work that allows to derive results in the genuine two-source scenario is the work by König and Terhal [KT08]. Using what is implicit in their work, and with some extra effort, it is possible to derive results in the one-bit output non-entangled two-source scenario (which hold against guessing entropy adver-

⁷We thank Ronald de Wolf [dW10] for generously allowing us to adapt his upper bound to our setting.

saries, but with worse performance than our results for the inner product extractor), and we give this result in detail in Chap. 4. Moreover, they show that any classical multi-bit extractor is secure against bounded storage adversaries, albeit with an exponential decay in the error parameter. This easily extends to the non-entangled two-source scenario, to give results in the spirit of Thm. 1.2. We have worked out the details and comparison to Thm. 1.2 in App. A. Note, however, that to our knowledge no previous work gives results in the entangled scenario.

1.5 Discussion and Open Problems

We have, for the first time, studied two-source extractors in the quantum world. Previously, only seeded extractors have been studied in the quantum setting. In the two-source scenario a new phenomenon appears: entanglement between the (otherwise independent) sources. We have formalized what we believe the strongest possible notion of quantum adversaries in this setting and shown that one of the best performing extractors, the DEOR-construction, remains secure. We also show that our results are tight in the one-bit output case.

Our results for the multi-bit output DEOR-construction allow to extract slightly less bits compared to what is possible classically. An interesting open question is whether it is possible to obtain matching parameters in the (non-entangled) quantum case. One might have to refine the analysis and not rely solely on communication complexity lower bounds. Alternatively, our quantum XOR-Lemma currently incurs a penalty exponential in either the length of the output or the length of the storage. Any improvement here also immediately improves all three main theorems. In particular, by removing the penalty entirely, Thm. 1.2 can be made essentially optimal (with respect to the classical case).

We show (see Sec. 4.3 that inner product based constructions are necessarily insecure in two reasonable models of entangled guessing entropy adversaries (and hence that bounded storage adversaries are the more appropriate model in the entangled case). It should be noted that it is possible that other extractor constructions (not based on inner product) could remain secure in this setting, and this subject warrants further exploration.

As pointed out, it is conceivable that entanglement could break the security of two-source extractors. Evidence for this is provided by the communication complexity separation in the entangled vs. non-entangled SMP-model, given in [GKRdW09]. A fascinating open problem is to turn this relational separation into an extractor that is secure against non-entangled quantum adversaries but completely broken when entanglement is present.

Our work leaves several other open questions. It would be interesting to see if other multi-source extractors remain secure against entangled adversaries, in particular the recent breakthrough construction by Bourgain [Bou05] which works for two sources with min-entropy $(1/2 - \alpha)n$ each for some small constant α , or the construction of Raz [Raz05], where one source is allowed to have

logarithmic min-entropy while the other has min-entropy slightly larger than $n/2$. Both extractors output $\Omega(n)$ almost uniform bits.

And lastly, it would be interesting to see other applications of secure multi-source extractors in the quantum world. One possible scenario is multi-party computation. To understand the connection between extractors and multi-party computation, consider a very simple scenario where we allow one (known) honest party to be left out of the computation. Here, the honest player could simply reveal his (weak) source, and all other players would apply a strong two-source extractor to this source and their own sources, resulting in private randomness for all but the honest party.⁸ This randomness may then be used in any traditional protocol requiring perfect randomness. Classically, Kalai et al. [KLR09] showed that sufficiently strong two-source extractors allow to perform multi-party communication with weak sources when at least two parties are honest. Perhaps similar results hold in the quantum setting.

1.6 Organization

In Chap. 2 we introduce our basic notation and definitions, and describe the DEOR construction. Here we also present one of our tools, the "quantum" XOR-Lemma. Chap. 3 presents extractors against quantum storage. Sec. 3.2 is dedicated to the one-bit output case and the connection to communication complexity and gives our tightness results. In Sec. 3.3 we deal with the multi-bit output case and prove Thms. 1.2 and 1.3. Chap. 4 deals with guessing entropy adversaries. We give our results in the non-entangled setting (partly based on [KT08]) and prove Thm. 1.4 in Sec. 4.2. We discuss the entangled setting in Sec. 4.3 and provide some impossibility results. App. A works out the results that can be derived from [KT08] in the case of multi-bit extractors against non-entangled bounded storage.

⁸Such a tool for obtaining private randomness from independent weak sources in the multi-party setting is a *network extractor protocol*, defined in [KLRZ08].

Chapter 2

Preliminaries and Tools

In this chapter we provide the necessary notation, describe the DEOR-extractor and present and prove our quantum XOR-Lemma.

2.1 Quantum Computation

A pure quantum state is a vector in some Hilbert space. Generally, a quantum system is in a mixed state - a probability distribution over pure states. Suppose a quantum system is in one of a number of states $|\psi_i\rangle$ with respective probabilities p_i . Then the behavior of the system is completely characterized by its *density matrix* (or *density operator*) $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$. Alternatively, any Hermitian positive semi-definite matrix of trace one is associated to at least one ensemble $\{\lambda_i, |v_i\rangle\}$, where λ_i and $|v_i\rangle$ are its eigenvalues and corresponding eigenvectors.

Any classical random variable Z can be modelled by the density matrix $\sum_{z \in Z} \Pr[Z = z] |z\rangle\langle z|$. Given a set of density matrices $\{\rho_z\}_{z \in Z}$ we denote by $Z\rho_Z$ the classical-quantum (cq) state $\sum_{z \in Z} \Pr[Z = z] |z\rangle\langle z| \otimes \rho_z$. When the distribution is clear from the context we write $p(z)$ instead of $\Pr[Z = z]$. For any random variable Z' on the domain of Z , we define $\rho_{Z'} := \sum_{z \in Z'} \Pr[Z' = z] \rho_z$. For any random variable Y , let $Y\rho_Z := \sum_{y \in Y} \Pr[Y = y] |y\rangle\langle y| \otimes \rho_{Z|Y=y}$. We denote by U_m the uniform distribution on m bits.

A POVM (Positive Operator Valued Measurement) on a Hilbert space is a collection $M = \{M_i\}$ of positive semi-definite operators satisfying completeness, $\sum M_i = I$. Applying a POVM M on density operator ρ results in answer i with probability $\text{Tr}(M_i\rho)$. We denote by $M(\rho)$ the resulting classical probability distribution on $\{i\}$. For example, given a cq-state $Z\rho_Z$, $M(\rho_Z)$ is defined by the marginal $\Pr[M(\rho_Z) = i | Z = z] = \text{Tr}(M_i\rho_z)$.

Trace Distance and Matrix Norms: The variational distance between two classical random variables X, Y is

$$|X - Y|_{\text{tr}} := \frac{1}{2} |X - Y|_1 = \max_T |\Pr[T(X) = 1] - \Pr[T(Y) = 1]|,$$

where the maximum ranges over all functions $T : \{0, 1\}^* \rightarrow \{0, 1\}$.

For a matrix A , we define $|A|_{\text{tr}} = \frac{1}{2} \|A\|_1 = \frac{1}{2} \text{Tr}(\sqrt{A^\dagger A})$ and $\|A\|_2 = \sqrt{\text{Tr}(A^\dagger A)}$. For density operators ρ, σ , we define the trace distance as the trace norm $|\rho - \sigma|_{\text{tr}}$. It is well known that $|\rho - \sigma|_{\text{tr}} = \max_M |M(\rho) - M(\sigma)|_{\text{tr}}$, where the maximum ranges over all POVMs, and can be restricted only to two-outcome POVMs.

When ρ and σ are classical (i.e., diagonal matrices), the two definitions coincide, justifying the use of the same notation $|\cdot|_{\text{tr}}$.

For more information, see e.g. [NC00].

2.2 Two-Source Extractors and the DEOR Construction

In this section we define extractors against classical storage, explain the DEOR-construction and give its parameters in this setting.

Two-Source Extractors: We begin with the definition of two-source extractors against *classical storage*.

Definition 2.1. A (k_1, k_2, ε) two-source extractor against (b_1, b_2) classical storage is a function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any independent n -bit weak sources X, Y with respective min-entropies k_1, k_2 , and any b_1 - and b_2 -bit random variables C_X, C_Y , such that C_X depends only on X and C_Y depends only on Y , we have

$$|E(X, Y)C_X C_Y - U_m C_X C_Y|_{\text{tr}} \leq \varepsilon.$$

The extractor is called *X-strong* if $|E(X, Y)C_Y X - U_m C_Y X|_{\text{tr}} \leq \varepsilon$. It is called *strong* if it is both *X-* and *Y-strong*.

We note that classically two-source extractors are defined without storage, i.e. $b_1 = b_2 = 0$. However, the problem of extracting randomness in the presence of storage can be reduced to the problem of extracting randomness from sources with slightly lower min-entropy. It is easy to show that with high probability over C_X the min-entropy of $X|C_X$ (and by our definition, of $X|C_X C_Y$) is reduced by roughly b_1 , and likewise for $Y|C_Y$ (see, e.g., [Ta-09] Lem. 3.1). An averaging argument then leads to the following claims:¹

Claim 2.2. If E is a $(k_1 - b_1 - \log \varepsilon^{-1}, k_2 - b_2 - \log \varepsilon^{-1}, \varepsilon)$ extractor, then E is a $(k_1, k_2, 3\varepsilon)$ extractor against (b_1, b_2) storage.

Claim 2.3. If E is a $(k_1, k_2 - b_2 - \log \varepsilon^{-1}, \varepsilon)$ X-strong extractor, then E is a $(k_1, k_2, 2\varepsilon)$ extractor against (b_1, b_2) storage.²

¹In fact, the claims hold even if we modify Def. 2.1 to use the weaker requirement that for any values c_X, c_Y , the random variables $(X|C_X = c_X, C_Y = c_Y)$ and $(Y|C_X = c_X, C_Y = c_Y)$ are independent.

²For X-strong extractors, the storage on X is irrelevant and the claim applies for any b_1 .

The DEOR Construction: The following (strong) extractor construction is due to Dodis et al. [DEOR04]. Every output bit is a linearly transformed inner product, namely $A_i x \cdot y$ for some full rank matrix A_i , where x and y are the n -bit input vectors. The matrices A_i have the additional property that every subset sum is also of full rank. This ensures that any XOR of some bits of the output is itself a transformed inner product.

Lemma 2.4 ([DEOR04]). *For all $n > 0$, there exist an efficiently computable set of $n \times n$ matrices A_1, A_2, \dots, A_n over $GF(2)$ such that for any non-empty set $S \subseteq [n]$, $A_S := \sum_{i \in S} A_i$ has full rank.*

Definition 2.5 (strong blender of [DEOR04]). *Let $n \geq m > 0$, and let $\{A_i\}_{i=1}^m$ be a set as above. The DEOR-extractor $E_D : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ is given by $E_D(x, y) = A_1 x \cdot y, A_2 x \cdot y, \dots, A_m x \cdot y$.*

Theorem 2.6 ([DEOR04, Theorem 1]). *E_D is a (k_1, k_2, ε) X-strong extractor³ provided $k_1 + k_2 \geq m + n - 2 + 2 \log \varepsilon^{-1}$.*

A rough idea of the proof is to use the XOR-Lemma (see Sec. 2.3) to reduce the security of E_D to that of the one-bit function $E_{IP}(x, y) = (\sum_{i \in S} A_i x) \cdot y$ for some subset S . This one-bit function can be described by a Boolean matrix M with the (x, y) -th entry equals to $E_{IP}(x, y)$. For sources that are uniform on their support, $E_{IP}(X, Y)$ is a good extractor if the submatrix $M_{\text{supp}(X), \text{supp}(Y)}$ is well balanced (i.e., has roughly the same number of ones and zeros), and in fact, it suffices to consider only such sources [CG88]. The proof then continues by giving a careful analysis of M .

Using the reduction of Claim 2.3 on the DEOR-extractor we get:

Corollary 2.7. *E_D is a (k_1, k_2, ε) X-strong extractor against (b_1, b_2) storage provided $k_1 + k_2 - b_2 \geq m + n + 1 + 3 \log \varepsilon^{-1}$.*

More Output Bits: A *seeded* extractor is a special case of a two-source extractor, where one of the sources is completely uniform (and usually much shorter). Again, we define the extractor with respect to classical storage.

Definition 2.8 ([Ta-09]). *A (k, ε) seeded extractor against b classical storage is a function $E : \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ such that for any n -bit weak source X with min-entropy k , and any b -bit random variable C_X , $|E(X, U_d)C_X - U_m C_X|_{\text{tr}} \leq \varepsilon$.*

As observed in [DEOR04], any X-strong two-source extractor E_B can be composed with a seeded extractor E_S by using the output of the two-source extractor as a seed: $E(x, y) := E_S(x, E_B(x, y))$. Informally, since $X, E_B(X, Y) \approx X, U$, then $E(X, Y) \approx E_S(X, U_d) \approx U_m$. A similar argument shows that we can compose also in the presence of classical storage (for details, see Sec. 3.3).

A seeded extractor with *optimal entropy loss*, i.e. $m = k + d - 2 \log \varepsilon^{-1} - O(1)$ was given by [RRV99].

³In [DEOR04], this is called a *strong blender*.

Theorem 2.9 ([RRV99, Theorem 4], and discussion thereafter). *There exists an explicit (k, ε) seeded extractor with seed length $d = O(\log n \cdot (\log^2 n + \log \varepsilon^{-1}))$ and output $m = k + d - 2 \log \varepsilon^{-1} - O(1)$.*

By a similar argument to Claim 2.3 (in particular, [Ta-09] Lem. 3.1), we get:

Corollary 2.10. *There exists an explicit (k, ε) seeded extractor against b storage with seed length $d = O(\log n \cdot (\log^2 n + \log \varepsilon^{-1}))$ and $m = k - b + d - 3 \log \varepsilon^{-1} - O(1)$ output bits.*

Composing this (k_1, ε) seeded extractor against b_1 storage with E_D , we obtain a two-source extractor with $m = 2k_1 - b_1 + k_2 - b_2 - n - 6 \log \varepsilon^{-1} - O(1)$, as long as the output of E_D is larger than $\Omega(\log^3 n + \log n \log \varepsilon^{-1})$. Similarly, we can compose the seeded extractor on the analogous Y-strong variant of E_D , and choose the better of the two results.

Theorem 2.11. *There exists an explicit (k_1, k_2, ε) two-source extractor against (b_1, b_2) storage with $m = \max(k_1, k_2) + k_1 - b_1 + k_2 - b_2 - n - 6 \log \varepsilon^{-1} - O(1)$, provided $k_1 + k_2 - \max(b_1, b_2) > n + \Omega(\log n \cdot (\log^2 n + \log \varepsilon^{-1}))$.*

2.3 Classical-Quantum XOR-Lemma

Vazirani's XOR-Lemma [Vaz87] relates the non-uniformity of a distribution to the non-uniformity of the characters of the distribution, i.e., the XOR of certain bit positions. For the DEOR-extractor it allows to reduce the multi-bit output case to the binary output case.

Lemma 2.12 (Classical XOR-Lemma [Vaz87, Gol95]). *For every m -bit random variable Z*

$$|Z - U_m|_1^2 \leq \sum_{0 \neq S \in \{0,1\}^m} |(S \cdot Z) - U_1|_1^2.$$

This lemma is not immediately applicable in our scenario, as we need to take into account *quantum* side information. For this, we need a slightly more general XOR-Lemma.

Lemma 2.13 (Classical-Quantum XOR-Lemma). ⁴ *Let $Z\rho_Z$ be an arbitrary cq-state, where Z is an m -bit classical random variable and ρ_Z is of dimension 2^d . Then*

$$|Z\rho_Z - U_m\rho_Z|_{\text{tr}}^2 \leq 2^{\min(d,m)} \cdot \sum_{0 \neq S \in \{0,1\}^m} |(S \cdot Z)\rho_Z - U_1\rho_Z|_{\text{tr}}^2.$$

Proof. Following the proof of the classical XOR-Lemma in [Gol95], we first relate $\|Z\rho_Z - U_m\rho_Z\|_1$ to $\|Z\rho_Z - U_m\rho_Z\|_2$, and then view $Z\rho_Z - U_m\rho_Z$ in the Hadamard (or Fourier) basis, giving us the desired result. We need the following simple claim.

Claim 2.14. *For any Boolean function f , $\|f(Z)\rho_Z - U_1\rho_Z\|_1 = \|\sum_z (-1)^{f(z)} p(z)\rho_z\|_1$.*

⁴We thank Thomas Vidick [Vid10] for pointing out that we can also have a bound in terms of m and not only d .

Proof. Denote $\rho_b = \sum_{z:f(z)=b} p(z)\rho_z$ for $b = 0, 1$. Then $\rho_Z = \rho_0 + \rho_1$ and

$$\begin{aligned} \|f(Z)\rho_Z - U_1\rho_Z\|_1 &= \left\| |0\rangle\langle 0| \otimes \rho_0 + |1\rangle\langle 1| \otimes \rho_1 - \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|) \otimes (\rho_0 + \rho_1) \right\|_1 \\ &= \frac{1}{2} \left\| |0\rangle\langle 0| \otimes (\rho_0 - \rho_1) + |1\rangle\langle 1| \otimes (\rho_1 - \rho_0) \right\|_1 \\ &= \|\rho_0 - \rho_1\|_1 = \left\| \sum_z (-1)^{f(z)} p(z)\rho_z \right\|_1. \end{aligned} \quad (2.1)$$

□

Let $\chi_S(z) = (-1)^{S \cdot z}$ for $S \in \{0, 1\}^m$. Denote $D = 2^d$, $M = 2^m$, and $\sigma_z = p(z)\rho_z - \frac{1}{M}\rho_Z$. Then

$$\begin{aligned} \|Z\rho_Z - U_m\rho_Z\|_1^2 &= \left\| \sum_z |z\rangle\langle z| \otimes \sigma_z \right\|_1^2 = \left\| (H^{\otimes m} \otimes I_D) \left(\sum_z |z\rangle\langle z| \otimes \sigma_z \right) (H^{\otimes m} \otimes I_D) \right\|_1^2 \\ &= \frac{1}{M^2} \cdot \left\| \sum_{z,y,S} |y\rangle\langle S| \otimes \chi_S(z)\chi_y(z)\sigma_z \right\|_1^2 \leq \frac{D}{M} \cdot \left\| \sum_{z,y,S} |y\rangle\langle S| \otimes \chi_S(z)\chi_y(z)\sigma_z \right\|_2^2, \end{aligned} \quad (2.2)$$

where H is the Hadamard transform.

Factor D: Using the fact that the $\|\cdot\|_2^2$ of a matrix is the sum of $\|\cdot\|_2^2$ of its $(D \times D)$ sub-blocks, together with $\chi_S(z)\chi_y(z) = \chi_{y+S}(z)$ and $\|\cdot\|_2 \leq \|\cdot\|_1$, (2.2) gives

$$\|Z\rho_Z - U_m\rho_Z\|_1^2 \leq \frac{D}{M} \sum_y \sum_S \left\| \sum_z \chi_{y+S}(z)\sigma_z \right\|_2^2 = D \sum_S \left\| \sum_z \chi_S(z)\sigma_z \right\|_2^2 \leq D \sum_S \left\| \sum_z \chi_S(z)\sigma_z \right\|_1^2. \quad (2.3)$$

Using Claim 2.14 with $f(Z) = S \cdot Z$, we get

$$\sum_{S \neq 0} \|(S \cdot Z)\rho_Z - U_1\rho_Z\|_1^2 = \sum_{S \neq 0} \left\| \sum_z \chi_S(z)p(z)\rho_z \right\|_1^2 = \sum_{S \neq 0} \left\| \sum_z \chi_S(z)\sigma_z \right\|_1^2 = \sum_S \left\| \sum_z \chi_S(z)\sigma_z \right\|_1^2, \quad (2.4)$$

where the second equality holds since χ_S is balanced, and the third since $\sum_z \sigma_z = 0$. Combining Eqs. (2.3) and (2.4) gives the desired result.

Factor M: Restarting from the next-to-last step of (2.2), using again $\chi_S(z)\chi_y(z) = \chi_{y+S}(z)$ and the triangle inequality, we obtain

$$\begin{aligned} \|Z\rho_Z - U_m\rho_Z\|_1^2 &\leq \frac{1}{M^2} \cdot \left(\sum_S \left\| \sum_y |y\rangle\langle S+y| \otimes \left(\sum_z \chi_S(z)\sigma_z \right) \right\|_1 \right)^2 \\ &\leq \frac{1}{M} \cdot \sum_S \left\| \sum_y |y\rangle\langle S+y| \otimes \left(\sum_z \chi_S(z)\sigma_z \right) \right\|_1^2 = M \cdot \sum_S \left\| \sum_z \chi_S(z)\sigma_z \right\|_1^2, \end{aligned}$$

where the last step follows from the observation that the matrices inside the norms are of the form $P \otimes B$ where P is a permutation matrix. In this case $\|P \otimes B\|_1 = \dim(P) \cdot \|B\|_1 = M \cdot \|B\|_1$. As before, combining this with Eq. (2.4) gives the desired bound. \square

We note that we do not have any example showing the dependence on $2^{\min(d,m)}$ is indeed necessary, or that the lemma is tight.

Chapter 3

Extractors Against Quantum Storage

3.1 Definition

We first formalize the different types of quantum storage.

Definition 3.1. For two random variables X, Y we say ρ_{XY} is a (b_1, b_2) entangled storage if it is generated by two non communicating parties, Alice and Bob, in the following way. Alice and Bob share an arbitrary entangled state. Alice receives $x \in X$, Bob receives $y \in Y$. They each apply any quantum operation on their qubits. Alice then stores b_1 of her qubits (and discards the rest), and Bob stores b_2 of his qubits, giving the state ρ_{xy} .

We denote by ρ_{XY}^A the state obtained when Alice stores her entire state, whereas Bob stores only b_2 qubits of his, and similarly for ρ_{XY}^B .

We say ρ_{XY} is (b_1, b_2) non-entangled storage if $\rho_{xy} = \rho_x \otimes \rho_y$ for all $x \in X, y \in Y$.

The security of the extractor is defined relative to the storage.

Definition 3.2. A (k_1, k_2, ε) 2-source extractor against (b_1, b_2) (entangled) quantum storage is a function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any independent n -bit weak sources X, Y with respective min-entropies k_1, k_2 , and any (b_1, b_2) (entangled) storage ρ_{XY} ,

$$|E(X, Y)\rho_{XY} - U_m\rho_{XY}|_{\text{tr}} \leq \varepsilon.$$

The extractor is called X-strong if $|E(X, Y)\rho_{XY}X - U_m\rho_{XY}X|_{\text{tr}} \leq \varepsilon$, and X-superstrong when ρ_{XY} is replaced by ρ_{XY}^A . It is called (super)strong if it is both X- and Y- (super)strong.

A note on the definition: A strong extractor is secure even if at the distinguishing stage one of the sources is completely exposed. A superstrong extractor is secure even if, in addition, the matching party's entire state is also given. Without entanglement, the two are equal, as the state can be completely reconstructed from the source. In the communication complexity setting the model of strong extractors corresponds to the SMP model where the referee also gets access to

one of the inputs, whereas the model of superstrong extractors corresponds to the one-way model, where one party also has access to its share of the entangled state.

To prove E is an extractor, it suffices to show it is either X-strong or Y-strong. All our proofs follow this route.

Flat sources: Classically, it suffices to consider only flat sources (i.e., sources that are uniformly distributed over their support) when analyzing the security of extractors. This comes from the well known fact that any source with min-entropy k is a convex combination of flat sources with min-entropy k . One can easily verify that also in the quantum setting we have that

$$|E(X, Y)\rho_{XY} - U_m\rho_{XY}|_{\text{tr}} \leq \max_{i,j} |E(X_i, Y_j)\rho_{X_i Y_j} - U_m\rho_{X_i Y_j}|_{\text{tr}},$$

where $X = \sum \alpha_i X_i$ and $Y = \sum \beta_j Y_j$ are convex combinations of flat sources. Therefore, in what follows we only consider such sources.

3.2 One Bit Extractor

3.2.1 Average Case Lower Bound for Inner Product

Cleve et al. [CvDNT98] give a lower bound for the worst case one-way quantum communication complexity of inner product with arbitrary prior entanglement. It is achieved by first reducing the problem of computing the inner product to that of transmitting one input over a quantum channel, and then using an extended Holevo bound. Nayak and Salzman [NS06] obtained an optimal lower bound by replacing Holevo with a more "mission-specific" bound:

Theorem 3.3 ([NS06], Thm 1.3 and discussion thereafter). *Let X be an n -bit random variable with min-entropy k , and suppose Alice wishes to convey X to Bob over a one-way quantum communication channel using b qubits. Let Y be the random variable denoting Bob's guess for X . Then*

1. $\Pr[Y = X] \leq 2^{-(k-b)}$, if the parties don't share prior entanglement, and
2. $\Pr[Y = X] \leq 2^{-(k-2b)}$.

Revisiting Cleve et al.'s reduction, we now show how to adapt it to flat sources, to the average case error and to the linearly transformed inner product. The main challenge is to carefully treat the error terms so as to not cancel out the (small) amplitude of the correct state.

Lemma 3.4. *Let X, Y be flat sources over n bits with min-entropies k_1, k_2 , and A, B full rank n by n matrices over $GF(2)$. Let P be a b qubit one-way protocol for $(AX) \cdot (BY)$ with success probability $\frac{1}{2} + \varepsilon$. Then*

- (a) $\varepsilon \leq 2^{-(k_1+k_2-2b-n+2)/2}$, if the parties share prior entanglement and

(b) $\varepsilon \leq 2^{-(k_1+k_2-b-n+2)/2}$ otherwise.

Proof. Let us first consider the case $A = B = I$. Assume w.l.o.g. Bob delays his operations until receiving the message from Alice and that in his first step he copies his input, leaving the original untouched throughout. Further assume Bob outputs the result in one of his qubits.

For a fixed x , denote the success probability of P by $\frac{1}{2} + \varepsilon_x$ (ε_x might be negative). Denote Bob's state after receiving the message as $|y\rangle|0\rangle|\sigma_x\rangle$, where σ_x is taken to contain Alice's message and Bob's prior entangled qubits as required by the protocol (if present). The rest of the protocol is now performed locally by Bob. We denote this computation P_B . After applying P_B , Bob's state is of the form

$$\alpha_{x,y}|y\rangle|x \cdot y\rangle|J_{x,y}\rangle + \beta_{x,y}|y\rangle|\bar{x} \cdot y\rangle|K_{x,y}\rangle,$$

and by assumption, $\mathbb{E}_y \beta_{x,y}^2 = \frac{1}{2} - \varepsilon_x$. Following the analysis in [CvDNT98], using *clean* computation, where the output is produced in a new qubit (the leftmost), gives the state

$$|z + x \cdot y\rangle|y\rangle|0\rangle|\sigma_x\rangle + \sqrt{2}\beta_{x,y}|M_{x,y,z}\rangle,$$

where $|M_{x,y,z}\rangle = \left(\frac{1}{\sqrt{2}}|z + \bar{x} \cdot y\rangle - \frac{1}{\sqrt{2}}|z + x \cdot y\rangle\right) P_B^\dagger |y\rangle|\bar{x} \cdot y\rangle|K_{x,y}\rangle$. Observe the following properties of M : 1. $|M_{x,y,0}\rangle = -|M_{x,y,1}\rangle$ 2. As $y \in Y$ varies, the states $|M_{x,y,z}\rangle$ are orthonormal. 3. Since P_B^\dagger does not affect the first n (so called input) qubits, $|M_{x,y,z}\rangle$ is orthogonal to states of the form $|a\rangle|y'\rangle \otimes |\cdot\rangle$ for all $a \in \{0, 1\}, y \in Y, y' \notin Y$.

We now use the following steps to transfer X from Alice to Bob:

1. Bob prepares the state $\sqrt{2^{-k_2-1}} \cdot \sum_{y \in Y, a \in \{0,1\}} (-1)^a |a\rangle|y\rangle$.
2. Alice and Bob execute the clean version of P .
3. Bob performs Hadamard on his first $n + 1$ qubits and measures in the computational basis.

After the second step, Bob's state is $|\psi\rangle = |v\rangle + \vec{e}$ where

$$|v\rangle = \sqrt{2^{-k_2-1}} \sum_{y \in Y, a \in \{0,1\}} (-1)^{a+x \cdot y} |a\rangle|y\rangle|0\rangle|\sigma_x\rangle \quad \vec{e} = \sqrt{2^{-k_2-1}} \sum_{y \in Y, a \in \{0,1\}} (-1)^a \sqrt{2}\beta_{x,y} |M_{x,y,a}\rangle.$$

By the properties of $|M_{x,y,z}\rangle$, $\|\vec{e}\| = 2\sqrt{\mathbb{E}_y \beta_{x,y}^2} = 2\sqrt{\frac{1}{2} - \varepsilon_x}$. Since $|v\rangle + \vec{e}$ and $|v\rangle$ are normalized states, we can easily derive $\langle v | (|v\rangle + \vec{e}) \rangle = 2\varepsilon_x$. Define

$$|\psi_0\rangle = H^{\otimes n+1} |1x\rangle \otimes |0\rangle|\sigma_x\rangle = \sqrt{2^{k_2-n}} |v\rangle + \sqrt{2^{-n-1}} \sum_{y \notin Y, a \in \{0,1\}} (-1)^{a+x \cdot y} |a\rangle|y\rangle|0\rangle|\sigma_x\rangle,$$

and note that the second term is orthogonal to both $|v\rangle$ and \vec{e} . It follows that $\langle \psi | \psi_0 \rangle = \sqrt{2^{k_2-n+2}} \varepsilon_x$. Applying the Hadamard in Step 3. does not affect the inner product, and so Bob will measure $|1x\rangle$ with probability $2^{k_2-n+2} \cdot \varepsilon_x^2$. Applying Thm. 3.3.1 and 3.3.2 along with Jensen's inequality now completes the proof.

For the general case where $A \neq I$ or $B \neq I$, we modify Step 3. of the transmission protocol. Instead of the Hadamard transform, Bob applies the inverse of the unitary transformation $|z\rangle|x\rangle \mapsto \sqrt{2^{-n-1}} \cdot \sum_{y,a} (-1)^{za+(Ax)\cdot(By)} |a\rangle|y\rangle$. It is easy to check that this gives the desired result. \square

3.2.2 The Extractor and Tightness Results

When the extractor's output is binary, distinguishing it from uniform is equivalent to computing the output on average. This was shown by Yao [Yao82] when the storage is classical and is trivially extended to the quantum setting. With this observation, reformulating Lem. 3.4 in the language of trace distance yields a one bit extractor.

Corollary 3.5. *The function $E_{IP}(x, y) = x \cdot y$ is a (k_1, k_2, ε) extractor against (b_1, b_2) (entangled) quantum storage provided*

$$(a) \text{ (entangled) } k_1 + k_2 - 2 \min(b_1, b_2) \geq n - 2 + 2 \log \varepsilon^{-1},$$

$$(b) \text{ (non-entangled) } k_1 + k_2 - \min(b_1, b_2) \geq n - 2 + 2 \log \varepsilon^{-1}.$$

Proof. With Yao's equivalence, Lem. 3.4.(a) immediately gives

$$|(AX \cdot Y)\rho_{XY}X - U\rho_{XY}X|_{\text{tr}} \leq 2^{-(k_1+k_2-2b_2-n+2)/2} \quad (3.1)$$

$$|(AX \cdot Y)\rho_{XY}Y - U\rho_{XY}Y|_{\text{tr}} \leq 2^{-(k_1+k_2-2b_1-n+2)/2} \quad (3.2)$$

for any full rank matrix A , and specifically for $A = I$. By the assumption on ε , E_{IP} is either Y-strong or X-strong. Repeating this argument with Lem. 3.4.(b) gives the non-entangled case. \square

Recall (see Def. 3.2 and discussion thereafter) that one-way communication corresponds to the model of *superstrong* extractors. It is not surprising then that Lem. 3.4 actually implies a superstrong extractor. By choosing ε in the above proof such that both inequalities (3.1) and (3.2) are satisfied, where we replace ρ_{xY} by ρ_{xY}^A to include Alice's complete state as well as Bob's entangled qubits and similarly for ρ_{Xy}^B , we obtain:

Corollary 3.6. *The function $E_{IP}(x, y) = x \cdot y$ is a (k_1, k_2, ε) superstrong extractor against (b_1, b_2) (entangled) quantum storage provided*

$$(a) \text{ (entangled) } k_1 + k_2 - 2 \max(b_1, b_2) \geq n - 2 + 2 \log \varepsilon^{-1},$$

$$(b) \text{ (non-entangled) } k_1 + k_2 - \max(b_1, b_2) \geq n - 2 + 2 \log \varepsilon^{-1}.$$

We now show that the parameters of all our extractors are *tight* up to an additive constant. For simplicity, assume first that the error ε is close to $1/2$, the sources are uniform and $b_1 = b_2 := b$. Cor. 3.5 then states that E_{IP} is an extractor as long as $b < n$ in the non-entangled case and $b < n/2$ in the entangled case. Indeed, in the non-entangled case it is trivial to compute the inner product in the SMP model (i.e., break the extractor) when $b \geq n$. With entanglement, $b \geq n/2$ suffices as demonstrated by the following protocol, adapted from a protocol by de Wolf [dW10].

Claim 3.7. *The inner product function for n bit strings is exactly computable in the SMP model with entanglement with $n/2 + 2$ qubits of communication from each party.¹*

Proof. Let $x, y \in \{0, 1\}^n$ be Alice and Bob's inputs. Since $x \cdot y = \frac{1}{2}(|x| + |y| - |x \oplus y|) \pmod{4}$, it suffices to show that the referee can compute $x \oplus y$ with $n/2$ qubits of communication from each party, or simply $x_1x_2 \oplus y_1y_2$ with one qubit of communication.

Denote the Pauli matrices $\sigma_{00} = I$, $\sigma_{01} = Z$, $\sigma_{10} = X$, $\sigma_{11} = ZX$. Given a shared EPR pair, Alice applies $\sigma_{x_1x_2}$ to her qubit and sends it to the referee, and Bob does the same with $\sigma_{y_1y_2}$. Note that applying $\sigma_{b_1b_2}$ to the first qubit has the same effect as applying it to the second qubit. Further, X is applied iff b_1 is 1 and Z is applied iff b_2 is 1. Since two applications of X (Z) cancel each other out, we have that X is applied to the first qubit iff $x_1 + y_1 = 1$ and Z is applied to the first qubit iff $x_2 + y_2 = 1$. The net effect on the EPR state is $\sigma_{x_1x_2 \oplus y_1y_2} \otimes I$. For each value of $x_1x_2 \oplus y_1y_2$ this gives one of the orthogonal (completely distinguishable) Bell states. \square

Showing that our results are tight for arbitrary ε is trickier. We show

Lemma 3.8. *If $E_{IP} = x \cdot y$ is a (k_1, k_2, ε) extractor against (b_1, b_2) (entangled) storage then*

- (a) (entangled) $k_1 + k_2 - 2 \min(b_1, b_2) > n - 9 + 2 \log \varepsilon^{-1}$,
- (b) (non-entangled) $k_1 + k_2 - \min(b_1, b_2) > n - 5 + 2 \log \varepsilon^{-1}$.

If E_{IP} is superstrong, then

- (a) (entangled) $k_1 + k_2 - 2 \max(b_1, b_2) > n - 9 + 2 \log \varepsilon^{-1}$,
- (b) (non-entangled) $k_1 + k_2 - \max(b_1, b_2) > n - 5 + 2 \log \varepsilon^{-1}$.

Proof. We give a slightly modified version of Proposition 10 in [CG88], taking into account quantum side information. We need the following theorem.

Theorem 3.9 ([CG88, Theorem 3]). *There exist independent random variables X, Y on l bits with min-entropy $l - 3$ each² such that $\Pr[X \cdot Y = 0] > \frac{1}{2} + 2^{-(l-1)/2}$.*

We start in the weak extractor setting with entanglement. We construct sources X, Y with min-entropy k_1, k_2 and (b_1, b_2) entangled quantum storage ρ_{XY} for which the error will be "large". Let $b = 2(\min(b_1, b_2) - 2)$, and let $\Delta = k_1 + k_2 - n$. If $\Delta \leq b$, we pick X to be uniform on the first k_1 bits and 0 elsewhere, Y uniform on the last k_2 bits and 0 elsewhere. The inner product of X, Y is then the inner product of at most b bits, and can be computed exactly using the SMP protocol in Claim 3.7 with $\min(b_1, b_2)$ qubits from each.

¹We thank Ronald de Wolf [dW10] for generously allowing us to adapt his upper bound to our setting.

²[CG88] prove the claim with slightly different parameters for arbitrary Boolean functions. Our modification is trivial.

In the case $\Delta > b$, we define $X = X_1X_2X_3X_4$ as follows: X_1 is uniform on b bits, X_2 is uniform on $k_1 - \Delta - 3$ bits, X_3 is the first $(\Delta + 6 - b, \Delta + 3 - b)$ source promised by Theorem 3.9 (for $l = \Delta + 6 - b$), and X_4 is constant 0^{n-k_1-3} . Analogously, $Y = Y_1Y_2Y_3Y_4$ is defined as: Y_1 is uniform on b bits, Y_2 is constant 0^{n-k_2-3} , Y_3 is the second $(\Delta + 6 - b, \Delta + 3 - b)$ source promised by Theorem 3.9, and Y_4 is uniform on $k_2 - \Delta - 3$ bits. It is easily verified that $H_\infty(X) \geq k_1$ and $H_\infty(Y) \geq k_2$. Finally, we set ρ_{XY} to be the entangled $(\min(b_1, b_2), \min(b_1, b_2))$ storage of the SMP protocol in Theorem 3.7 allowing us to compute $x_1 \cdot y_1$ exactly, and M the measurement strategy of the referee. Applying Theorem 3.9,

$$\Pr[M(\rho_{XY}) = X \cdot Y] = \Pr[X_1 \cdot Y_1 = X \cdot Y] = \Pr[X_3 \cdot Y_3 = 0] > \frac{1}{2} + 2^{-(\Delta+5-b)/2}$$

and $|(X \cdot Y)\rho_{XY} - U\rho_{XY}|_{\text{tr}} > 2^{-(k_1+k_2-b-n+5)/2}$.

In the non-entangled case, we simply set $b = \min(b_1, b_2)$ and replace the SMP protocol with a trivial protocol for IP on b bits.³

In the superstrong case with entanglement, assume w.l.o.g. that $b_1 > b_2$ and choose $b = b_1/2$. We then let ρ_{xy} be the entangled state that appears in the superdense coding protocol for X_1 . Thus, exposing Bob's state allows us to compute $X_1 \cdot Y_1$ exactly. Without entanglement, we set $b = b_1$ and have Alice send X_1 to Bob. \square

3.3 Many Bit Extractor

Here we prove our main Theorems 1.2 and 1.3. First, using our quantum XOR-Lemma 2.13, we obtain results in the *strong* case.

Lemma 3.10. *E_D is a (k_1, k_2, ε) X-strong extractor against (b_1, b_2) (entangled) quantum storage provided*

- (a) (entangled) $k_1 + k_2 - 2b_2 \geq 2m + n - 2 + 2 \log \varepsilon^{-1}$,
- (b) (non-entangled) $k_1 + k_2 - b_2 \geq 2m + n - 2 + 2 \log \varepsilon^{-1}$.

Proof. Recall that $E_D(x, y) = A_1x \cdot y, A_2x \cdot y, \dots, A_mx \cdot y$ (see Def. 2.5). For $0 \neq S \in \{0, 1\}^m$, let $A_S = \sum_{i:S_i=1} A_i$ and note that $S \cdot E(x, y) = A_Sx \cdot y$. By the XOR-Lemma 2.13,

$$|E(X, Y)\rho_{XY}X - U_m\rho_{XY}X|_{\text{tr}} \leq \sqrt{2^m \sum_{S \neq 0} |(A_SX \cdot Y)\rho_{XY}X - U_1\rho_{XY}X|_{\text{tr}}^2}.$$

The result then follows by Ineq. (3.1) in the proof of Cor. 3.5 and its non-entangled analogue. \square

In a similar way, we also obtain a *Y-strong* extractor with analogous parameters. Following [DEOR04], we now apply a seeded extractor against quantum storage (see Def. 3.11) to the output

³In fact, this shows our non-entangled extractor is tight even for *classical* storage.

of an X-strong (Y-strong) extractor to obtain a two-source extractor with more output bits (see Lem. 3.12).

Definition 3.11 ([Ta-09]). *A function $E : \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ is a (k, ε) seeded extractor against b quantum storage if for any n -bit source X with min-entropy k and any b qubit quantum storage ρ_X ,*

$$|E(X, U_d)\rho_X - U_m\rho_X|_{\text{tr}} \leq \varepsilon.$$

Lemma 3.12. *Let $E_B : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^d$ be a (k_1, k_2, ε) X-strong extractor against (b_1, b_2) (entangled) quantum storage, and let $E_S : \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ and $E(x, y) = E_S(x, E_B(x, y))$.*

- (a) *(entangled) If E_S is a (k_1, ε) seeded extractor against $b_1 + b_2$ quantum storage then E is a $(k_1, k_2, 2\varepsilon)$ extractor against (b_1, b_2) entangled quantum storage.*
- (b) *(non-entangled) If E_S is a (k_1, ε) seeded extractor against b_1 quantum storage then E is a $(k_1, k_2, 2\varepsilon)$ extractor against (b_1, b_2) non-entangled quantum storage.*

Proof. Part (a): $|E_B(X, Y)\rho_{XY}X - U_d\rho_{XY}X|_{\text{tr}} \leq \varepsilon$ and so $|E_S(X, E_B(X, Y))\rho_{XY} - E_S(X, U_d)\rho_{XY}|_{\text{tr}} \leq \varepsilon$. But $|E_S(X, U_d)\rho_{XY} - U_m\rho_{XY}|_{\text{tr}} \leq \varepsilon$ by definition of E_S . The result follows from the triangle inequality. For part (b) note that when the storage is non-entangled, $|E_S(X, U_d)\rho_X\rho_Y - U_m\rho_X\rho_Y|_{\text{tr}} = |E_S(X, U_d)\rho_X - U_m\rho_X|_{\text{tr}}$, and it suffices to require that E_S be a seeded extractor against only b_1 quantum storage. \square

A seeded extractor with almost optimal min-entropy loss is given in [DPVR09]. Their extractor is secure against guessing entropy sources, and so trivially against quantum storage [KT08] (see Chap. 4 for details). We reformulate the seeded extractor in terms of Def. 3.11.

Corollary 3.13 ([DPVR09, Corrolary 5.3]). *There exists an explicit (k, ε) seeded extractor against b quantum storage with seed length $d = O(\log^3(n/\varepsilon))$ and $m = d + k - b - 8 \log(k - b) - 8 \log \varepsilon^{-1} - O(1)$ output bits.*

The proofs of Thms. 1.3 and 1.2 now follow by composing the explicit extractors of Lem. 3.10 and Cor. 3.13 as in Lem. 3.12.

Proof of Theorem 1.3: E_D is an X-strong extractor against entangled storage with $\frac{1}{2}(k_1 + k_2 - 2b_2 - n - 2 \log \varepsilon^{-1})$ almost uniform output bits. This is larger than $O(\log^3(n/\varepsilon))$ when $k_1 + k_2 - 2b_2 > n + \Omega(\log^3(n/\varepsilon))$, allowing us to compose it with the seeded extractor secure against $b_1 + b_2$ storage of Cor. 3.13 on the source X , obtaining $m = \frac{1}{2}(k_1 + k_2 - 2b_2 - n - 2 \log \varepsilon^{-1}) + (k_1 - b_1 - b_2) - 8 \log(k_1 - b_1 - b_2) - 8 \log \varepsilon^{-1} - O(1)$. Similarly, E_D is a Y-strong extractor, and can be composed with the seeded extractor on the source Y . Choosing the better of the two, we prove the desired result.⁴ \square

⁴We slightly sacrifice the parameters in the formulation of the theorem to simplify the result.

Proof of Theorem 1.2: E_D is an X -strong extractor against non-entangled storage with $\frac{1}{2}(k_1 + k_2 - b_2 - n - 2 \log \varepsilon^{-1})$ almost uniform output bits. This is larger than $O(\log^3(n/\varepsilon))$ when $k_1 + k_2 - b_2 > n + \Omega(\log^3(n/\varepsilon))$. Composing with the seeded extractor secure against b_1 storage of Cor. 3.13 on the source X gives $m = \frac{1}{2}(k_1 + k_2 - b_2 - n - 2 \log \varepsilon^{-1}) + (k_1 - b_1) - 8 \log(k_1 - b_1) - 8 \log \varepsilon^{-1} - O(1)$, and similarly for Y . \square

Chapter 4

Extractors Against Quantum Knowledge

4.1 Guessing Entropy

In the classical world, a standard measure for the randomness of X is its *min-entropy*, defined $H_\infty(X) := -\log \max_x \Pr[X = x]$.

In the presence of a quantum adversary, we would like to consider sources X with min-entropy k relative to the adversary's quantum information. Such a generalization of min-entropy is non-trivial and given in [Ren05]. We follow the approach of König and Terhal [KT08], and focus on the probability of guessing X given the adversary's state. For cq-states, the two definitions were shown to be equivalent [KRS09].

Definition 4.1. *Let $X\rho_X$ be an arbitrary cq-state. The guessing entropy of X given ρ_X is*

$$H_g(X \leftarrow \rho_X) := -\log \max_M \mathbb{E}_{x \leftarrow X} [\text{Tr}(M_x \rho_x)],$$

where the maximum ranges over all POVM $M = \{M_x\}_{x \in X}$.

Treating $M(\rho_X)$ as a classical probability distribution over the support of X , the above can be perhaps more easily understood as $H_g(X \leftarrow \rho_X) = -\log \max_M \Pr[M(\rho(X)) = X]$. When ρ_X is trivial, i.e. of dimension one, the guessing entropy reduces to the classical min-entropy.

The following claim implies that a high entropy source X is hard to guess even given some bounded storage ρ_X , and sets the relation between bounded storage adversaries and guessing entropy adversaries, which is discussed in detail in the following sections.

Claim 4.2. [KT08, Proposition 2'] $H_g(X \leftarrow \rho_X) \geq H_\infty(X) - \log \dim \rho_X$

4.2 Non-entangled Adversaries

We first define two-source extractors secure against non-entangled guessing entropy adversaries. We note that *seeded* extractors against guessing entropy adversaries were extensively studied in the literature [Ren05, KT08, FS08, DPVR09, TSSR10].

Recall that in the non-entangled case the bounded storage is given by $\rho_X \otimes \rho_Y$ (see Def. 3.1). Here, we place a limit not on the amount of storage, but on the amount of information, in terms of guessing entropy, the adversaries have on their respective sources. That is, we require that the guessing entropy of X (Y) given ρ_X (ρ_Y) be high. We refer to the state $\rho_X \otimes \rho_Y$ as *quantum knowledge*, or if ρ_x, ρ_y are classical for every x, y , as *classical knowledge*.

Definition 4.3. A (k_1, k_2, ε) two-source extractor against quantum knowledge is a function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any independent sources X, Y and quantum knowledge $\rho_X \otimes \rho_Y$ with guessing entropies $H_g(X \leftarrow \rho_X) \geq k_1$, $H_g(Y \leftarrow \rho_Y) \geq k_2$, we have

$$|E(X, Y)\rho_X\rho_Y - U_m\rho_X\rho_Y|_{\text{tr}} \leq \varepsilon.$$

The extractor is called *X-strong* if $|E(X, Y)\rho_Y X - U_m\rho_Y X|_{\text{tr}} \leq \varepsilon$. It is called *strong* if it is both *X-strong* and *Y-strong*.

By Claim. 4.2, we can view adversaries with bounded quantum storage as a special case of general adversaries. In particular, a $(k_1 - b_1, k_2 - b_2, \varepsilon)$ extractor against quantum knowledge is trivially a (k_1, k_2, ε) extractor against *non-entangled* (b_1, b_2) storage.

4.2.1 One-Bit Output

König and Terhal [KT08] show that every classical one-bit output strong seeded extractor is also a strong extractor against quantum knowledge with roughly the same parameters. They reduce the "quantum security" of the extractor to the "classical security", *irrespective* of the entropy of the seed. Informally, $|E(X, Y)\rho_X Y - U_1\rho_X Y|_{\text{tr}}$ is small if the statement is also true when ρ_X is classical. We give a version of their Lem. 2 with slightly improved parameters. The lemma shows that it suffices to prove security of an extractor with respect only to classical knowledge obtained by performing a Pretty Good Measurement (PGM) [HW94] on arbitrary quantum knowledge. For a cq-state $Z\rho_Z$, a PGM is a POVM $\mathcal{E} = \{\mathcal{E}_z\}_{z \in Z}$ such that $\mathcal{E}_z = p(z)\rho_Z^{-1/2} \rho_z \rho_Z^{-1/2}$.

Lemma 4.4. Let $Z\rho_Z$ be a cq-state, and f be a Boolean function. Then¹

$$|f(Z)\rho_Z - U\rho_Z|_{\text{tr}} \leq \sqrt{\frac{1}{2} |f(Z)\mathcal{E}(\rho_Z) - U\mathcal{E}(\rho_Z)|_{\text{tr}}},$$

where $\mathcal{E} = \{\mathcal{E}_z\}_{z \in Z}$ is a Pretty Good Measurement, $\mathcal{E}_z = p(z)\rho_Z^{-1/2} \rho_z \rho_Z^{-1/2}$.

¹ $\mathcal{E}(\rho_Z)$ is a classical probability distribution and the trace distance $|f(Z)\mathcal{E}(\rho_Z) - U\mathcal{E}(\rho_Z)|_{\text{tr}}$ reduces to the classical variational distance.

Proof. We need the following lemma.

Lemma 4.5 ([Ren05, Lemma 5.1.3]). *Let S be a Hermitian operator and let σ be a nonnegative operator. Then $|S|_{\text{tr}} \leq \frac{1}{2} \sqrt{\text{Tr}(\sigma) \text{Tr}(\sigma^{-1/2} S \sigma^{-1/2} S)}$.*

Denote $\rho = \rho_Z$, $\rho_b = \sum_{z:f(z)=b} p(z) \rho_z$ for $b = 0, 1$. Further define (informally) a POVM M for guessing f from ρ_Z by first applying \mathcal{E} to get z and then computing $f(z)$. Then

$$\begin{aligned} \Pr[M(\rho_Z) = f(Z)] &= \sum_z p(z) \sum_{z':f(z')=f(z)} \text{Tr}(\mathcal{E}_{z'} \rho_z) \\ &= \text{Tr} \left(\sum_{f(z')=f(z)} \rho^{-1/2} (p(z') \rho_{z'}) \rho^{-1/2} (p(z) \rho_z) \right) \\ &= \text{Tr}(\rho^{-1/2} \rho_0 \rho^{-1/2} \rho_0 + \rho^{-1/2} \rho_1 \rho^{-1/2} \rho_1), \end{aligned}$$

and similarly $\Pr[M(\rho_Z) \neq f(Z)] = \text{Tr}(\rho^{-1/2} \rho_0 \rho^{-1/2} \rho_1 + \rho^{-1/2} \rho_1 \rho^{-1/2} \rho_0)$. Hence

$$|\Pr[M(\rho_Z) = f(Z)] - \Pr[M(\rho_Z) \neq f(Z)]| = \text{Tr}(\rho^{-1/2} (\rho_0 - \rho_1) \rho^{-1/2} (\rho_0 - \rho_1)). \quad (4.1)$$

By Eq. (2.1), $|f(Z) \rho_Z - U \rho_Z|_{\text{tr}} = |\rho_0 - \rho_1|_{\text{tr}}$, and by Lem. 4.5, setting $S = \rho_0 - \rho_1$, $\sigma = \rho$,

$$|\rho_0 - \rho_1|_{\text{tr}} \leq \frac{1}{2} \sqrt{\text{Tr}(\rho^{-1/2} (\rho_0 - \rho_1) \rho^{-1/2} (\rho_0 - \rho_1))}. \quad (4.2)$$

Combining Eq. (4.1) with Eq. (4.2) gives

$$|f(Z) \rho_Z - U \rho_Z|_{\text{tr}} \leq \sqrt{\frac{1}{4} |\Pr[M(\rho_Z) = f(Z)] - \Pr[M(\rho_Z) \neq f(Z)]|}.$$

Finally,

$$|\Pr[M(\rho_Z) = f(Z)] - \Pr[M(\rho_Z) \neq f(Z)]| \leq 2 |f(Z) M(\rho_Z) - U M(\rho_Z)|_{\text{tr}} \leq 2 |f(Z) \mathcal{E}(\rho_Z) - U \mathcal{E}(\rho_Z)|_{\text{tr}},$$

as the left hand side describes a trivial strategy to guess f from $M(\rho)$, giving the desired result. \square

Corollary 4.6. *If E is a classical one-bit output (k_1, k_2, ε) two-source extractor, then it is a $(k_1 + \log \varepsilon^{-1}, k_2 + \log \varepsilon^{-1}, \sqrt{3\varepsilon/2})$ two-source extractor against quantum knowledge.*

Proof. By Lem. 4.4, $|E(X, Y) \rho_X \rho_Y - U \rho_X \rho_Y|_{\text{tr}} \leq \sqrt{\frac{1}{2} |E(X, Y) \mathcal{E}(\rho_X \rho_Y) - U \mathcal{E}(\rho_X \rho_Y)|_{\text{tr}}}$. A direct calculation shows that for every x, y , $\mathcal{E}(\rho_x \otimes \rho_y) = \mathcal{E}_1(\rho_x) \otimes \mathcal{E}_2(\rho_y)$, where $\mathcal{E}_1, \mathcal{E}_2$ are Pretty Good Measurements on states $X \rho_X, Y \rho_Y$ respectively. In other words, $\mathcal{E}(\rho_X \otimes \rho_Y)$ induces a classical distribution $C_X \otimes C_Y$. Thus

$$|E(X, Y) \rho_X \rho_Y - U \rho_X \rho_Y|_{\text{tr}} \leq \sqrt{\frac{1}{2} |E(X, Y) C_X C_Y - U C_X C_Y|_{\text{tr}}}, \quad (4.3)$$

where $H_g(X \leftarrow C_X) \geq H_g(X \leftarrow \rho_X)$, and the same for Y .

By the definition of (classical) guessing entropy, one can easily show that a classical (k_1, k_2, ε) two-source extractor is a $(k_1 + \log \varepsilon^{-1}, k_2 + \log \varepsilon^{-1}, 3\varepsilon)$ extractor against *classical knowledge* (for details see Proposition 1 in [KT08]). Ineq. (4.3) then gives the desired parameters against quantum knowledge. \square

By a similar argument and following the proof of Theorem 1 in [KT08], we get

Corollary 4.7. *If E is a classical one-bit output (k_1, k_2, ε) X-strong extractor, then it is a $(k_1, k_2 + \log \varepsilon^{-1}, \sqrt{\varepsilon})$ X-strong extractor against quantum knowledge.*

In particular, by Ineq. (3.1) in the proof of Cor. 3.5, inner product is a *classical* X-strong extractor with error $\varepsilon \leq 2^{-(k_1+k_2-n+2)/2}$. Plugging this into Cor. 4.7 we obtain

Corollary 4.8. *The function $E_{IP_A}(x, y) = Ax \cdot y$, for any full rank matrix A , is a (k_1, k_2, ε) X-strong (Y-strong) extractor against quantum knowledge provided that $k_1 + k_2 \geq n - 2 + 6 \log \varepsilon^{-1}$.*

4.2.2 Multi-Bit Output

We now show how to apply the results in the one-bit case, together with our XOR-Lemma 2.13, to show security in the multi-bit case, proving Thm. 1.4. We repeat the steps performed in Sec. 3.3 in the setting of non-entangled guessing entropy adversaries to obtain a multi-bit extractor against quantum knowledge. In exactly the same fashion as in the proof of Lem. 3.10 we use the XOR-Lemma 2.13 to reduce the security of E_D to the strong one bit case of Cor. 4.8.

Lemma 4.9. *E_D is a (k_1, k_2, ε) X-strong (Y-strong) extractor against quantum knowledge provided that $k_1 + k_2 \geq 6m + n - 2 + 6 \log \varepsilon^{-1}$.*

Proof. By the XOR-Lemma 2.13 and Cor. 4.8,

$$|E(X, Y)\rho_{YX} - U_m\rho_{YX}|_{\text{tr}} \leq \sqrt{2^m \sum_{S \neq 0} |(A_S X \cdot Y)\rho_{YX} - U_1\rho_{YX}|_{\text{tr}}^2} \leq 2^m \cdot 2^{-(k_1+k_2-n+2)/6}.$$

□

To obtain our final result, we now compose our strong extractor with a seeded extractor against quantum knowledge.

Lemma 4.10. *Let $E_B : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^d$ be a (k_1, k_2, ε) X-strong extractor against quantum knowledge and let $E_S : \{0, 1\}^n \times \{0, 1\}^d \rightarrow \{0, 1\}^m$ be a (k_1, ε) seeded extractor against quantum knowledge². Then $E(x, y) = E_S(x, E_B(x, y))$ is a $(k_1, k_2, 2\varepsilon)$ extractor against quantum knowledge.*

Proof. Immediate from the extractor definitions and the triangle inequality. □

Corollary 4.11 ([DPVR09, Corollary 5.3]). *There exists an explicit (k, ε) seeded extractor against quantum knowledge with seed length $d = O(\log^3(n/\varepsilon))$ and $m = d + k - 8 \log k - 8 \log \varepsilon^{-1} - O(1)$.*

²For a formal definition see [DPVR09].

Proof of Theorem 1.4: E_D is an X -strong extractor against quantum knowledge with $\frac{1}{6}(k_1+k_2-n-6\log \varepsilon^{-1})-O(1)$ output bits. This is larger than $O(\log^3(n/\varepsilon))$ when $k_1+k_2 > n+\Omega(\log^3(n/\varepsilon))$. Composing with the seeded extractor of Cor. 4.11 on the source X gives $m = \frac{1}{6}(k_1+k_2-n-6\log \varepsilon^{-1})+k_1-8\log k_1-8\log \varepsilon^{-1}-O(1)$, and similarly for Y . \square

4.3 Entangled Adversaries

Following our bounded storage definition (Def. 3.1), we denote the entangled adversaries' state by ρ_{XY} , but place no limitation on the dimension of $\rho_X = \text{Tr}_Y(\rho_{XY})$ or $\rho_Y = \text{Tr}_X(\rho_{XY})$. Again, we refer to this state as *quantum knowledge*.

One (seemingly natural) way to define the model of security against such adversaries is to require the guessing entropy of each source given the corresponding adversary's storage to be high.

Definition 4.12 (Attempt 1). *A (k_1, k_2, ε) two-source extractor against entangled quantum knowledge is a function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any independent sources X, Y and entangled quantum knowledge ρ_{XY} , such that $H_g(X \leftarrow \rho_X) \geq k_1$, $H_g(Y \leftarrow \rho_Y) \geq k_2$, we have $|E(X, Y)\rho_{XY} - U_m\rho_{XY}|_{\text{tr}} \leq \varepsilon$.*

This definition, however, is too strong: it is easy to see that no extractor can be secure against such adversaries. This follows from the observation that by sharing a random string $r_1 r_2$ (which is a special case of shared entanglement) and having the first adversary store $r_1 \oplus x, r_2$ and the other store $r_1, r_2 \oplus y$, we keep the guessing entropy of X (resp. Y) given the adversary's storage unchanged yet we can recover x and y completely from the combined storage.

Hence we are naturally lead to consider the weaker requirement that the guessing entropy of each source given the combined storage of *both* adversaries is high.

Definition 4.13 (Attempt 2). *A (k_1, k_2, ε) two-source extractor against entangled quantum knowledge is a function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any independent sources X, Y and entangled quantum knowledge ρ_{XY} , such that $H_g(X \leftarrow \rho_{XY}) \geq k_1$, $H_g(Y \leftarrow \rho_{XY}) \geq k_2$, we have $|E(X, Y)\rho_{XY} - U_m\rho_{XY}|_{\text{tr}} \leq \varepsilon$.*

We now observe that the inner product is not secure under this definition, indicating that it might still be too strong.

Claim 4.14. *E_{IP} is not a (k_1, k_2, ε) two-source extractor against entangled quantum knowledge for any $\varepsilon < 1/2$, $k_1, k_2 \leq n - O(1)$.*

Proof. We show sources X, Y and entangled quantum knowledge ρ_{XY} s.t. $H_g(X \leftarrow \rho_{XY}), H_g(Y \leftarrow \rho_{XY}) \geq n - O(1)$ but $X \cdot Y$ is exactly computable from ρ_{XY} .

Let X, Y be uniform n -bit sources, R shared randomness, and W_X, W_Y the Hamming weight mod 4 of X, Y respectively, and say Alice stores $X \oplus R, W_X$ and Bob stores $Y \oplus R, W_Y$. Their joint state allows to compute $x \cdot y$ exactly, since $x \cdot y = \frac{1}{2}((|x| + |y| - |x \oplus y|) \bmod 4)$. However,

$$2^{-H_g(X \leftarrow \rho_{XY})} = \mathbb{E}_{W_X, W_Y} \left(\max_M \Pr[M(X \oplus Y) = X \cdot Y \mid W_X = w_X, W_Y = w_Y] \right),$$

and any possible combination (w_X, w_Y) occurs with probability $1/16 - o(1)$. Any strategy to guess X from $X \oplus Y$ for a fixed w_X, w_Y is also a good strategy (up to a constant factor) to guess completely random X from $X \oplus Y$, which is possible with probability at most 2^{-n} . Hence, $2^{-H_g(X \leftarrow \rho_{XY})} \lesssim 2^{-n}/16$ and similarly for Y . \square

By further weakening the model, we may require that the guessing entropy of each source be high given both the combined storage *and* the other source. Intuitively, this prevents us from using the randomness of Y to "hide" information on X in ρ_{XY} . Indeed, our previous example fails in this scenario, and we do not know whether this model is viable.

Definition 4.15 (Attempt 3). A (k_1, k_2, ε) two-source extractor against entangled quantum knowledge is a function $E : \{0, 1\}^n \times \{0, 1\}^n \rightarrow \{0, 1\}^m$ such that for any independent sources X, Y and entangled quantum knowledge ρ_{XY} , such that $H_g(X \leftarrow \rho_{XY}Y) \geq k_1, H_g(Y \leftarrow \rho_{XY}X) \geq k_2$, we have $|E(X, Y)\rho_{XY} - U_m\rho_{XY}|_{\text{tr}} \leq \varepsilon$.

It is important to note that this definition, as well as the first two, generalizes *entangled quantum storage*. Similarly to the non-entangled setting, this is a direct consequence of Claim. 4.2, and the following slightly more general claim (which is needed for the third definition only):

Claim 4.16. $H_g(X \leftarrow \rho_{XY}Y) \geq H_\infty(X) - \log \dim \rho_{XY}$

Proof. Let $M = \{M_x\}$ be the best possible POVM on $\rho_{XY}Y$ for guessing X . Since the second register is classical, for a fixed y there exists a measurement on ρ_{Xy} , $M^y = \{E_x^y\}_{x \in X}$ that produces the same distribution as M on $\rho_{Xy}y$. That is, for every x ,

$$\text{Tr}(M_x \cdot (|y\rangle\langle y| \otimes \rho_{Xy})) = \text{Tr}(M_x^y \cdot \rho_{Xy}).$$

(Formally, viewing M_x as a block matrix with blocks of size $\dim \rho_{XY}$, then M_x^y is simply the y -th block on the diagonal.) Thus

$$2^{-H_g(X \leftarrow \rho_{XY}Y)} = \mathbb{E}_{y \leftarrow Y} \mathbb{E}_{x \leftarrow X} \text{Tr}(M_x \cdot (|y\rangle\langle y| \otimes \rho_{Xy})) = \mathbb{E}_{y \leftarrow Y} \mathbb{E}_{x \leftarrow X} \text{Tr}(M_x^y \cdot \rho_{Xy}) \leq \mathbb{E}_{y \leftarrow Y} 2^{-H_g(X \leftarrow \rho_{Xy})}.$$

The result now follows from Claim 4.2. \square

Thus, any $(k_1 - b_1 - b_2, k_2 - b_1 - b_2, \varepsilon)$ extractor against *entangled quantum knowledge* (Attempts 2,3) is a (k_1, k_2, ε) extractor against (b_1, b_2) *entangled quantum storage*.

Appendix A

Many Bit Extractors Against Quantum Storage from Classical Storage

König and Terhal [KT08] prove that any (classical) seeded extractor is secure against *non-entangled* quantum storage, albeit with exponentially larger (in the storage size) error. Their proof is also valid for X-strong (Y-strong) two-source extractors.

Their Lemma 5 essentially shows that every (k_1, k_2, ε) X-strong extractor has error $4 \cdot 2^{3b_2} \cdot \varepsilon$ against (b_1, b_2) quantum storage (for any b_1), assuming $H_\infty(X) \geq k_1$ and $H_g(Y \leftarrow \rho_Y) \geq k_2 + \log \varepsilon^{-1}$. Recall $H_g(Y \leftarrow \rho_Y) \geq H_\infty(Y) - b_2$. Adapted to our definitions, their result is

Lemma A.1 ([KT08, Lemma 5]). *Let E be a (k_1, k_2, ε) X-strong extractor. Then E is a $(k_1, k_2 + b_2 + \log \varepsilon^{-1}, 4 \cdot 2^{3b_2} \varepsilon)$ X-strong extractor against (b_1, b_2) non-entangled storage.*

In particular, this shows that E_D is an X-strong extractor with $m = k_1 + k_2 - 10b_2 - n - 4 - 3 \log \varepsilon^{-1}$. For comparison, our Lem. 3.10 gives $m = \frac{1}{2}(k_1 + k_2 - b_2 - n + 2 - 2 \log \varepsilon^{-1})$, which is better when the storage is large, say, $b_2 \geq k_2/19$.

For completeness, we derive an alternate version of Thm. 1.2 based on Lem. A.1, by composing the extractor above with the seeded extractor of [DPVR09].

Theorem A.2. *The DEOR-construction is a (k_1, k_2, ε) extractor against (b_1, b_2) non-entangled storage with $m = (1 - o(1)) \max(k_1 - 9b_2, k_2 - 9b_1) + k_1 - b_1 + k_2 - b_2 - n - 11 \log \varepsilon^{-1} - O(1)$ output bits provided $k_1 + k_2 - 10 \max(b_1, b_2) > n + \Omega(\log^3(n/\varepsilon))$.*

Here too we are able to extract more bits than guaranteed by Thm. 1.2 when the storage is symmetric and constitutes a small fraction ($< 1/19$) of the min-entropy. In particular, the storage must be at least ten times smaller than the min-entropy, whereas no such restriction exist in Thm. 1.2.

We note that it is not immediately possible to obtain an analogue of Lem. A.1 for weak two-source extractors. The proof relates the security of an extractor with respect to quantum side information, to its security with respect to classical side information. In the weak extractor setting, it thus suffices to consider classical side information of the form $\mathcal{F}(\rho_X \otimes \rho_Y)$ for some specific POVM \mathcal{F} given in the proof. The problem with this approach is that generally $\mathcal{F}(\rho_X \otimes \rho_Y)$ might induce a random variable C_{XY} correlated with both X and Y , breaking the independence assumption (i.e., when conditioning on values of C_{XY} , X and Y might not be independent) and rendering the classical extractor insecure. It is not inconceivable that \mathcal{F} does have the property $\mathcal{F}(\rho_X \otimes \rho_Y) = C_X \otimes C_Y$, but we leave this open.

Bibliography

- [Bel64] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1:195–200, 1964.
- [Bou05] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(1):1–32, 2005.
- [BT10] A. Ben-Aroya and A. Ta-Shma. Better short-seed extractors against quantum knowledge. *CoRR*, abs/1004.3737, 2010.
- [BW92] C. H. Bennett and S. J. Wiesner. Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69(20):2881–2884, 1992.
- [CG88] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal of Computing*, 17(2):230–261, 1988.
- [CvDNT98] R. Cleve, W. van Dam, M. Nielsen, and A. Tapp. Quantum entanglement and the communication complexity of the inner product function. In *Quantum Computing and Quantum Communications, First NASA International Conference*, pages 61–74. 1998.
- [DEOR04] Y. Dodis, A. Elbaz, R. Oliveira, and R. Raz. Improved randomness extraction from two independent sources. In *Proc. 8th International Workshop on Randomization and Computation*, pages 334–344. 2004.
- [DO03] Y. Dodis and R. Oliveira. On extracting private randomness over a public channel. In *Proc. 7th International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 252–263. 2003.
- [DPVR09] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan’s extractor in the presence of quantum side information. *CoRR*, abs/0912.5514, 2009.
- [DV10] A. De and T. Vidick. Near-optimal extractors against quantum storage. In *Proc. 42nd ACM Symp. on Theory of Computing*. 2010. To appear.

- [dW10] R. de Wolf, 2010. Personal communication.
- [FS08] S. Fehr and C. Schaffner. Randomness extraction via delta-biased masking in the presence of a quantum attacker. In *Theory of Cryptography, Fifth Theory of Cryptography Conference*, pages 465–481. 2008.
- [GKdW06] D. Gavinsky, J. Kempe, and R. de Wolf. Strengths and weaknesses of quantum fingerprinting. In *IEEE Conference on Computational Complexity*, pages 288–298. 2006.
- [GKK⁺08] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM Journal of Computing*, 38(5):1695–1708, 2008.
- [GKRdW09] D. Gavinsky, J. Kempe, O. Regev, and R. de Wolf. Bounded-error quantum state identification and exponential separations in communication complexity. *SIAM Journal of Computing*, 39(1):1–24, 2009.
- [Gol95] O. Goldreich. Three xor-lemmas - an exposition. *Electronic Colloquium on Computational Complexity (ECCC)*, 2(56), 1995.
- [HW94] P. Hausladen and W. K. Wootters. A 'pretty good' measurement for distinguishing quantum states. 41(12):2385–2390, 1994.
- [ILL89] R. Impagliazzo, L. A. Levin, and M. Luby. Pseudo-random generation from one-way functions (extended abstracts). In *Proc. 21st ACM Symp. on Theory of Computing*, pages 12–24. 1989.
- [KLR09] Y. T. Kalai, X. Li, and A. Rao. 2-source extractors under computational assumptions and cryptography with defective randomness. In *Proc. 50th Annual Symposium on Foundations of Computer Science*, pages 617–626. 2009.
- [KLRZ08] Y. T. Kalai, X. Li, A. Rao, and D. Zuckerman. Network extractor protocols. In *Proc. 49th Annual Symposium on Foundations of Computer Science*, pages 654–663. 2008.
- [KMR05] R. König, U. M. Maurer, and R. Renner. On the power of quantum memory. *IEEE Transactions on Information Theory*, 51(7):2391–2401, 2005.
- [KRS09] R. König, R. Renner, and C. Schaffner. The operational meaning of min- and max-entropy. *IEEE Transactions on Information Theory*, 55(9):4337–4347, 2009.
- [KT08] R. T. König and B. M. Terhal. The bounded-storage model in the presence of a quantum adversary. *IEEE Transactions on Information Theory*, 54(2):749–762, 2008.

- [NC00] M. A. Nielsen and I. L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 1 edition, 2000.
- [NS06] A. Nayak and J. Salzman. Limits on the ability of quantum states to convey classical messages. *Journal of the ACM*, 53(1):184–206, 2006.
- [Raz05] R. Raz. Extractors with weak random seeds. In *Proc. 37th ACM Symp. on Theory of Computing*, pages 11–20. 2005.
- [Ren05] R. Renner. *Security of Quantum Key Distribution*. Ph.D. thesis, ETH Zurich, September 2005. Available at <http://arxiv.org/abs/quant-ph/0512258>.
- [RK05] R. Renner and R. König. Universally composable privacy amplification against quantum adversaries. In *Theory of Cryptography, Second Theory of Cryptography Conference*, pages 407–425. 2005.
- [RRV99] R. Raz, O. Reingold, and S. P. Vadhan. Error reduction for extractors. In *Proc. 40th Annual Symposium on Foundations of Computer Science*, pages 191–201. 1999.
- [Sha02] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.
- [SV84] M. Santha and U. V. Vazirani. Generating quasi-random sequences from slightly-random sources (extended abstract). In *Proc. 25th Annual Symposium on Foundations of Computer Science*, pages 434–440. 1984.
- [Ta-09] A. Ta-Shma. Short seed extractors against quantum storage. In *Proc. 41st ACM Symp. on Theory of Computing*, pages 401–408. 2009.
- [Tre01] L. Trevisan. Extractors and pseudorandom generators. *Journal of the ACM*, 48(4):860–879, 2001.
- [TSSR10] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner. Leftover hashing against quantum side information. In *IEEE International Symposium on Information Theory*. 2010. To appear.
- [Vaz87] U. V. Vazirani. Strong communication complexity or generating quasirandom sequences from two communicating semi-random sources. *Combinatorica*, 7(4):375–392, 1987.
- [Vid10] T. Vidick, 2010. Personal communication.

- [Yao82] A. C.-C. Yao. Theory and applications of trapdoor functions (extended abstract). In *Proc. 23rd Annual Symposium on Foundations of Computer Science*, pages 80–91. 1982.
- [Zuc90] D. Zuckerman. General weak random sources. In *Proc. 31st Annual Symposium on Foundations of Computer Science*, pages 534–543. 1990.