Maximisation of Extractable Randomness in Quantum Random Number Generator

J. Y. Haw, S. M. Assad, A. M. Lance, N. H. Y. Ng, V. Sharma, P. K. Lam, and T. Symul

1Centre for Quantum Computation and Communication Technology, Department of Quantum Science, The Australian National University, Canberra, ACT 0200, Australia
2QuintessenceLabs, Unit 1 Lower Ground, 15 Denison Street, Deakin, ACT, 2600 Australia
3Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, 117543 Singapore

(Dated: November 18, 2014)

The generation of random numbers via quantum processes is an efficient and reliable method to obtain true indeterministic random numbers that are of vital importance to cryptographic communication and large scale computer modelling. However, in realistic scenarios, the raw output of a quantum random number generator is inevitably tainted with classical technical noise. The integrity of the device could be compromised if this noise is tampered with, or even controlled by some malicious party. To safeguard against this, we propose and experimentally demonstrate an approach that produces side-information independent randomness that is quantified by min-entropy conditioned on this classical noise. We present a method for maximising the conditional min-entropy of the number sequence generated from a given quantum to classical noise ratio (QCNR). The detected photo-current in our experiment is shown to have a real-time random number generation rate of 14 Mbps/MHz. Integrating this figure across the spectral response of the detection system shows the potential to deliver more than 70 Gbps of random numbers in our experimental setup.

PACS numbers: 03.67.Dd, 42.50.Ex

I. INTRODUCTION

Randomness is a vital resource in many information and communications technology applications, such as computer simulations, statistics, gaming, and cryptography. For applications that are not concerned with the security and uniqueness of randomness, a sequence with uniformly distributed numbers mostly suffices. Such sequences can be generated using a pseudo-random number generator (PRNG) that works via certain deterministic algorithm. Although PRNGs can offer highly unbiased random numbers, they cannot be used for applications that require information security for two reasons: Firstly, PRNG generated sequences are unpredictable only under limitations of computational power, since PRNGs are inherently based on deterministic algorithms. Secondly, the random seeds, which are required to define the initial state of a PRNG, limit the amount of entropy in the random number sequences they generate. This compromises the security of an encryption protocol.

For cryptographic applications, a random sequence is required to be truly unpredictable and to have maximum entropy. To achieve this, intensive efforts have been devoted to developing high-speed hardware RNGs that generate randomness via physical noise. Hardware RNGs are attractive alternatives because they provide fresh randomness based on physical processes that are apparently unpredictable (i.e. uncorrelated with any existing information either with past settings or side information). Moreover, they also provide a solution to the problem of having insufficient entropy. Due to the deterministic nature of classical physics, however, some of these hardware generators may only be truly random under practical assumptions that cannot be validated. RNGs that rely on quantum processes, on the other hand, can have guaranteed indeterminism and entropy, since quantum processes are inherently unpredictable. Examples of such processes include quantum phase fluctuations, spontaneous emission noise, photon arrival times, stimulated Raman scattering, photon polarization state, vacuum fluctuations, and even mobile phone cameras. These QRNGs resolve both shortcomings of the PRNGs. However, despite the fact that they rely on entropy ultimately guaranteed by the laws of quantum physics, measurements on quantum systems are often tainted by classical noise. We quantify the amount of quantum randomness to the amount of classical noise using a quantum-to-classical-noise ratio (QCNR). When QCNR is low, both the quality and the security of the random sequence generated may be compromised. Recently, certifiable randomness based on violation of fundamental inequalities have been proposed and demonstrated as well. Although certifiable randomness, which holds independence even from the internal structure of the generator can in principle be achieved with such systems, it is very challenging experimentally to achieve high generation rate even for state-of-the-art technology.

To address this issue, Gabriel et al. took into account potential eavesdropping on the classical noise by considering the channel capacity of their QRNG. Their setup exhibited a good QCNR clearance and was able to extract approximately 3 bits per sample of guaranteed randomness out of 5 bits of digitization (~60%). More recently, Ma et al. proposed a framework for QRNG entropy evaluation. By using min-entropy as the quantifier for randomness, they extracted a higher rate...
of random bits of 6.7 bit per sample from 8 bits (~84%), where the quantum contribution of the randomness was obtained by inferring the QCNR.

In the process of generating random bits via measuring continuous-variable systems, an analog-to-digital converter (ADC) is commonly used to discretized measurement outcomes. It has been speculated [12] that the freedom of choosing the ADC range could be exploited to optimize extractable randomness. Meanwhile, in [30][31], the choice of dynamical ADC range was justified by experimental observations. Thus, a systematic approach to determine the dynamical ADC range in extracting maximum amount of secure randomness is very much required.

In this work, we propose a new generic framework where the dynamical ADC range can be appropriately chosen to deliver maximum randomness. By quantifying randomness via min-entropy conditioned on classical noise, we show that QCNR is not the sole limiting factor in generating secure random bits. In fact, by carefully optimizing the dynamical ADC range, one can extract a non-zero amount of secure randomness even when the classical noise is larger than the quantum noise. By applying this method to our continuous variable (CV) QRNG based on homodyne measurement of vacuum state [22], we demonstrate that the setup is capable of delivering more than 70 Gbps of secure random bits.

This paper is structured as follows. In Sec I we present the modelling of our CV-QRNG and the quantification of entropy via (conditional) min-entropy. We outline the procedure of optimizing the dynamical ADC range under different operating conditions and experimental paramaters. In Sec. III we analyze and configure our CV-QRNG based on our framework. We then review and discuss the randomness extraction in our CV-QRNG, ending with a brief concluding remarks in Sec. IV.

II. ENTROPY QUANTIFICATION

The main goal of entropy evaluation of a secure QRNG is to quantify the amount of randomness available in the measurement outcome $M$, conditioned upon side-information which might be accessible, controllable or correlated with an eavesdropper. The concept of side-information-independent randomness is well established in information theory, commonly known as privacy amplification or randomness extraction, both in the classical and quantum setting [22][35]. However, to the best of our knowledge, such considerations have not been fully taken into account in most of the existing QRNG literature, where the total entropy (quantum and classical) of the system was assumed to be fully available to the user. The security aspect of randomness generation only started to get considerable attention recently in the framework of QRNG [13][19][22][25][26]. In our work, we consider randomness conditioned on classical noise, where the classical noise could either be known by the adversary due to monitoring or controlling of the classical noise. In order to guarantee the security of the random bits generated, we resort to the min-entropy, which is a quantity for quantifying extractable randomness in a single-shot scenario [20][25][26]. In particular, the conditional min-entropy tells us the amount of (almost) uniform and independent bits that one can extract from a random source. Our goal here is to quantify the maximum amount of independent randomness by optimizing the measurement settings. More precisely, we maximize the conditional min-entropy of the measurement outcomes given the freedom of manipulating the dynamical ADC range and digitisation bits.

A. Characterization of noise and measurement

We first discuss the model of the probability distribution we deploy in our CV-QRNG. Following our previous work [22], homodyne measurement is performed on a vacuum state, which has a symmetric Gaussian probability density function (PDF) $P_Q(q)$ in the phase space. Together with the assumption of Gaussian electronic noise with PDF $P_E(e)$, which could be known fully by an adversary, the resulting measured signal’s PDF $P_M(m)$ is the convolution of the quantum noise $q \in Q$ and the classical noise $e \in E$,

$$P_M(m) = \frac{1}{\sqrt{2\pi\sigma_M}} e^{-\frac{m^2}{2\sigma_M^2}},$$

(1)

which is a Gaussian distribution with variance $\sigma_M^2 = \sigma_Q^2 + \sigma_E^2$, with $m = q + e$. Here QCNR is defined as the log of the ratio between the variances of the quantum signal and the electronic noise, i.e. QCNR=$10\log_{10}(\sigma_Q^2/\sigma_E^2)$. The sampling is performed over a $n$-bit analogue-to-digital converter (ADC) with dynamical ADC range $[-R+\delta/2, R-3\delta/2]$. Upon measurement, the sampled signal is discretised over $2^n$ bins with bin width $\delta = R/2^{n-1}$. The resulting probability distribution of...
The conditional PDF over the distribution conditioned on the classical noise.

We see that an appropriate choice of dynamical ADC range for a given QCNR and digitisation bit \( n \) should not be naively optimized over the measured distribution \( P_M(m_i) \), but over the distribution conditioned on the classical noise.

The conditional PDF \( P_{M|E}(m|e) \) is given by

\[
P_{M|E}(m|e) = \frac{1}{\sqrt{2\pi\sigma^2_Q}} e^{-\frac{(m-e)^2}{2\sigma^2_Q}}.
\]  

In other words, the conditional PDF between the measured signal \( M \) and the electronic noise \( E \) is given by the PDF of the quantum signal. By setting \( \sigma^2_Q = 1 \), we normalize all the relevant quantities by the quantum noise.

From Eq. (2), the discretised conditional probability distribution is thus

\[
P_{M_{\text{dis}}|E}(m_i|e)
= \begin{cases}
    \int_{-\infty}^{m_i+\delta/2} P_{M|E}(m|e) \, dm, & i = i_{\text{min}}, \\
    \int_{m_i-\delta/2}^{m_i+\delta/2} P_{M|E}(m|e) \, dm, & i_{\text{min}} < i < i_{\text{max}}, \\
    \int_{R-3\delta/2}^{R} P_{M|E}(m|e) \, dm, & i = i_{\text{max}}.
\end{cases}
\]  

With these we are now ready to discuss how \( R \) should be chosen under two different definitions of min-entropy, namely \textit{worst-case} min-entropy and \textit{average} min-entropy.

### B. Worst-case conditional min-entropy

The min-entropy for variable \( X \) with distribution \( P_X(x_i) \), in unit of bits, is defined as \[34]\:

\[
H_{\text{min}}(X) = -\log_2 \left[ \max_{x_i \in X} P_X(x_i) \right].
\]  

Operationally, this corresponds to entropy associated with the maximum guessing probability for an eavesdropper about \( X \). It also tells us about how much (almost) uniform randomness can be extracted out of the distribution \( P_X(x_i) \) \[30\]. To obtain a lower bound for the randomness in our entropy source, we first look into the worst-case min-entropy conditioned on classical side information \( K \), which is defined as \[37\]:

\[
H_{\text{min}}(X|K) = -\log_2 \left[ \max_{k_j \in \text{supp}(P_K)} \max_{x_i \in X} P_X(x_i|k_j) \right],
\]  

where the support \( \text{supp}(f) \) is the set of values \( x_i \) such as \( f(x_i) > 0 \). In the case of Gaussian distributions, the support of the probability distribution will be \( \mathbb{R} \). Following Eq. (4), upon discretization of the measured signal \( M \), the worst-case min-entropy conditioned on classical noise \( E \) is

\[
H_{\text{min}}(M_{\text{dis}}|E) = -\log_2 \left[ \max_{e \in \mathbb{R}} \max_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e) \right].
\]  

Here we assumed that from eavesdropper’s perspective, the electronic noise is known fully with arbitrary precision. Performing the integration in Eq. (4), the maximization over \( M_{\text{dis}} \) in Eq. (7) becomes

\[
\max_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e)
= \max_{e} \left[ \frac{1}{2} \left( 1 - \text{erf} \left( \frac{e+R-\delta/2}{\sqrt{2}} \right) \right) \right],
\]  

\[
= \max_{e} \left[ \frac{1}{2} \left( \text{erf} \left( \frac{e-R+\delta/2}{\sqrt{2}} \right) + 1 \right) \right],
\]
where \( \text{erf}(x) = \frac{2}{\sqrt{\pi}} \int_{0}^{x} e^{-t^2} dt \) is the error function. We note that we have \( \text{Max}_{e} \text{Max}_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e) = 1 \), achieved when \( e \to -\infty \) or \( e \to \infty \). This results in \( H_{\text{min}}(M_{\text{dis}}|E) = 0 \) [See inset of Fig. 3(a)]. Indeed it is intuitive to see that in the case where the electronic noise \( e \) takes on an extremely large positive value, the outcome of \( M_{\text{dis}} \) is almost certain to be \( m_{i_{\text{max}}} \) with large probability. However, this scenario happens with a very small probability. Hence for practical purposes, one could bound the maximum excursion of \( e \), for example \(-5\sigma_E \leq e \leq 5\sigma_E \), which is valid for 99.9999% of the time. With this bound on the electronic noise, we now have

\[
H_{\text{min}}(M_{\text{dis}}|E) = -\log_2 \left[ \text{Max}_{e \in [e_{\text{min}}, e_{\text{max}}]} \text{Max}_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e) \right] 
= \text{Max}_{e \in [e_{\text{min}}, e_{\text{max}}]} \text{Max}_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e) \left\{ \frac{1}{2} \left[ \text{erf} \left( \frac{e_{\text{max}} - R + 3\delta/2}{\sqrt{2}} \right) + 1 \right] ; \text{erf} \left( \frac{\delta}{2\sqrt{2}} \right) \right\}, \tag{10}
\]

which can be optimized by choosing \( R \) such that

\[
\frac{1}{2} \left[ \text{erf} \left( \frac{e_{\text{max}} - R + 3\delta/2}{\sqrt{2}} \right) + 1 \right] = \text{erf} \left( \frac{\delta}{2\sqrt{2}} \right). \tag{11}
\]

This optimized worst case min-entropy \( H_{\text{min}}(M_{\text{dis}}|E) \) is directly related to the extractable secure bits independent of the classical noise. As shown in Fig. 3(a), when Eq. (10) is not optimized with respect to \( R \), the saturation in the first (last) bin for \( e_{\text{min}}/e_{\text{max}} = \pm 10\sigma_E \) becomes the peaks of the conditional probability distribution, hence compromising the attainable min-entropy. By choosing the optimal value for \( R \) via Eq. (11), as depicted in Fig. 3(b), the peaks at the first and last bins will always be lower than or equal to the probability within the dynamical range. Thus, by allowing the dynamical ADC range to be chosen freely, one could obtain the lowest possible conditional probability distribution, and hence producing the highest amount of possible secure random bits per sample for a given QCNR and \( n \)-bit ADC. Eq. (10) can be further generalized to take into account the DC offset of the device \( \Delta \), which could be due to intrinsic offset of the electronic signal, or even a deliberate constant offset induced by the eavesdropper over the sampling period [See Appendix A].

In Fig. 3(a), we show the extractable secure random bits for different digitisation \( n \) under the confidence interval of \( 5\sigma_E \leq |e + \Delta| \leq 20\sigma_E \). At high QCNR regime, the electronic noise contribution does not compromise the extractable bits too much. As the electronic noise gets more and more comparable to the quantum noise, although more bits have to be discarded, one can still extract a decent amount of secure random bits. More surprisingly, even if the QCNR goes below 0, that is, classical noise becomes larger than quantum noise, in princi-
Figure 4. (a) Optimized $H_{\text{min}}(M_{\text{dis}}|E)$ and (b) Normalized $H_{\text{min}}(M_{\text{dis}}|E)$ as a function of QCNR for different $n$-bit ADC. Shaded areas: $5\sigma_{E} \leq |e + \Delta| \leq 20\sigma_{E}$. The extractable bits is robust against the excursion of the electronic noise, especially when the QCNR is large. A non-zero amount of secure randomness is extractable even when the classical noise is larger than the quantum noise. The extractable secure randomness per bit increases as the digitisation bits $n$ is increased.

Example, one can still obtain a non-zero amount of random bits independent of classical noise. From Fig. 4(b), we notice extractable secure randomness per bit increases as we increase the digitisation bits $n$. This interplay between the digitisation bit $n$ and signal-to-noise-ratio QCNR is further explored in Fig. 5, where normalized $H_{\text{min}}(M_{\text{dis}}|E)$ is plotted against $n$ for several values of QCNR. We can see that for higher ratios of quantum to classical noise, a less amount of digitisation bits is required to achieve a certain value of secure randomness per bit. In other words, even if QCNR could not be improved further, one can achieve a higher ratio of secure randomness per bit simply by increasing $n$.

C. Average conditional min-entropy

As described in Section 11B without a bound on the range of electronic noise, one cannot extract any secure randomness. However, if we assume that an eavesdropper can only listen to, but has no control over the electronic noise we can compute the average of Eq. (5) over the classical side information $E$ instead. This leads us to the average guessing probability of $M_{\text{dis}}$ given $E_{\text{dis}}$.

$$p_{\text{guess}}(M_{\text{dis}}|E_{\text{dis}}) = \left[ \sum_{e_{j} \in E_{\text{dis}}} p_{E_{\text{dis}}}(e_{j}) \max_{m_{i} \in M_{\text{dis}}} p_{M_{\text{dis}}|E_{\text{dis}}}(m_{i}|e_{j}) \right],$$

which denotes the probability of correctly predicting the value of discretized measured signal $M_{\text{dis}}$ using the optimal strategy, given access to discretized electronic noise $E_{\text{dis}}$. Here $p_{E_{\text{dis}}}(e_{j})$ is the discretized probability distribution of the electronic noise. The extractable secure randomness from our device is then quantified by the average conditional min-entropy

$$H_{\text{min}}(M_{\text{dis}}|E_{\text{dis}}) = -\log_{2} p_{\text{guess}}(M_{\text{dis}}|E_{\text{dis}}).$$

We now proceed to find the optimized average conditional min-entropy. Again, we assume that the eavesdropper can measure the full spectrum of the electronic noise, with arbitrary precision. This gives the eavesdropper maximum power, including an infinite ADC range $R_{e} \rightarrow \infty$ and infinitely small binning $\delta_{e} \rightarrow 0$. As detailed in Appendix B under these limits, Eq. (13) takes the form of

$$H_{\text{min}}(M_{\text{dis}}|E) = \lim_{\delta_{e} \rightarrow 0} H_{\text{min}}(M_{\text{dis}}|E_{\text{dis}})$$

$$= -\log_{2} \left[ \int_{-\infty}^{\infty} p_{E}(e) \max_{m_{i} \in M_{\text{dis}}} p_{M_{\text{dis}}|E}(m_{i}|e) de \right].$$

Figure 5. Normalized worst-case conditional min-entropy $H_{\text{min}}(M_{\text{dis}}|E)$ as a function of $n$-bit ADC for different QCNR values. Here, $|\Delta| = 0, |e| \leq 5\sigma_{E}$. The interplay between the QCNR and digitisation bits $n$ was shown, where one could improve the rate of secure randomness per bit by either increasing the QCNR or $n$. Inset: Zoom in for $H_{\text{min}}(M_{\text{dis}}|E)/n \geq 0.85$ (Dashed line). Even when the classical noise is more dominant compared to the quantum noise (QCNR = −3dB), surprisingly, within our framework, 85% of randomness per bit could be recovered by having at least ~22 bits of digitisation.
The full expression of Eq. (14) is shown in Eq. (B7). Optimized result for average min-entropy $\tilde{H}_{\text{min}}(M_{\text{dis}}|E)$ with the corresponding dynamical ADC range $R$ is depicted in Table I. Similar to worst-case min-entropy case discussed in Sec. II.B, one can still obtain a significant amount of random bits even if the quantum signal to classical noise ratio is 1 (0 dB). On the contrary, a conventional unoptimized QRNG requires high operating QCNR to access high bitrate regime. When QCNR $\rightarrow \infty$, the measured signal does not depend on the electronic noise and the result coincides with that of the worst-case conditional min-entropy. In fact, the worst-case conditional min-entropy [Eq. (7)] is the lower bound for the average conditional min-entropy [Eq. (13)]. In the absence of side-information $E$, both entropies will reduce to the usual min-entropy Eq. (5). Compared to the worst-case min-entropy, the average conditional min-entropy is more robust against degradation of QCNR, hence it allows one to extract more secure random bits for a given QCNR. This is because the eavesdropper is limited to monitoring in this scenario, which is a reasonable assumption for most noise sources. Calibration of the classical noise is therefore required, and necessary in evaluating this quantity.

Table I. Optimised $\tilde{H}_{\text{min}}(M_{\text{dis}}|E)$ (and $R$) for 8 and 16 bits ADC

| QCNR (dB) | $n = 8$ | $n = 16$ |
|-----------|---------|---------|
| $\infty$  | 7.03 (2.45) | 14.36 (3.90) |
| 20        | 6.93 (2.59) | 14.28 (4.09) |
| 10        | 6.72 (2.93) | 14.11 (4.55) |
| 0         | 6.11 (4.33) | 13.57 (6.48) |
| $\infty$  | 0       | 0       |

III. EXPERIMENTAL IMPLEMENTATION

A. Physical setup and characterization

As depicted in Fig. 6, our CV-QRNG setup consists of homodyne detection of the quantum vacuum fluctuation, followed by post-processing. A 1550 nm fibre coupled laser (NP Photonic Rock) operating at 60mW serves as the local oscillator of the homodyning setup. The light is sent into one port of a 50/50 beam splitter, while the second input port is physically blocked. The outputs from the beam splitter are then optically coupled to two identical 3 GHz broadband photo-detectors with a common mode rejection ratio of 30 dB, where the sum and the difference of the signals were recorded. The electronic output is further split and mixed down at 1.375 GHz and 1.625 GHz, where technical noise is less prominent and the laser is shot-noise limited [Fig. 7(a)]. With an appropriately chosen dynamical ADC range parameter $R$, maximal amount of secure entropy from two sideband regions (Channel 0: 1.25-1.50 GHz) and (Channel 1: 1.50-1.75 GHz) are recorded by two 16-bit ADC (National Instrument 5762) at the rate of 250 MSamples per second. Before the digitisation, the signals are sent to two low pass filters with cut-off frequency of 125MHz to prevent aliasing of the signals so that the correlations between the sampling points can be minimized \cite{38}. Finally, the data processing is performed using a National Instrument field-programmable gate array (FPGA).

The average QCNR clearances for Channel 0 (Ch 0) and Channel 1 (Ch 1) are 13.52 dB and 13.32 dB respectively. Taking into account the intrinsic DC offsets, which is $-0.02\sigma_Q$ for both channels, we quantify our conditional min-entropies using the method described in Sec. II. For our ADC with 16 bits of digitisation, the worst-case conditional min-entropies are 13.76 bits (Ch 0) and 13.75 bits (Ch 1), while the average conditional min-entropies are 14.19 bits for both channels. Here, since we assumed a trusted detection scenario, we evaluate our entropy with average conditional min-entropy and set $R$ as $4.32\sigma_Q$ according to Eq. (B8).

B. Upper bound of extractable min-entropy

The extractable randomness of our QRNG is limited by the sampling rate and the digitisation bit, which is defined by Nyquist’s theorem on maximum data rate $C$, 

$$C = 2H\log_2 V,$$  \hspace{1cm} (15)
where $H$ is the bandwidth of the spectrum and $V = 2^6$ is the quantization level. For our 16 bits ADC, the shot-noise limited and technical noise free bandwidth is around 2.5 GHz out of 3 GHz. With an average of 10 dB of QCNR clearance, one can extract 14.11 bits out of 16 bits (Table I). Combining this with Eq. 15, our setup can deliver up to 70 Gbps extractable random bits.

One can further improve the maximum bit rate by increasing the digitisation bit $n$, which is ultimately upper bounded by the photon-number within a given detection time window. For our setup powered 1550 nm fibre coupled laser of 60 mW, a detection bandwidth of 3 GHz corresponds to $1.6 \times 10^8$ photons per sampling. Given a perfect photon number resolving detector, the maximal entropy is given by $\log_2(1.6 \times 10^8) \approx 27.2$ bits [25]. In principle, one could send more power to extract more random bits, however this bound can only increase logarithmically with laser intensity.

C. Randomness Extraction

It is commonly the case that QRNGs are not ideal sources of randomness, in the sense that the distribution is often biased, while uniform randomness is required for application purposes. In our situation, the quantum vacuum state measured by our CV-QRNG exhibits a Gaussian distribution. To generate ideal randomness, post-processing of the raw outputs is necessary to produce shorter, yet almost uniformly distributed random strings. Ad-hoc algorithms such as the Von Neumann extractor, XOR corrector and least significant bit (LSB) operation are widely used [11, 15, 16, 30, 31, 39]. These methods, although simple in principle, might fail to produce randomness at all if non-negligible correlations exist among the raw bits [40].

From an information-theoretic standpoint, universal hashing functions are desirable candidates for randomness extraction [25, 32]. These functions act to recombine bits within a sample according to a short yet randomly chosen seed, and map them to shorter, almost-uniform random strings. They have been used in recent development of QRNGs [19, 21, 25, 26, 41], with functions such as the Toeplitz-hashing matrix. These constructions, however, suffer from large entropy loss and often require long seed [42]. Meanwhile, recent implementations of a information theoretic randomness extractor called Trevisan’s construction [25, 33, 35] has received considerable attention. This particular construction of a strong extractor has been proven secure against quantum side-information, furthermore it requires a relatively short seed. Despite so, the complexity of the algorithm imposes a very stringent limit on the extraction speed (0.7 kb/s [25] and 150 kb/s [39]).

Other attractive alternatives for secure randomness extraction are cryptographic hashing functions [13, 40]. These hashing functions are well suited in many cryptographic applications and settings where the adversary is assumed to be computationally bounded. They also enjoy high throughput due to efficient hardware implementation. Previously, cryptographic hashing extractors have been deployed in [11, 13, 18, 23], with functions such as SHA-512 and Whirlpool. Most of the implementation keep exactly min-entropy number of bits, which might not be fully secure (see Appendix C).

Here, we demonstrate randomness extraction with Advanced Encryption Standard (AES) [47] cryptographic hashing algorithm of 128 bits (See Appendix D). Since a detailed crypto-analysis of our framework is non-trivial and beyond the scope of this paper, we keep only half of our conditional min-entropy [13, 48] to obtain an almost perfectly uniform output. The final real-time guaranteed secure random number generation rate of our CV-QRNG is 3.55 Gbps. If all the available bandwidth from our detector (~ 2.5 GHz) could be sampled, we will be able to achieve up to 35Gbps/s (c.f. Sec. III B). This corresponds to a rate of 14 Mbps/ MHz in term of bits per bandwidth. Our random numbers consistently pass the standard statistical tests (NIST [49], DieHard [50]) and the results are available on ANU Quantum Random Number Server (http://qrng.anu.edu.au).

IV. CONCLUDING REMARKS

In this work, we have proposed a generic framework for secure random numbers generation, taking into account the existence of classical side-information which in principle could be manipulated or predicted by an eavesdropper. If the eavesdropper is assumed to have...
full control over the classical noise, for example, the detectors’ noise could be originating from pre-established values by a malicious vendor, the worst-case conditional min-entropy should then be used to quantify available secure randomness. Meanwhile, if we restrict the third party to passive eavesdropping, one can use the average conditional min-entropy instead to quantify extractable randomness. By using the dynamical ADC range as a free parameter, we have shown that quantum signal-to-classical noise ratio is not the sole decisive factor in generating secure random bits. Surprisingly, one can still extract a non-zero amount of secure randomness even when the classical noise is comparable to the quantum noise. This is done simply by optimising the dynamical ADC range via conditional min-entropies. Such an approach not only provides a rigorous justification for choosing the suitable ADC parameter, but also largely increases the range of QCNR for which true randomness can be extracted, thus relaxing the condition of high QCNR clearance in conventional CV-QRNGs. We also notice that we can increase the min-entropy per bit simply by increasing the number of digitization bits. We apply these observations to analyse the amount of randomness produced by our CV-QRNG setup. Efficient cryptographic hashing functions are then deployed to extract randomness, quantified by average conditional min-entropy.

We note several possible extensions of our work. For instance, one can apply entropy smoothing \[11,12\] on the worst-case min-entropy to tighten the analysis. Our framework can also be generalized to encapsulate potential quantum side information by considering the analysis described in \[20\]. A detailed crypto-analysis of our framework could also increase the final throughput of the QRNG \[45\]. Lastly, a hybrid of information-theoretic provable and cryptographic randomness extractor is also an interesting avenue to be explored in the construction of a high speed, side-information (classical and quantum) proof QRNG \[42\].

To conclude, this work allows the maximisation of extractable high quality randomness without compromising both the integrity and the speed of a QRNG. In fact, within our framework, when the QRNG is appropriately calibrated, the generated random numbers are secure even if the electronic noise is full known. This is of practical importance, given the fact that QRNGs play a decisive role in the implementation of cryptographic protocols such as quantum key distribution. From a practical point of view, our method also relaxes the QCNR requirement on the detector, thus allowing QRNGs that are more cost-effective and smaller in size. As such, we believe that our work paves the way towards a reliable, high bit rate and environmental-immuned QRNG \[53\] for information security.

In the completion of our work, we notice several recent papers considering the security aspects of QRNGs \[54–57\].

ACKNOWLEDGEMENT

This research was conducted by the Australian Research Council Centre of Excellence for Quantum Computation and Communication Technology (project number CE110001027). N. Ng is funded by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant “Random numbers from quantum processes” (MOE2012-T3-1-009).

Appendix A: Optimised conditional min-entropy with DC offset

![Figure 8. Model of the n bits ADC, with analog input in the ADC dynamical range \((-R, R)\) and bin width \(\delta = R/2^{n-1}\). Offset of the distribution is modeled by another reference frame \(m'\) centered at offset \(\Delta\). In original frame \(m\), the lowest and highest bins are now centered around \(-R - \Delta\) and \(R - \delta - \Delta\).](image)

In a realistic scenario, the mean of the measured signal’s probability distribution is often non-zero. It is possible that such an offset might be induced by a malicious party over the sampling period. The model is depicted in Fig. 8, where the offset \(\Delta\) of the distribution is captured by another reference frame \(m'\) centered at \(\Delta\). In this model, Eq. 1 can now be rewritten as

\[
P(\Delta)_{\text{Mdis}}(m|e) = \begin{cases} 
\int_{-\infty}^{-R-\Delta + \delta/2} P_{M'|E}(m'|e)dm', & i = i_{\text{min}}, \\
\int_{m_{\text{min}}-\Delta + \delta/2}^{m_{\text{max}}-\Delta - \delta/2} P_{M'|E}(m'|e)dm', & i_{\text{min}} < i < i_{\text{max}}, \\
\int_{R-3\delta/2 - \Delta}^{R-\Delta + \delta/2} P_{M'|E}(m'|e)dm', & i = i_{\text{max}}.
\end{cases}
\]  
(A1)

Following the steps in Sec. II and bounding \(\Delta\), we finally arrives at the generalization of Eq.(10),

\[
H_{\text{min}}(M_{\text{dis}}|E) = -\log_2 \max(c_1, c_2),
\]  
(A2)

Here \(c_1 = \frac{1}{2} \left[ \text{erf} \left( \frac{\epsilon_{\text{max}}+\Delta_{\text{max}}-R+3\delta/2}{\sqrt{2}} \right) + 1 \right]\) and \(c_2 = \text{erf} \left( \frac{\delta}{2\sqrt{2}} \right)\), and \(\text{erf}(x)\) is the error function. The results are tabulated in Table I and II.
By invoking Theorem 1, there exists $\delta$ in the infinite dynamical ADC range and digitization bits, i.e. $e_{\text{eavesdropper}}$ possesses a device with order to achieve the lower bound of the average conditional

$$\text{Theorem: For any continuous function } f(x), \text{ we make use of the mean value theorem, stated as below: Theorem 1 Mean Value Theorem: For any continuous function } f(x) \text{ on an interval } [a, b], \text{ there exists some } \bar{x} \in [a, b] \text{ such that,}$$

$$\int_a^b f(x) dx = (b-a)f(\bar{x}) \tag{B3}$$

By invoking Theorem 1, there exists $\bar{e}_j \in [e_j - \delta_e/2, e_j + \delta_e/2]$ such that Eq. (B2) can be written as

$$P_{E_{\text{disc}}}(e_j) = P_E(\bar{e}_j)\delta_e. \tag{B4}$$

Substituting this back to Eq. (12), we ended up with

$$p_{\text{guess}}(M_{\text{dis}}|E_{\text{dis}}) = \sum_{e_j \in E_{\text{dis}}} P_E(\bar{e}_j)\delta_e \max_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E_{\text{dis}}}(m_i|e_j). \tag{B5}$$

Now together with the assumption of infinite binning $\delta_e \to 0$, the sum becomes an integral,

$$p_{\text{guess}}(M_{\text{dis}}|E) = \lim_{b_e \to 0} p_{\text{guess}}(M_{\text{dis}}|E_{\text{dis}}) = \int_{-\infty}^{\infty} P_E(e) \max_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e) de. \tag{B6}$$

Together with Eq. 8, we finally arrive at

$$p_{\text{guess}}(M_{\text{dis}}|E) = \left[ \int_{-\infty}^{e_1} P_E(e) \max_{m_i \in M_{\text{dis}}} P_{M_{\text{dis}}|E}(m_i|e) de \right]$$

$$= \frac{1}{2} \left[ \int_{-\infty}^{e_1} P_E(e) \left[ 1 - \text{erf} \left( \frac{e + R - \delta_e/2}{\sqrt{2}} \right) \right] de + \text{erf} \left( \frac{e_1}{\sqrt{2}\sigma_E} \right) \text{erf} \left( \frac{\delta_e}{\sqrt{2} \sigma_E} \right) \right]$$

$$+ \int_{e_1}^{\infty} P_E(e) \left[ \text{erf} \left( \frac{e - R + 3\delta_e/2}{\sqrt{2}} \right) + 1 \right] de,$$

where $e_1$ and $e_2$ are chosen to satisfy the maximization upon $M_{\text{dis}}$ for a given $R$. The optimal $R$ is then determined numerically. This result could be easily generalised to take into account a DC offset with the steps described above.
in Appendix A giving

\[
p_{\text{guess}}(M_{\text{dis}}|E) = \frac{1}{2} \left( \int_{-\infty}^{\epsilon_1} P_E(e - \Delta) \left[ 1 - \operatorname{erf} \left( \frac{e + \Delta + R - \delta/2}{\sqrt{2}} \right) \right] de + \left[ \operatorname{erf} \left( \frac{\epsilon_2 - \Delta}{\sqrt{2}\sigma_E} \right) - \operatorname{erf} \left( \frac{\epsilon_1 - \Delta}{\sqrt{2}\sigma_E} \right) \right] \operatorname{erf} \left( \frac{\delta}{2\sqrt{2}} \right) + \int_{\epsilon_2}^{\infty} P_E(e - \Delta) \left[ \operatorname{erf} \left( \frac{e + \Delta - R + 3\delta/2}{\sqrt{2}} \right) + 1 \right] de \right) .
\]

\begin{equation}
(B8)
\end{equation}

**Appendix C: Notes on Leftover Hash Lemma**

From an information-theoretic standpoint, the most prominent advantage of universal hashing functions described in Sec. 11 is that the randomness of the output guaranteed unconditionally by Leftover Hash Lemma (LHL). More specifically, LHL states that for any \( \epsilon > 0 \), if the output of a universal hashing function has length

\[
l \leq t - 2\log_2(1/\epsilon),
\]

where \( t \) denotes the (conditional) min-entropy, then the output will be \( \epsilon \)-statistically close to a perfectly uniform distribution. Moreover, a universal hashing function constructs a strong extractor, where the output string is also independent of the seed of the function \[23, 32].

On the other hand, for a strong cryptographic extractor, the output is \( \epsilon \)-computationally indistinguishable from the uniform distribution (See Ref. \[44, 45\] for formal definitions). It is shown in Refs. \[45, 46, 51\] that LHL could be generalized to take into account almost universal functions (functions statistically \( \xi \)-close to being universal hashing functions), where the min-entropy \( t \) in LHL as expressed in Eq. (C1) is replaced by \( \min(t, \log_2(1/\xi)) \). Under suitable parameter constraints and operating modes, an \( \epsilon \)-cryptographic extractor can be treated as a \( \xi \)-almost universal function, and hence a strong randomness extractor.

As we can see, in both cases, it is necessary to discard some bits with respect to the security parameter \( \epsilon \) or \( \epsilon' \) to ensure the security and uniformity of the output.

**Appendix D: AES hashing and auto-correlation**

In our QRNG, randomness extraction is performed with Advanced Encryption Standard (AES)\[47\] cryptographic hashing algorithm of 128-bit, seeded with a 128-bit secret initialization vector. Before hashing, our data is collected from two channels into a block consisting of eight 16-bit digitized samples. Four most significant bits of the samples are discarded before randomness extraction to ensure low autocorrelation among consecutive samples [Fig. 9] before hashing \[45\]. The resulting output was concatenated with partial raw data from the previous run, forming a 128-bit block for cryptographic hashing. Since the complete crypto-analysis of the cryptographic hashing is intricate, we discard half of the output to ensure uniformity of the generated random sequence. We further strengthened our security by renewing the seed of our AES extractor with part of these discarded bits. Thus, the final real-time throughput of our QRN is 3.55 Gbps.

[1] M. Stipcevic, in MIPRO, 2011 Proceedings of the 34th International Convention (IEEE, 2011), pp. 1474–1479.

[2] P. Xu, Y. Wong, T. Horiuchi, and P. Abshire, Electronics Letters 42, 1346 (2006).
