Semidefinite programming hierarchies for quantum adversaries

Mario Berta
Institute for Quantum Information and Matter
Caltech, Pasadena

Omar Fawzi
École Normale Supérieure de Lyon
Université de Lyon

Volkher B. Scholz
Institute for Theoretical Physics
ETH Zurich

***Abstract*—Randomness extractors are an important building block for classical and quantum cryptography. However, for many applications it is crucial that the extractors are quantum-proof, i.e., that they work even in the presence of quantum adversaries. In general, quantum-proof extractors are poorly understood and we would like to argue that in the same way as Bell inequalities (multi prover games) and communication complexity, the setting of randomness extractors provides a operationally useful framework for studying the power and limitations of a quantum memory compared to a classical one. We start by recalling how to phrase the extractor property as a quadratic program with linear constraints. We then construct a semidefinite programming (SDP) relaxation for this program that is tight for some extractor constructions. Moreover, we show that this SDP relaxation is even sufficient to certify quantum-proof extractors. This gives a unifying approach to understand the stability properties of extractors against quantum adversaries. We analyze the limitations of this SDP relaxation and propose a converging hierarchy of SDPs that gives increasingly tight characterizations of quantum-proof extractors.**

Finally, we discuss more generally how to quantize quadratic optimization programs with linear constraints and develop a converging semidefinite programming hierarchy. We consider two examples other than randomness extractors. For the quadratic program corresponding to the winning probability in a two-prover game (also known as a Bell inequality), our quantization captures the entangled value of the game. If instead the quadratic program is a maximization of the success probability over encoding strategies for a given channel, the quantization corresponds to optimizing over entanglement assisted encoding strategies.

Randomness extractors— A randomness extractor is a procedure to distill from a weakly random system as much (almost) uniform random bits as possible. Such objects are essential in many cryptographic protocols, in particular in quantum key distribution and device independent randomness expansion [2], [10], [24], [27], [36]. In this context, the process of transforming a partly private string into one that is almost uniformly random from the adversary's point of view is called privacy amplification [3], [4]. Even though we take a cryptographic point of view in this submission, we should mention that randomness extractors are very useful combinatorial objects in particular in the study of the computational power of randomness (see [35] for a survey). More precisely, a randomness extractor is described by a family of functions $\mathrm{Ext} = \{f_s\}_{s \in D}$ where $f_s : N \to M$. We use $N = 2^n$ to denote the input system (consisting of strings of n bits), $M = 2^m$ (bit-strings of length m) to denote the output system, and $D = 2^d$ (d bits) to denote the seed system that labels the functions f_s. Note that in a slight abuse of notation, we use the same letter for the actual set of inputs/outputs as well as its size. We say that Ext is a (k, ϵ)-extractor if for any random variable X taking values in N,

$$\begin{aligned} &H_{\min}(X) := -\log p_{\text{guess}}(X) \geq k \\ &\implies f_{U_D}(X) \text{ is } \epsilon\text{-close to } U_M\,, \end{aligned} \tag{1}$$

where U_D is uniformly distributed on D and independent of X and U_M denotes the uniform distribution over M. As mentioned in the equation, the min-entropy $H_{\min}(X)$ is defined by the maximum probability of success in guessing a source X with only the knowledge of the distribution p of X. In this case, we simply have $H_{\min}(X) = -\log \max p(x)$. To quantify the distance between distributions, we use the total variation distance.[1] Equation (1) can thus be more explicitly written as

$$\forall x \in N,\ p(x) \leq 2^{-k} \implies \frac{1}{D} \sum_{\substack{s \in D \\ y \in M}} \left| \sum_{\substack{x: y= \\ f_s(x)}} p(x) - \frac{1}{M} \right| \leq \epsilon\,. \tag{2}$$

Even though the concept was already present in [3], [4], the definition of randomness extractors was formulated in [23]. The typical example of a family $\{f_s\}_s$ of functions that satisfy this condition are randomly chosen functions. In fact, one can show [26], [30] that choosing D functions f_s independently at random among all the functions from N to M satisfies equation (2) with the following parameters

$$m = k - 2\log(1/\epsilon) - O(1) \quad \text{and} \tag{3}$$

$$d = \log(n-k) + 2\log(1/\epsilon) + O(1)\,. \tag{4}$$

We even know that these parameters cannot be improved except for additive constants [26]. Probabilistic constructions are interesting, but for applications we usually want the functions f_s to be efficiently computable. The most famous example of an explicit extractor is given by two-universal hash functions [3], [4], [15]. However, this construction has a seed size d of the order of n, very far from the $\log n$ achieved by probabilistic constructions (4). Constructing efficiently computable extractors that match the parameters of

[1] It is more convenient here to use simply the ℓ_1-norm between the distributions, ignoring the $\frac{1}{2}$ factor in the usual definition of the total variation distance.

randomly chosen functions has been the subject of a large body of research. Starting with the work of Nisan and Ta-Shma [22] and followed by Trevisan's breakthrough result [34], there has been a lot of progress in achieving polylogarithmic seed size, and there are now many intricate constructions that come close to the parameters in (3) (see the review articles [29], [35]).

Quantum-proof randomness extractors— For applications in classical and quantum cryptography (see, e.g., [19], [27]) and for constructing device independent randomness amplification and expansion schemes (see, e.g., [9], [11], [21]) it is important to find out if extractor constructions also work when the input source is correlated to another (possibly quantum) system Q. That is, we would like that for all classical-quantum input density matrices $\rho_{QN} = \sum_{x \in N} \rho(x) \otimes |x\rangle\langle x|$ acting on QN with conditional min-entropy

$$H_{\min}(N|Q)_\rho := -\log p_{\text{guess}}(N|Q)_\rho \geq k\,, \tag{5}$$

where $p_{\text{guess}}(N|Q)$ denotes the maximal probability of guessing the system N given Q, the output is uniform and independent of Q,[2]

$$\frac{1}{D} \sum_{\substack{s \in D \\ y \in M}} \left\| \sum_{\substack{x: y = \\ f_s(x)}} \rho(x) - \frac{1}{M} \sum_{x \in N} \rho(x) \right\|_1 \leq \epsilon\,. \tag{6}$$

As observed in [18, Proposition 1], if we restrict the system Q to be classical with respect to some basis $\{|e\rangle\}_{e \in Q}$ then every (k, ϵ)-extractor as in (2) is also a $\big(k + \log(1/\epsilon), 2\epsilon\big)$-extractor in the sense of (6). That is, even when the input source is correlated to a classical system Q, every extractor construction still works (nearly) equally well for extracting randomness. However, if Q is quantum no such generic reduction is known and extractor constructions that also work for quantum Q are called quantum-proof.[3] Examples of (approximately) quantum-proof extractors include:

- Spectral (k, ϵ)-extractors are quantum-proof $(k, 2\sqrt{\epsilon})$-extractors [8, Theorem 4]. This includes in particular two-universal hashing [27], [33], two-wise independent permutations [31], as well as sample and hash based constructions [17].
- One-bit output (k, ϵ)-extractors are quantum-proof $(k + \log(1/\epsilon), 3\sqrt{\epsilon})$-extractors [18, Theorem 1].
- (k, ϵ)-extractors constructed along Trevisan [34] are quantum-proof $\big(k + \log(1/\epsilon), 3\sqrt{\epsilon}\big)$-extractors [13, Theorem 4.6] (see also [1]).

We emphasize that all these stability results are specifically tailored proofs that make use of the structure of the particular extractor constructions. In contrast to these findings it was shown by Gavinsky *et al.* [14, Theorem 1] that there exists a valid (though contrived) extractor for which the decrease in the quality of the output randomness has to be at least $\epsilon \mapsto \Omega(m\epsilon)$.[4] As put forward by Ta-Shma [32, Slide 84], this then raises the question if the separation found by Gavinsky *et al.* is maximal, that is: is every (k, ϵ)-extractor a quantum-proof $\big(O(k + \log(1/\epsilon)), O(m\sqrt{\epsilon})\big)$-extractor or does there exists an extractor that is not quantum-proof with a large separation, say $\epsilon \mapsto (2^m\epsilon)^{\Omega(1)}$? We note that such a stability result would make every extractor with reasonable parameters (approximately) quantum-proof. However, for reasons discussed later it is unclear if such a generic quantum-proof reduction is possible and small sets of randomly chosen functions are interesting candidates to study this possibly large classical/quantum separation.

Our results about extractors— (technical details can be found in [6]):

- We write the extractor condition (2) as a quadratic optimization program. The optimal value for this program denoted as $\mathrm{C}(\mathrm{Ext}, k)$ is the smallest error ϵ such that Ext is a (k, ϵ)-extractor. We then construct a semidefinite programming (SDP) relaxation for this program whose optimal value is denoted $\mathrm{SDP}(\mathrm{Ext}, k)$. This program gives an efficiently computable procedure to certify that a family of functions Ext is a (k, ϵ)-extractor for $\epsilon = \mathrm{SDP}(\mathrm{Ext}, k)$.
- We show that this certification procedure gives us much more: it certifies that Ext is a *quantum-proof* $(k, \sqrt{2}\epsilon)$-extractor. Thus, we give a general efficient method for proving that an extractor is quantum-proof. This technique can recover in a unified way many of the currently known methods for constructing quantum-proof extractors. In particular, we can show that constructions based on two-universal hashing [28], [33] are quantum-proof, and that any extractor with entropy deficit $n - k$ or output size m small is quantum-proof [5] (for $m = 1$ this was known before [18]). This latter result is a basic building block for showing that Trevisan based extractors are quantum-proof [13], and the extension from $m = 1$ to general small m could lead to more efficient implementations of short seed quantum-proof extractors [20].
- We consider the limitations of this SDP relaxation. Even though $\mathrm{SDP}(\mathrm{Ext}, k)$ is a tight bound on $\mathrm{C}(\mathrm{Ext}, k)$ for many extractor constructions, there can be a large gap between these two values. In particular, if $\mathrm{Ext}_{\text{rand}}$ is given by a small number of randomly chosen functions, then $\mathrm{C}(\mathrm{Ext}_{\text{rand}}, k) \ll \mathrm{SDP}(\mathrm{Ext}_{\text{rand}}, k)$. This shows that the method we propose cannot be used to prove that a small set of randomly chosen functions define good extractors. This means that other techniques would be needed to determine whether $\mathrm{Ext}_{\text{rand}}$ is a quantum-proof extractor or not. To go in this direction, we propose a hierarchy of SDPs that gives increasingly tight characterizations of the quantum-proof extractor condition (6) at the cost of

[2] Other notions for weaker quantum adversaries have also been discussed in the literature, e.g., in the bounded storage model (see [13, Section 1] for a detailed overview).

[3] Note that the dimension of Q is unbounded and that it is a priori unclear if there exist any extractor constructions that are quantum-proof (even with arbitrarily worse parameters).

[4] Since the quality of the output randomness of Gavinsky *et al.*'s construction is bad to start with, the decrease $\epsilon \mapsto \Omega(m\epsilon)$ for quantum Q already makes the extractor fail completely in this case.

increasing dimension.

Quadratic programs and their non-commutative versions—(technical details will be publicly available soon [7]). Our methods are not restricted to study randomness extractors but also allow to analyze general quadratic optimization programs with linear constraints. That is, for A_{ij} a real-valued symmetric matrix, $i,j \in \{1,\dots N\}$, we want to optimize expressions of the form $\sum_{ij} A_{ij}x_ix_j$ such that the variables $x_i \in \mathbb{R}$ satisfy linear constraints $g_k \in \mathbb{R}[x_1,\dots,x_N]$:

$$p[A,\mathcal{G}] := \max \quad \left|\sum_{ij} A_{ij}x_ix_j\right| \tag{7}$$

$$\text{subject to} \quad \mathcal{G} := \{g_1, g_2, \dots\}\,. \tag{8}$$

As discussed above the extractor condition (2) is exactly of this form, but more generally quadratic optimization problems appear frequently in graph theory where we might think of A_{ij} as the adjacency matrix of a graph and indexes i,j labeling vertices.[5] Another class of problems of the form (7)-(8) are the classical value of two-prover games (also known as a Bell inequalities). As an interesting special case we would like to mention classical channel coding with fixed message length. Here we want to send k possible classical messages over a classical channel $p(y|x)$ and maximize the average success probability:

$$p_{\text{success}} := \max \quad \frac{1}{k}\sum_{i,x,y} e(i,x)p(y|x)d(y,i) \tag{9}$$

$$\text{subject to} \quad e(i,x) \geq 0 \quad \forall i,\ \sum_x e(i,x) = 1 \tag{10}$$

$$d(y,i) \geq 0 \quad \forall i,\ \sum_i d(y,i) = 1\,, \tag{11}$$

where $e(i,x)$ and $d(y,i)$ correspond to the encoder and decoder, respectively. In this case, we have the matrix $A_{xi,yi'} = p(y|x)\delta_{ii'}/k$.

Now our general aim is to take the quadratic optimization program (7)-(8) and to quantize it by allowing for non-commuting variables. That is, the variables $X_1,\dots,X_N$ are allowed to be arbitrary free variables, with no commutation relation to be assumed. We define $\mathcal{A}(N)_X = \mathcal{A}\{X_1,\dots,X_N\}$ to be the free complex algebra generated by the set $\{X_1,\dots,X_N\}$, and its elements are expressed as complex linear combination of products of arbitrary length. The algebra $\mathcal{A}_X$ caries a natural involution $* : \mathcal{A}_X \to \mathcal{A}_X$, obtained by reversing the order and complex conjugation of the linear coefficients. Furthermore, we introduce a partial order on the free algebra by saying that $A \in \mathcal{A}_X \geq 0$ if there exists an element $B \in \mathcal{A}_X$ with $A = B^*B$. Now, in order to arrive at a meaningful expression to optimize, the product of two positive numbers is replaced by a general bilinear form mapping elements of the free algebras to the complex numbers, $\omega : \mathcal{A}_X \times \mathcal{A}_X \to \mathbb{C}$. We find that the non-commutative version of the optimization problem (7)-(8) can be written as (cf. [25])

$$p^{NC}[A,\mathcal{G}] = \max \quad \left|\sum_{ij} A_{ij}\omega(X_i,X_j)\right| \tag{12}$$

$$\text{subject to} \quad \omega : \mathcal{A}_X \times \mathcal{A}_X \to \mathbb{C}\,, \tag{13}$$

$$\omega \text{ self-polar and normalized} \tag{14}$$

$$\omega \text{ is linear constrained by } g_k \in \mathcal{G} \tag{15}$$

$$\exists C:\ C\mathbb{1} \geq X_i \geq -C\mathbb{1}\,, \tag{16}$$

where self-polar is defined in [7], and we have to assume for technical reasons the existence of some constant $C > 0$ as in (16). Starting from (12)-(16) and inspired by Navascues *et al.* [25] we then develop a converging semidefinite programming hierarchy

$$p^{NC}[A,\mathcal{G}] = \mathrm{SDP}_\infty[A] \leq \ldots \leq \mathrm{SDP}_2[A] \leq \mathrm{SDP}_1[A]\,. \tag{17}$$

For randomness extractors we take the quadratic optimization $p[A,\mathcal{G}] = \mathrm{C}(\mathrm{Ext},k)$ and find that the corresponding non-commutative quadratic optimization $p^{NC}[A,\mathcal{G}] =: Q(\mathrm{Ext},k)$ exactly quantifies quantum-proof extractors as defined in (6). Moreover, the first level of the hierarchy recovers the semidefinite programming (SDP) relaxation as mentioned above: $\mathrm{SDP}_1[A] = \mathrm{SDP}(\mathrm{Ext},k)$. We emphasize that these results do not follow from [25] since the definition of self-polar forms as used in (14) gives a potentially tighter hierarchy as the one studied by Navascues *et al.* (and this was important for the properties we showed about quantum-proof extractors). For two-prover games (Bell inequalities) with classical value $p[A,\mathcal{G}]$ the non-commutative optimization $p^{NC}[A,\mathcal{G}]$ becomes the corresponding quantum value. In particular, for the channel coding example as in (9)-(11) we find that $p^{NC}[A,\mathcal{G}] = p^E_{\text{success}}$, the entanglement-assisted success probability. Interestingly, there exist channels with $p^E_{\text{success}} > p_{\text{success}}$ [12]. It would be interesting to explore this gap more systematically [16].

[5] Examples of this form include densest subgraph problems, vertex expanders, randomness condensers, etc. [7].

References

[1] A. Ben-Aroya and A. Ta-Shma. Better short-seed quantum-proof extractors. *Theoretical Computer Science*, 419:17–25, 2012.

[2] C. H. Bennett and G. Brassard. Quantum cryptography: Public key distribution and coin tossing. In *Proc. International Conference on Computers, Systems and Signal Processing*, 1984.

[3] C. H. Bennett, G. Brassard, C. Crépeau, and U. M. Maurer. Generalized privacy amplification. *Information Theory, IEEE Transactions on*, 41:1915–1923, 1995.

[4] C. H. Bennett, G. Brassard, and J.-M. Robert. Privacy amplification by public discussion. *SIAM journal on Computing*, 17:210–229, 1988.

[5] M. Berta, O. Fawzi, and V. B. Scholz. Quantum-proof randomness extractors via operator space theory. *arXiv preprint arXiv:1409.3563*, 2014.

[6] M. Berta, O. Fawzi, and V. B. Scholz. Semidefinite programming hierarchies for randomness extractors. Extended Abstract, 2015.

[7] M. Berta, O. Fawzi, and V. B. Scholz. Quantization of quadratic programs, hierachies of semidefinite programs and applications to quantum information theory. 2015 - in preparation.

[8] M. Berta, O. Fawzi, V. B. Scholz, and O. Szehr. Variations on classical and quantum extractors. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 1474–1478, 2014.

[9] K.-M. Chung, Y. Shi, and X. Wu. Physical randomness extractors. *arXiv preprint arXiv:1402.4797*, 2014.

[10] R. Colbeck and A. Kent. Private randomness expansion with untrusted devices. *Journal of Physics A: Mathematical and Theoretical*, 44:095305, 2011.

[11] M. Coudron and H. Yuen. Infinite randomness expansion and amplification with a constant number of devices. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC '14, pages 427–436. ACM, 2014.

[12] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter. Improving zero-error classical communication with entanglement. *Phys. Rev. Lett.*, 104:230503, 2010.

[13] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41:915–940, 2012.

[14] D. Gavinsky, J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf. Exponential separations for one-way quantum communication complexity, with applications to cryptography. In *Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, STOC '07, pages 516–525. ACM, 2007.

[15] J. Håstad, R. Impagliazzo, L. A. Levin, and M. Luby. A pseudorandom generator from any one-way function. *SIAM Journal on Computing*, 28:1364–1396, 1999.

[16] B. Hemenway, C. A. Miller, Y. Shi, and M. Wootters. Optimal entanglement-assisted one-shot classical communication. *Phys. Rev. A*, 87:062301, 2013.

[17] R. Konig and R. Renner. Sampling of min-entropy relative to quantum knowledge. *Information Theory, IEEE Transactions on*, 57:4760–4787, 2011.

[18] R. Konig and B. Terhal. The bounded-storage model in the presence of a quantum adversary. *Information Theory, IEEE Transactions on*, 54:749–762, 2008.

[19] R. Konig, S. Wehner, and J. Wullschleger. Unconditional security from noisy quantum storage. *Information Theory, IEEE Transactions on*, 58:1962–1984, 2012.

[20] W. Mauerer, C. Portmann, and V. B. Scholz. A modular framework for randomness extraction based on trevisan's construction. *arXiv preprint arXiv:1212.0520*, 2012.

[21] C. A. Miller and Y. Shi. Robust protocols for securely expanding randomness and distributing keys using untrusted quantum devices. *arXiv preprint arXiv:1402.0489*, 2014.

[22] N. Nisan and A. Ta-Shma. Extracting randomness: a survey and new constructions. *Journal of Computer and System Sciences*, 58:148–173, 1999.

[23] N. Nisan and D. Zuckerman. Randomness is linear in space. *Journal of Computer and System Sciences*, 51:43 – 52, 1996.

[24] S. Pironio, A. Acín, S. Massar, A. B. de La Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, et al. Random numbers certified by bell's theorem. *Nature*, 464:1021–1024, 2010.

[25] S. Pironio, M. Navascués, and A. Acín. Convergent Relaxations of Polynomial Optimization Problems with Noncommuting Variables. *Siam Journal on Optimization*, 20:2157–2180, 2010.

[26] J. Radhakrishnan and A. Ta-Shma. Bounds for dispersers, extractors, and depth-two superconcentrators. *SIAM Journal on Discrete Mathematics*, 13:2–24, 2000.

[27] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, 2005.

[28] R. Renner and R. König. Universally composable privacy amplification against quantum adversaries. In J. Kilian, editor, *Theory of Cryptography*, volume 3378 of *Lecture Notes in Computer Science*, pages 407–425. Springer Berlin Heidelberg, 2005.

[29] R. Shaltiel. Recent developments in explicit constructions of extractors. *Bulletin of the EATCS*, 77:67–95, 2002.

[30] M. Sipser. Expanders, randomness, or time versus space. *Journal of Computer and System Sciences*, 36:379 – 383, 1988.

[31] O. Szehr. Decoupling theorems. Master's thesis, ETH Zurich, 2011.

[32] A. Ta-Shma. Extractors against classical and quantum adversaries. *Tutorial QCrypt*, 2013.

[33] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner. Leftover hashing against quantum side information. *Information Theory, IEEE Transactions on*, 57:5524–5535, 2011.

[34] L. Trevisan. Construction of extractors using pseudo-random generators (extended abstract). In *Proceedings of the thirty-first annual ACM symposium on Theory of computing*, STOC '99, pages 141–148. ACM, 1999.

[35] S. P. Vadhan. *Pseudorandomness*. Now Publishers Inc., Hanover, MA, USA, 2012.

[36] U. Vazirani and T. Vidick. Certifiable quantum dice: Or, true random number generation secure against quantum adversaries. In *Proceedings of the Forty-fourth Annual ACM Symposium on Theory of Computing*, STOC '12, pages 61–76, New York, NY, USA, 2012. ACM.

Semidefinite programming hierarchies for randomness extractors

Mario Berta,[1] Omar Fawzi,[2,3] and Volkher B. Scholz[4]

[1]*Institute for Quantum Information and Matter, Caltech, Pasadena, CA 91125, USA*
[2]*Department of Computing and Mathematical Sciences, Caltech, Pasadena, CA 91125, USA*
[3]*ENS de Lyon, UMR 5668 LIP - CNRS - UCBL - INRIA - Université de Lyon, 69364 Lyon, France*
[4]*Institute for Theoretical Physics, ETH Zurich, 8093 Zürich, Switzerland*

Extended abstract for the QCrypt submission *Semidefinite programming hierarchies for quantum adversaries* by the same authors.

I. PRELIMINARIES

A. Quantum information

In quantum theory, a system is described by an inner-product space, that we denote here by letters like N, M, Q.[1] Note that we use the same symbol Q to label the system, the corresponding inner-product space and also the dimension of the space. Let $\mathrm{Mat}_Q(S)$ be the vector space of $Q \times Q$ matrices with entries in S. Whenever S is not specified, it is assumed to be the set of complex numbers $\mathbb{C}$, i.e., we write $\mathrm{Mat}_Q(\mathbb{C}) =: \mathrm{Mat}_Q$. The state of a system is defined by a positive semidefinite operator ρ_Q with trace 1 acting on Q. The set of states on system Q is denoted by $\mathcal{S}(Q) \subset \mathrm{Mat}_Q(\mathbb{C})$. The inner-product space of a composite system QN is given by the tensor product of the inner-product spaces $Q \otimes N =: QN$. From a joint state $\rho_{QN} \in \mathcal{S}(QN)$, we can obtain marginals on the system Q by performing a partial trace of the N system $\rho_Q := \mathrm{Tr}_N[\rho_{QN}]$. The state ρ_{QN} of a system QN is called quantum-classical (with respect to some basis) if it can be written as $\rho_{QN} = \sum_x \rho(x) \otimes |x\rangle\langle x|$ for some basis $\{|x\rangle\}$ of N and some positive semidefinite operators $\rho(x)$ acting on Q with $\sum_x \mathrm{Tr}[\rho(x)] = 1$. We denote the maximally mixed state on system N by υ_N.

To measure the distance between two states, we use the trace norm $\|A\|_1 := \mathrm{Tr}[\sqrt{A^* A}]$, where A^* is the conjugate transpose of A. In the special case when A is diagonal, $\|A\|_1$ becomes the familiar ℓ_1 norm of the diagonal entries. Moreover, the Hilbert-Schmidt norm is defined as $\|A\|_2 := \sqrt{\mathrm{Tr}[A^* A]}$, and when A is diagonal this becomes the usual ℓ_2 norm. Another important norm we use is the operator norm, or the largest singular value of A, denoted by $\|A\|_\infty$. When A is diagonal, this corresponds to the familiar ℓ_∞ norm of the diagonal entries. For a probability distribution P_N on the set N, $\|P_N\|_{\ell_\infty}$ corresponds to the optimal probability with which P_N can be guessed successfully. We write

$$H_{\min}(N)_P := -\log \|P_N\|_{\ell_\infty} \, , \tag{1}$$

the min-entropy of P_N. More generally, the conditional min-entropy of N given Q is used to quantify the uncertainty in the system N given the system Q. The conditional min-entropy is defined as

$$H_{\min}(N|Q)_\rho := -\log \min_{\sigma_Q \in \mathcal{S}(Q)} \left\| (\mathbb{1}_N \otimes \sigma_Q^{-1/2}) \rho_{NQ} (\mathbb{1}_N \otimes \sigma_Q^{-1/2}) \right\|_\infty \, , \tag{2}$$

with generalized inverses. Note that in the special case where the system Q is trivial, we have $H_{\min}(N)_\rho = -\log \|\rho_N\|_\infty$.

B. Semidefinite programming

Semidefinite programs (SDP) are a large class of optimization problems that can be efficiently solved. Even if one is not explicitly interested in solving it numerically, a semidefinite program often has appealing properties such as strong duality. Semidefinite programming has been extensively used in various contexts in quantum information.

We use a formulation of semidefinite programs sometimes called vector programs. For some fixed values $\alpha_{x,x'}$, $\beta_{x,x',k}$ and γ_k, the optimization program can be written as follows:

---

[1] In the following all spaces are assumed to be finite-dimensional.

$$\text{maximize} \quad \sum_{x,x'} \alpha_{x,x'} \vec{a}_x \cdot \vec{a}_{x'} \tag{3}$$

$$\text{subject to} \quad \sum_{x,x'} \beta_{x,x',k} \vec{a}_x \cdot \vec{a}_{x'} \leq \gamma_k \quad \text{for all } k \tag{4}$$

Here the optimization is over all vector $\vec{a}_x$ (of arbitrary finite dimension) that satisfy the constraints stated above. Note that we can always assume that the dimension of the vectors $\vec{a}_x$ is bounded by the number of vectors, i.e., the size of the set x runs over.

II. QUADRATIC PROGRAMS FOR RANDOMNESS EXTRACTORS

It is useful to see the definition of extractors using the following optimization program:

$$\underline{\text{Error for extractor Ext} = \{f_s\}}$$

$$\mathrm{C}(\mathrm{Ext}, k) := \text{maximize} \quad \frac{1}{D} \sum_{s,y} \sum_{x} \left(\delta_{f_s(x)=y} - \frac{1}{M} \right) p(x) \beta_{s,y} \tag{5}$$

$$\text{subject to} \quad 0 \leq p(x) \leq 2^{-k} \tag{6}$$

$$\sum_x p(x) = 1 \tag{7}$$

$$-1 \leq \beta_{s,y} \leq 1 \tag{8}$$

Definition II.1. Ext *is a* (k, ε)*-extractor if and only if* $\mathrm{C}(\mathrm{Ext}, k) \leq \varepsilon$.

To relate this to the definition given in the introduction, it suffices to observe that the optimal choice for $\beta_{s,y}$ is the sign of $\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) p(x)$ so the objective function becomes $\frac{1}{D} \sum_{s,y} \left|\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) p(x)\right|$. The conditions (6) and (7) ensure that the input distribution has min-entropy at least k.

To simplify the program (5) we note that this function is convex in the distribution p and so the maximum is attained in the extreme points of the feasible region. These are simply the distributions that are uniform over a set of size at least 2^k. So we can equivalently write

$$\mathrm{C}(\mathrm{Ext}, k) = \max \left\{ \sum_{s,y} \left| \frac{1}{KD} \sum_{x \in L} \delta_{f_s(x)=y} - \frac{1}{MD} \right| : L \subseteq N, L \geq 2^k \right\} , \tag{9}$$

where again in a slight abuse of notation, we use the letter L for the actual set as well as its size. As the expression being maximized is the ℓ_1 norm between two probability distributions, we can write it as:

$$\mathrm{C}(\mathrm{Ext}, k) = 2 \cdot \max \left\{ \frac{1}{KD} \sum_{x \in L, (y,s) \in R} \delta_{f_s(x)=y} - \frac{R}{MD} : L \subseteq N, L \geq 2^k, R \subseteq M \times D \right\} . \tag{10}$$

This allows us to interpret $\mathrm{C}(\mathrm{Ext}, k)$ in graph-theoretic terms. For that we introduce a bipartite graph with left vertex set N and right vertex set $M \times D$, and there is an edge between vertices x and (y, s) if and only if $f_s(x) = y$. By writing $E(L, R)$ for the set of edges with one endpoint in L and the other endpoint in R, this expression is simply

$$\mathrm{C}(\mathrm{Ext}, k) = 2 \cdot \max \left\{ \frac{E(L, R)}{2^k D} - \frac{R}{MD} : L \subseteq N, L \geq 2^k, R \subseteq M \times D \right\} . \tag{11}$$

Written in this way, we see that the optimization in $\mathrm{C}(\mathrm{Ext}, k)$ is a kind of bipartite densest subgraph problem. Algorithms for a slightly different problem known as the densest K-subgraph problem have been extensively studied, see e.g., [6, 12]. The best known approximation algorithms for this problem achieve a factor of N^α for some constant α, but even ruling out constant factor approximations is only known using quite strong assumptions [1].

We can similarly write a program for the error of Ext against potentially quantum adversaries:

Error for extractor $\mathrm{Ext} = \{f_s\}$ against quantum adversaries

$$\mathrm{Q}(\mathrm{Ext}, k) := \text{maximize} \quad \frac{1}{D} \sum_{s,y} \sum_{x} \left(\delta_{f_s(x)=y} - \frac{1}{M} \right) \mathrm{Tr}\left[\rho(x) B_{s,y}\right] \tag{12}$$

$$\text{subject to} \quad 0 \leq \rho(x) \leq 2^{-k}\sigma \tag{13}$$

$$\sum_x \mathrm{Tr}[\rho(x)] = 1 \tag{14}$$

$$\mathrm{Tr}[\sigma] = 1 \tag{15}$$

$$\|B_{s,y}\|_\infty \leq 1 \tag{16}$$

Here the maximization is understood over all $\rho(x)$ of arbitrary dimension. Unlike for SDPs for which one can give an upper bound on the dimension of the vector of an optimal solution, no such bound is know in this setting. In fact, we do not even know if the quantity Q is computable.

Definition II.2. Ext *is a quantum-proof* (k, ε)*-extractor if and only if* $\mathrm{Q}(\mathrm{Ext}, k) \leq \varepsilon$.

To see that this definition coincides with the definition given in the introduction, observe that for fixed $\rho(x)$, the maximum over $B_{s,y}$ of the quantity $\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) \mathrm{Tr}\left[\rho(x) B_{s,y}\right]$ is $\| \sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) \rho(x) \|_1$. The constraints on $\rho(x)$ and σ ensure that the state $\sum_x \rho(x) \otimes |x\rangle\langle x|$ has conditional min-entropy at least k.

III. SEMIDEFINITE RELAXATIONS FOR RANDOMNESS EXTRACTORS

A. A relaxation for the extractor condition

Motivated by the fact that the two quantities $\mathrm{C}(\mathrm{Ext}, k)$ and $\mathrm{Q}(\mathrm{Ext}, k)$ are generally difficult to understand, we introduce a SDP that, as we show later, provides a relaxation for both of these quantities. For $\mathrm{Ext} = \{f_s\}_{s \in D}$ and fixed k, we define:

SDP relaxation for error of $\mathrm{Ext} = \{f_s\}$

$$\mathrm{SDP}(\mathrm{Ext}, k) := \text{maximize} \quad \frac{1}{D} \sum_{s,y,x} \left(\delta_{f_s(x)=y} - \frac{1}{M} \right) \vec{a}_x \cdot \vec{b}_{s,y} \tag{17}$$

$$\text{subject to} \quad 0 \leq \vec{a}_x \cdot \vec{a}_{x'} \leq 2^{-k} \cdot q(x) \tag{18}$$

$$q(x) \leq 2^{-k} \tag{19}$$

$$\sum_x q(x) = 1 \tag{20}$$

$$\sum_{x,x'} \vec{a}_x \cdot \vec{a}_{x'} \leq 1 \tag{21}$$

$$\|b_{s,y}\|_2 \leq 1 \tag{22}$$

We maximize over all possible dimensions of the vectors $\vec{a}_x$ and $\vec{b}_x$. Moreover, the Cauchy-Schwarz inequality implies that the optimal choice for $\vec{b}_{s,y}$ is

$$\frac{\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) \vec{a}_x}{\| \sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) \vec{a}_x \|_2} , \tag{23}$$

and thus the objective function of the SDP relaxation becomes

$$\frac{1}{D} \sum_{s,y} \left\| \sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M} \right) \vec{a}_x \right\|_2 , \tag{24}$$

subject to the constraints on the vectors $\vec{a}_x$ stated in (17). By simply plugging $\vec{a}_x = p(x)$, $q(x) = p(x)$ and $\vec{b}_{s,y} = \beta_{s,y}$, we see that this SDP gives an upper bound on the extractor program (5).

Proposition III.1. *For any* Ext *and* k*,* $\mathrm{C}(\mathrm{Ext}, k) \leq \mathrm{SDP}(\mathrm{Ext}, k)$*. In other words, if* $\mathrm{SDP}(\mathrm{Ext}, k) \leq \varepsilon$*, then* Ext *is a* (k, ε)*-extractor.*

This gives a computationally efficient criterion for certifying that an extractor is good. As we show in Section III C, this method can certify that many important constructions are good extractors. However, this technique does in general not give a tight characterization of extractors and there can be a large gap between the values $\mathrm{C}(\mathrm{Ext}, k)$ and $\mathrm{SDP}(\mathrm{Ext}, k)$ as we will see in Section III D.

B. A relaxation for the error against quantum adversaries

A very interesting property about the SDP (17) is that it also gives an upper bound on the error of an extractor against quantum adversaries. This means that if an extractor satisfies the stronger property $\mathrm{SDP}(\mathrm{Ext}, k) \leq \varepsilon$ then it is not only a (k, ε)-extractor but also a quantum proof $(k, \sqrt{2}\varepsilon)$-extractor.

Theorem III.2. *For any* Ext *and* k*, we have*

$$\mathrm{C}(\mathrm{Ext}, k) \leq \mathrm{Q}(\mathrm{Ext}, k) \leq \sqrt{2} \cdot \mathrm{SDP}(\mathrm{Ext}, k) \ . \tag{25}$$

Proof. Let $\rho = \sum_x \rho(x) \otimes |x\rangle\langle x|$ be a quantum state on QN with $H_{\min}(N|Q)_\rho \geq k$. By the definition of the conditional min-entropy, this implies that there exists $\sigma \in \mathcal{S}(Q)$ such that $\rho(x) \leq 2^{-k}\sigma$ for all $x \in N$. We now define the average state $\bar{\rho} = \sum_x \rho(x)$ and $\omega = \frac{\bar{\rho}+\sigma}{2}$, as well as the vectors $\vec{a}_x$ as the list of entries of the matrix $\frac{1}{\sqrt{2}}\omega^{-1/4}\rho(x)\omega^{-1/4}$. This is so that we have $\vec{a}_x \cdot \vec{a}_{x'} = \frac{1}{2}\mathrm{Tr}[\omega^{-1/2}\rho(x)\omega^{-1/2}\rho(x')]$. As the trace of the product of two positive semidefinite operators is nonnegative, we have $\vec{a}_x \cdot \vec{a}_{x'} \geq 0$. Moreover, we have

$$\vec{a}_x \cdot \vec{a}_{x'} = \frac{1}{2}\mathrm{Tr}[\omega^{-1/2}\rho(x)\omega^{-1/2}\rho(x')] \leq \frac{1}{2}\mathrm{Tr}[\omega^{-1/2}\rho(x)\omega^{-1/2}2^{-k}\sigma] \tag{26}$$

$$\leq \frac{1}{2} \cdot 2^{-k}\mathrm{Tr}[\omega^{-1/2}\rho(x)\omega^{-1/2}2\omega] \leq 2^{-k}\mathrm{Tr}[\rho(x)] \ . \tag{27}$$

We set $q(x) = \mathrm{Tr}[\rho(x)]$. Note that we have $q(x) = \mathrm{Tr}[\rho(x)] \leq 2^{-k}\mathrm{Tr}[\sigma] = 2^{-k}$ and $\sum_x q(x) \leq 1$. We can also write

$$\sum_{x,x'} \vec{a}_x \cdot \vec{a}_{x'} = \frac{1}{2}\mathrm{Tr}[\omega^{-1/2}\bar{\rho}\omega^{-1/2}\bar{\rho}] \leq \frac{1}{2}\mathrm{Tr}[\omega^{-1/2}\bar{\rho}\omega^{-1/2}2\omega] \leq 1 \ . \tag{28}$$

We now analyze the objective function. We use the following Hölder-type inequality for operators $\|\alpha\beta\gamma\|_1 \leq \||\alpha|^4\|_1^{1/4}\||\beta|^2\|_1^{1/2}\||\gamma|^4\|_1^{1/4}$, see e.g., [7, Corollary IV.2.6]. The error the extractor makes on input ρ is given by

$$\frac{1}{D}\sum_{s,y}\left\|\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right)\rho(x)\right\|_1$$

$$\leq \frac{1}{D}\sum_{s,y}\|\omega\|_1^{1/4}\left\|\left(\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right)\omega^{-1/4}\rho(x)\omega^{-1/4}\right)^2\right\|_1^{1/2}\|\omega\|_1^{1/4} \tag{29}$$

$$= \frac{1}{D}\sum_{s,y}\sqrt{\mathrm{Tr}\left[\sum_{x,x'}\left(\delta_{f_s(x)=y} - \frac{1}{M}\right)\left(\delta_{f_s(x')=y} - \frac{1}{M}\right)\omega^{-1/2}\rho(x)\omega^{-1/2}\rho(x')\right]} \tag{30}$$

$$= \frac{1}{D}\sum_{s,y}\sqrt{\sum_{x,x'}\left(\delta_{f_s(x)=y} - \frac{1}{M}\right)\left(\delta_{f_s(x')=y} - \frac{1}{M}\right)2 \cdot \vec{a}_x \cdot \vec{a}_{x'}} \tag{31}$$

$$= \frac{\sqrt{2}}{D}\sum_{s,y}\left\|\sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M}\right)\vec{a}_x\right\|_2 \ . \tag{32}$$

This proves that the error the extractor makes in the presence of quantum adversaries is upper bounded by $\sqrt{2} \cdot \mathrm{SDP}(\mathrm{Ext}, k)$. □

C. Applications

We now give several applications of the SDP relaxation. We show that many results about quantum-proof extractors can be shown with the SDP quantity. First, let us consider general results that do not use the structure of the functions in Ext but simply the extractor's parameters. We know the advantage obtained by a quantum adversary compared to a classical one can by bounded by a function of the number of output bits m or the min-entropy deficit $n-k$ [3] (for $m=1$ this was first shown in [15]). In particular, if m or $n-k$ are small, then the quantum advantage cannot be large. We show that this is actually a property of the SDP.

Theorem III.3. *For any* Ext *and* k*, we have for any* $\varepsilon > 0$*,*

$$\mathrm{SDP}\big(\mathrm{Ext}, k+\log(1/\varepsilon)\big) \leq \sqrt{2^m}\sqrt{\mathrm{C}(\mathrm{Ext},k)+\varepsilon} \tag{33}$$

$$\mathrm{SDP}(\mathrm{Ext}, k) \leq 3K_G 2^{n-k}\mathrm{C}(\mathrm{Ext}, k-1)\ , \tag{34}$$

where $K_G \leq 1.8$ *is Grothendieck's constant.*

Proof. As Ext is usually clear from the context, we use $\mathrm{C}(k)$ and $\mathrm{SDP}(k)$ for $\mathrm{C}(\mathrm{Ext},k)$ and $\mathrm{SDP}(\mathrm{Ext},k)$. To prove (33), we consider an optimal solution for $\mathrm{SDP}(k+\log(1/\varepsilon))$. Define $p(x,x') = \vec{a}_x \cdot \vec{a}_{x'}$, with $\bar{p}(x) = \sum_{x'} p(x,x')$. Now consider the set $S_\varepsilon = \{x \in N : \bar{p}(x) \leq \varepsilon q(x)\}$. Then $\sum_{x\in S_\varepsilon} \bar{p}(x) \leq \varepsilon \sum_{x \in S_\varepsilon} q(x) \leq \varepsilon$. Using the fact that $\vec{a}_x$ define a feasible solution for $\mathrm{SDP}(k+\log(1/\varepsilon))$, we have for $x \notin S_\varepsilon$, $p(x,x') \leq 2^{-(k+\log(1/\varepsilon))}q(x) \leq 2^{-k}\bar{p}(x)$. We can then write using the Cauchy Schwarz inequality,

$$\frac{1}{D}\sum_{s,y}\left\|\sum_x \left(\delta_{f_s(x)=y} - 2^{-m}\right)\vec{a}_x\right\|_2 \leq \sqrt{\frac{1}{D}\sum_{s,y}\left\|\sum_x \left(\delta_{f_s(x)=y} - 2^{-m}\right)\vec{a}_x\right\|_2^2}\sqrt{2^m}\ . \tag{35}$$

We now look at the expression $\frac{1}{D}\sum_{s,y}\left\|\sum_x \left(\delta_{f_s(x)=y} - 2^{-m}\right)\vec{a}_x\right\|_2^2$ which equals

$$\begin{aligned}
&\frac{1}{D}\sum_{s,y}\sum_{x,x'}\left(\delta_{f_s(x)=y} - 2^{-m}\right)\cdot\left(\delta_{f_s(x')=y} - 2^{-m}\right)p(x,x') && (36)\\
&\leq \frac{1}{D}\sum_{s,y}\sum_{x}\left|\sum_{x'}\left(\delta_{f_s(x)=y} - 2^{-m}\right)\cdot\left(\delta_{f_s(x')=y} - 2^{-m}\right)p(x,x')\right| && (37)\\
&\leq \frac{1}{D}\sum_{s,y}\sum_{x}\left|\sum_{x'}\left(\delta_{f_s(x')=y} - 2^{-m}\right)p(x,x')\right|\ . && (38)
\end{aligned}$$

We separate the sum into $x \in S_\varepsilon$ and $x \notin S_\varepsilon$ and get

$$\begin{aligned}
&\frac{1}{D}\sum_{s,y}\sum_{x}\left|\sum_{x'}\left(\delta_{f_s(x')=y} - 2^{-m}\right)p(x,x')\right| && (39)\\
&= \frac{1}{D}\sum_{s,y}\sum_{x}\bar{p}(x)\left|\sum_{x'}\left(\delta_{f_s(x')=y} - 2^{-m}\right)\frac{p(x,x')}{\bar{p}(x)}\right| && (40)\\
&= \sum_{x\in S_\varepsilon}\bar{p}(x)\frac{1}{D}\sum_{s,y}\left|\sum_{x'}\left(\delta_{f_s(x')=y} - 2^{-m}\right)\frac{p(x,x')}{\bar{p}(x)}\right| && (41)\\
&\quad + \sum_{x\notin S_\varepsilon}\bar{p}(x)\frac{1}{D}\sum_{s,y}\left|\sum_{x'}\left(\delta_{f_s(x')=y} - 2^{-m}\right)\frac{p(x,x')}{\bar{p}(x)}\right| \leq \varepsilon + \mathrm{C}(k)\ , && (42)
\end{aligned}$$

which proves (33).

We now prove the inequality (34). For that, we simply upper bound $\mathrm{SDP}(\mathrm{Ext},k)$ by forgetting several constraints and then apply Grothendieck's inequality (Theorem A.1). Observe first that for any feasible vectors $\vec{a}_x$ for the SDP, we have $\|\vec{a}_x\|_2^2 \leq 2^{-k} q(x) \leq 2^{-2k}$.

$$\mathrm{SDP}(\mathrm{Ext}, k) \leq \max \left\{ \frac{1}{D} \sum_{s,y,x} \left(\delta_{f_s(x)=y} - 2^{-m} \right) \vec{a}_x \cdot \vec{b}_{s,y} : \|\vec{a}_x\|_2 \leq 2^{-k}, \|\vec{b}_{s,y}\|_2 \leq 1 \right\} \tag{43}$$

$$\leq K_G \max \left\{ \frac{1}{D} \sum_{s,y,x} \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x b_{s,y} : |a_x| \leq 2^{-k}, |b_{s,y}| \leq 1 \right\} \tag{44}$$

$$= K_G \max \left\{ \frac{1}{D} \sum_{s,y} \left| \sum_x \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x \right| : |a_x| \leq 2^{-k} \right\} . \tag{45}$$

We partition the set of $x \in N$ into $\{x : a_x \geq 0\}$ and $\{x : a_x < 0\}$ and write

$$\left| \sum_x \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x \right| \leq \left| \sum_{x:a_x \geq 0} \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x \right| \tag{46}$$

$$+ \left| \sum_{x:a_x < 0} \left(\delta_{f_s(x)=y} - 2^{-m} \right) (-a_x) \right| . \tag{47}$$

Let us write $\alpha_+ := \sum_{x:a_x \geq 0} a_x$. If $\alpha_+ \geq 1$, then we define $p_+(x) = \frac{\max\{a_x, 0\}}{\alpha_+}$. Observing that $\alpha_+ \leq 2^{n-k}$, we have

$$\frac{1}{D} \sum_{s,y} \left| \sum_{x:a_x \geq 0} \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x \right| = \alpha_+ \cdot \frac{1}{D} \sum_{s,y} \left| \sum_x \left(\delta_{f_s(x)=y} - 2^{-m} \right) p_+(x) \right| \tag{48}$$

$$\leq \alpha_+ \mathrm{C}(k + \log(\alpha_+)) \leq 2^{n-k} \mathrm{C}(k) , \tag{49}$$

where we have used the abbreviation $\mathrm{C}(k) = \mathrm{C}(\mathrm{Ext}, k)$. Otherwise (if $\alpha_+ < 1$), we define $p_+(x) = \max\{a_x, 0\} + (1 - \alpha_+)2^{-n}$. We get

$$\frac{1}{D} \sum_{s,y} \left| \sum_{x:a_x \geq 0} \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x \right| \tag{50}$$

$$= \frac{1}{D} \sum_{s,y} \left| \sum_x \left(\delta_{f_s(x)=y} - 2^{-m} \right) (p_+(x) - (1 - \alpha_+)2^{-n}) \right| \tag{51}$$

$$\leq \frac{1}{D} \sum_{s,y} \left| \sum_x \left(\delta_{f_s(x)=y} - 2^{-m} \right) p_+(x) \right| + (1 - \alpha_+) \frac{1}{D} \sum_{s,y} \left| \sum_x \left(\delta_{f_s(x)=y} - \frac{1}{M} \right) 2^{-n} \right| \tag{52}$$

$$\leq \mathrm{C}(k-1) + (1 - \alpha_+)\mathrm{C}(n) . \tag{53}$$

With a similar argument for the set $\{x : a_x < 0\}$, we reach the bound

$$\frac{1}{D} \sum_{s,y} \left| \sum_x \left(\delta_{f_s(x)=y} - 2^{-m} \right) a_x \right| \tag{54}$$

$$\leq \max\{2 \cdot 2^{n-k} \mathrm{C}(k), \mathrm{C}(k-1) + \mathrm{C}(n) \tag{55}$$

$$+ 2^{n-k}\mathrm{C}(k), 2\mathrm{C}(k-1) + (1 - \alpha_+ - \alpha_-)\mathrm{C}(n)\} \leq 3 \cdot 2^{n-k} \mathrm{C}(k-1) . \tag{56}$$

Finally, we get $\mathrm{SDP}(k) \leq 3K_G 2^{n-k} \mathrm{C}(k-1)$. □

Some specific constructions are also known to be quantum-proof, in particular constructions based on two-universal hash functions [20–22]. This type of construction is captured by spectral extractors [5]. For an extractor $\mathrm{Ext} = \{f_s\}_{s \in D}$ we define the linear maps [Ext] and τ that map vectors of dimension N to vectors of dimension DM as follows:

$$[\mathrm{Ext}] \left(\sum_x p(x) |x\rangle\langle x|_N \right) = \frac{1}{D} \cdot \sum_{s,y} \sum_x \delta_{f_s(x)=y} p(x) |y\rangle\langle y|_M \otimes |s\rangle\langle s|_D \tag{57}$$

$$\tau \left(\sum_x p(x) |x\rangle\langle x|_N \right) = \left(\sum_x p(x) \right) \upsilon_M \otimes \upsilon_D . \tag{58}$$

Note that we used a quantum notation and identified vectors with diagonal matrices. A spectral (k,ε)-extractor is then defined via the largest eigenvalue bound

$$\lambda_1\Big([\mathrm{Ext}]^* \cdot [\mathrm{Ext}] - \tau^* \cdot \tau\Big) \leq 2^{k-m-d}\varepsilon \ , \tag{59}$$

where $*$ refers to the adjoint of a linear map. We prove next that for spectral extractor, there can be at most a quadratic gap between $\mathrm{C}(\mathrm{Ext},k)$ and $\mathrm{SDP}(\mathrm{Ext},k)$.

Theorem III.4. *Let* $\mathrm{Ext}_{\mathrm{spec}} = \{f_s\}_{s\in D}$ *be a spectral* (k,ε)*-extractor as defined in* (59)*. Then, we have*

$$\mathrm{SDP}(\mathrm{Ext}_{\mathrm{spec}},k) \leq \sqrt{\varepsilon} \ . \tag{60}$$

The proof can be found in Appendix B. Another class of extractors that are quantum-proof are Trevisan based constructions [2, 10]. These are particularly important to understand because they are the only known quantum-proof constructions with short seed $d = O(\mathrm{poly}(\log n))$. Trevisan's construction can be thought of as a composition of one-bit output extractors cleverly interleaved by slightly reusing the seed. Specifically, the construction is based on a family of subsets $S_1,\ldots,S_m \subset \{1,\ldots,d\}$ such that for each i we have

$$|S_i| = l \quad \text{and} \quad \sum_{j<i} 2^{|S_i\cap S_j|} \leq r(m-1) \ , \tag{61}$$

for some $r>0$. Such a family $\{S_i\}_{i\in\{1,\ldots,m\}}$ is also called weak (l,r)-design. Now, take a one-bit output extractor $\mathrm{Ext}_{\mathrm{one}} = \{g_t\}_{t\in\{0,1\}^l}$ with $g_t : N \to \{0,1\}$, and a weak (l,r)-design as defined in (61). Trevisan then defines a m-bit output extractor

$$\mathrm{Ext}_{\mathrm{Trev}} = \{f_s\}_{s\in D} \quad \text{with} \quad f_s : N \to M \tag{62}$$

$$f_s(x) := g_{s|S_1}(x) \circ g_{s|S_1}(x) \circ \cdots \circ g_{s|S_m}(x) \ , \tag{63}$$

where $s|S_i$ denotes the l-bits of s that correspond to the position indexed by the set S_i, and $\circ$ means concatenation.[2] The basic idea of the proof is to bound the quality of $\mathrm{Ext}_{\mathrm{Trev}}$ as a function of the quality of $\mathrm{Ext}_{\mathrm{one}}$. Then (using Theorem III.3) one can relate the quality of $\mathrm{Ext}_{\mathrm{one}}$ against quantum adversaries to its quality against classical adversaries. We give (in the Appendix) a concise proof of this result using our notation in terms of the quantum program (12).

Theorem III.5. *Let* $\{S_i\}_{i\in\{1,\ldots,m\}}$ *be a weak* (l,r)*-design as defined in* (61)*, and* $\mathrm{Ext}_{\mathrm{one}} = \{g_t\}_{t\in\{0,1\}^l}$ *be a one-bit output extractor. Then, we have for Trevisan's extractor* $\mathrm{Ext}_{\mathrm{Trev}} = \{f_s\}_{s\in D}$ *as defined in* (62)–(63)*,*

$$\mathrm{Q}(\mathrm{Ext}_{\mathrm{Trev}},k) \leq m\cdot \mathrm{Q}(\mathrm{Ext}_{\mathrm{one}},k-r(m-1)) \tag{64}$$

$$\leq 2m\cdot\sqrt{\mathrm{C}(\mathrm{Ext}_{\mathrm{one}},k-r(m-1)-\log(1/\varepsilon))+\varepsilon} \ , \tag{65}$$

for any $\varepsilon > 0$.

D. Gap between C and SDP

In this section, we show that there can be a large gap between the value C and SDP. In fact, we show that SDP cannot be used to prove that randomly chosen functions are good randomness extractors. Random functions are good extractors with essentially optimal parameters. In other words, for a family of functions $\mathrm{Ext}_{\mathrm{rand}} = \{f_s\}_{s\in D}$ chosen at random, we have with very high probability that

$$\mathrm{C}(\mathrm{Ext}_{\mathrm{rand}},k) \leq \varepsilon \quad \text{for} \quad m = k - 2\log(1/\varepsilon) - O(1) \tag{66}$$

$$d = \log(n-k) + 2\log(1/\varepsilon) + O(1) \ . \tag{67}$$

In contrast to this, we find that the SDP relaxation for random constructions can become very large for sufficiently small min-entropy k.

[2] Actual parameters for Trevisan based extractor constructions are, e.g, discussed in detail in [10, Section 5].

Theorem III.6. *Let* $\mathrm{Ext} = \{f_s\}_{s\in D}$ *be a family of functions such that*

$$\gamma_1 \frac{DN^2}{M} \leq \sum_{x,x',s} \delta_{f_s(x)=f_s(x')} \leq \gamma_2 \frac{DN^2}{M} \ , \tag{68}$$

and $k \leq \log\left(\gamma_1 \frac{N}{M}\right)$. *Then, we have*

$$\mathrm{SDP}(\mathrm{Ext}, k) \geq \frac{1}{2}\sqrt{\frac{M}{\gamma_2 D}} \ . \tag{69}$$

When the functions f_s are chosen at random, then the condition (68) is satisfied with very high probability for constant values of γ_1 and γ_2 (see Proposition B.1 for a proof). Hence, we find that for instance if $k = n/2$, $m = n/4$ and $d = O(\log n)$, with high probability $\mathrm{SDP}(\mathrm{Ext}_{\mathrm{rand}}, k) \gg 2$, whereas we have with very high probability $\mathrm{C}(\mathrm{Ext}_{\mathrm{rand}}, k) \leq \frac{1}{n}$. As clearly $\mathrm{Q}(\mathrm{Ext}, k) \leq 2$, this also shows that Q can be much smaller than SDP.

Moreover we can show that for Trevisan's extractor, we cannot replace $\mathrm{Q}(\mathrm{Ext}_{\mathrm{Trev}})$ with $\mathrm{SDP}(\mathrm{Ext}_{\mathrm{Trev}}, k)$ in general in Theorem III.5. This is because if the one-bit extractors $\{g_t\}$ in Trevisan's construction are chosen at random, then it is possible to show that the condition (68) is satisfied with high probability for constant values of γ_1 and γ_2 (see Proposition B.1 for a proof).

Proof of Theorem III.6. Use $\vec{a}_x = \alpha^{-1/2} \cdot \sum_{s,y} \delta_{f_s(x)=y} |s\rangle|y\rangle$, $\alpha = \sum_{x,x'} \sum_{s,y} \delta_{f_s(x)=y} \delta_{f_s(x')=y}$. By definition the normalization condition $\sum_{x,x'} \vec{a}_x \cdot \vec{a}_{x'} \leq 1$ is satisfied. Moreover, for any fixed x, x', we have

$$\vec{a}_x \cdot \vec{a}_{x'} = \frac{1}{\alpha} \sum_{s,y} \delta_{f_s(x)=y} \delta_{f_s(x')=y} \leq \frac{D}{\alpha} \leq \frac{1}{\gamma_1} \frac{M}{N^2} \leq \frac{1}{\gamma_1} \frac{M}{N} q(x) \ , \tag{70}$$

where we used the lower bound on γ_1 and we choose $q(x) = 1/N$. Now if $k \leq \log\left(\gamma_1 \frac{N}{M}\right)$, the min-entropy condition for the vectors is satisfied. Now let us analyze the objective function by choosing $\vec{b}_{s,y} = |s\rangle|y\rangle$. We find

$$\frac{1}{D} \sum_{s,y} \sum_{x} \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) \vec{a}_x \cdot \vec{b}_{s,y} = \frac{1}{D} \sum_{s,y} \sum_{x} \left(\delta_{f_s(x)=y} - \frac{1}{M}\right) \alpha^{-1/2} \delta_{f_s(x)=y} \tag{71}$$

$$= \frac{1}{D\alpha^{1/2}} \sum_{s,x} \left(1 - \frac{1}{M}\right) = \frac{N}{\alpha^{1/2}} \left(1 - \frac{1}{M}\right) \geq \frac{1}{2}\sqrt{\frac{M}{\gamma_2 D}} \ , \tag{72}$$

which proves the claim. □

IV. HIERARCHY OF SDPS FOR Q

Given the SDP relaxation for extractors against quantum adversaries (Theorem III.2), it is natural to ask whether we can add positive definite conditions such that the upper bound improves or even becomes equal to the it. Note that a similar question has been studied in the case of two-player games, also called bipartite Bell inequalities in the physics literature (see the review article [8]). Here, the task to bound the entangled value of the game. Again, this value can be upper bounded in terms of an SDP, and the goal is to add more and more constraints to ensure a better and better upper bound (see [11, 17] for two complementary approaches in this direction). It turns out that similar thing can be done for extractors and that even a unified discussion is possible. We refer to the full version in preparation [4] for further details and only say a few words and state the levels of the SDP hierarchy (in order motivate further studies in this direction).

Our approach is motivated from the construction of Lasserre's Sum-of-Squares hierarchies for constrained optimization problems [16]. Given the extractor program (5), which is itself of such a form, this is not surprising. Assuming that the constraints single out a closed convex set, the underlying idea is to construct a positive measure supported on this set while simultaneously maximizing the expectation value of the objective function with respect to this measure. Since the objective value is thus linear in the measure, the optimum is attained at a point measure defining the optimal solution. In order to construct such measures, we have to define its moments, i.e., expectation values of monomials. It is convenient to arrange them into matrix form, and this matrix (indexed by monomials) then has to be positive semidefinite. Thus the task is reduced to constructing positive semidefinite matrices, further satisfying constraints imposed by the structure of the convex set - which again can be phrased as positive semidefinite constraints.

A fixed level n is then defined by only considering the constraints originating from monomials of power at most n. The idea for the quantum-proof extractor program (12) that we want to upper bound is quite similar, but rather than considering commuting variables, the "measure" now has to define a functional on non-commuting operators - or more precisely a state on an operator algebra. The general idea is very alike the non-commutative polynomial optimization techniques by Navascues [19], however for extractors a few more constraints have to be added (again, we refer to the full version for details). In comparison to the classical case, though, these matrices are now indexed by non-commuting monomials, or equivalently, by elements of the free algebra spanned by the variables in the objective function. Expressing a feasible matrix of the n^{th}-level of the hierarchy as a Gram matrix then leads to vectors c_w indexed by "words" $w = (i_1, i_2, \ldots, i_n)$ - strings of indices corresponding to variables given by the input x and output $\tilde{y} := (s, y)$ of length at most n. We denote the set of such words by S_n. To simplify notation, it is useful to introduce for a fixed pair of indices (i, j) the matrices

$$C[(i),(j)] := \sum_{r,s,k,l \in S_n} \vec{c}_{r^* \circ (i) \circ s} \cdot \vec{c}_{k^* \circ (j) \circ l} \, |r\rangle\langle s| \otimes |k\rangle\langle l| \,, \tag{73}$$

where (i, j) label the variables, i.e., (i) and (j) are words of length one. Here, we abbreviated the operations corresponding to reversing the order of words, $r^* := (r_1, \ldots, r_p)^* = (r_p, \ldots, r_1)$ as well as to concatenation, $r \circ s := (r_1, \ldots, r_p, s_1, \ldots, s_q)$. We also allow the indices to take the value $\emptyset$, which we interpret as "no variable" and thus define $r \circ \emptyset := r$. For n even, the program then reads as follows:

n^{th}-level SDP relaxation for error of $\text{Ext} = \{f_s\}$

$$\begin{aligned} \text{SDP}_n(\text{Ext}, k) := \text{maximize} \quad & \frac{1}{D} \sum_{s,y,x} \left(\delta_{f_s(x)=y} - \frac{1}{M} \right) \vec{c}_{(x)} \cdot \vec{c}_{(\tilde{y})} \\ \text{subject to} \quad & \forall\, (i),(j) \in S_1, \ \forall x \ : \\ & \quad 0 \le C[(x),(j)] \le 2^{-k}\, C[(\emptyset),(j)] \\ & \quad 0 \le C[(i),(x)] \le 2^{-k}\, C[(i),(\emptyset)] \\ & \sum_x C[(\emptyset),(x)] \le C[(\emptyset),(\emptyset)] \text{ and } \sum_x C[(x),(\emptyset)] \le C[(\emptyset),(\emptyset)] \\ & \sum_{x,x'} C[(x),(x')] \le C[(\emptyset),(\emptyset)] \\ & \text{all vectors have norm bounded by one} \end{aligned} \tag{74}$$

The dependence on n clearly enters through the dimension of the matrices $C[(i),(j)]$, which are indexed by words of length up to n. Note that in particular for the level $n = 0$ the matrices $C[(i),(j)]$ in (73) become scalars again and we get back the SPD relaxation (17) - including an additional dummy vector indexed by the empty word.

However generally the programs are defined for bigger and bigger sets of matrices and therefore vectors, providing more and more constraints and thus reducing the optimal value of the program. Finally, we note that similarly as for two-player games, the hierarchies of the SDP do not converge to extractors against finite-dimensional quantum adversaries but instead they converge to extractors against infinite-dimensional quantum adversaries. Whether these two cases can be different is a wide open question connected to a major open problem in operator algebra theory: Connes' embedding problem [9, 13, 14, 18].

Acknowledgments We acknowledge discussions with Matthias Christandl, Fabian Furrer, Patrick Hayden, Christopher Portmann, Renato Renner, Oleg Szehr, Marco Tomamichel, Thomas Vidick, Stephanie Wehner, Reinhard Werner, and Andreas Winter. Part of this work was done while OF and VBS where visiting the Institute for Quantum Information and Matter at Caltech and we would like to thank John Preskill and Thomas Vidick for their hospitality. Most of this work was done while OF was at ETH Zurich supported by the the European Research Council grant No. 258932. Additional funding was provided by the EU under the project "Randomness and Quantum Entanglement" (RAQUEL). VBS was supported by an ETH postdoctoral fellowship.

Appendix A: Useful Lemmas

Theorem A.1 (Grothendieck's inequality)**.** *For any real matrix* $\{A_{ij}\}$*, we have*

$$\max\left\{\sum_{i,j} A_{ij}\vec{a}_i\cdot\vec{b}_j : \|\vec{a}_i\|_2\le 1, \|\vec{b}_j\|_2\le 1\right\} \tag{A1}$$

$$\le K_G\cdot\max\left\{\sum_{i,j} A_{ij}a_ib_j : a_i,b_j\in\mathbb{R}, |a_i|\le 1, |b_j|\le 1\right\}\ . \tag{A2}$$

Theorem A.2 (Chernoff bound)**.** *Let* $X_i\in\{0,1\}$ *be independent and identically distributed random variables, and* $\mu := \mathbf{E}\{\sum_i X_i\}$*. Then, we have*

$$\mathbf{P}\left\{\sum_i X_i \ge (1+\delta)\mu\right\} \le \left(\frac{e^{\delta}}{(1+\delta)^{(1+\delta)}}\right)^{\mu} \qquad \textit{for any } \delta > 0 \tag{A3}$$

$$\mathbf{P}\left\{\sum_i X_i \le (1-\delta)\mu\right\} \le \left(\frac{e^{-\delta}}{(1-\delta)^{(1-\delta)}}\right)^{\mu} \qquad \textit{for any } 0<\delta<1\ . \tag{A4}$$

Appendix B: Missing Proofs

Proof of Theorem III.4. We start with the expression $\frac{1}{D}\sum_{s,y}\|\sum_x\left(\delta_{f_s(x)=y}-\frac{1}{M}\right)\vec{a}_x\|_2$ for the SDP, where the vectors $\vec{a}_x$ fulfill the conditions stated in (17). Using Cauchy-Schwarz, we may bound

$$\frac{1}{D}\sum_{s,y}\left\|\sum_x\left(\delta_{f_s(x)=y}-\frac{1}{M}\right)\vec{a}_x\right\|_2 \le \left(\frac{1}{D}\sum_{s,y}\left\|\sum_x\left(\delta_{f_s(x)=y}-\frac{1}{M}\right)\vec{a}_x\right\|_2^2\right)^{1/2} 2^{m/2}\ . \tag{B1}$$

We now take a closer look at the expression in the brackets. Expanding the norm squared gives rise to the expression

$$\frac{1}{D}\sum_{s,y}\left(\sum_x\left(\delta_{f_s(x)=y}-\frac{1}{M}\right)\vec{a}_x\right)\cdot\left(\sum_{x'}\left(\delta_{f_s(x')=y}-\frac{1}{M}\right)\vec{a}_{x'}\right) \tag{B2}$$

$$\begin{aligned}
&=\frac{1}{D}\sum_{s,y}\left(\sum_x\delta_{f_s(x)=y}\vec{a}_x\right)\cdot\left(\sum_{x'}\delta_{f_s(x')=y}\vec{a}_{x'}\right)\\
&\quad-\frac{1}{D}\sum_{s,y}\frac{1}{M}\sum_{x,x'}\delta_{f_s(x)=y}\vec{a}_x\cdot\vec{a}_{x'}\\
&\quad-\frac{1}{D}\sum_{s,y}\frac{1}{M}\sum_{x,x'}\delta_{f_s(x')=y}\vec{a}_x\cdot\vec{a}_{x'}\\
&\quad+\frac{1}{D}\frac{1}{M^2}\sum_{s,y}\sum_{x,x'}\vec{a}_x\cdot\vec{a}_{x'}\ .
\end{aligned} \tag{B3}$$

Let us examine the cross terms:

$$\frac{1}{D}\sum_{s,y}\frac{1}{M}\sum_{x,x'}\delta_{f_s(x)=y}\vec{a}_x\cdot\vec{a}_{x'} = \frac{1}{D}\sum_s\frac{1}{M}\sum_{x,x'}\vec{a}_x\cdot\vec{a}_{x'}\ , \tag{B4}$$

since for each fixed pair $s,x\in D\times N$ there is exactly one $y\in M$ such that $f_s(x)=y$. The second cross term evaluates analogously to the same value, which is also equal to the fourth term in the expansion of the norm, and hence we are left with

$$\frac{1}{D}\sum_{s,y}\left(\sum_x\delta_{f_s(x)=y}\vec{a}_x\right)\cdot\left(\sum_{x'}\delta_{f_s(x')=y}\vec{a}_{x'}\right)-\frac{1}{D}\sum_{s,y}\frac{1}{M}\left(\sum_x\vec{a}_x\right)\cdot\frac{1}{M}\left(\sum_{x'}\vec{a}_{x'}\right)\ . \tag{B5}$$

Introducing the maps ψ_s and τ from $\ell_2(N)$ to $\ell_2(M)$,

$$\psi_s : \vec{e}_x \mapsto \sum_y \delta_{f_s(x)=y} \vec{e}_y \quad \text{and} \quad \tau : \vec{e}_x \mapsto \frac{1}{M} \sum_y \vec{e}_y \tag{B6}$$

this may be written as

$$\frac{1}{D} \sum_s \psi_s(\vec{a}) \cdot \psi_s(\vec{a}) - \tau(\vec{a}) \cdot \tau(\vec{a}) \ , \tag{B7}$$

where the dot now means taking the scalar product in the Hilbert space $\ell_2(M) \otimes \mathcal{H}$ and we set $\vec{a} = \sum_x \vec{e}_x \otimes \vec{a}_x \in \ell_2 \otimes \mathcal{H}$. However, this is up to a factor of $\frac{1}{D}$ exactly the defining expression of a spectral extractor. Hence we may bound

$$\frac{1}{D} \sum_s \psi_s(\vec{a}) \cdot \psi_s(\vec{a}) - \tau(\vec{a}) \cdot \tau(\vec{a}) \leq 2^k \frac{\varepsilon}{M} \|\vec{a}\|^2 \ . \tag{B8}$$

The last norm evaluates to

$$\|\vec{a}\|^2 = \sum_x \vec{a}_x \cdot \vec{a}_x \leq 2^{-k} \sum_x q(x) = 2^{-k} \ , \tag{B9}$$

and comparison with (B1) gives the desired bound. □

Proof of Theorem III.5. Consider a feasible solution of (12) given by $\rho(x), \sigma, B_{s,y}$ all acting on a Hilbert space Q. The objective function can be written as

$$\begin{aligned}
&\frac{1}{2^d} \sum_{s,y,x} \left(\delta_{f_s(x)=y} - 2^{-m}\right) \mathrm{Tr}[\rho(x)\, B_{s,y}] \\
&= \frac{1}{2^d} \sum_{s,x} \sum_{y \in \{0,1\}} \left(\sum_{t=0}^{m-1} \frac{1}{2^{m-t-1}} \prod_{k=1}^{t+1} \delta_{f_s(x)_k = y_k} - \frac{1}{2^{m-t}} \prod_{k=1}^{t} \delta_{f_s(x)_k = y_k} \right) \mathrm{Tr}[\rho(x)\, B_{s,y}] && \text{(B10)} \\
&= \sum_{t=0}^{m-1} \frac{1}{2^d} \sum_{s,x} \sum_{y_1, y_2, \dots y_{t+1}} \prod_{k=1}^{t} \delta_{f_s(x)_k = y_k} \left(\delta_{f_s(x)_{t+1} = y_{t+1}} - \frac{1}{2} \right) \mathrm{Tr}[\rho(x)\, C_{s,y_1,y_2,\dots,y_{t+1}}] \ , && \text{(B11)}
\end{aligned}$$

where we defined

$$C_{s,y_1,\dots,y_t,y_{t+1}} := \frac{1}{2^{m-t-1}} \sum_{y_{t+2},\dots,y_m \in \{0,1\}} B_{s,y_{t+2},\dots,y_m} \ . \tag{B12}$$

We now start using the particular structure of the extractor in (63). From now, we fix the value of t and the dependence on t of many variables are omitted to lighten the notation. The seed s can be specified by $a = s|S_{t+1} \in \{0,1\}^l$ and $b = s|S_{t+1}^c \in \{0,1\}^{d-l}$ where S_{t+1}^c is the complement on S_{t+1} in the set $\{1, \dots, d\}$. We will thus interchangeably use s and (a, b). Using this notation with the structure of f_s, we obtain

$$\begin{aligned}
&\frac{1}{2^d} \sum_{s,y,x} \left(\delta_{f_s(x)=y} - 2^{-m}\right) \mathrm{Tr}[\rho(x)\, B_{s,y}] \\
&= \sum_{t=0}^{m-1} \frac{1}{2^d} \sum_{\substack{x \\ a \in \{0,1\}^l \\ b \in \{0,1\}^{d-l}}} \sum_{y_1, y_2, \dots y_{t+1}} \delta_{h_{x,b}(a) = y_1 \dots y_t} \left(\delta_{g_a(x) = y_{t+1}} - \frac{1}{2} \right) \mathrm{Tr}[\rho(x)\, C_{a,b,y_1,y_2,\dots,y_{t+1}}] && \text{(B13)} \\
&= \sum_{t=0}^{m-1} \frac{1}{2^l} \sum_{\substack{x \\ a \in \{0,1\}^l}} \sum_{z \in \{0,1\}} \left(\delta_{g_a(x) = z} - \frac{1}{2} \right) \frac{1}{2^{d-l}} \sum_{b \in \{0,1\}^{d-l}} \mathrm{Tr}[\rho(x)\, C_{a,b,h_{x,b}(a),z}] && \text{(B14)}
\end{aligned}$$

where $h_{x,b}(a)$ represents the first t bits of $f_s(x)$. Note that for a fixed x and b, the outcome of this function only depends on the bits of s that belong to one of the sets $S_1, \dots, S_t$. In particular, the first bit of $h_{x,b}$ only depends on the substring of a corresponding to indices in $S_1 \cap S_{t+1}$. Thus, for any x, b, the function $h_{x,b}$ belongs to the family $\mathcal{F}_t$

of functions $h : \{0,1\}^l \to \{0,1\}^t$ for which the j-th bit h^j of h is a function $h^j : \{0,1\}^{S_j \cap S_{t+1}} \to \{0,1\}$. Thus, for any x, b only $\sum_{j=1}^{t} 2^{|S_j \cap S_{t+1}|} \leq r(m-1)$ bits are sufficient to fully describe the function $h_{x,b}$. As a result, $|\mathcal{F}_t| \leq 2^{r(m-1)}$.

Let us define new positive operators on a larger $Q \otimes H \otimes G$ system as

$$\hat{\rho}(x) := \frac{1}{2^{d-l}} \sum_{\substack{b \in \{0,1\}^{d-l} \\ h \in \mathcal{F}_t}} \rho(x) \otimes \delta_{h=h_{x,b}} |h\rangle\langle h|_H \otimes |b\rangle\langle b|_G \tag{B15}$$

$$\hat{\sigma} := \frac{1}{|\mathcal{F}_t| 2^{d-l}} \sum_{\substack{b \in \{0,1\}^{d-l} \\ h \in \mathcal{F}_t}} \sigma \otimes |h\rangle\langle h|_H \otimes |b\rangle\langle b|_G \tag{B16}$$

$$\hat{C}_{a,z} := \sum_{b, h \in \mathcal{F}_t} C_{a,b,h(a),z} \otimes |h\rangle\langle h| \otimes |b\rangle\langle b| \; . \tag{B17}$$

Note that $\hat{\sigma}$ as well as $\sum_x \hat{\rho}(x)$ have unit trace and $\|\hat{C}_{a,z}\|_\infty \leq 1$. In addition,

$$\hat{\rho}_x \leq \frac{1}{2^{d-l}} \sum_{\substack{b \in \{0,1\}^{d-l} \\ h \in \mathcal{F}_t}} \rho(x) \otimes |h\rangle\langle h|_H \otimes |b\rangle\langle b|_G \leq |\mathcal{F}_t| 2^{-k} \hat{\sigma} \leq 2^{-k+r(m-1)} \hat{\sigma} \; , \tag{B18}$$

where we used the fact that $\rho(x) \leq 2^{-k}\sigma$. This shows that the newly defined operators $\hat{\rho}(x), \hat{\sigma}, \hat{C}_{a,z}$ satisfy the constraints of (12) for the extractor $\mathrm{Ext}_{\mathrm{one}}$ with min-entropy $k - r(m-1)$. Looking at the value of the objective function for this solution, we obtain

$$\frac{1}{2^l} \sum_{a,z,x} \left(\delta_{g_a(x)=z} - \frac{1}{2} \right) \mathrm{Tr}[\hat{\rho}(x) \hat{C}_{a,z}] = \frac{1}{2^l} \sum_{a,z,x} \left(\delta_{g_a(x)=z} - \frac{1}{2} \right) \mathrm{Tr}[\hat{\rho}(x) \hat{C}_{a,z}] \tag{B19}$$

$$= \frac{1}{2^l} \sum_{a,z,x} \left(\delta_{g_a(x)=z} - \frac{1}{2} \right) \frac{1}{2^{d-l}} \sum_b \mathrm{Tr}[\rho(x) C_{a,h(a),z}] \; , \tag{B20}$$

which is exactly the t-th term in the sum in (B14). To relate $\mathrm{Q}(\mathrm{Ext}_{\mathrm{one}}, k - r(m-1))$ to $\mathrm{C}(\mathrm{Ext}_{\mathrm{one}}, k - r(m-1) - \log(1/\varepsilon)) + \varepsilon$, we use Theorem III.2 and Theorem III.3. □

Proposition B.1. *Suppose the functions $f_s : N \to M$ from the family $\{f_s\}_{s \in D}$ are chosen at random with $f_s(x)$ and $f_{s'}(x')$ uniformly distributed and independent whenever $x \neq x'$. Then, we have for $N \geq 16$ that*

$$\mathbf{P}\left\{ \left| \sum_{x,x',s} \delta_{f_s(x)=f_s(x')} - \left(DN + \frac{DN(N-1)}{M} \right) \right| \geq \frac{1}{2} \frac{DN(N-1)}{M} \right\} \leq \frac{1}{16} \; . \tag{B21}$$

This of course includes the case when the functions f_s are chosen uniformly and independently, but also the case of Trevisan's construction where the one-bit extractor is a randomly chosen function.

Proof of Proposition B.1. We start by separating the cases $x = x'$ and $x \neq x'$,

$$\sum_{x,x',s} \delta_{f_s(x)=f_s(x')} = DN + \sum_{s, x \neq x'} \delta_{f_s(x)=f_s(x')} \; . \tag{B22}$$

We compute the expectation over the choice of f:

$$\mathop{\mathbf{E}}_{f} \left\{ \sum_{s, x \neq x'} \delta_{f_s(x)=f_s(x')} \right\} = DN(N-1) \frac{1}{M} \; , \tag{B23}$$

simply using the fact then for $x \neq x'$, $f_s(x)$ and $f_s(x')$ are independently chosen. We now would like to show that with high probability this random variable is close to its expectation. For that we compute the second moment

$$\mathop{\mathbf{E}}_{g}\left\{\left(\sum_{s,x\neq x'}\delta_{f_s(x)=f_s(x')}\right)^2\right\} \tag{B24}$$

$$= \sum_{s_1,s_2,x_1\neq x_2,x_1'\neq x_2'} \mathbf{P}\left\{f_{s_1}(x_1)=f_{s_1}(x_1'), f_{s_2}(x_2)=f_{s_2}(x_2')\right\} \tag{B25}$$

$$= \sum_{s_1,s_2,x_1\neq x_2,x_1'\neq x_2',\{x_1,x_1'\}\neq\{x_2,x_2'\}} \mathbf{P}\left\{f_{s_1}(x_1)=f_{s_1}(x_1'), f_{s_2}(x_2)=f_{s_2}(x_2')\right\} \tag{B26}$$

$$+ \sum_{s_1,s_2,x_1\neq x_2,x_1'\neq x_2',\{x_1,x_1'\}=\{x_2,x_2'\}} \mathbf{P}\left\{f_{s_1}(x_1)=f_{s_1}(x_1'), f_{s_1}(x_2)=f_{s_1}(x_2')\right\} \tag{B27}$$

$$\leq D^2N(N-1)(N(N-1)-2)\frac{1}{M^2} \tag{B28}$$

$$+ 2\sum_{s_1,s_2,x_1\neq x_2} \mathbf{P}\left\{f_{s_1}(x_1)=f_{s_1}(x_1')\right\} \tag{B29}$$

$$= D^2N(N-1)(N(N-1)-2)\frac{1}{M^2} + 2D^2N(N-1)\frac{1}{M}\ . \tag{B30}$$

As a result the variance is at most

$$\mathbf{Var}\left\{\sum_{s,x\neq x'}\delta_{f_s(x)=f_s(x')}\right\} \tag{B31}$$

$$\leq D^2N(N-1)(N(N-1)-2)\frac{1}{M^2} + 2D^2N(N-1)\frac{1}{M} - \left(DN(N-1)\frac{1}{M}\right)^2 \tag{B32}$$

$$\leq 2D^2N(N-1)\frac{1}{M}\ . \tag{B33}$$

Using Chebyshev's inequality gives with a standard deviation $\sigma \leq \sqrt{2}D\sqrt{N(N-1)/M}$ we have

$$\mathbf{P}\left\{\left|\sum_{s,x\neq x'}\delta_{f_s(x)=f_s(x')} - \frac{DN(N-1)}{M}\right| \geq 4\sigma\right\} \leq \frac{1}{16}\ . \tag{B34}$$

But $4\sigma \leq 4\sqrt{2}D\sqrt{N(N-1)/M} \leq \frac{1}{2}\frac{DN(N-1)}{M}$ for $N \geq 16$. □

[1] N. Alon, S. Arora, R. Manokaran, D. Moshkovitz, and O. Weinstein. Inapproximability of densest k-subgraph from average case hardness. *Manuscript, available at www.csc.kth.se/˜rajsekar/papers/dks.pdf*, 2011.

[2] A. Ben-Aroya and A. Ta-Shma. Better short-seed quantum-proof extractors. *Theoretical Computer Science*, 419:17–25, 2012. DOI: 10.1016/j.tcs.2011.11.036.

[3] M. Berta, O. Fawzi, and V. B. Scholz. Quantum-proof randomness extractors via operator space theory. *arXiv preprint arXiv:1409.3563*, 2014.

[4] M. Berta, O. Fawzi, and V. B. Scholz. Quantization of quadratic programs, hierachies of semi-definite programs and applications to quantum information theory. 2015 - in preparation.

[5] M. Berta, O. Fawzi, V. B. Scholz, and O. Szehr. Variations on classical and quantum extractors. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 1474–1478, 2014. DOI: 10.1109/ISIT.2014.6875078.

[6] A. Bhaskara, V. Guruswami, M. Charikar, A. Vijayaraghavan, and Y. Zhou. Polynomial integrality gaps for strong SDP relaxations of Densest k-subgraph. In *SODA '12: Proceedings of the Twenty-Third Annual ACM-SIAM Symposium on Discrete Algorithms*, 2012. Available online: http://dl.acm.org/citation.cfm?id=2095116.2095150.

[7] R. Bhatia. *Matrix Analysis*. Springer, 1997.

[8] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86:419–478, 2014. DOI: 10.1103/RevModPhys.86.419.

[9] A. Connes. Classification of injective factors. Cases II_1, II_∞, III_λ, $\lambda \neq 1$. *Annals of Mathematics. Second Series*, 104:73–115, 1976.

[10] A. De, C. Portmann, T. Vidick, and R. Renner. Trevisan's extractor in the presence of quantum side information. *SIAM Journal on Computing*, 41:915–940, 2012. DOI: 10.1137/100813683.

[11] A. C. Doherty, Y.-C. Liang, B. Toner, and S. Wehner. The Quantum Moment Problem and Bounds on Entangled Multi-prover Games. In *2008 23rd Annual IEEE Conference on Computational Complexity*, pages 199–210. IEEE, 2008. DOI: 10.1109/CCC.2008.26.

[12] U. Feige and M. Seltser. On the densest k-subgraph problem. *Algorithmica*, 29:2001, 1997.

[13] M. Junge, M. Navascues, C. Palazuelos, D. Perez-Garcia, V. B. Scholz, and R. F. Werner. Connes' embedding problem and Tsirelson's problem. *Journal of Mathematical Physics*, 52:012102, 2011. DOI: 10.1063/1.3514538.

[14] E. Kirchberg. On non-semisplit extensions, tensor products and exactness of group C *-algebras. *Inventiones mathematicae*, 112:449–489, 1993.

[15] R. Konig and B. Terhal. The bounded-storage model in the presence of a quantum adversary. *Information Theory, IEEE Transactions on*, 54:749–762, 2008. DOI: 10.1109/TIT.2007.913245.

[16] J. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11:796–817, 2001.

[17] M. Navascues, S. Pironio, and A. Acin. Bounding the set of quantum correlations. *Physical Review Letters*, 98:010401, 2007. DOI: 10.1103/PhysRevLett.98.010401.

[18] N. Ozawa. About the Connes embedding conjecture: algebraic approaches. *Japanese Journal of Mathematics*, 8:147–183, 2013.

[19] S. Pironio, M. Navascués, and A. Acín. Convergent Relaxations of Polynomial Optimization Problems with Noncommuting Variables. *Siam Journal on Optimization*, 20:2157–2180, 2010.

[20] R. Renner. *Security of Quantum Key Distribution*. PhD thesis, ETH Zurich, 2005. DOI: 10.1142/S0219749908003256.

[21] R. Renner and R. König. Universally composable privacy amplification against quantum adversaries. In J. Kilian, editor, *Theory of Cryptography*, volume 3378 of *Lecture Notes in Computer Science*, pages 407–425. Springer Berlin Heidelberg, 2005. DOI: 10.1007/978-3-540-30576-7_22.

[22] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner. Leftover hashing against quantum side information. *Information Theory, IEEE Transactions on*, 57:5524–5535, 2011. DOI: 10.1109/TIT.2011.2158473.