THEORY OF COMPUTING, Volume 8 (2012), pp. 461–486
www.theoryofcomputing.org

Two-Source Extractors Secure Against Quantum Adversaries*

Roy Kasher† Julia Kempe‡

Received: August 17, 2010; published: September 19, 2012.

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 (1988) two-source inner product extractor and its multi-bit variant by Dodis et al. (RANDOM'04). 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.

ACM Classification: F.1.2, F.2.0, F.2.2, F.2.3

AMS Classification: 68Q10, 68Q12, 68Q17, 68Q25

Key words and phrases: extractors, quantum information

1 Introduction and results

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

*A preliminary version of this paper has appeared in RANDOM'10 [21].

†Supported by JK's ERC Starting Grant QUCO.

‡Supported by the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract Number 015848, by an Alon Fellowship of the Israeli Higher Council of Academic Research, by an Individual Research Grant of the Israeli Science Foundation, by a European Research Council (ERC) Starting Grant and by the Wolfson Family Charitable Trust.

© 2012 Roy Kasher and Julia Kempe
© Licensed under a Creative Commons Attribution License
DOI: 10.4086/toc.2012.v008a021

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} [30, 5, 37]. 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., [30]). 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., [31]). The other main approach, which is the focus of the current work, is to use several independent weak sources (*seedless extractors*) (e. g., [5, 35, 10, 4, 27] 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 [29, 22, 28] proved that the pairwise independent extractor of Impagliazzo et al. [19, 17] is also good against quantum adversaries, and with the same parameters. König and Terhal [24] 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. [14] 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 [32] gave an extractor with a short (polylogarithmic) seed extracting a polynomial fraction of the min-entropy. His result was improved by De and Vidick [8] extracting almost all of the min-entropy. Both constructions are based on Trevisan's extractor [34].

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. In the classical setting this 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 is known to yield several astonishing effects with no counterpart in the classical world, e. g., non-local correlations [1] and superdense coding [3].

We note that the example of Gavinsky et al. can also be viewed as an example in the two-source

[1] Such an effect appears also in *strong seeded* extractors and has been discussed in more detail by König and Terhal [24].

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 (i. e., with equally sized sources of non-full entropy) 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?

Our results In this paper, we answer all these questions affirmatively. We focus on the inner-product based two-source extractor of Dodis et al. [10] (DEOR-extractor). Given two independent weak sources X and Y with the same length n and min-entropies k_1 and k_2 satisfying $k_1 + k_2 \gtrapprox n$, 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 (e. g., [4, 27]).

A first conceptual step in this paper is to define the model of quantum adversaries and of security in the two-source scenario (see Definitions 2.1 and 2.2): Each adversary gets access to an independent weak source X (respectively, 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 $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ 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. [10] 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*

$$m = (1 - o(1)) \max\left(k_1 - \frac{b_1}{2}, k_2 - \frac{b_2}{2}\right) + \frac{k_1 - b_1 + k_2 - b_2 - n}{2} - 9\log \varepsilon^{-1} - O(1)$$

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.

[2] In the setting of seeded extractors with one source, this type of adversary was called *quantum encoding* by Ta-Shma [32].

Theorem 1.3. *The DEOR-construction is a (k_1,k_2,ε) extractor against (b_1,b_2) entangled storage with*

$$m=(1-o(1))\max(k_1-b_2,k_2-b_1)+\frac{k_1-2b_1+k_2-2b_2-n}{2}-9\log\varepsilon^{-1}-O(1)$$

output bits, provided $k_1+k_2-2\max(b_1,b_2)>n+\Omega(\log^3(n/\varepsilon))$.

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., [32, Lemma 3.1]). Using this correspondence, the DEOR-extractor [10] 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 $k_1+k_2-\max(b_1,b_2)>n+\Omega(\log n\cdot(\log^2 n+\log\varepsilon^{-1}))$. Hence the conditions under which one can extract randomness are essentially the same for DEOR and for our Theorem 1.2. The amount of random bits we can extract is somewhat less than that 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*. In this case, the entropy of the source is measured by the (in)ability of the adversary to guess it given some side information (see discussion below and Section 5 for details). 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*

$$m=(1-o(1))\max(k_1,k_2)+\frac{k_1+k_2-n}{6}-9\log\varepsilon^{-1}-O(1)$$

output bits, provided $k_1+k_2>n+\Omega(\log^3(n/\varepsilon))$.

Strong extractors The extractor in Theorems 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 Corollary 3.5, Lemma 4.1, Corollary 5.7 and Lemma 5.8). 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.[3] See Section 2 for details and a discussion of the subtleties in defining a strong extractor in the entangled case, and Sections 3, 4 and 5 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 (see Lemma 3.7).

Proof ideas and tools To show both of our results, we first focus on the simplest case of one-bit outputs. In this case the DEOR-extractor [10] 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

[3] This is called a *strong blender* in [10].

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,[4] 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., [15, 13]).

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. [15] showed an exponential saving in communication in the entangled SMP model, compared to the non-entangled case.[5] 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 we can tap into known results on the quantum communication complexity of the inner product problem (IP). Cleve et al. [6] and Nayak and Salzman [25] 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* (with respect to weak randomness) problems. With some extra work we can adapt the communication lower bounds to 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 Chor and Goldreich [5] 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 Cleve et al. [6] give an $n/2$ lower bound for the message length for Alice and Bob. It turns out this bound is tight,[6] which essentially leads to the factor 2 separation in our results for the entangled vs. non-entangled case (see Section 3).

To show our results for the case of multi-bit extractors, we use the nice properties of the DEOR construction (and its precursors [35, 11]). 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 the 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 Section 2). This way we obtain results

[4] The connection between extractors and communication complexity has been long known, see e. g., [35].

[5] This 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.

[6] We thank Ronald de Wolf [9] for generously allowing us to adapt his upper bound to our setting.

for multi-bit extractors, which even hold in the case of strong extractors. Following Dodis et al. [10], 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 De et al. [7]) to extract even more bits. See Section 4 for details.

Guessing entropy One can weaken the requirement of bounded storage, and instead only place a lower bound on the *guessing entropy* of the source given the adversary's storage, leading to the more general definition of extractors secure against guessing entropy. 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). Guessing entropy is arguably a more natural defintion of "limited knowledge," and applies also when a priori bound on the storage dimension is not known. The resulting extractors are more generally applicable, and secure also in the bounded storage setting (see Section 5). A convenient side effect is that we no longer have to worry about two parameters (min-entropy and storage) instead only working with one (guessing entropy).

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 [28, 24, 12, 7, 33], where several constructions secure against bounded guessing entropy were shown.[7]

In the case of *non-entangled* two-source extractors, we can show (based on [24]) that any classical *one-bit* output two-source extractor remains secure against bounded guessing entropy adversaries, albeit with slightly worse parameters. Moreover, our XOR-Lemma allow us to prove security of the DEOR-extractor against guessing entropy adversaries even in the multi-bit case (Theorem 1.4, see Section 5 for the details).[8]

In the *entangled* adversaries case, one natural way to define the model is to require the guessing entropy of each source given the corresponding adversary's storage to be high. 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_1r_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 (respectively, Y) relative to the adversary's storage unchanged yet we can recover x and y completely from the combined storage.

Hence we are naturally led to consider the weaker requirement that the guessing entropy of each source given the combined storage of *both* adversaries is high. We now observe that already the DEOR one-bit extractor (where the output is simply the inner product) is not secure under this definition, indicating that this definition is still too strong. To see this, consider uniform n-bit sources X, Y, and say Alice stores $x \oplus r$, and Bob stores $y \oplus r$, where r is a shared random string. Obviously, their joint state does not help in guessing X (or Y), hence the guessing entropy of the sources is still n; but their joint state does give $x \oplus y$. If, in addition, Alice also stores the Hamming weight $|x|$ mod 4 and Bob $|y|$ mod 4, the guessing entropy is barely affected, and indeed one can easily show it is $n - O(1)$. However, their information now suffices to compute $x \cdot y$ exactly, since $x \cdot y = (1/2)((|x| + |y| - |x \oplus y|) \bmod 4)$. Hence

[7] Renner [28] deals with the notion of *conditional min-entropy*, which was shown to be equivalent to guessing entropy [23].

[8] We are grateful to Thomas Vidick for pointing out that our XOR-Lemma allows us to obtain results also in this setting.

the inner product is insecure in this model even for very high guessing entropies, even though it is secure against a fair amount of bounded storage.

In light of this, it is not clear if and how entangled guessing entropy sources can be incorporated into the model, and hence we only consider bounded storage adversaries in the entangled case. We note that it is also unclear how to define extractors for dependent random variables in the classical scenario, and just as there are constructions for certain kinds of dependent weak sources, it might be possible to give constructions for certain kinds of entanglement.

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 [29, 22, 28, 24, 32, 8, 7, 2] 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 [24]. Using what is implicit in their work, and with some extra effort, it is possible to obtain results in the one-bit output non-entangled two-source scenario (which hold against guessing entropy adversaries, but with worse performance than our results for the inner product extractor), and we give this result in detail in Section 5. Moreover, König and Terhal [24] 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 Theorem 1.2. We have worked out the details and comparison to Theorem 1.2 in Section 6. Note, however, that to our knowledge no previous work gives results in the entangled scenario.

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, Theorem 1.2 can be made essentially optimal (with respect to the classical case).

We have shown 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 the 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 [15]. 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 [4] which works for two sources with min-entropy $(1/2-\alpha)n$ each for some small constant α, or the construction of Raz [27], 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 application of secure multi-source extractors in the quantum world. One possible scenario is multi-party computation. Classically, Kalai et al. [20] show 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.

Structure of the paper In Section 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. Section 3 is dedicated to the one-bit output case and the connection to communication complexity and gives our tightness results. In Section 4 we deal with the multi-bit output case and prove our main result, Theorems 1.2 and 1.3. In Section 5 we present our results against non-entangled guessing entropy adversaries (partly based on [24]) and prove Theorem 1.4. Section 6 works out the results that can be derived from [24] in the case of multi-bit extractors against non-entangled bounded storage.

2 Preliminaries and tools

In this section we provide the necessary notation, formalize Definition 1.1, describe the DEOR-extractor and present and prove our quantum XOR-Lemma. For background on quantum information see, e. g., [26].

Notation Given a classical random variable Z and a set of density matrices $\{\rho_z\}_{z\in Z}$ we denote by $Z\rho_Z$ the 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. For matrix norms, we define

$$\|A\|_{\mathrm{tr}} = \mathrm{Tr}\left(\sqrt{A^\dagger A}\right) \quad \text{and} \quad \|A\|_{\mathrm{F}} = \sqrt{\mathrm{Tr}(A^\dagger A)}\,.$$

Both norms are invariant under unitary transformations and satisfy

$$\|A\|_{\mathrm{F}} \leq \|A\|_{\mathrm{tr}} \leq \sqrt{D}\,\|A\|_{\mathrm{F}}\,,$$

where D is the dimension of A. Lastly, we define the trace *distance* between two matrices:[9]

$$|A-B|_{\mathrm{tr}} = \frac{1}{2}\,\|A-B\|_{\mathrm{tr}}\,.$$

Extractors against quantum storage We first formalize the different types of quantum storage.

Definition 2.1. For two random variables X,Y we say ρ_{XY} is a (b_1,b_2) *entangled storage*[10] if it is generated by two non-communicating parties, Alice and Bob, in the following way. Alice and Bob initially share an arbitrary entangled state. Alice receives $x \in X$, Bob receives $y \in Y$. They each apply an arbitrary quantum operation on their qubits. Alice then traces out all but b_1 of her qubits, and Bob b_2 of his qubits, giving the state ρ_{xy}.

We denote by ρ^A_{XY} the state obtained when Alice keeps her entire state, whereas Bob traces all but b_2 of his qubits, and similarly for ρ^B_{XY}.

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 2.2. A (k_1,k_2,ε) two-source extractor against (b_1,b_2) (entangled) quantum storage is a function $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ 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},

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

For ease of notation we'll omit the tensor sign when dealing with the uniform distribution; it should always be assumed to be completely independent.

The extractor is called *X-strong* if

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

X-superstrong when ρ_{XY} is replaced by ρ^A_{XY}, and similarly for Y. 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 equivalent (and correspond to an extractor against (k_1,b_2) storage), as the state can be completely reconstructed from the source. Further note that 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 that it is either X-strong or Y-strong. All our proofs follow this route.

[9] This corresponds to the classical statistical distance and is commonly used in quantum information theory.

[10] Technically, the storage is fully defined by $XY\rho_{XY}$, but adversaries only see partial traces, e. g., $X\rho XY$ or ρXY.

Flat sources It is well known that any source with min-entropy k is a convex combination of flat sources (i. e., sources that are uniformly distributed over their support) with min-entropy k. In what follows we will therefore only consider such sources in our analysis of extractors, as one can easily verify that for every sources X,Y and quantum storage ρ_{XY},

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

where $X = \sum \alpha_i X_i$ and $Y = \sum \beta_j Y_j$ are convex combinations of flat sources.

The DEOR construction The following (strong) extractor construction is due to Dodis et al. [10]. Every output bit is the 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. Here $x \cdot y := \sum_{j=1}^{n} x_j y_j \pmod 2$. 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 the inner product.

Lemma 2.3 ([10])**.** *For all $n > 0$, there exist an efficiently computable set of $n \times n$ matrices $A_1, A_2, \ldots, A_n$ 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 2.4 (strong blender of [10])**.** Let $n \geq m > 0$, and let $\{A_i\}_{i=1}^{m}$ be a set as above. The DEOR-extractor $E_D : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ is given by $E_D(x,y) = A_1 x \cdot y, A_2 x \cdot y, \ldots, A_m x \cdot y$.

The XOR-Lemma Vazirani's XOR-Lemma [35] 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 2.5 (Classical XOR-Lemma [35, 16])**.** *For every m-bit random variable Z*

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

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 2.6 (Classical-Quantum XOR-Lemma)**.** *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*[11]

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

Proof. Following the proof of the classical XOR-Lemma in [16], we first relate $\|Z\rho_Z - U_m\rho_Z\|_{\mathrm{tr}}$ to $\|Z\rho_Z - U_m\rho_Z\|_{\mathrm{F}}$, 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 2.7. *For any Boolean function f, $\left\|f(Z)\rho_Z - U_1\rho_Z\right\|_{\mathrm{tr}} = \left\|\sum_z (-1)^{f(z)} p(z)\rho_z\right\|_{\mathrm{tr}}$.*

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

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

$$\begin{aligned}\big\|f(Z)\rho_Z - U_1\rho_Z\big\|_{\mathrm{tr}} &= \Big\||0\rangle\langle 0|\otimes\rho_0 + |1\rangle\langle 1|\otimes\rho_1 - \frac{1}{2}(|0\rangle\langle 0| + |1\rangle\langle 1|)\otimes(\rho_0+\rho_1)\Big\|_{\mathrm{tr}} \\ &= \frac{1}{2}\Big\||0\rangle\langle 0|\otimes(\rho_0-\rho_1) + |1\rangle\langle 1|\otimes(\rho_1-\rho_0)\Big\|_{\mathrm{tr}} \\ &= \big\|\rho_0-\rho_1\big\|_{\mathrm{tr}} = \Big\|\sum_z (-1)^{f(z)}p(z)\rho_z\Big\|_{\mathrm{tr}}. \end{aligned} \tag{2.1}$$

□

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

$$\begin{aligned}\big\|Z\rho_Z - U_m\rho_Z\big\|_{\mathrm{tr}}^2 &= \Big\|\sum_z |z\rangle\langle z|\otimes\sigma_z\Big\|_{\mathrm{tr}}^2 = \Big\|(H^{\otimes m}\otimes I_D)\Big(\sum_z |z\rangle\langle z|\otimes\sigma_z\Big)(H^{\otimes m}\otimes I_D)\Big\|_{\mathrm{tr}}^2 \\ &= \frac{1}{M^2}\cdot\Big\|\sum_{z,y,S}|y\rangle\langle S|\otimes\chi_S(z)\chi_y(z)\sigma_z\Big\|_{\mathrm{tr}}^2 \le \frac{D}{M}\cdot\Big\|\sum_{z,y,S}|y\rangle\langle S|\otimes\chi_S(z)\chi_y(z)\sigma_z\Big\|_{\mathrm{F}}^2, \end{aligned} \tag{2.2}$$

where H is the Hadamard transform.

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

$$\big\|Z\rho_Z - U_m\rho_Z\big\|_{\mathrm{tr}}^2 \le \frac{D}{M}\sum_y\sum_S\Big\|\sum_z\chi_{y+S}(z)\sigma_z\Big\|_{\mathrm{F}}^2 = D\sum_S\Big\|\sum_z\chi_S(z)\sigma_z\Big\|_{\mathrm{F}}^2 \le D\sum_S\Big\|\sum_z\chi_S(z)\sigma_z\Big\|_{\mathrm{tr}}^2, \tag{2.3}$$

where for the equality we used the fact that $\sum_S\big\|\sum_z\chi_{y+S}(z)\sigma_z\big\|_{\mathrm{F}}^2$ does not depend on y. Using Claim 2.7 with $f(Z) = S\cdot Z$, we get

$$\sum_{S\neq 0}\big\|(S\cdot Z)\rho_Z - U_1\rho_Z\big\|_{\mathrm{tr}}^2 = \sum_{S\neq 0}\Big\|\sum_z\chi_S(z)p(z)\rho_z\Big\|_{\mathrm{tr}}^2 = \sum_{S\neq 0}\Big\|\sum_z\chi_S(z)\sigma_z\Big\|_{\mathrm{tr}}^2 = \sum_S\Big\|\sum_z\chi_S(z)\sigma_z\Big\|_{\mathrm{tr}}^2, \tag{2.4}$$

where the second equality holds since χ_S is balanced, and the third since $\sum_z\sigma_z = 0$. Combining equations (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 Cauchy-Schwarz inequality, we obtain

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

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 $\|P\otimes B\|_{\mathrm{tr}} = \dim(P)\cdot\|B\|_{\mathrm{tr}} = M\cdot\|B\|_{\mathrm{tr}}$. As before, combining this with equation (2.4) gives the desired bound. □

3 Communication complexity and one-bit extractors

3.1 Average case lower bound for inner product

Cleve et al. [6] 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 transmitting bits over a quantum channel to that of jointly computing the inner product, and then using an extended Holevo bound. Nayak and Salzman [25] obtained an optimal lower bound by replacing Holevo with a more "mission-specific" bound.

Theorem 3.1 ([25], Theorem 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*

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

(b) $\Pr[Y = X] \leq 2^{-(k-2b)}$.

Revisiting the reduction of Cleve et al., we now show how to adapt it to flat sources, to the average case error and to the 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 3.2. *Let X,Y be flat sources over n bits with min-entropies k_1,k_2, and A,B be 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 $1/2+\varepsilon$. Then*

(a) $\varepsilon \leq 2^{-(k_1+k_2-b-n+2)/2}$*, if the parties don't share prior entanglement, and*

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

Proof. Let us first consider the case $A = B = I$. Assume without loss of generality that 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.

Cleve's idea is to 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, where the output resides in Bob's left-most qubit.
3. Bob performs the Hadamard transform on each of his first $n+1$ qubits and measures in the computational basis.

For a fixed x, denote the success probability of P by $1/2+\varepsilon_x$ (ε_x might be negative). After receiving the message, each term of Bob's state is of the form $|z\rangle|y\rangle|0\rangle|u_x\rangle$, where $|z\rangle$ is a place holder for the output and $|u_x\rangle$ 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 then performed locally by Bob. We denote this computation by P_B. After applying P_B, each term becomes

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

and by assumption, $\mathbb{E}_y \beta_{x,y}^2 = 1/2 - \varepsilon_x$. Following the analysis of Cleve et al. [6], using *clean* computation produces the state

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

where

$$|M_{x,y,z}\rangle = \left(\frac{1}{\sqrt{2}} |z + \overline{x \cdot y}\rangle - \frac{1}{\sqrt{2}} |z + x \cdot y\rangle \right) P_B^\dagger |y\rangle |\overline{x \cdot y}\rangle |K_{x,y}\rangle \,.$$

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 $a \in \{0,1\}, y \in Y, y' \notin Y$.

It follows that after the second step, Bob's complete state is $|\psi\rangle = |v\rangle + |e\rangle$ where

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

By the properties of $|M_{x,y,z}\rangle$,

$$\big| |e\rangle \big| = 2\sqrt{\mathbb{E}_y \beta_{x,y}^2} = 2\sqrt{\frac{1}{2} - \varepsilon_x} \,.$$

Since $|v\rangle + |e\rangle$ and $|v\rangle$ are normalized states, we can easily derive $\langle v | (|v\rangle + |e\rangle) = 2\varepsilon_x$. Define

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

and note that the second term is orthogonal to both $|v\rangle$ and $|e\rangle$. It follows that $\langle \psi | \psi_0 \rangle = \sqrt{2^{k_2-n+2}} \varepsilon_x$. Applying the Hadamard transform 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 Theorem 3.1(a) and 3.1(b) 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. □

3.2 One bit extractor

When the extractor's output is binary, distinguishing it from the uniform distribution is equivalent to computing the output on average. This was shown by Yao [36] when the storage is classical and is trivially extended to the quantum setting.

Fact 3.3. *For every cq-state $Z\rho_Z$ and Boolean function f,*

$$|f(Z)\rho_Z - U_1\rho_Z|_{\mathrm{tr}} \leq \varepsilon \iff \max_M \Pr[M(\rho_Z) = f(Z)] \leq \frac{1}{2} + \varepsilon\,,$$

where $M(\rho_Z)$ is the (classical) distribution of measurement outcomes of POVM M, the maximum ranges over all Boolean POVM, and the probability is over Z and the measurement M.

With this observation, reformulating Lemma 3.2 in the language of trace distance yields a one bit extractor.

Corollary 3.4. *The function $E_{\mathrm{IP}}(x,y) = x \cdot y$ is a (k_1,k_2,ε) extractor against (b_1,b_2) (entangled) quantum storage provided*

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

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

Proof. With Yao's equivalence, Lemma 3.2(b) immediately gives

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

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

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 Lemma 3.2(a) gives the non-entangled case. □

Recall (see Definition 2.2 and discussion thereafter) that one-way communication corresponds to the model of *superstrong* extractors. It is not surprising then that Lemma 3.2 actually gives a superstrong extractor. By choosing ε in the above proof of Corollary 3.4 such that both inequalities (3.1) and (3.2) are satisfied, where we replace ρ_{xY} by ρ^A_{xY} to include Alice's complete state as well as Bob's entangled qubits and similarly for ρ^B_{Xy}, we obtain the following corollary.

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

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

(b) (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 $b_1 = b_2 =: b$. Corollary 3.4 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 [9].

Claim 3.6. *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.*

Proof. Let $x,y \in \{0,1\}^n$ be Alice and Bob's inputs. Since $x \cdot y = (1/2)((|x|+|y|-|x \oplus y|) \bmod 4)$, it suffices to show that the referee can compute $x \oplus y$ with $n/2$ qubits of communication from each party, or simply $(x_1 \oplus y_1, x_2 \oplus y_2) \in \{0,1\}^2$ 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 $\sigma_{x_1 x_2}$ to her qubit and sends it to the referee, and Bob does the same with $\sigma_{y_1 y_2}$. Note that applying $\sigma_{b_1 b_2}$ to the first qubit has the same effect as applying it to the second qubit, up to global phase. 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 $x_1 \oplus y_1 = 1$ and Z is applied to the first qubit iff $x_2 \oplus y_2 = 1$. The net effect on the EPR state (up to global phase) is $\sigma_{x_1 \oplus y_1, x_2 \oplus y_2} \otimes I$. For each value of $(x_1 \oplus y_1, x_2 \oplus y_2)$ this gives one of the orthogonal (completely distinguishable) Bell states. □

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

Lemma 3.7. *If $E_{\mathrm{IP}} = x \cdot y$ 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. Our proof follows the lines of Proposition 10 of Chor and Goldreich [5] with adaptations to the case of quantum side information. We need the following theorem.

Theorem 3.8 ([5, Theorem 3])**.** *There exist independent random variables X, Y on ℓ bits with min-entropy $\ell - 3$ each*[12] *such that* $\Pr[X \cdot Y = 0] > 1/2 + 2^{-(\ell-1)/2}$.

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 3.6 with $\min(b_1, b_2)$ qubits from each.

In the case $\Delta > b$, we define $X = X_1 X_2 X_3 X_4$ 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 3.8 (for $\ell = \Delta + 6 - b$), and X_4 is constant 0^{n-k_1-3}. Analogously, $Y = Y_1 Y_2 Y_3 Y_4$, such that the lengths $|Y_i| = |X_i|$ are 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 3.8, and Y_4 is uniform on $k_2 - \Delta - 3$ bits. It is easily verified that $H_\infty(X) \geq k_1$

[12] The theorem in [5] is stated with slightly different parameters and for arbitrary Boolean functions. Our modification is trivial.

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 Claim 3.6 allowing us to compute $x_1 \cdot y_1$ exactly, and M the measurement strategy of the referee. Applying Theorem 3.8,

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

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.[13]

In the superstrong case with entanglement, assume without loss of generality 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. □

4 Many bit extractors

Here we prove our main theorems, Theorems 1.2 and 1.3. First, using our quantum XOR-Lemma, Lemma 2.6, we obtain results in the *strong* case.

Lemma 4.1. *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 $E_D(x,y) = A_1x \cdot y, A_2x \cdot y, \ldots, A_mx \cdot y$ (see Definition 2.4). 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 2.6,

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

The result then follows by inequality (3.1) in the proof of Corollary 3.4 and its non-entangled analogue. □

In a similar way, we also obtain a *Y-strong* extractor with analogous parameters. Following Dodis et al. [10], we now apply a seeded extractor against quantum storage (see Definition 4.2) to the output of an X-strong (Y-strong) extractor to obtain a two-source extractor with more output bits.

Definition 4.2 ([32])**.** A function $E : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ is a (k,ε) *seeded* extractor against b qubit quantum storage[14] if for any n-bit source X with min-entropy k and any b qubit quantum storage ρ_X,

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

[13] In fact, this shows that our non-entangled extractor is tight even for *classical* storage.

[14] In terms of Definition 2.2, this corresponds to a (k,d,ε) two-source extractor against $(b,0)$ quantum storage where the second source is of length d.

Lemma 4.3. *Let $E_B : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^d$ be a (k_1,k_2,ε) X-strong extractor against (b_1,b_2) (entangled) quantum storage, let $E_S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be some function and define $E(x,y) = E_S(x,E_B(x,y))$.*

(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. For part (a),

$$|E_B(X,Y)\rho_{XY}X - U_d\rho_{XY}X|_{\mathrm{tr}} \leq \varepsilon \quad \text{and so} \quad |E_S(X,E_B(X,Y))\rho_{XY} - E_S(X,U_d)\rho_{XY}|_{\mathrm{tr}} \leq \varepsilon\,.$$

But $|E_S(X,U_d)\rho_{XY} - U_m\rho_{XY}|_{\mathrm{tr}} \leq \varepsilon$ by definition of E_S. The result follows from the triangle inequality. For part (b) note that when the storage is non-entangled,

$$\left|E_S(X,U_d)\rho_X\rho_Y - U_m\rho_X\rho_Y\right|_{\mathrm{tr}} = \left|E_S(X,U_d)\rho_X - U_m\rho_X\right|_{\mathrm{tr}}\,,$$

and it suffices to require that E_S be a seeded extractor against only b_1 quantum storage. □

A seeded extractor with almost optimal min-entropy loss is given by De et al. [7]. Their extractor is secure against quantum guessing entropy sources, and so trivially against quantum storage [24] (see Section 5 for details). We reformulate the seeded extractor in terms of Definition 4.2.

Corollary 4.4 ([7, 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 Theorems 1.2 and 1.3 now follow by composing the explicit extractors of Lemma 4.1 and Corollary 4.4 as in Lemma 4.3.

Proof of Theorem 1.2. E_D is an X-strong extractor against non-entangled storage with

$$\frac{k_1+k_2-b_2-n-2\log\varepsilon^{-1}}{2}$$

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 Corollary 4.4 on the source X gives

$$m = \frac{k_1+k_2-b_2-n-2\log\varepsilon^{-1}}{2} + (k_1-b_1) - 8\log(k_1-b_1) - 8\log\varepsilon^{-1} - O(1)\,.$$

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.[15] □

[15] We slightly sacrifice the parameters in the formulation of the theorem to simplify the result.

Proof of Theorem 1.3. E_D is an X-strong extractor against entangled storage with

$$\frac{k_1+k_2-2b_2-n-2\log\varepsilon^{-1}}{2}$$

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 Corollary 4.4 on the source X, obtaining

$$m = \frac{k_1+k_2-2b_2-n-2\log\varepsilon^{-1}}{2} + (k_1-b_1-b_2) - 8\log(k_1-b_1-b_2) - 8\log\varepsilon^{-1} - O(1),$$

and similarly for Y. □

5 Guessing entropy adversaries

In previous sections, we considered extractors in the presence of quantum adversaries with limited storage. A stronger notion of quantum adversary was also studied in the literature [28, 24, 12, 7, 33].

Definition 5.1 ([24])**.** Let $X\rho_X$ be an arbitrary cq-state. The guessing entropy of X given ρ_X is

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

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

Considering the probability distribution on the support of X induced by measuring with M on ρ_X (which we denote by $M(\rho_X)$), the above can be perhaps more easily understood as

$$H_g(X \leftarrow \rho_X) = -\log\max_M \Pr[M(\rho_X) = X].$$

Renner [28] considered sources with high *conditional min-entropy*, rather than *guessing entropy*. The two were shown to be equivalent [23].

We can now define two-source extractors secure against non-entangled guessing entropy adversaries. Recall that in the non-entangled case the bounded storage is given by $\rho_X \otimes \rho_Y$ (see Definition 2.1). 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 5.2. A (k_1, k_2, ε) two-source extractor against quantum knowledge is a function $E : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^m$ 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 $|E(X,Y)\rho_X\rho_Y - U_m\rho_X\rho_Y|_{\mathrm{tr}} \leq \varepsilon$.

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

It was shown that $H_g(X \leftarrow \rho_X) \geq H_\infty(X) - \log\dim(\rho_X)$ [24]. Thus, 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.

One-bit output case König and Terhal [24] 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, $|E(X,Y)\rho_{XY} - U_1\rho_{XY}|_{\mathrm{tr}}$ is small if the statement is also true when ρ_X is classical. We give a version of their Lemma 2 with slightly better parameters (smaller multiplicative constant on the right hand side). 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) [18] 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 5.3. *Let $Z\rho_Z$ be a cq-state, and f be a Boolean function. Then*[16]

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

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

Proof. We need the following lemma.

Lemma 5.4 ([28, Lemma 5.1.3])**.** *Let S be a Hermitian operator and let σ be a nonnegative operator. Then*

$$|S|_{\mathrm{tr}} \le \frac{1}{2}\sqrt{\mathrm{Tr}(\sigma)\,\mathrm{Tr}(\sigma^{-1/2}S\sigma^{-1/2}S)}.$$

Let $\rho = \rho_Z$, $\rho_b = \sum_{z:f(z)=b} p(z)\rho_z$ 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

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

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

$$\big|\Pr[M(\rho_Z) = f(Z)] - \Pr[M(\rho_Z) \ne f(Z)]\big| = \mathrm{Tr}\big(\rho^{-1/2}(\rho_0-\rho_1)\rho^{-1/2}(\rho_0-\rho_1)\big). \tag{5.1}$$

By equation (2.1), $|f(Z)\rho_Z - U\rho_Z|_{\mathrm{tr}} = |\rho_0 - \rho_1|_{\mathrm{tr}}$, and by Lemma 5.4, setting $S = \rho_0 - \rho_1$, $\sigma = \rho$,

$$\big|\rho_0 - \rho_1\big|_{\mathrm{tr}} \le \frac{1}{2}\sqrt{\mathrm{Tr}\big(\rho^{-1/2}(\rho_0-\rho_1)\rho^{-1/2}(\rho_0-\rho_1)\big)}. \tag{5.2}$$

Combining equation (5.1) with equation (5.2) gives

$$\big|f(Z)\rho_Z - U\rho_Z\big|_{\mathrm{tr}} \le \sqrt{\frac{1}{4}\big|\Pr[M(\rho_Z) = f(Z)] - \Pr[M(\rho_Z) \ne f(Z)]\big|}.$$

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

Finally,

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

where the right most expression describes a trivial strategy to guess f from $M(\rho)$, giving the desired result. □

Corollary 5.5. *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 Lemma 5.3,

$$\left|E(X,Y)\rho_X\rho_Y-U\rho_X\rho_Y\right|_{\mathrm{tr}} \leq \sqrt{\frac{1}{2}\left|E(X,Y)\mathcal{E}(\rho_X\rho_Y)-U\mathcal{E}(\rho_X\rho_Y)\right|_{\mathrm{tr}}}.$$

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, $\mathcal{E}_1(\rho_X)$ and $\mathcal{E}_2(\rho_Y)$ induce independent classical distributions, which we denote by C_X and C_Y. Thus

$$\left|E(X,Y)\rho_X\rho_Y-U\rho_X\rho_Y\right|_{\mathrm{tr}} \leq \sqrt{\frac{1}{2}\left|E(X,Y)C_XC_Y-UC_XC_Y\right|_{\mathrm{tr}}}, \tag{5.3}$$

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 [24]). Inequality (5.3) then gives the desired parameters against quantum knowledge. □

By a similar argument and following the proof of Theorem 1 in [24], we get

Corollary 5.6. *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.*

The multi-bit output case We now show how to apply the results in the one-bit case, together with our XOR-Lemma 2.6, to show security in the multi-bit case, proving Theorem 1.4.

By inequality (3.1) in the proof of Corollary 3.4, the inner product is a *classical* X-strong extractor with error $\varepsilon\leq 2^{-(k_1+k_2-n+2)/2}$. Plugging this into Corollary 5.6 we obtain

Corollary 5.7. *The function $E_{\mathrm{IP}_A}(x,y)=Ax\cdot y$, for any full rank matrix A, is a (k_1,k_2,ε) X-strong (Y-strong) extractor against quantum knowledge provided that $k_1+k_2\geq n-2+6\log\varepsilon^{-1}$.*

We now repeat the steps performed in Section 4 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 Lemma 4.1 we use the XOR-Lemma 2.6 to reduce the security of E_D to the strong one-bit case of Corollary 5.7.

Lemma 5.8. *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 2.6 and Corollary 5.7,

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

□

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

Lemma 5.9. *Let $E_B : \{0,1\}^n \times \{0,1\}^n \to \{0,1\}^d$ be a (k_1,k_2,ε) X-strong extractor against quantum knowledge and let $E_S : \{0,1\}^n \times \{0,1\}^d \to \{0,1\}^m$ be a (k_1,ε) seeded extractor against quantum knowledge.*[17] *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 5.10 ([7, Corrolary 5.3])**.** *There exists an explicit (k,ε) seeded extractor against quantum knowledge with seed length $d = O(\log^3(n/\varepsilon))$ and $m = d + k - 8\log k - 8\log \varepsilon^{-1} - O(1)$.*

Proof of Theorem 1.4. E_D is an X-strong extractor against quantum knowledge with

$$\frac{k_1 + k_2 - n - 6\log \varepsilon^{-1}}{6} - O(1)$$

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 Corollary 5.10 on the source X gives

$$m = \frac{k_1 + k_2 - n - 6\log \varepsilon^{-1}}{6} + k_1 - 8\log k_1 - 8\log \varepsilon^{-1} - O(1),$$

and similarly for Y. □

6 Many bit extractors against quantum storage from classical storage

König and Terhal [24] 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 that $H_g(Y \leftarrow \rho_Y) \geq H_\infty(Y) - b_2$. Adapted to our definitions, their result is

Lemma 6.1 ([24, 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.*

[17] For a formal definition see [7].

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 Lemma 4.1 gives $m = (1/2)(k_1 + k_2 - b_2 - n + 2 - 2\log\varepsilon^{-1})$, which is better when the storage is large, say, $b_2 \geq k_2/19$.

For completeness, we derive an alternate version of Theorem 1.2 based on Lemma 6.1, by composing the extractor above with the seeded extractor of De et al. [7].

Theorem 6.2. *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 Theorem 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 Theorem 1.2.

We note that it is not immediately possible to obtain an analogue of Lemma 6.1 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.

Acknowledgments

The authors would like to thank Nir Bitansky, Ashwin Nayak, Oded Regev, Amnon Ta-Shma, Thomas Vidick and Ronald de Wolf for valuable discussions. We are especially indebted to Ronald de Wolf for allowing us to use his exact protocol for IP in the SMP model with entanglement, and to Thomas Vidick for pointing out how to replace 2^d with 2^m in our XOR-Lemma, which allowed us to prove Theorem 1.4.

References

[1] JOHN STEWART BELL: On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3):195–200, 1964. 462

[2] AVRAHAM BEN-AROYA AND AMNON TA-SHMA: Better short-seed quantum-proof extractors. *Theoretical Computer Science*, 419:17 – 25, 2012. [doi:10.1016/j.tcs.2011.11.036, arXiv:1004.3737] 467

[3] CHARLES H. BENNETT AND STEPHEN J. WIESNER: Communication via one- and two-particle operators on Einstein-Podolsky-Rosen states. *Phys. Rev. Lett.*, 69(20):2881–2884, 1992. [doi:10.1103/PhysRevLett.69.2881] 462

[4] JEAN BOURGAIN: More on the sum-product phenomenon in prime fields and its applications. *Internat. J. Number Theory*, 1(1):1–32, 2005. [doi:10.1142/S1793042105000108] 462, 463, 468

[5] BENNY CHOR AND ODED GOLDREICH: Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM J. Comput.*, 17(2):230–261, 1988. Preliminary version in FOCS'85. [doi:10.1137/0217015] 462, 465, 475

[6] RICHARD CLEVE, WIM VAN DAM, MICHAEL NIELSEN, AND ALAIN TAPP: Quantum entanglement and the communication complexity of the inner product function. In *1st NASA Internat. Conf. Quantum Computing and Quantum Communications (QCQC'98)*, pp. 61–74. Springer, 1999. To appear in TCS. [doi:10.1007/3-540-49208-9_4] 465, 472, 473

[7] ANINDYA DE, CHRISTOPHER PORTMANN, THOMAS VIDICK, AND RENATO RENNER: Trevisan's extractor in the presence of quantum side information. *SIAM J. Comput.*, 41(4):915–940, 2012. [doi:10.1137/100813683, arXiv:0912.5514] 466, 467, 477, 478, 481, 482

[8] ANINDYA DE AND THOMAS VIDICK: Near-optimal extractors against quantum storage. In *Proc. 42nd STOC*, pp. 161–170. ACM Press, 2010. [doi:10.1145/1806689.1806713] 462, 467

[9] RONALD DE WOLF: 2010. personal communication. 465, 474

[10] YEVGENIY DODIS, ARIEL ELBAZ, ROBERTO OLIVEIRA, AND RAN RAZ: Improved randomness extraction from two independent sources. In *Proc. 8th Internat. Workshop on Randomization and Computation (RANDOM'04)*, pp. 334–344. Springer, 2004. [doi:10.1007/978-3-540-27821-4_30] 462, 463, 464, 466, 470, 476

[11] YEVGENIY DODIS AND ROBERTO OLIVEIRA: On extracting private randomness over a public channel. In *Proc. 7th Internat. Workshop on Randomization and Computation (RANDOM'03)*, pp. 252–263, 2003. [doi:10.1007/978-3-540-45198-3_22] 465

[12] SERGE FEHR AND CHRISTIAN SCHAFFNER: Randomness extraction via δ -biased masking in the presence of a quantum attacker. In *5th Theory of Cryptography Conf. (TCC'08)*, pp. 465–481, 2008. [doi:10.1007/978-3-540-78524-8_26] 466, 478

[13] DMITRY GAVINSKY, JULIA KEMPE, AND RONALD DE WOLF: Strengths and weaknesses of quantum fingerprinting. In *Proc. 21st IEEE Conf. on Computational Complexity (CCC'06)*, pp. 288–298. IEEE Comp. Soc. Press, 2006. [doi:10.1109/CCC.2006.39] 465

[14] DMITRY GAVINSKY, JULIA KEMPE, IORDANIS KERENIDIS, RAN RAZ, AND RONALD DE WOLF: Exponential separation for one-way quantum communication complexity, with applications to cryptography. *SIAM J. Comput.*, 38(5):1695–1708, 2008. Preliminary version in STOC'07. [doi:10.1137/070706550] 462

[15] DMITRY GAVINSKY, JULIA KEMPE, ODED REGEV, AND RONALD DE WOLF: Bounded-error quantum state identification and exponential separations in communication complexity. *SIAM J. Comput.*, 39(1):1–24, 2009. Preliminary version in STOC'06. [doi:10.1137/060665798] 465, 467

[16] ODED GOLDREICH: Three XOR-lemmas - an exposition. In ODED GOLDREICH, editor, *Studies in Complexity and Cryptography: Miscellanea on the Interplay between Randomness and Computation*, volume 6650 of *Lecture Notes in Computer Science*, pp. 248–272. Springer, 2011. ECCC. [doi:10.1007/978-3-642-22670-0_22] 470

[17] JOHAN HÅSTAD, RUSSELL IMPAGLIAZZO, LEONID A. LEVIN, AND MICHAEL LUBY: A pseudorandom generator from any one-way function. *SIAM J. Comput.*, 28(4):1364–1396, 1999. Preliminary version in STOC'89. [doi:10.1137/S0097539793244708] 462

[18] PAUL HAUSLADEN AND WILLIAM K. WOOTTERS: A 'pretty good' measurement for distinguishing quantum states. *J. Modern Optics*, 41(12):2385–2390, 1994. [doi:10.1080/09500349414552221] 479

[19] RUSSELL IMPAGLIAZZO, LEONID A. LEVIN, AND MICHAEL LUBY: Pseudo-random generation from one-way functions (extended abstract). In *Proc. 21st STOC*, pp. 12–24. ACM Press, 1989. [doi:10.1145/73007.73009] 462

[20] YAEL TAUMAN KALAI, XIN LI, AND ANUP RAO: 2-source extractors under computational assumptions and cryptography with defective randomness. In *Proc. 50th FOCS*, pp. 617–626. IEEE Comp. Soc. Press, 2009. [doi:10.1109/FOCS.2009.61] 468

[21] ROY KASHER AND JULIA KEMPE: Two-source extractors secure against quantum adversaries. In *Proc. 14th Internat. Workshop on Randomization and Computation (RANDOM'10)*, pp. 656–669. Springer, 2010. [doi:10.1007/978-3-642-15369-3_49] 461

[22] ROBERT KÖNIG, UELI MAURER, AND RENATO RENNER: On the power of quantum memory. *IEEE Trans. Inform. Theory*, 51(7):2391–2401, 2005. [doi:10.1109/TIT.2005.850087] 462, 467

[23] ROBERT KÖNIG, RENATO RENNER, AND CHRISTIAN SCHAFFNER: The operational meaning of min- and max-entropy. *IEEE Trans. Inform. Theory*, 55(9):4337–4347, 2009. [doi:10.1109/TIT.2009.2025545] 466, 478

[24] ROBERT T. KÖNIG AND BARBARA M. TERHAL: The bounded-storage model in the presence of a quantum adversary. *IEEE Trans. Inform. Theory*, 54(2):749–762, 2008. [doi:10.1109/TIT.2007.913245] 462, 466, 467, 468, 477, 478, 479, 480, 481

[25] ASHWIN NAYAK AND JULIA SALZMAN: Limits on the ability of quantum states to convey classical messages. *J. ACM*, 53(1):184–206, 2006. Preliminary versions in FOCS'99 and CCC'02. [doi:10.1145/1120582.1120587] 465, 472

[26] MICHAEL A. NIELSEN AND ISAAC L. CHUANG: *Quantum Computation and Quantum Information*. Cambridge University Press, 1st edition, 2000. 468

[27] RAN RAZ: Extractors with weak random seeds. In *Proc. 37th STOC*, pp. 11–20. ACM Press, 2005. [doi:10.1145/1060590.1060593] 462, 463, 468

[28] RENATO RENNER: *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, 2005. [arXiv:quant-ph/0512258v2] 462, 466, 467, 478, 479

[29] RENATO RENNER AND ROBERT KÖNIG: Universally composable privacy amplification against quantum adversaries. In *2nd Theory of Cryptography Conf. (TCC'05)*, pp. 407–425. Springer, 2005. [doi:10.1007/978-3-540-30576-7_22] 462, 467

[30] MIKLOS SANTHA AND UMESH V. VAZIRANI: Generating quasi-random sequences from semi-random sources. *J. Comput. System Sci.*, 33(1):75–87, 1986. Preliminary version in FOCS'84. [doi:10.1016/0022-0000(86)90044-9] 462

[31] RONEN SHALTIEL: *Recent Developments in Explicit Constructions of Extractors*, volume 1 of *Current Trends in Theoretical Computer Science:The Challenge of the New Century*, chapter 13, pp. 189–228. World Scientific, 2004. Preliminary version in Bull. EATCS, v.77, 2002. [doi:10.1142/9789812562494_0013] 462

[32] AMNON TA-SHMA: Short seed extractors against quantum storage. *SIAM J. Comput.*, 40(3):664–677, 2011. Preliminary version in STOC'09. [doi:10.1137/09076787X] 462, 463, 464, 467, 476

[33] MARCO TOMAMICHEL, CHRISTIAN SCHAFFNER, ADAM SMITH, AND RENATO RENNER: Leftover hashing against quantum side information. *IEEE Trans. Inform. Theory*, 57(8):5524–5535, 2011. Preliminary version in ISIT'10. [doi:10.1109/TIT.2011.2158473] 466, 478

[34] LUCA TREVISAN: Extractors and pseudorandom generators. *J. ACM*, 48(4):860–879, 2001. Preliminary version in STOC'99. [doi:10.1145/502090.502099] 462

[35] UMESH V. VAZIRANI: Strong communication complexity or generating quasirandom sequences from two communicating semi-random sources. *Combinatorica*, 7(4):375–392, 1987. Preliminary version in STOC'85. [doi:10.1007/BF02579325] 462, 465, 470

[36] ANDREW C. YAO: Theory and applications of trapdoor functions (extended abstract). In *Proc. 23rd FOCS*, pp. 80–91. IEEE Comp. Soc. Press, 1982. [doi:10.1109/SFCS.1982.45] 474

[37] DAVID ZUCKERMAN: General weak random sources. In *Proc. 31st FOCS*, pp. 534–543. IEEE Comp. Soc. Press, 1990. [doi:10.1109/FSCS.1990.89574] 462

AUTHORS

Roy Kasher
Google, Inc.[18]
New York, NY
kasroy@gmail.com

Julia Kempe
Blavatnik School of Computer Science
Tel Aviv University

and

CNRS, LIAFA
Université de Paris 7
Paris, France
kempe@lri.fr
http://www.cs.tau.ac.il/~kempe/

ABOUT THE AUTHORS

ROY KASHER completed his M. Sc. at Tel Aviv University in 2010 under the supervision of Julia Kempe and Amnon Ta-Shma, and this work was written while a student there. He currently works for Google in New York.

JULIA KEMPE completed her Ph. D. in Mathematics on noiseless quantum computation at the University of California, Berkeley, under the supervision of Elwyn Berlekamp and K. Birgitta Whaley, and in Computer Science on quantum walks at the École Nationale Supérieure des Télécommunications, France, under the supervision of Gérard Cohen. She is currently Directeur de Recherche (Senior Researcher) at the CNRS in Paris, France and Associate Professor (on leave) at the Computer Science Department of Tel Aviv University, Israel. Her research interests lie in quantum computation and quantum information.

[18]This work was done while the author was a graduate student at the Blavatnik School of Computer Science, Tel Aviv University, Tel Aviv, Israel.