Randomness of post-selected data

Le Phuc Thinh,[1] Gonzalo de la Torre,[2] Jean-Daniel Bancal,[1] Nicolas Brunner,[3] and Valerio Scarani[1,4]

[1]*Centre for Quantum Technologies, National University of Singapore, 3 Science Drive 2, Singapore 117543*
[2]*ICFO-Institut de Ciencies Fotoniques, Mediterranean Technology Park, 08860 Castelldefels (Barcelona), Spain*
[3]*Département de Physique Théorique, Université de Genève, 1211 Genève, Switzerland*
[4]*Department of Physics, National University of Singapore, 2 Science Drive 3, Singapore 117542*

I. INTRODUCTION

Sources of true randomness have numerous applications, be they in algorithms, sampling, gambling or cryptography [1] where the last two demand sources that can be certified as being uncorrelated to any outside process or variable, i.e. private randomness. Quantum physics offers this opportunity when a Bell inequality is violated [2].

The number of random bits which can be obtained is characterized by the min-entropy which can be bounded by the violation of some Bell inequality or the observed statistics [3–6]. To obtain a truely private and random sequence for use in the aforementioned applications, we must apply a randomness extractor to the outcomes of Bell tests, which requires an independent random seed. Generally, the longer the initial string, the longer the needed seed and the computational time to output the result; in fact, it is an active research direction to construct randomness extractor with short seed length [1].

In a practical Bell experiment, because of the inefficiencies of the source and the detectors, the recorded data will mostly be populated with the no detection events $\varnothing$. From physics, we know that these no-detection events carry little or no randomness: for instance, they may be associated to the source not having emitted any pair in a given time. However, in a device-independent certification, one cannot simply ignore those events, because this opens the detection loophole. Here we provide a method to quantify the randomness present in a subset of events (for instance, double detections) which takes into account the whole observed statistics (including the no-detection events) and thus does not open the detection loophole. A priori, there are two natural scenario we can consider, depending on whether we reveal to Eve the runs that have been kept or discarded. For simplicity, in this work we focus only on the scenario where Eve is allowed to learn the runs we have kept, and we assume i.i.d.. For several physically-motivated models of the observed statistics, we show that one can indeed vindicate the idea that most of the randomness is present in the double-detection events. In other words, the current analysis allows us to extract *more randomness from the outcomes of Bell tests* than a previous analysis [3], and also permit hashing the post-selected subset of the original data thereby reducing the needed seed length, and also the time required to compute the final output.

II. THE GUESSING PROBABILITY FOR POST-SELECTED DATA

For concreteness we consider a bipartite Bell test where each party has two measurement settings, which has three possible outcomes 0,1 and $\varnothing$. The set $\{0,1,\varnothing\}^2$ of joint outcomes is partitioned into K, the set of outcomes that is kept and D, the set of outcomes that is discarded. As mentioned, Eve is allowed to know whether we have kept or discarded the event, even though this additional information may not be intentionally broadcasted to Eve[7].

Claim: If the raw data of a Bell test satisfy the original model $p(ab|xy) = \sum_\lambda p(\lambda)p(ab|xy\lambda)$ then the randomness extractable from the post-selected outcomes of each pair of settings xy is the min-entropy of the guessing probability

$$G_{xy} = \frac{1}{p(K|xy)} \sum_\lambda p(\lambda) \max_{ab\in K} p(ab|xy\lambda). \tag{1}$$

The main idea of the proof is to describe the model of the post-selected data[8], namely $p(ab|xyK)$, in terms of the original model by correctly renormalizing $p(\lambda)$ and $p(ab|xy\lambda)$ given the (revealed) information of keep or discard.

The optimal guessing probability for Eve given an observed $p(ab|xy)$ can be obtained by optimizing (1) over all decompositions respecting $p(ab|xy)$ which can be relaxed to an SDP and implemented efficiently using the techniques in [3].

III. NUMERICAL TESTS

We computed lower bounds on the extractable randomness from the outcomes of the first setting (in terms of asymptotic rate) for various situations. Every situation is a Bell-like experiment in which a source creates some

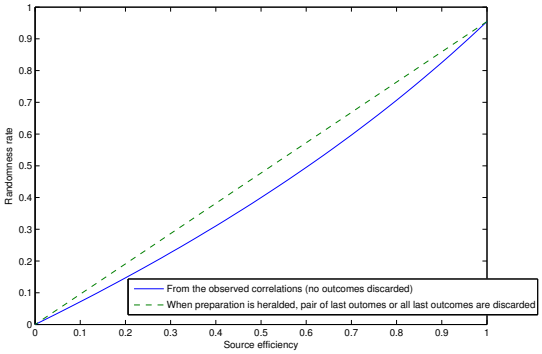

FIG. 1: Measurement of a singlet with 99% visibility, which is produced with finite probability.

bipartite state with efficiency ν which is detected by the parties with one of two possible measurements. The detectors might have a finite efficiency η, allowing for the observation of a third outcome (absence of any detection). Concretely, if $p(ab|xy)$ for $a, b, x, y \in \{0, 1\}$ is the ideal correlation (when $\eta = \nu = 1$) obtained from some choice of state and measurements, then

$$q(ab|xy) = \begin{cases} \nu\eta^2 p(ab|xy) & \text{if } ab \in \{0,1\}, \\ \nu(1-\eta)^2 + (1-\nu) & \text{if } ab = \varnothing\varnothing, \\ 2\nu\eta(1-\eta) & \text{otherwise} \end{cases} \tag{2}$$

is the correlation when we have imperfect detectors and imperfect source. For a given choice of state, measurements, η and ν, we compute the randomness that can be extracted:

- From the complete string of outcomes.
- From the complete string of outcomes upon heralding of a successful preparation of the state.
- From the string of outcomes in which the double occurence of $\varnothing$ are removed (i.e. $\varnothing\varnothing$).
- From the string of outcomes in any occurence of a no-detection event $\varnothing$ is removed (i.e. $0\varnothing$, $1\varnothing$, $\varnothing\varnothing$, $\varnothing 0$, $\varnothing 1$).

Removing some outcomes reduces the length of the considered data. To facilitate the comparison with cases where no outcome is discarded, we choose to renormalize the randomness rate obtained on post-selected data with respect to total number of runs.

A. Perfect detectors, imperfect source

In Figure 1, we show the result when a singlet with visibility 99% is produced, the detection efficiency is $\eta = 1$ and ν varies. In this case, the lower bound on the randomness computed from the full statistics is lower than what can be certified after removing double no-detections from the data. Thus, *performing this post-processing on the data allows one to certify a larger amount of randomness than that which is apparent at first sight.* In fact, after discarding double no-detections, the same amount of randomness that could be certified if the source was heralded is recovered by our analysis. Thus, the effect of no-detections seems to be cleanly identified and removed by our analysis[9].

B. Imperfect detectors, perfect source

In Figure 2, we consider the case where the source efficiency is $\nu = 1$ and η varies, for either measurement on a singlet with 99% visibility, or on partially entangled states (using then Eberhard's choice of measurements). In this case, no-detection events come solely from the imperfection of the detectors, and the source can be considered to be already heralded. This time, the largest amount of randomness is certified when either all outcomes are taken into account, or double no-detection events are discarded. In particular, it seems that the same amount of randomness can be extracted from the full string of outcomes as from the (smaller set of) data in which $\varnothing\varnothing$ is removed. This indicate that *in a heralded Bell test with finite detection efficiency, experimentalists can indeed discard double no-detection events, provided that they keep a count of how many of them appear (they keep their statistics).* Note also that in this figure, no randomness can be extracted when the detection efficiency is lower than either 82.8% or 66.6%, as is expected.

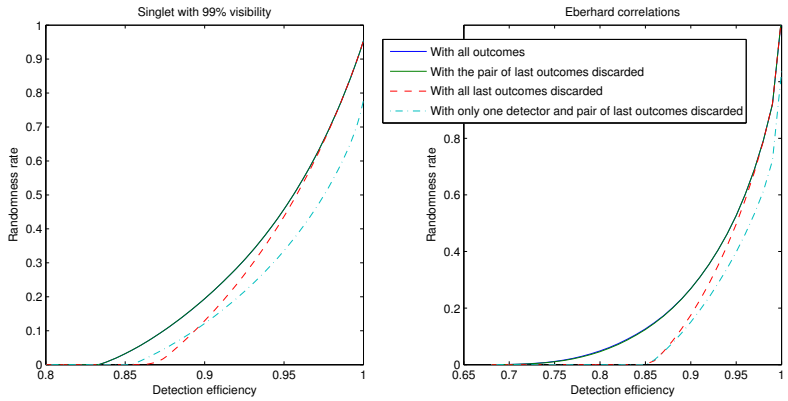

FIG. 2: Measurement of the singlet with 99% visibility (left), or of partially entangled state following Eberhard (right).

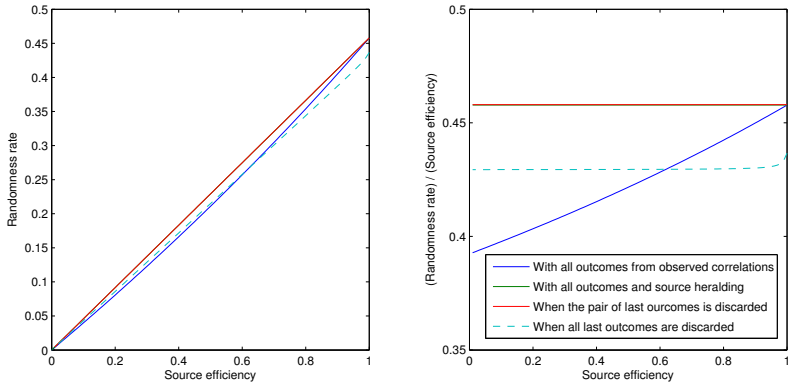

FIG. 3: Measurement of a singlet with 99% visibility and 95% detection efficiency η. Plot on the right is the same as on the left, but rescaled with respect to the source efficiency ν. The first two curves are superposed.

C. Imperfect detectors, imperfect source

In Figure 3, we considere a case where no-detection events can come either from the source or from the detectors. Namely, we set $\eta = 95\%$ and vary ν, for measurements on a singlet with visibility 99%. This time, the largest amount of randomness is obtained only by discarding the double no-detection events. In fact, *as much randomness as if the source was heralded can again be certified here.* Discarding all rounds where at least one party observed a no-detection leads to a lower randmness rate. For all cases considered so far, discarding double no-detections always produces the best result.

IV. CONCLUSION AND FUTURE WORK

In this work, we have studied the impact of post-selection on the randomness extractable from Bell tests. A natural follow up is to extend our results beyond i.i.d. scenarios and to the situation where we do not reveal to Eve which data is kept. Since post-selection is one possible processing one could do on the outcomes of Bell tests, our work also raise the question: what kind of post-processing will reveal the optimal amount of randomness from Bell tests?

Remarks: This work is an ongoing collaboration with Stephano Pironio (Laboratoire d'Information Quantique, Université Libre de Bruxelles). A working draft is available upon request.

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

[2] R. Colbeck, Ph.D. thesis, University of Cambridge (2006), arXiv:0911.3814.
[3] J-D. Bancal, L. Sheridan and V. Scarani, New J. Phys. 16, 033011 (2014).
[4] 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., Nature **464**, 1021 (2010).
[5] S. Pironio and S. Massar, Phys. Rev. A **87**, 012336 (2013).
[6] U. Vazirani and T. Vidick (2012), arXiv:1111.6054.
[7] This scenario is the most paranoid one can get, since we allow Eve to know everything except the random numbers which we are trying to generate.
[8] Post-selected data is the subset of the raw data after discarding outcomes in D.
[9] This is possible and can be seen analytically because of the special form of the observed correlation in this case.