Robust randomness generation on quantum computers

Mario Berta[1,2,3] and Fernando Brandão[1,2]

[1] *AWS Center for Quantum Computing, Pasadena, CA 91125, USA*
[2] *IQIM, California Institute of Technology, Pasadena, CA 91125, USA*
[3] *Department of Computing, Imperial College London, London SW7 2AZ, UK*

This technical document is accompanying the Amazon Braket Jupyter notebook carrying the same name. Here, we provide the necessary formal security definitions as well as full security and complexity proofs. As our main technical result, we construct and implement a quantum-proof two-source extractor with quasi-linear runtime for the efficient distillation of random bits from two weak sources of randomness generated by noisy quantum processing units. For realistic parameters, the construction works for input sizes from around 10^2 to 10^7 bits. Crucially, the output bits remain information-theoretically random even when the adversary has knowledge of the noise occurring on the quantum processing units.

CONTENTS

I. OVERVIEW

The classical theory of pseudorandonmess [22] allows to distil information-theoretically secure random bits from two independent sources of randomness whenever they have sufficiently high entropy [14]. It is thereby crucial for reasonable high throughput rates that the corresponding classical algorithms — called two-source extractors — have complexity linear in the input size [6, 8].[1]

Following the Amazon Braket Jupyter notebook [2], we propose to program two separate quantum processing units from different suppliers in Amazon Braket to supply two streams of weak

[1] The underlying process to create these weak sources of randomness can either be classical or quantum.

randomness. With current noisy quantum technologies, each source on its own then typically has some bias and is thus not particularly close to being fully random. As a consequence, any knowledge of the noise occurring renders the randomness insecure. However, under minimal physical assumptions, employing classical post-processing based on strong two-source extractors results in random output bits that are information-theoretically secure. The output then even remains random when revealing one of the two imperfect input streams. As such, the two suppliers of the quantum processing units can only break the randomness generation scheme when collaborating.[2]

The quantum part of the randomness generation protocol consists of the following most simple circuit:

- Initialize n physical qubits in the zero state $|0\rangle^{\otimes n}$
- Apply the tensored Hadamard gate $H^{\otimes n}$ leading to n physical qubit states $\frac{1}{\sqrt{2}}\big(|0\rangle + |1\rangle\big)$
- Measure each qubit in the computational basis

Doing this on two separate quantum processing units leads to two independent n-bit strings that are somewhat but not fully random due to the bias present in the quantum hardware. One then needs to model the noise occurring via benchmarking and find lower bounds on the entropy of the sources — conditioned on knowledge of the noise model.

For the classical post-processing, we implement a strong two-source extractor based on modified Toeplitz matrices [8] with quasi-linear complexity $O(n \log n)$ in the input size n. For sources with linear entropy rate $n \cdot \alpha_i$ for $i = 1, 2$ the m output bits of the extractor are information-theoretically random with security parameter $\varepsilon \in (0, 1]$ for

$$m = \lceil n \cdot (\alpha_1 + \alpha_2 - 1) + 1 - 2 \log (1/\varepsilon) \rceil \ . \tag{1}$$

This allows the robust generation of randomness on quantum computers as long as we have $\alpha_1 + \alpha_2 - 1 > 0$ for the sources — which is easy to achieve given the fidelities of state-of-the-art quantum processing units as provided in Amazon Braket.

An implementation of the proposed protocol together with high level explanations can be found in the Amazon Braket Jupyter notebook [2]. In the remainder of this note, we give in Section II formal security definitions, discuss in Section III the security and complexity of our randomness extractor, and then discuss in Section IV how to model noisy quantum processing units to find lower bounds on the entropy present in the raw bits of randomness generated. We conclude in Section V with a comparison of our work with previous results in the literature.

II. SECURITY DEFINITIONS

For the theory of pseudo-randomness the relevant entropy measure is the so-called min-entropy. We use the standard definition of the quantum conditional min-entropy [13, Definition 1 & Theorem 1], enabling us to work with classical sources of randomness correlated to quantum systems.

Definition 1 (Min-entropy)**.** *For classical-quantum states $\rho_{NQ} = \sum_x p_x |x\rangle\langle x|_N \otimes \rho_Q^x$ with probability distribution $\{p_x\}$ and ensemble of quantum states $\{\rho_Q^x\}$, the conditional min-entropy of N given Q is defined as*

$$H_{\min}(N|Q)_\rho = -\log \max_{\substack{0 \le M^x \le 1 \\ \sum_x M^x = 1}} \sum_{x \in X} p_x \, Tr\left[M_Q^x \rho_Q^x\right] \ . \tag{2}$$

[2] One still has to trust the encryption in the Amazon Braket.

For trivial systems Q, we use the notation $H_{\min}(N)_P$ for the min-entropy of the probability distribution P_N.

The standard definition of two-source randomness extractors is as follows [17, Definition 1].

Definition 2 (Two-source extractor)**.** *Let $n_1, n_2 \in \mathbb{N}$, $k_1 \in [0, n_1], k_2 \in [0, n_2]$, $m \in \mathbb{N}$, and $\varepsilon \in [0, 1]$. A $(n_1, k_1, n_2, k_2, m, \varepsilon)$ two-source extractor is defined as a function $Ext : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ such that for independent N_1, N_2 with $H_{\min}(N_1) \geq k_1$ and $H_{\min}(N_2) \geq k_2$, we have*

$$\frac{1}{2} \left\| Ext(N_1, N_2) - U_M \right\|_1 \leq \varepsilon \quad \text{with } U_M \text{ the uniform random variable on } m \text{ bits} \tag{3}$$

and $\| \cdot \|_1$ denoting the total variational distance. The function Ext is defined to be strong in the input $i = 1, 2$ when

$$\frac{1}{2} \left\| Ext(N_1, N_2) - U_M \circ N_i \right\|_1 \leq \varepsilon \,. \tag{4}$$

We refer to n_1, n_2 as the input length of the first and second source, resp., to $m \in \mathbb{N}$ as the output length, and to $\varepsilon \in [0, 1]$ as the security parameter.

Following Arnon-Friedman *et al.* [1], this security criteria extends to adversaries holding quantum information $Q_1 Q_2$ about the sources N_1 and N_2, respectively. In their work, Arnon-Friedman *et al.* discuss general Markov sources (also see [4] for alternative models), but for us it is sufficient to restrict to product sources — as first mentioned in [11].

Definition 3. *A classical-classical-quantum-quantum state $\rho_{N_1 N_2 Q_1 Q_2}$ is a product source if*

$$\rho_{N_1 N_2 Q_1 Q_2} = \rho_{N_1 Q_1} \otimes \rho_{N_2 Q_2} \,. \tag{5}$$

The standard, composable security criteria [16] in the presence of quantum adversaries is then as follows [1, Definition 8].

Definition 4 (Quantum-proof two-source extractor)**.** *Let $n_1, n_2 \in \mathbb{N}$, $k_1 \in [0, n_1], k_2 \in [0, n_2]$, $m \in \mathbb{N}$, and $\varepsilon \in [0, 1]$. A product quantum-proof $(n_1, k_1, n_2, k_2, m, \varepsilon)$ two-source extractor is defined as a function $Ext : \{0,1\}^{n_1} \times \{0,1\}^{n_2} \to \{0,1\}^m$ such that for product sources $\rho_{N_1 N_2 Q_1 Q_2}$ with $H_{\min}(N_1|Q_1)_\rho \geq k_1$ and $H_{\min}(N_2|Q_2)_\rho \geq k_2$, we have*

$$\frac{1}{2} \left\| \rho_{Ext(N_1,N_2) Q_1 Q_2} - \tau_M \otimes \rho_{N_1 N_2} \right\|_1 \leq \varepsilon \tag{6}$$

with $\rho_{Ext(N_1,N_2)Q} = (Ext(N_1, N_2) \otimes \mathcal{I}_Q)(\rho_{N_1 N_2 Q})$, τ_M the fully mixed state on $\mathbb{C}^{2^m}$, and $\| \cdot \|_1$ denoting the metric induced by the Schatten one-norm. The function Ext is defined to be product quantum-proof strong in the $i = 1, 2$ input when

$$\frac{1}{2} \left\| \rho_{Ext(N_1,N_2) N_i Q_1 Q_2} - \tau_M \otimes \rho_{N_i Q_1 Q_2} \right\|_1 \leq \varepsilon \,. \tag{7}$$

Moreover, when above criteria is only known to hold for classical $Q_1 Q_2$, then the function Ext is called product classical-proof two-source extractor.

III. RANDOMNESS EXTRACTOR CONSTRUCTION

We employ a randomness extractor construction based on modified Toeplitz matrices [8]. The novelty in our analysis is to prove the product quantum-proof strong property (Section III A), as well as to give an explicit implementation with complexity quasi-linear in the input size (Section III B).

A. Security proof

Proposition 1. *For $n, m \in \mathbb{N}$ and $y = (y_{1-m}, \ldots, y_0, \ldots, y_{n-m-1}) \in \{0,1\}^{n-1}$, let*

$$T(y) = \begin{pmatrix} y_0 & y_1 & \cdots & y_{n-m-1} \\ y_{-1} & y_0 & \cdots & y_{n-m-2} \\ \vdots & \vdots & \vdots & \vdots \\ y_{1-m} & y_{2-m} & \cdots & y_{n-2m} \end{pmatrix} \tag{8}$$

be the $m \times (n-m)$ normal Toeplitz matrix, i.e., $T(y) = y_{j-i}$. Then, the function $Ext_T : \{0,1\}^n \times \{0,1\}^{n-1} \to \{0,1\}^m$ defined via the matrix-vector multiplication

$$(x, y) \mapsto z = Ext_T(x, y) = x(T(y)|1_m)^T \tag{9}$$

is a product quantum-proof $(n, k_1, n-1, k_2, m, \varepsilon)$ two-source extractor whenever

$$m \leq \lfloor (k_1 + k_2 - n) + 1 - 2\log(1/\varepsilon) \rfloor \ . \tag{10}$$

Moreover, Ext_T is then also product quantum-proof strong in the source of size $\{0,1\}^{n-1}$.

Notice that these are the exact same parameters as in the case of trivial quantum systems Q, which is in contrast to other constructions that are known to be quantum-proof (see, e.g., [1, 6]). Namely, the threshold for our construction to work is that the difference between $(k_1 + k_2)$ and n is positive.[3] Then, roughly this difference can be extracted.

In other words, for fixed desired output size m and linear min-entropy rates $k_i = \alpha \cdot n$ for $i = 1, 2$, we need two input bit strings with

$$n = \left\lfloor \frac{m - 1 + 2\log(1/\varepsilon)}{2\alpha - 1} \right\rfloor \ . \tag{11}$$

We refer to the Amazon Braket Jupyter notebook [2] for a numerical example that showcases the excellent performance for small input sizes.

Proof of Proposition 1. The following argument can be seen as a fully quantum version of the development presented in [8, Section VII]. For $\rho_{N_2Q_2} = \sum_y p_y |y\rangle\langle y|_{N_2} \otimes \rho^y_{Q_2}$, we estimate

$$\begin{aligned}
&\|(\mathrm{Ext}_{N_1N_2 \to MN_2} \otimes \mathcal{I}_{Q_1Q_2})(\rho_{N_1Q_1} \otimes \rho_{N_2Q_2}) - \tau_M \otimes \rho_{Q_1} \otimes \rho_{N_2Q_2}\|_1 \\
&= \sum_y p_y \left\| (\mathrm{Ext}^y_{N_1 \to M} \otimes \mathcal{I}_{Q_1Q_2}) \left(\rho_{N_1Q_1} \otimes \rho^y_{Q_2} \right) - \tau_M \otimes \rho_{Q_1} \otimes \rho^y_{Q_2} \right\|_1 \qquad (12) \\
&\leq 2^{\frac{m}{2}} \cdot \sum_y p_y \Bigg(\mathrm{Tr}\Big[\big((\rho_{Q_1} \otimes \rho^y_{Q_2})^{-1/4} \big((\mathrm{Ext}^y_{N_1 \to M} \otimes \mathcal{I}_{Q_1Q_2})(\rho_{N_1Q_1} \otimes \rho^y_{Q_2}) \\
&\qquad\qquad - \tau_M \otimes \rho_{Q_1} \otimes \rho^y_{Q_2} \big) (\rho_{Q_1} \otimes \rho^y_{Q_2})^{-1/4} \big)^2 \Big] \Bigg)^{\frac{1}{2}} \qquad (13) \\
&\leq 2^{\frac{m}{2}} \cdot \sqrt{ \sum_y \mathrm{Tr}\left[p_y \rho^y_{Q_2} \otimes \left(\rho^{-1/4}_{Q_1} \left(\left(\mathrm{Ext}^y_{N_1 \to M} \otimes \mathcal{I}_{Q_1} \right) (\rho_{N_1Q_1}) - \tau_M \otimes \rho_{Q_1} \right) \rho^{-1/4}_{Q_1} \right)^2 \right] } \ , \qquad (14)
\end{aligned}$$

[3] Proposition 1 can be extended to general Markov sources by employing the steps in [1, Section 4.2] together with [1, Lemma 6]. This will only occur in a relatively small parameter loss.

where we used the Hoelder type inequality from [19, Lemma 4] in the first inequality and Jensen's inequality for the square root function in the second inequality. Now, by the definition of the conditional min-entropy in the form of [13, Theorem 1] we have for all $y = \{1, \ldots, n-1\}$ and some quantum state σ_{Q_2} that

$$p_y \rho_{N_2}^y \leq 2^{-H_{\min}(N_2|Q_2)_\rho} \cdot \sigma_{Q_2} \,. \tag{15}$$

Thus, we have

$$\begin{aligned}
&\|(\mathrm{Ext}_{N_1N_2\to MN_2} \otimes \mathcal{I}_{Q_1Q_2})(\rho_{N_1Q_1} \otimes \rho_{N_2Q_2}) - \tau_M \otimes \rho_{Q_1} \otimes \rho_{N_2Q_2}\|_1 \\
&\leq 2^{\frac{1}{2}(m-k_2)} \cdot 2^{\frac{n-1}{2}} \cdot \sqrt{\sum_y \frac{1}{2^{n-1}} \cdot \mathrm{Tr}\left[\left(\rho_{Q_1}^{-1/4}\left(\left(\mathrm{Ext}^x_{N_1\to M} \otimes \mathcal{I}_{Q_1}\right)(\rho_{N_1Q_1}) - \tau_M \otimes \rho_{Q_1}\right)\rho_{Q_1}^{-1/4}\right)^2\right]} \qquad (16) \\
&= 2^{-\frac{1}{2}(k_2-n-m+1)} \cdot \underbrace{\sqrt{\mathrm{Tr}\left[\left(\rho_{Q_1}^{-1/4}\left(\left(\mathrm{Ext}_{N_1N_2\to MN_2} \otimes \mathcal{I}_{Q_1}\right)(\rho_{N_1Q_1} \otimes \tau_{N_2}) - \tau_{MN_2} \otimes \rho_{Q_1}\right)\rho_{Q_1}^{-1/4}\right)^2\right]}}_{\leq (*)} \cdot \qquad (17)
\end{aligned}$$

Notice that for the $(*)$ term the input on N_2 is now uniform — it is acting as a so-called seed — and this directly allows to apply that modified Toeplitz matrices define quantum-proof two-universal hash functions [8, Appendix B.B] (also see [21] and references therein). As such, we have $(*) \leq 2^{-\frac{k_1}{2}}$ and we arrive at

$$\|(\mathrm{Ext}_{N_1N_2\to MN_2} \otimes \mathcal{I}_{Q_1Q_2})(\rho_{N_1Q_1} \otimes \rho_{N_2Q_2}) - \tau_M \otimes \rho_{Q_1} \otimes \rho_{N_2Q_2}\|_1 \leq 2^{-\frac{1}{2}(k_1+k_2-n-m+1)} \,. \tag{18}$$

Equivalently, we can choose the output size as

$$m = \lfloor (k_1 + k_2 - n) + 1 - 2\log(1/\varepsilon) \rfloor \,, \tag{19}$$

which is what we set out to prove. □

B. Implementation

The a priori complexity of the function $z = \mathrm{Ext}_T(x, y) = x(T(y)|1_m)^T$ is $O(n^2)$, which is good for block sizes up to $n = 10^4$. To go to higher block sizes up to $n = 10^7$, we now show that the complexity can be brought down to the quasi-linear complexity $O(n \log(n))$ by means of the Discrete Fourier Transform (DFT) via the Fast Fourier Transform (FFT) denoted as F. This is achieved in several steps:

- Recall the definitions $y = (y_{1-m}, \ldots, y_0, \ldots, y_{n-m-1}) \in \{0,1\}^{n-1}$ and

$$T(y) = \begin{pmatrix} y_0 & y_1 & \cdots & y_{n-m-1} \\ y_{-1} & y_0 & \cdots & y_{n-m-2} \\ \vdots & \vdots & \vdots & \\ y_{1-m} & y_{2-m} & \cdots & y_{n-2m} \end{pmatrix} \,. \tag{20}$$

- We rewrite the transpose of the function of interest as

$$z^T = (T(y)|1_m)x^T = T(y)\underline{x}^T + \overline{x}^T \tag{21}$$

for $\underline{x} = (x_0, x_1, \ldots, x_{n-m-1}) \in \{0,1\}^{n-m}$ and $\overline{x} = (x_{n-m}, x_{n-m+1}, \ldots, x_{n-1}) \in \{0,1\}^m$. The addition has complexity $O(m)$ and as such it remains to analyse the matrix-vector multiplication $T(y)\underline{x}^T$.

- We notice that the normal Toeplitz matrix $T(y)$ can be completed to a $(n-1)\times(n-1)$ square Toeplitz matrix $\underline{T}(y)$ as $\underline{T}(y)_{ij} = y_{j-i}$ via the extended vector $\underline{y} = (y_{2-n}, \ldots, y_0, \ldots, y_{n-2}) \in \{0,1\}^{2n-3}$ with entries

$$y_i = \begin{cases} y_i & \text{for } i \in \{1-m, \ldots, 0, \ldots, n-m-1\} \\ 0 & \text{else.} \end{cases} \tag{22}$$

- We further embed the square Toeplitz matrix $\underline{T}(y)$ into the $(2n-3)\times(2n-3)$ circulant matrix $C(y)$ with the first row

$$(y_0, y_1, \ldots, y_{n-2}|y_{2-n}, \ldots, y_{-1}) \tag{23}$$

and all other entries populated according to the rule of circulant matrices. In particular, this leads to the first column given by the transpose of the vector

$$\overline{y} = (y_0, y_{-1}, \ldots, y_{2-n}|y_{n-2}, \ldots, y_1)\,. \tag{24}$$

We note that the upper left block of the $(2n-3)\times(2n-3)$ circulant matrix $C(y)$ is given by the $(n-1)\times(n-1)$ square Toeplitz matrix $\underline{T}(y)$, in which again the upper left block is given by the $m\times(n-m)$ normal Toeplitz matrix $T(y)$.

- The matrix-vector multiplication $T(y)\underline{x}^T$ is then rewritten as

$$T(y)\underline{x}^T = (1_m|0_{2n-m-3})C(y)(\underline{x}, 0, \ldots, 0)^T\,, \tag{25}$$

where 0_{2n-m-3} denotes the $(0_{2n-m-3})\times(0_{2n-m-3})$ zero matrix and the number of zeroes in the last term is $n+m-3$. Since the left multiplication with $(1_m|0_{2n-m-3})$ just corresponds to throwing away the last $2n-m-3$ values, it remains to analyse the complexity of the matrix-vector multiplication $C(y)(\underline{x}, 0, \ldots, 0)^T$.

- Since $C(y)$ is a circulant matrix, it is taken by the DFT to diagonal form via

$$FC(y)F^{-1} = \mathrm{diag}\,(F\,[\overline{y}]) \tag{26}$$

and we can write

$$C(y)(\underline{x}, 0, \ldots, 0)^T = F^{-1}\left[F\left[\overline{y}^T\right] * F\left[(\underline{x}, 0, \ldots, 0)^T\right]\right]\,, \tag{27}$$

where $*$ denotes the Hadamard product.

- Since the Hadamard product has complexity $O(n)$ and the DFT is implemented via the FFT in complexity $O(n\log n)$, this leaves an overall complexity of $O(n\log n)$ for computing the output as

$$z = \mathrm{Ext}_T(x,y) = F^{-1}\big[F[\overline{y}] * F[(\underline{x}, 0, \ldots, 0)]\big]\begin{pmatrix} 1_m \\ 0_{2n-m-3}\end{pmatrix} + \overline{x}\,. \tag{28}$$

IV. MODELLING NOISY QUANTUM DEVICES

A. Setting

In case the quantum processing units perform perfectly for the protocol described in Section I, there is no need for classical post-processing by means of randomness extractors. However, whenever the quantum processing units are noisy and faulty, we need to model the effective state preparation and measurements appropriately to find a lower bound on the min-entropy of the raw bits of randomness created before the classical post-processing. Importantly, we do not directly analyse the min-entropy of the sources, but rather consider the min-entropy of the sources conditioned on all the information that is available in principle given the laws of quantum mechanics [7]. This ensures that the output bits become truly unpredictable, even when an adversary has access to the noise occurring on the quantum processing units. That is, the output is random even conditioned on the noise and is thus exclusively and freshly generated from the quantum resources.

Ideally, the circuit of our quantum random number generator creates n copies of the pure state

$$\psi_A = |\psi\rangle\langle\psi|_A \text{ with } |\psi\rangle_A = \frac{1}{\sqrt{2}}\left(|0\rangle_A + |1\rangle_A\right) \tag{29}$$

followed by the n-fold application of the projective measurement $\mathcal{M} = \{|0\rangle\langle 0|_A, |1\rangle\langle 1|_A\}$. However, in practice the state to start with is rather given by some imperfect mixed state ρ_A only close to ψ_A and the measurement is given by some imperfect two-outcome POVM $\mathcal{N} = \{N, 1-N\}$ only close to $\mathcal{M}$ for some small approximation parameter. Hence, restricting to one qubit for now, the imperfect post-measurement probability distribution $Q_X = \mathcal{N}_{A\to X}(\rho_A)$ is only close to the desired uniform distribution $(0.5, 0.5) = \mathcal{M}_{A\to X}(\psi_A)$.

At first, we might then think of directly using Q_X with the corresponding min-entropy $H_{\min}(X)_Q$ as the input for randomness extraction. However, this fails to take into account that some information about Q has leaked via the noise into the environment. In order to endow the adversary with the maximal amount of leakage available, we need to purify the imperfect input on R and write the imperfect POVM via its Naimark extension as a projective measurement on a combined larger system AE. The min-entropy of the post-measurement distribution X conditioned on these purifying quantum systems RE is then the relevant measure of randomness. This is detailed in the following sections by means of concrete error models.

B. Characterization

For a given quantum processing unit, the supplier typically publishes some noise specification with it. This includes both the noise characterization of state preparation as well as the read-out measurements. If such specifications are not available, there is a wide range of theoretical benchmarking tools available (see [12] and references therein). However, when no information at all on the inner workings of the quantum processing unit are available, it is in general inefficient to benchmark the device properly. The reason being that it is a priori unclear what part of the noise observed stems from noisy state preparation and what part from the noisy read-out measurements. Nevertheless, different methods from so-called self-consistent tomography are available (see [3, 18] and references therein). These techniques are beyond the scope of our work, but luckily we do not need a full characterization of the statistics of the source. Rather, for the randomness extractor to work, we only need to provide a lower bound on the min-entropy of the source. That is, it is sufficient to give a conservative upper bound on the noise strength.

For the state preparation step via Hadamard gates, the typical noise in quantum architectures is symmetric, depolarizing noise of strength $\lambda \in [0, 1]$. This takes the intended pure state vector

$|\psi\rangle_A = \frac{1}{\sqrt{2}}\left(|0\rangle_A + |1\rangle_A\right)$ to some quantum state

$$\rho_A(\lambda) = \frac{1}{2}\Big(|0\rangle\langle 0|_A + (1-\lambda)|0\rangle\langle 1|_A + (1-\lambda)|1\rangle\langle 0|_A + |1\rangle\langle 1|_A\Big)\,. \tag{30}$$

Now, the sensitive part for randomness generation is the noise affecting the measurement device. As such, we present in Appendix A some simple statistical methods on how to benchmark measurement devices with only noisy state preparation available. We note that instead of the ideal measurement device given by the POVM $\mathcal{M} = \{|0\rangle\langle 0|_A, |1\rangle\langle 1|_A\}$, the typical noise measurement device is described by the POVM $\mathcal{N}(\mu) = \{1_A - \mu|1\rangle\langle 1|_A, \mu|1\rangle\langle 1|_A\}$ with some bias $\mu \in (0,1)$ to reading-out the basis state $|0\rangle\langle 0|_A$. To keep the presentation simple, we proceed with this form $\mathcal{N}(\mu) = \{1_A - \mu|1\rangle\langle 1|_A, \mu|1\rangle\langle 1|_A\}$, leading to the post-measurement probability distribution

$$\left(q_0 = 1 - \frac{\mu}{2}, q_1 = \frac{\mu}{2}\right) \text{ instead of the uniform distribution } \left(p_0 = \frac{1}{2}, p_1 = \frac{1}{2}\right). \tag{31}$$

Notice that this form on its own is independent of the parameter λ, as depolarizing noise is uniform.

C. Bounding min-entropy

The imperfect distribution observed then has min-entropy

$$H_{\min}(X)_Q = 1 - \log\mu \leq 1\,. \tag{32}$$

However, as argued previously this is not the relevant quantity for secure randomness extraction. Rather, we need to model the conditional min-entropy of the post-measurement probability distribution X given the purifying quantum registers RE of the state $\rho_A(\lambda)$ and measurement process $\mathcal{N}_{A\to X}(\mu)$, respectively.

The detailed modelling is as follows: The mixed state $\rho_A(\lambda)$ is extended to the pure state $\rho_{AR}(\lambda)$ via its eigenvalues and corresponding eigenvectors. The imperfect measurement $\mathcal{N}(\mu)$ can, e.g., be written as a projective measurement on AE_1 with

$$\mathcal{N} = \{|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_{E_1}, |1\rangle\langle 1|_A \otimes |1\rangle\langle 1|_{E_1}\} \tag{33}$$

acting on the additional input state $\sigma_{E_1}(\mu) = (1-\mu)|0\rangle\langle 0|_{E_1} + \mu|1\rangle\langle 1|_A$. The latter is then again purified as

$$|\sigma(\mu)\rangle_{E_1E_2} = \sqrt{1+\mu}|0\rangle_{E_1}\otimes|0\rangle_{E_2} - \sqrt{\mu}|1\rangle_{E_1}\otimes|1\rangle_{E_2}\,. \tag{34}$$

Consequently, the relevant classical-quantum state to consider becomes

$$\omega_{XRE_2}(\lambda,\mu) = \big(\mathcal{N}_{AE_1\to X}\otimes\mathcal{I}_{RE_2}\big)\big(\rho_{AR}(\lambda)\otimes\sigma_{E_1E_2}(\mu)\big) \tag{35}$$

with corresponding conditional min-entropy $H_{\min}(X|RE_2)_\omega$. Note that we have

$$H_{\min}(X|RE_2)_\omega \leq H_{\min}(X)_Q \tag{36}$$

with the gap typically being strict. Furthermore, in contrast to $H_{\min}(X)_Q$, the term $H_{\min}(X|RE_2)_\omega$ now also depends on λ. Even though the conditional min-entropy as given in Definition 1 does not have a closed form expression, the quantity $H_{\min}(X|RE_2)_\omega$ is computed efficiently via a semi-definite program (sdp). This is done for numerical examples in the accompanying Amazon Braket Jupyter notebook [2].

It is then also easily checked by inspection that the conditional min-entropy is monotone in the noise parameters λ, μ and hence it is in practice indeed sufficient to choose a conservative estimate for the noise strength λ, μ in state preparation and read-out measurement. As we restricted ourselves to the practically relevant setting of single qubit errors so far, the n qubit setting then just becomes

$$H_{\min}(X|RE_2)_{\omega^{\otimes n}} = n \cdot H_{\min}(X|RE_2)_{\omega} \tag{37}$$

with the linear min-entropy rate $\alpha = H_{\min}(X|RE_2)_{\omega}$ — as the conditional min-entropy is additive on tensor product states [13].

For the benchmarking of correlated multiple qubit errors, we have to compute the block spd for the respective $H_{\min}(X|RE_2)_{\omega^n}$ directly. Modern benchmarking schemes promise to find good bounds on $H_{\min}(X|RE_2)_{\omega^n}$ that potentially scale efficiently up to moderate sizes of n for realistic settings with weak multiple qubit correlations. We leave that for further investigation and refer for a starting point to [12] and references therein.

V. COMPARISON WITH PREVIOUS WORK

Our approach to randomness generation is in the so-called device-dependent approach, where we do characterize the quantum processing units to some degree. There is a large body of work developing methods for settings that are (semi) device-independent — where the inner workings of the quantum technologies employed can be self-tested. There is typically a trade-off in practicability and efficiency versus the level of device independence and we refer to the review article [9] for further discussions.

Often specialised quantum hardware is used for quantum randomness generation. When it comes to employing quantum processing units for randomness generation, it is interesting to compare our methods with the recent results of Cambridge Quantum Computing (CQC) [6]. Their work is also implemented in Amazon Braket [10], but is different in the sense that it operates in the semi device-independent setting. The quantum part of the protocol is then more involved, but appealing as the semi device-independent setting requires fewer assumptions on the characterization of the underlying hardware. On the other hand, their results come with the caveat of larger input sizes n needed for their scheme to produce non-zero output. When it comes to the classical post-processing, compared to [6] we use a slightly improved quantum-proof randomness extractor that is more flexible and allows for smaller input sizes n. However, CCQ's implementation is more efficient and secure, relying on the Number Theoretic Transform (NTT) instead of the DFT in our case. Amongst other things, this allows to go to larger input sizes $n > 10^7$ which is important to make their approach efficient.

Further improvements in classical post-processing seem possible, such as allowing lower quality entropy sources (e.g., based on [17]) or additional steps based on short seeded quantum-proof randomness extractors [5, 20]. All of this however, comes with the bottleneck of required implementations in basically linear runtime. This seems to require further results on the theory of algorithms behind randomness extractors [15].

ACKNOWLEDGEMENTS

We thank Steven T. Flammia for discussions on benchmarking noisy quantum measurement devices and suggesting the ideas depicted in Appendix A.

Appendix A: Benchmarking measurement devices

Our goal is to characterize noisy measurement devices when only noisy state preparation is available, e.g., affected by depolarizing noise of some strength λ as in Section IV. Note that this is reversed from the usual quantum state tomography problem, when the measurement devices are assumed to be (at least somewhat) reliable.

If the preparation of qubit states ρ and the sub-sequent measurement $\mathcal{M} = \{|0\rangle\langle 0|_A, |1\rangle\langle 1|_A\}$ are both perfect, then measuring different states $\rho_1, \rho_2, \ldots, \rho_K$ leads to the measurement result zero with probability

$$p_i = \mathrm{tr}\left[\rho_i |0\rangle\langle 0|\right] \quad \text{for each } i = 1, \ldots, K. \tag{A1}$$

Measuring each state ρ_i a total of N times, in the large N limit the fraction of zeroes becomes equal to p_i and we can write

$$\underbrace{\begin{pmatrix} \vec{\rho}_1^T \\ \vec{\rho}_2^T \\ \vdots \\ \vec{\rho}_K^T \end{pmatrix}}_{=A} \cdot \underbrace{\begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}}_{=\overrightarrow{M}} = \underbrace{\begin{pmatrix} p_1 \\ p_2 \\ \vdots \\ p_K \end{pmatrix}}_{=\vec{p}} \tag{A2}$$

with $\vec{\rho}_i$ denoting the vectorization of the density matrices representing the quantum states ρ_i, and $\overrightarrow{M}$ the vectorization of the matrix representation of the measurement operator $|0\rangle\langle 0|$.

Now, when we are restricted to noisy state preparation and finite statistics, we observe the following differences:

- We can only prepare different qubit states $\rho_1(\varepsilon), \rho_2(\varepsilon), \ldots, \rho_K(\varepsilon)$ each affected by some state preparation noise with parameter $\varepsilon \in [0, 1]$. For example, in Section IV the parameter ε corresponds to the depolarizing noise of some strength λ.

- The entries in the vector $\vec{p}(\delta)$ can only be populated with the finite statistical values

$$p_i(\delta) = \frac{\#(\text{of zeroes})}{N} \tag{A3}$$

  sampling from the underlying probability distribution (corresponding to the maximum likelihood estimate for p_i).

- The vectorized measurement operator $\overrightarrow{M}(\varepsilon, \delta)$ is not specified, but we only know that it has the general structure of a qubit POVM element, i.e., coming from a complex two times two matrix $0 \leq M(\varepsilon, \delta) \leq 1$. For example, in Section IV we worked with the specific Ansatz $\overrightarrow{M}^T = (1, 0, 0, 1 - \mu)$ for the bias parameter μ.

Hence, we have a linear system of equations for the unknown qubit POVM element $\overrightarrow{M}(\varepsilon, \delta)$ as

$$\underbrace{\begin{pmatrix} \vec{\rho}_1^T(\varepsilon) \\ \vec{\rho}_2^T(\varepsilon) \\ \vdots \\ \vec{\rho}_K^T(\varepsilon) \end{pmatrix}}_{=A(\varepsilon)} \cdot \overrightarrow{M}(\varepsilon, \delta) = \underbrace{\begin{pmatrix} p_1(\delta) \\ p_2(\delta) \\ \vdots \\ p_K(\delta) \end{pmatrix}}_{=\vec{p}(\delta)} \tag{A4}$$

Going forward we choose $K = 4$ as A and $A(\varepsilon)$ then conveniently become square matrices (the general case can be handled similarly). Denoting the differences

$$\Delta A = A - A(\varepsilon), \quad \Delta\vec{p} = \vec{p} - \vec{p}(\delta), \quad \Delta\overrightarrow{M} = \overrightarrow{M} - \overrightarrow{M}(\varepsilon, \delta)\,, \tag{A5}$$

standard estimates on the stability of the solution of systems of linear equations give the error propagation bound

$$\frac{\|\Delta\overrightarrow{M}\|}{\|\overrightarrow{M}\|} \leq \frac{\kappa(A)}{1 - \kappa(A) \cdot \frac{\|\Delta A\|}{\|A\|}} \left(\frac{\|\Delta\vec{p}\|}{\|\vec{p}\|} + \frac{\|\Delta A\|}{\|A\|} \right), \tag{A6}$$

where $\kappa(A) = \|A\| \cdot \|A^{-1}\|$ denotes the condition number of the matrix A. By inspection this bound is non-trivial as long as $\|\Delta A\| \leq \frac{1}{\|A^{-1}\|}$, which corresponds to small deviations ΔA while having non-singular A.[4]

The term $\|\Delta A\|$ on the rhs of (A6) is upper bounded by the error model for noisy state preparation. For the finite statistics maximum likelihood estimate $\vec{p}(\delta)$, we can pool the $N = N_1 N_2$ measurement results into N_1 groups of N_2 measurement results each, leading to $(1 - \alpha) \cdot 100\%$ confidence intervals for the entries as[5]

$$p_i(\delta) \pm \frac{z_{1-\alpha/2} \cdot s_{N_1 - 1}}{\sqrt{N_1}}\,, \tag{A7}$$

where $z_{1-\alpha/2}$ denotes the $(1 - \alpha/2)$-quantile of the standard normal distribution and we have the bias-corrected standard deviation

$$s^2_{N_1-1} = \frac{\sum_{j=1}^{N_1} \left(\frac{\#(\text{of zeroes})_j}{N_2} - \frac{\sum_{i=1}^{N_1} \#(\text{of zeroes})_i}{N_1 N_2} \right)^2}{N_1 - 1}\,. \tag{A8}$$

The confidence intervals from (A7) for the entries $p_i(\delta)$ of $\vec{p}(\delta)$ then give upper bounds on $\|\Delta\vec{p}\|$ and trough (A6) upper bounds on $\|\Delta\overrightarrow{M}\|$. In turn, such upper bounds on $\|\Delta\overrightarrow{M}\|$ allow to give lower bounds on the relevant min-entropy as discussed Section IV.

Finally, for $K \neq 4$ setting with non-square A similar bounds can be derived, but then featuring the pseudo-inverse A^+ instead of the inverse A^{-1}.

[1] R. Arnon-Friedman, C. Portmann, and V. B. Scholz. *"Quantum-Proof Multi-Source Randomness Extractors in the Markov Model"*. In *Proceedings TQC*, volume 61, pages 1–34, (2016).

[2] M. Berta. *"Robus randomness generation on quantum computers"*. Available on Amazon Braket, (2021). Available online: tobeinserted.

[3] R. Blume-Kohout, J. K. Gamble, E. Nielsen, J. Mizrahi, J. D. Sterk, and P. Maunz. *"Robust, self-consistent, closed-form tomography of quantum logic gates on a trapped ion qubit "*, (2013). arXiv: 1310.4492.

[4] K.-M. Chung, X. Li, and X. Wu. *"Multi-Source Randomness Extractors Against Quantum Side Information, and their Applications "*. Preprint, arXiv: 1411.2315 .

[5] A. De, C. Portmann, T. Vidick, and R. Renner. *"Trevisan's Extractor in the Presence of Quantum Side Information"*. SIAM Journal of Computing **41**(4): 915–940 (2012).

[4] In practice, one has the freedom to slightly shift the base point A by including some of the error parameter ε in order to control the condition number $\kappa(A)$.

[5] This is a good approximation as long as $N_1 \geq 40$.

[6] C. Foreman, S. Wright, A. Edgington, M. Berta, and F. J. Curchod. *"Practical Randomness and Privacy Amplification"*, (2020). arXiv: 2009.06551.

[7] D. Frauchiger, R. Renner, and M. Troyer. *"True randomness from realistic quantum devices"*. Preprint, arXiv: 1311.4547 .

[8] M. Hayashi and T. Tsurumaru. *"More Efficient Privacy Amplification With Less Random Seeds via Dual Universal Hash Function"*. IEEE Transactions on Information Theory **62**(4): 2213–2232 (2016).

[9] M. Herrero-Collantes and J. C. Garcia-Escartin. *"Quantum random number generators"*. Review of Modern Physics **89**(1): 015004 (2017).

[10] D. Jones. *"Quantum-Proof Cryptography with IronBridge, TKET and Amazon Braket"*, (2020). Available online: https://medium.com/cambridge-quantum-computing/quantum-proof-cryptography-with-ironbridge-tket-and-amazon-braket-e8e96777cacc.

[11] R. Kasher and J. Kempe. *"Two-Source Extractors Secure Against Quantum Adversaries"*. Theory of Computing **8**: 461–486 (2012).

[12] M. Kliesch and I. Roth. *"Theory of Quantum System Certification"*. PRX Quantum **2**(1): 010201 (2021).

[13] R. König, R. Renner, and C. Schaffner. *"The Operational Meaning of Min- and Max-Entropy"*. IEEE Transactions on Information Theory **55**(9): 4337–4347 (2009).

[14] X. Li. *"Improved Two-Source Extractors, and Affine Extractors for Polylogarithmic Entropy"*. In *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 168–177, (2016).

[15] W. Mauerer, C. Portmann, and V. B. Scholz. *"A modular framework for randomness extraction based on Trevisan's construction"*. Preprint, arXiv: 1212.0520 .

[16] C. Portmann and R. Renner. *"Cryptographic Security of Quantum Key Distribution"*, (2014). arXiv: 1409.3525.

[17] R. Raz. *"Extractors with Weak Random Seeds"*. In *Proceedings STOC*, pages 11–20, (2005).

[18] C. J. Stark. *"Global Completability with Applications to Self-Consistent Quantum Tomography"*. Communications in Mathematical Physics **348**(1): 1–25 (2016).

[19] M. Tomamichel, C. Schaffner, A. Smith, and R. Renner. *"Leftover Hashing Against Quantum Side Information"*. IEEE Transactions on Information Theory **57**(8): 5524–5535 (2011).

[20] L. Trevisan. *"Extractors and Pseudorandom Generators"*. Journal of the ACM **48**(4): 860–879 (2001).

[21] T. Tsurumaru and M. H. and. *"Dual Universality of Hash Functions and Its Applications to Quantum Cryptography"*. IEEE Transactions on Information Theory **59**(7): 4700–4717 (2013).

[22] S. P. Vadhan. *Pseudorandomness*. volume 7 of *Foundations and Trends in Theoretical Computer Science*, Now Publishers (2012).