Entropy certification of a realistic QRNG based on single-particle entanglement

Sonia Mazzucchi,[1] Nicolò Leone,[2] Stefano Azzini,[2] Lorenzo Pavesi,[2] and Valter Moretti[1]

[1] *Department of Mathematics and TIFPA-INFN, University of Trento, Italy*
[2] *Nanoscience Laboratory, Department of Physics, University of Trento, Italy*

(Dated: August 2, 2021)

In single-particle entanglement (SPE) two degrees of freedom of a single particle are entangled. SPE is a resource that can be exploited both in quantum communication protocols and in experimental tests of noncontextuality based on the Kochen-Specker theorem. SPE can be certified via a test of quantum contextuality based on Bell inequalities. Experiments of Bell-like inequality violation by single particle entangled systems may be affected by an analogue of the locality loophole in this context, due to the presence of unavoidable non-idealities in the experimental devices which actually produce unwanted correlations between the two observables that are simultaneously measured. This issue is tackled here by quantitatively analyzing the behaviour of realistic devices in SPE experiments with photons. In particular, we show how it is possible to provide a semi-device independent randomness certification of realistic quantum random number generators based on Bell inequality violation by SPE states of photons. The analysis is further enlarged to encompass, with a Markovian model, memory effects due to dead time, dark counts and afterpulsing affecting single photon detectors, in particular when not dealing with coincidence measurements. An unbiased estimator is also proposed for quantum transition probabilities out of the collection of experimental data.

I. INTRODUCTION

Entanglement is one of the fundamental features of quantum theory since it is able to produce correlations that do not have an analogue in classical physics. Initially pointed out as a source of paradoxes [1, 2], it has recently gained a relevant role in the blooming ares of quantum information and quantum computing [3–6]. From the mathematical point of view, the very definition of entanglement (or non-separability) relies upon the tensor product structure of the Hilbert space associated to the states of a quantum system and, in the simple case of a bipartite system, it is related to the notion of Schmidt rank [7]. From the physical point of view, we can distinguish between two kinds of entanglement. In *inter-particle entanglement* non-classical correlations are shared between the degrees of freedom of two different particles, while *intra-particle* or *single-particle entanglement* (SPE) involves independent degrees of freedom of a single particle such as momentum and polarization of a photon [8, 9] or momentum and spin of a massive particle [10, 11].

Analogously to interparticle entanglement, even in the case of SPE the violation of Bell-type inequalities highlights the presence of correlations between outcomes of measurements that cannot be described by means of a realistic non-contextual hidden variable theory. To this regard, several experiments of Bell inequality violation by single-particle entangled states have been recently proposed [11–13].

Usually, *Bell tests* are understood as tests of *local* realism when *two* (or more) spatially separated parties of a quantum system are considered. Here, instead, there is a single particle, made of two independent subsystems, with an entangled state. In this case, the violation of the Bell inequality is related with contextuality rather than non-locality.

In addition, SPE can be a resource in quantum information, as shown in [14–17]. Advantages and disadvantages of inter-particle entanglement and SPE are discussed in [17], where the experimental setup of fig. 1 is explicitly considered. We refer also to [16] for a review of the theory and few applications of SPE. Briefly, a clear advantage of SPE over inter-particle entanglement is the ease of production and the possibility of use of cheap sources such as an attenuated laser or a lamp. Moreover, in an experiment of Bell inequalities violation, the estimate of the quantum transition probabilities does not require coincidence detection. These facts allow to raise significantly the efficiency of production of meaningful events contributing to the detection statistics. This feature and the robustness under environment-induced decoherence effects [14] make SPE a resource in quantum information [15]. On the other hand, experiments of Bell inequality violation by SPE states necessarily rely on the fair sampling assumption, since the critical detection efficiency necessary to close the detection loophole [18] is valid only in the case of inter-particle entanglement.

In addition, tests of quantum contextuality on SPE states are affected by issues arising from non-idealities of the experimental devices that, in fact, produce unwanted correlations between the outcomes of measurements of observables related to independent degrees of freedom. Such effects are in principle absent in the case of Bell tests on interparticle entangled states due to the space-like separation of the two independent components, but they have to be taken into account in the case of SPE states.

In this paper, we present a quantitative analysis of this issue in the case of optical experiments on violation of the Clauser-Horne-Shimony-Holt (CHSH) inequality by SPE states of photons, in the lines of the experiment described in [17]. One of the main results of the paper is inequality (52), which provides an upper analytical bound for the threshold that the observed value of the CHSH parameter

S (see Eq.(4)) has to exceed in order to actually certify entanglement, taking into account the non ideal behavior of the beam splitters that are employed in the experiment.

This analysis plays a relevant role in the derivation of a realistic bound for the device independent guessing probability in a quantum random number generator (QRNG) based on CHSH inequality violation via single particle entangled photons [19]. As extensively discussed in [20–23], entanglement and Bell-inequality violation provide a powerful tool for randomness certification in QRNG based on a Bell test. More precisely, the min-entropy H_∞[24] of the raw data in a Bell experiment with measured CHSH parameter S is bounded from below by [20]:

$$H_\infty \geq -\log_2\left(\frac{1}{2}+\frac{1}{2}\sqrt{2-\frac{S^2}{4}}\right) \tag{1}$$

The bound (1) is *device-independent*, i.e., it does not depend on a particular modelling of the physical system realizing the QRNG. On the other hand, it requires a loophole-free Bell test which is not straightforward to realize in practical real-world implementations. As remarked in [25], the requirement of device-independence is too demanding and can be replaced by weaker assumptions adopted in modification of the original protocol presented in [20]. To this end, Bell-inequality violation by single particle entanglement represents a good trade-off between ease of experimental implementation and security. Indeed, as shown in the proof-of-principle experiment described in [19], a QRNG based on SPE can be practically realized and a robust estimate of the corresponding min-entropy can be presented, relying upon a rather small set of assumptions on the features of the optical components. To this end, the second main result of the present paper is inequality (59), which modifies (1) providing an analytical bound for the min-entropy produced by a QRNG based on SPE, such as the one reported in [19].

The paper is organized as follows. In section II a detailed theoretical analysis of the realized experiment [17] on CHSH inequality violation for SPE states of photons is presented in the case where all optical components do not deviate from the ideal behaviour. Section III tackles the issues arising from realistic beam splitters and mirrors. In particular, it considers the case where reflectance and transmittance depend explicitly on the polarization of the incoming photon, causing unwanted correlations between the degrees of freedom of "momentum" and "polarization". In addition, the presence of losses is taken into account. Section IV deals with the application of the previous results to the entropy certification of a semi-device-independent QRNG based on Bell inequality violation. In section V we present a Markovian model for the description of the memory effects due to the presence of detectors dead time and afterpulsing as well as a technique for the construction of an unbiased estimator for the quantum transition amplitudes. Finally, section VI concludes the paper.

II. BELL INEQUALITY VIOLATION BY SPE STATES

Let us consider a quantum system with two independent degrees of freedoms, denoted by A and B, each of them with two values, in such a way that the associated Hilbert space can be represented as the tensor product $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$. On each system we perform the measurement of observables $O^{\mathbf{a}} : \mathcal{H}_A \to \mathcal{H}_A$ and $O^{\mathbf{b}} : \mathcal{H}_B \to \mathcal{H}_B$ with two possible outcomes $x \in \{+1, -1\}$ and $y \in \{+1, -1\}$ respectively. In practice, such observables can be written in the form $O^{\mathbf{a}} = \mathbf{a}\cdot\boldsymbol{\sigma} = a_1\sigma_1 + a_2\sigma_2 + a_3\sigma_3$ and $O^{\mathbf{b}} = \mathbf{b}\cdot\boldsymbol{\sigma} = b_1\sigma_1 + b_2\sigma_2 + b_3\sigma_3$ for suitable unit vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$ and σ_i, with $i = 1, 2, 3$, are the Pauli matrices. We shall denote $P^{\mathbf{a}}_{+1}, P^{\mathbf{a}}_{-1}$ and $P^{\mathbf{b}}_{+1}, P^{\mathbf{b}}_{-1}$ the projectors belonging to the projection-valued measure (PVM) associated to $O^{\mathbf{a}} = P^{\mathbf{a}}_{+1} - P^{\mathbf{a}}_{-1}$ and $O^{\mathbf{b}} = P^{\mathbf{b}}_{+1} - P^{\mathbf{b}}_{-1}$, that can be represented as $P^{\mathbf{a}}_{\pm 1} = \frac{1}{2}(I \pm \mathbf{a}\cdot\boldsymbol{\sigma})$ and $P^{\mathbf{b}}_{\pm 1} = \frac{1}{2}(I \pm \mathbf{b}\cdot\boldsymbol{\sigma})$. Denoting by ρ a state of the compound system, we shall focus on the probabilities

$$P(x, y|\rho, \mathbf{a}, \mathbf{b}) := \mathrm{Tr}[\rho P^{\mathbf{a}}_x \otimes P^{\mathbf{b}}_y], \qquad x, y \in \{+1, -1\}. \tag{2}$$

The Bell inequality deals with two pairs of observables $O^{\mathbf{a}}_i = \mathbf{a}_i \cdot \boldsymbol{\sigma}$ and $O^{\mathbf{b}}_j = \mathbf{b}_j \cdot \boldsymbol{\sigma}$, with $i, j = 0, 1$, on the subsystems A and B respectively. We henceforth assume that the four vectors $\mathbf{a}_0, \mathbf{a}_1, \mathbf{b}_0, \mathbf{b}_1$ are given. If ρ is a quantum state on $\mathcal{H}_A \otimes \mathcal{H}_B$, we shall denote by $E(\mathbf{a}_i, \mathbf{b}_j)$ the expected value of $O^{\mathbf{a}}_i \otimes O^{\mathbf{b}}_j$

$$\begin{aligned} E(\mathbf{a}_i, \mathbf{b}_j) = \mathrm{Tr}[\rho O^{\mathbf{a}}_i \otimes O^{\mathbf{b}}_j] = P(+1, +1|\rho, \mathbf{a}_i, \mathbf{b}_j) \\ + P(-1, -1|\rho, \mathbf{a}_i, \mathbf{b}_j) - P(-1, +1|\rho, \mathbf{a}_i, \mathbf{b}_j) \\ - P(+1, -1|\rho, \mathbf{a}_i, \mathbf{b}_j), \end{aligned} \tag{3}$$

while $S(\mathbf{a}_0, \mathbf{a}_1, \mathbf{b}_0, \mathbf{b}_1)$ will indicate the CHSH parameter

$$\begin{aligned} S(\mathbf{a}_0, \mathbf{a}_1, \mathbf{b}_0, \mathbf{b}_1) = E(\mathbf{a}_0, \mathbf{b}_0) + E(\mathbf{a}_0, \mathbf{b}_1) + E(\mathbf{a}_1, \mathbf{b}_0) \\ - E(\mathbf{a}_1, \mathbf{b}_1). \end{aligned} \tag{4}$$

The CHSH inequality,

$$|S(\mathbf{a}_0, \mathbf{a}_1, \mathbf{b}_0, \mathbf{b}_1)| \leq 2\,, \tag{5}$$

must hold whenever the quantities $E(\mathbf{a}_i, \mathbf{b}_j)$ are interpreted as the expectation values of products of random variables X_i, Y_j, $i, j = 1, 2$, defined on the same probability space. This physically corresponds to deal with a realistic non-contextual hidden variable theory where, physically speaking, X_i, Y_j play the role of *classical variables.*

The experimental implementation described in [17] exploits momentum and polarization degrees of freedom of a single photon, which can be described as a 2-qubit system. We shall denote as $\{|0\rangle, |1\rangle\}$ the vectors of an orthonormal basis in the momentum Hilbert space $\mathcal{H}_M$ corresponding to two particular propagation directions of the photon in the experimental setup. Analogously $\{|H\rangle, |V\rangle\}$ denote the vectors of the orthonormal basis in the polarization

Hilbert space $\mathcal{H}_P$ corresponding respectively to horizontal and vertical polarization. Eventually, the corresponding orthonormal basis in the tensor product Hilbert space will be made of the four vectors $\{|0H\rangle, |1H\rangle, |0V\rangle, |1V\rangle\}$, where henceforth $|XY\rangle = |X\rangle \otimes |Y\rangle$. From now on, unless otherwise stated, all the operators will be always represented by matrices in the above mentioned basis. As illustrated in [17], the experimental setup can be divided in three stages: (I) generation, (II) preparation, and (III) detection stage, as illustrated in Fig.1.

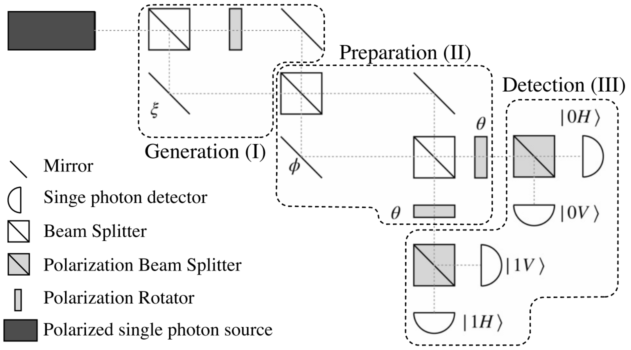

Figure 1: Scheme of the setup used in [17] to create momentum and polarization entangled photons and to perform the CHSH test of quantum contextuality

. The grey dotted line represents the optical path of the photons. The black dotted lines show the three main parts of the setup: Generation (I), Preparation (II) and Detection (III). The angle ξ is the angle used to correct phase mismatch in the SPE state, while the angles ϕ and θ are respectively, the rotation angles of the momentum and polarization states.

The generation stage (I) has the role of transforming the state of a polarized single photon into the SPE entangled state $|\Psi_+\rangle = \frac{1}{\sqrt{2}}(|0H\rangle + i|1V\rangle)$. We consider the polarization of single photons entering the setup to be set to vertical, in such a way that the state vector of the photons can be represented as $|0V\rangle$. Next, a beam splitter (BS), whose action on the vectors of $\mathcal{H}_M$ can be represented by the matrix

$$V_{BS} = \begin{pmatrix} \sqrt{0.5} & i\sqrt{0.5} \\ i\sqrt{0.5} & \sqrt{0.5} \end{pmatrix}, \qquad (6)$$

produces a superposition of momentum states of the form $\frac{1}{\sqrt{2}}(|0\rangle + i|1\rangle) \otimes |V\rangle$; finally a half-wave plate rotates the polarization of photons with propagation direction $|0\rangle$, eventually producing the SPE state $|\Psi_+\rangle$. The i phase term is compensated by properly setting the phase ξ, which is controlled by varying the relative positions of the mirrors. The preparation stage (II) consists of a Mach Zehnder interferometer (MZI) followed by two half-wave plates, one in each output port of the MZI. The MZI acts as a momentum-qubit gate $U_{\mathbf{a}} : \mathcal{H}_M \to \mathcal{H}_M$. In the ideal case of a lossless balanced beam splitter, the unitary operator $U_{\mathbf{a}}$ can be constructed out of the composition $U_{\mathbf{a}} = V_{BS} V_{mir} V(\phi) V_{BS}$, where V_{BS} is given by (6) and

$$V(\phi) = \begin{pmatrix} e^{i\phi} & 0 \\ 0 & 1 \end{pmatrix}, \; V_{mir} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \qquad (7)$$

Remark II.1 *The last operator satisfies*

$$V_{mir} = V_{mir}^{-1} = V_{mir}^{\dagger} \,,$$

and represents the action of the mirrors between the two beam splitters (we assume they do not affect the polarization in any way). Notice that V_{mir} commutes with V_{BS} and that $V_{mir} \otimes I$ does the same with the analog operators (11), which describe more realistic polarization-dependent and lossy beam splitters that we shall consider in the rest of the paper. Therefore we have:

$$U_{\mathbf{a}} = V_{mir} V_{BS} V(\phi) V_{BS} \,.$$

In addition, since

$$V_{mir} P_{\pm}^{M} V_{mir} = P_{\mp}^{M} \,,$$

the net final effect of V_{mir}, when composed with the projectors $P_{\pm}^{M}$, is just to flip $P_{\pm}^{M}$ to $P_{\mp}^{M}$ without affecting the final results, since we did not choose one of the two possible one-to-one correspondence between the two possible labels ± 1 and the momentum eigenstates $|0\rangle$ and $|1\rangle$. For this reason we shall henceforth omit V_{mir} in the rest of computations and assume

$$U_{\mathbf{a}} = V_{BS} V(\phi) V_{BS} \,.$$

A more realistic mirror will be considered later in Section III.

By explicit computation, we get

$$U_{\mathbf{a}} = \begin{pmatrix} \frac{e^{i\phi}-1}{2} & \frac{i+ie^{i\phi}}{2} \\ \frac{i+ie^{i\phi}}{2} & \frac{1-e^{i\phi}}{2} \end{pmatrix} = i e^{i\phi/2} \begin{pmatrix} \sin(\phi/2) & \cos(\phi/2) \\ \cos(\phi/2) & -\sin(\phi/2) \end{pmatrix}.$$

The last formula shows how the unitary operator $U_{\mathbf{a}}$ associated to the MZI is related to the phase shift ϕ in one arm of the Mach-Zehnder interferometer, where $\phi = \frac{2\pi\Delta L}{\lambda}$, ΔL is the path difference in the two arms and λ the photon wavelength [8]. The angle ϕ determines the vector $\mathbf{a}$ in the Bloch sphere associated to the 1-particle observable $O^{\mathbf{a}} = \mathbf{a} \cdot \boldsymbol{\sigma}$ related to the momentum degree of freedom, where $\boldsymbol{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ are the associated Pauli matrices. In particular, the orthogonal projectors $\{P_x^{\mathbf{a}}\}_{x=\pm 1}$ associated to the eigenvectors of $O^{\mathbf{a}}$ are obtained by applying the unitary map $U_{\mathbf{a}}$ to the projectors $\{P_{+1}^{M} = |0\rangle\langle 0|, P_{-1}^{M} = |1\rangle\langle 1|\}$ associated to the standard basis of $\mathcal{H}_M$, i.e.

$$P_x^{\mathbf{a}} = U_{\mathbf{a}}^{\dagger} P_x^{M} U_{\mathbf{a}}, \qquad x = \pm 1. \qquad (8)$$

Next, two half-wave plates, one in each output port of the MZI (Fig. 1), with the fast axis rotated by the same amount ϑ, perform a rotation in the polarization space by an angle $\theta = 2\vartheta$ with respect to the vertical direction. This transformation in the qubit space $\mathcal{H}_P$ can be described in terms of a unitary map $U_{\mathbf{b}}$. The angle θ determines the vector $\mathbf{b}$ in the Bloch sphere, i.e. the 1-particle observable $O^{\mathbf{b}} = \mathbf{b} \cdot \boldsymbol{\sigma}$ related to the polarization degrees of freedom. As above, the orthogonal projectors $\{P_y^{\mathbf{b}}\}_{y=\pm 1}$ associated to the eigenvectors of $O^{\mathbf{b}}$ can be obtained as

$$P_y^{\mathbf{b}} = U_{\mathbf{b}}^\dagger P_y^P U_{\mathbf{b}}, \qquad y = \pm 1, \tag{9}$$

where $\{P_{+1}^P = |H\rangle\langle H|, P_{-1}^P = |1\rangle\langle 1|\}$ are the projectors associated to the basis $\{|H\rangle, |V\rangle\}$ of $\mathcal{H}_P$. The net action of the preparation stage can be described by means of the unitary operator $U_{\mathbf{a},\mathbf{b}} := U_{\mathbf{a}} \otimes U_{\mathbf{b}}$ in the space $\mathcal{H}_M \otimes \mathcal{H}_P$ transforming the state of an incoming photon, i.e. a density matrix ρ in $\mathcal{H}_M \otimes \mathcal{H}_P$, to the prepared state $(U_{\mathbf{a}} \otimes U_{\mathbf{b}})\rho(U_{\mathbf{a}} \otimes U_{\mathbf{b}})^\dagger$.

The final detection stage (III) consists of two Polarizing Beam splitters (PBS) and four single-photon-avalanche-diodes (SPAD) that identify the four measurement channels and correspond to the PVM made of the orthogonal projectors

$$|0H\rangle\langle 0H|, \quad |0V\rangle\langle 0V|, \quad |1H\rangle\langle 1H|, \quad |1V\rangle\langle 1V| \tag{10}$$

Clearly, the projectors in (10) can be written as $P_x^M \otimes P_y^P$ where the superscrips M and P stands for momentum and polarizations, while $x, y \in \{+1, -1\}$. Actually $\{P_{+1}^M, P_{-1}^M\}$ (resp. $\{P_{+1}^P, P_{-1}^P\}$) are the PVM of the momentum observable $O^M = \sigma_z$ (resp. Polarization observable $O^P = \sigma_z$).

For any pair of unit vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$, the PVM associated to the joint measurement of the two commuting observables $\mathbf{a} \cdot \boldsymbol{\sigma} \otimes I$ and $I \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$ is given by

$$P_x^{\mathbf{a}} \otimes P_y^{\mathbf{b}} = U_{\mathbf{a}}^\dagger P_x^M U_{\mathbf{a}} \otimes U_{\mathbf{b}}^\dagger P_y^P U_{\mathbf{b}} = U_{\mathbf{a},\mathbf{b}}^\dagger (P_x^M \otimes P_y^P) U_{\mathbf{a},\mathbf{b}},$$

with $x, y = \pm 1$. Given a state ρ of the compound system, the corresponding quantum transition probabilities (2) are equal to

$$P(x, y|\rho, \mathbf{a}, \mathbf{b}) = \mathrm{Tr}[\rho P_x^{\mathbf{a}} \otimes P_y^{\mathbf{b}}] = \mathrm{Tr}[U_{\mathbf{a},\mathbf{b}} \rho U_{\mathbf{a},\mathbf{b}}^\dagger (P_x^M \otimes P_y^P)]$$

III. BELL INEQUALITY VIOLATION BY SPE STATES USING REALISTIC OPTICAL ELEMENTS

A. Issues with polarization-dependent transmittance and reflectance.

At this point it is worthwhile to stress that the correlations coefficients (3) and the CHSH parameter (4) refer to measurements of pair of commuting observables of form $\mathbf{a} \cdot \boldsymbol{\sigma} \otimes I$ and $I \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$ respectively. In the usual context of inter-particle entanglement this means that each observable is referred to a different particle of the entangled pair, while in the case of SPE *each observable refers to a different degree of freedom of the same particle* (momentum or polarization in the specific experimental setting described above). In particular, it is worth emphasising that this is a necessary condition for ruling out a description of CHSH violation in terms of a non-contextual hidden variables theory. Furthermore, protocols for device-independent entropy certification in quantum random number generators [20] based on CHSH inequality violation rely upon the tensor product form $O^{\mathbf{a}} = \mathbf{a} \cdot \boldsymbol{\sigma} \otimes I$ and $O^{\mathbf{b}} = I \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$ of the couple of measured observables. In the case where the observables $O^{\mathbf{a}}$ and $O^{\mathbf{b}}$ refer to space-like separated systems, this condition is naturally fulfilled, provided the Bell test is not affected by the locality loophole. In the case of SPE, the situation is more complicated. In the particular experimental implementation described above, the product form of the PVM $\{P_x^{\mathbf{a}} \otimes P_y^{\mathbf{b}}\}_{x,y=\pm 1}$ relies on the product form of the rotation operator $U_{\mathbf{a}} \otimes U_{\mathbf{b}}$. As discussed above, this is obtained from the composition of $U_{\mathbf{a}} \otimes I$, realized by the Mach-Zehnder interferometer and acting only on the momentum degree of freedom, and of $I \otimes U_{\mathbf{b}}$, realized by two half-wave plates and acting only on the polarization degree of freedom. While this last stage does not present significant issues, in practical experimental implementations the beam splitters (BSs) employed in the MZI present a few non-idealities that must be analyzed since they do not allow to represent the action of the interferometer in terms of a unitary operator of product form $U_{\mathbf{a}} \otimes I$. As we shall see shortly, this implies that the effectively measured observables are not of the product form $A \otimes B$, giving rise to a problem analogous to the locality loophole when dealing with couples of entangled particles. In particular, in realistic BSs reflectance and transmittance for the vertically polarized component differ from those for the horizontally polarized one. More specifically, the matrix representing the operator corresponding to the BS in the basis $\{|0H\rangle, |1H\rangle, |0V\rangle, |1V\rangle\}$ can be written as

$$U_{BS}^{real} = \begin{pmatrix} t_H & ir_H & 0 & 0 \\ ir_H & t_H & 0 & 0 \\ 0 & 0 & t_V & ir_V \\ 0 & 0 & ir_V & t_V \end{pmatrix} \tag{11}$$

where $|t_H|^2 + |r_H|^2 \leq 1$ and $|t_V|^2 + |r_V|^2 \leq 1$. Actually, the operator U_{BS}^{real} has the product form $V \otimes I$ if and only if

$$t_H = t_V, \quad \text{and} \quad r_H = r_V. \tag{12}$$

Hence, if conditions (12) are not fulfilled, the rotation operator describing the action of the preparation stage $U_{\mathbf{a},\mathbf{b}}^{real} = U_{BS}^{real}(V(\phi) \otimes I)U_{BS}^{real}$ (with $V(\phi)$ defined in (7)) can no longer be written as a tensor product $U_M \otimes U_P$, for a suitable pair of unitary operators $U_M : \mathcal{H}_M \to \mathcal{H}_M$ and $U_P : \mathcal{H}_P \to \mathcal{H}_P$. In other words, the probabilities of clicks of the four detectors in the final detection stage, namely

$$\mathrm{Tr}[U_{\mathbf{a},\mathbf{b}}^{real} \rho (U_{\mathbf{a},\mathbf{b}}^{real})^\dagger \, P_x^M \otimes P_y^P] \tag{13}$$

cannot be written in the form

$$\mathrm{Tr}[\rho (P_x^{\tilde{\mathbf{a}}} \otimes P_y^{\tilde{\mathbf{b}}})] \tag{14}$$

for a suitable pair of unit vectors $\tilde{\mathbf{a}}, \tilde{\mathbf{b}} \in \mathbb{R}^3$ (also different from $\mathbf{a}, \mathbf{b}$). To take this kind of non-idealities into account, we provide a bound for the difference between the real probabilities (13) and the idealized ones (14) corresponding to tensor product observables. Furthermore, in order to handle the general case of a lossy beam splitter, we have also to consider that a fraction of the photons passing through the Mach Zehnder interferometer performing the rotation in the momentum Hilbert space can actually be either scattered or absorbed. This means that the matrix (11) as well as the rotation operators $U^{real}_{\mathbf{a},\mathbf{b}}$ constructed out of it are not unitary and the observed statistics of detection outcomes refers only to the photons that aren't lost by the optical elements. Hence, the actual detection probabilities are computed as

$$P^{real}(x, y|\rho, \mathbf{a}, \mathbf{b}) = \frac{\mathrm{Tr}[U^{real}_{\mathbf{a},\mathbf{b}}\rho(U^{real}_{\mathbf{a},\mathbf{b}})^\dagger\, P^M_x \otimes P^P_y]}{\mathrm{Tr}[U^{real}_{\mathbf{a},\mathbf{b}}\rho(U^{real}_{\mathbf{a},\mathbf{b}})^\dagger]}, \quad (15)$$

where

$$U^{real}_{\mathbf{a},\mathbf{b}} = (I \otimes U_{\mathbf{b}})U^{real}_{BS}(V(\phi) \otimes I)U^{real}_{BS}. \quad (16)$$

If the losses for the vertically polarized component are comparable to the losses for the horizontally polarized one, i.e., when

$$t^2_H + r^2_H \sim t^2_V + r^2_V \quad (17)$$

then the denominator in (15) is equal to $t^2_H + r^2_H = t^2_V + r^2_V$ for any choice of $\rho, \mathbf{a}, \mathbf{b}$. This allows us to compute the detection probabilities (15) in terms of the following expression

$$P^{real}(x, y|\rho, \mathbf{a}, \mathbf{b}) = \mathrm{Tr}[\tilde{U}^{real}_{\mathbf{a},\mathbf{b}}\rho(\tilde{U}^{real}_{\mathbf{a},\mathbf{b}})^\dagger\, P^M_x \otimes P^P_y], \quad (18)$$

with :

$$\tilde{U}^{real}_{\mathbf{a},\mathbf{b}} = (I \otimes U_{\mathbf{b}})\tilde{U}^{real}_{BS}(V(\phi) \otimes I)\tilde{U}^{real}_{BS}, \quad (19)$$

where the effective unitary operator $\tilde{U}^{real}_{BS}$ is defined as

$$\tilde{U}^{real}_{BS} = \begin{pmatrix} \tilde{t}_H & i\tilde{r}_H & 0 & 0 \\ i\tilde{r}_H & \tilde{t}_H & 0 & 0 \\ 0 & 0 & \tilde{t}_V & i\tilde{r}_V \\ 0 & 0 & i\tilde{r}_V & \tilde{t}_V \end{pmatrix},$$

$$\tilde{t}_H = \frac{t_H}{\sqrt{t^2_H + r^2_H}}, \quad \tilde{r}_H = \frac{r_H}{\sqrt{t^2_H + r^2_H}},$$
$$\tilde{t}_V = \frac{t_V}{\sqrt{t^2_V + r^2_V}}, \quad \tilde{r}_V = \frac{r_V}{\sqrt{t^2_V + r^2_V}}. \quad (20)$$

In order to estimate the difference between the detection probabilities (18) and the ideal ones (14) associated to observables of product form, in the next two sub-sections we compute the unitary operator $U^{ideal}_{\mathbf{a}} : \mathcal{H}_M \otimes \mathcal{H}_P \to \mathcal{H}_M \otimes \mathcal{H}_P$ of product form which minimizes the Hilbert-Schmidt distance from the operator $U^{real}_{\mathbf{a}} : \mathcal{H}_M \otimes \mathcal{H}_P \to \mathcal{H}_M \otimes \mathcal{H}_P$ defined as $U^{real}_{\mathbf{a}} = \tilde{U}^{real}_{BS}(V(\phi) \otimes I)\tilde{U}^{real}_{BS}$ when varying the parameters of the factors entering the expression of $U^{ideal}_{\mathbf{a}}$ (see Eq. (25) and Remark III.1 below).

B. Equally lossy polarization channels

We first consider the general case where the two BS included in the MZI present different values of t_H, t_V, r_H, r_V. Let us assume for the time being that condition (17) is still satisfied. This condition will be relaxed in the next sub-section. Under this assumption the operator $U^{real}_{\mathbf{a}}$ can be written as

$$U^{real}_{\mathbf{a}(\phi)} = V^{BS}_1(V(\phi) \otimes I)V^{BS}_2, \quad (21)$$

where

$$V^{BS}_k = \begin{pmatrix} t_{H,k} & ir_{H,k} & 0 & 0 \\ ir_{H,k} & t_{H,k} & 0 & 0 \\ 0 & 0 & t_{V,k} & ir_{V,k} \\ 0 & 0 & ir_{V,k} & t_{V,k} \end{pmatrix}, \; k = 1, 2. \quad (22)$$

By condition (17) we can restrict ourselves to the case where

$$t^2_{H,k} + r^2_{H,k} = t^2_{V,k} + r^2_{V,k} = 1, \; k = 1, 2 \quad (23)$$

since, if this condition is not fulfilled, the coefficients t_H, r_H, t_V, r_v, can be replaced by the corresponding normalized coefficients $\tilde{t}_H, \tilde{r}_H, \tilde{t}_V, \tilde{r}_V$ as in (20). Hence, we can introduce the notation

$$V^{BS}_k = \begin{pmatrix} \cos\alpha^H_k & i\sin\alpha^H_k & 0 & 0 \\ i\sin\alpha^H_k & \cos\alpha^H_k & 0 & 0 \\ 0 & 0 & \cos\alpha^V_k & i\sin\alpha^V_k \\ 0 & 0 & i\sin\alpha^V_k & \cos\alpha^V_k \end{pmatrix}, \quad k = 1, 2. \quad (24)$$

We shall consider the difference between $U^{real}_{\mathbf{a}(\phi)}$ and a product operator $U^{ideal}_{\mathbf{a}(\phi)}$ of the form

$$U^{ideal}_{\mathbf{a}(\phi)}(u, v) = (U(u) \otimes I)(V(\phi) \otimes I)(U(v) \otimes I) \quad (25)$$

where $u, v \in [0, 2\pi]$ and

$$U(\theta) = \begin{pmatrix} \cos\theta & i\sin\theta \\ i\sin\theta & \cos\theta \end{pmatrix}, \quad (26)$$

and compute the values of u, v that minimize the Hilbert-Schmidt norm of the difference operator

$$R_{\mathbf{a}}(u, v) = U^{real}_{\mathbf{a}} - U^{ideal}_{\mathbf{a}}(u, v)$$

adopting the shortened notation $\mathbf{a} \equiv \mathbf{a}(\phi)$.

Remark III.1 *The operator $U^{ideal}_{\mathbf{a}}(u, v)$ corresponds to an initial setup where the two BS in the preparation stage of Fig. 1 are ideal, generally non-balanced, and different to each other if $u \neq v$. The use of this unitary operator is equivalent to a final measurement of the momentum observable*

$$\mathbf{a}'(\phi, u, v) \cdot \boldsymbol{\sigma} = (U^{ideal}_{\mathbf{a}(\phi)}(u, v))^\dagger \sigma_z U^{ideal}_{\mathbf{a}(\phi)}(u, v)$$

for a unit vector $\mathbf{a}' = \mathbf{a}'(\phi, u, v)$ fulfilling the identity above, instead of the initially chosen observable $\mathbf{a}(\phi)\cdot\boldsymbol{\sigma}$. In other words, when taking the said non-idealities into account but assuming (17), our setup is viewed to measure the observable

$$\mathbf{a}'(\phi, u_0, v_0)\cdot\boldsymbol{\sigma}\otimes\mathbf{b}\cdot\boldsymbol{\sigma}\,, \tag{27}$$

where (u_0, v_0) minimizes $||R_{\mathbf{a}}(u,v)||_{HS}$, instead of measuring $\mathbf{a}(\phi)\cdot\boldsymbol{\sigma}\otimes\mathbf{b}\cdot\boldsymbol{\sigma}$.

The operator $U^{real}_{\mathbf{a}(\phi)}$ can be written in the block form

$$U^{real}_{\mathbf{a}(\phi)} = e^{i\phi/2}\begin{pmatrix} U(\theta_H, \hat{n}_H) & 0 \\ 0 & U(\theta_V, \hat{n}_V)\end{pmatrix} \tag{28}$$

where

$$U(\theta_H, \hat{n}_H) = e^{i\theta_H\,\hat{n}_H\cdot\boldsymbol{\sigma}} = \cos\theta\, I_{2\times 2} + i\sin\theta_H\,\hat{n}_H\cdot\boldsymbol{\sigma}$$
$$U(\theta_V, \hat{n}_V) = e^{i\theta_V\,\hat{n}_V\cdot\boldsymbol{\sigma}} = \cos\theta_V\, I_{2\times 2} + i\sin\theta_V\,\hat{n}_V\cdot\boldsymbol{\sigma}$$

$$\begin{aligned}\cos\theta_H &= \cos\phi/2\cos(\alpha_1^H + \alpha_2^H),\\ \sin\theta_H\hat{n}_H &= (\cos\phi/2\sin(\alpha_1^H + \alpha_2^H),\\ &\quad \sin\phi/2\sin(\alpha_1^H - \alpha_2^H), \sin\phi/2\cos(\alpha_1^H - \alpha_2^H))\\ \cos\theta_V &= \cos\phi/2\cos(\alpha_1^V + \alpha_2^V),\\ \sin\theta_V\hat{n}_V &= (\cos\phi/2\sin(\alpha_1^V + \alpha_2^V),\\ &\quad \sin\phi/2\sin(\alpha_1^V - \alpha_2^V), \sin\phi/2\cos(\alpha_1^V - \alpha_2^V)).\end{aligned}$$

Analogously, $U^{ideal}_{\mathbf{a}}(u,v) = e^{i\phi/2}e^{i\tilde{\theta}(u,v)\,\hat{n}(u,v)\cdot\boldsymbol{\sigma}}\otimes I$, with

$$\begin{aligned}\cos\tilde{\theta}(u,v) &= \cos\phi/2\cos(u+v),\\ \sin\tilde{\theta}(u,v)\hat{n}(u,v) &= (\cos\phi/2\sin(u,v), \sin\phi/2\sin(u-v),\\ &\quad \sin\phi/2\cos(u-v)).\end{aligned} \tag{29}$$

Hence

$$\begin{aligned}\|R_{\mathbf{a}}\|^2_{HS} = \mathrm{Tr}[R_{\mathbf{a}}R^\dagger_{\mathbf{a}}] = 4\big(2-&\cos^2(\phi/2)(\cos(\alpha_1^H+\alpha_2^H-u-v)\\ &+\cos(\alpha_1^V+\alpha_2^V-u-v))\\ -\sin^2(\phi/2)(\cos(\alpha_1^H-\alpha_2^H-u+v)&+\cos(\alpha_1^V-\alpha_2^V-u+v))\big)\end{aligned} \tag{30}$$

and this quantity attains its minimal value for $u = \frac{\alpha_1^H+\alpha_1^V}{2}$, and $v = \frac{\alpha_2^H+\alpha_2^V}{2}$. In particular, in this case the operator $R_{\mathbf{a}}R^\dagger_{\mathbf{a}}$ is a multiple of the identity, i.e. $R_{\mathbf{a}}R^\dagger_{\mathbf{a}} = cI_{4\times 4}$ with

$$\begin{aligned}c = 2 - 2\cos^2(\phi/2)\cos\left(\frac{\alpha_1^H-\alpha_1^V}{2}+\frac{\alpha_2^H-\alpha_2^V}{2}\right)\\ - 2\sin^2(\phi/2)\cos\left(\frac{\alpha_1^H-\alpha_1^V}{2}-\frac{\alpha_2^H-\alpha_2^V}{2}\right)\end{aligned} \tag{31}$$

Hence $\|R_{\mathbf{a}}\| = \sqrt{c}$. A uniform bound on the norm of $R_{\mathbf{a}}R^+_{\mathbf{a}}$ that is independent of $\mathbf{a}$, hence of ϕ, is given by

$$\begin{aligned}e :=& \sup_{\phi\in[0,2\pi]}\|R_{\mathbf{a}}(\phi)R_{\mathbf{a}}(\phi)^\dagger\|\\ &= 2 - 2\min\Bigg\{\cos\left(\frac{\alpha_1^V-\alpha_1^H}{2}+\frac{\alpha_2^V-\alpha_2^H}{2}\right),\\ &\qquad \cos\left(\frac{\alpha_1^V-\alpha_1^H}{2}-\frac{\alpha_2^V-\alpha_2^H}{2}\right)\Bigg\}\end{aligned} \tag{32}$$

Remark III.2 *The optimization technique implemented in this section actually allows to obtain an even sharper bound. Indeed, it provides a value that coincides with the one attainable by the more general technique based on the comparison between the rotation operator $U^{real}_{\mathbf{a}(\phi)} = V_1^{BS}(V(\phi)\otimes I)V_2^{BS}$ associated to the MZ with a general unitary operator of product form $U_M\otimes V_P$, with U_M, V_P unitary operators on $\mathcal{H}_M$ and $\mathcal{H}_P$ respectively. More precisely, if $V_1^{ideal}(x)$ and $V_2^{ideal}(y)$ are the unitary operators defined by (26), we have:*

$$\begin{aligned}&\max_{\phi\in[0,2\pi]}\min_{u,v\in[0,2\pi]}\|U^{real}_{\mathbf{a}(\phi)}-V_1^{ideal}(u)(V(\phi)\otimes I)V_2^{ideal}(v)\|_2\\ &= \max_{\phi\in[0,2\pi]}\min_{\zeta\in[0,2\pi],U_M,U_P\in SU(2)}\|U^{real}_{\mathbf{a}(\phi)}-e^{i\zeta}U_M\otimes U_P\|_2\,.\end{aligned} \tag{33}$$

A detailed proof of this result is given in appendix.

We can now compute a uniform bound (independent of $\mathbf{a}, \mathbf{b}$) for the difference between the detection probabilities (15) and the ideal ones (14) associated to tensor product observables.

Indeed, by expanding the expression $U^{real}_{\mathbf{a},\mathbf{b}} = (I\otimes U_{\mathbf{b}})(U^{ideal}_{\mathbf{a}} + R)$ we obtain:

$$\begin{aligned}&P^{real}(x,y|\rho,\mathbf{a},\mathbf{b}) = P^{ideal}(x,y|\rho,\mathbf{a},\mathbf{b})\\ &+\mathrm{Tr}[U^{ideal}_{\mathbf{a}}\rho R^\dagger_{\mathbf{a}}P^M_x\otimes P^P_{y,\mathbf{b}}]+\mathrm{Tr}[R_{\mathbf{a}}\rho(U^{ideal}_{\mathbf{a}})^\dagger P^M_x\otimes P^P_{y,\mathbf{b}}]\\ &\qquad+\mathrm{Tr}[R_{\mathbf{a}}\rho R^\dagger_{\mathbf{a}}P^M_x\otimes P^P_{y,\mathbf{b}}]\end{aligned} \tag{34}$$

where $P^M_x\otimes P^P_{y,\mathbf{b}} = (I\otimes U_{\mathbf{b}})^\dagger P^M_x\otimes P^P_y(I\otimes U_{\mathbf{b}})$ and $P^{ideal}(x,y|\rho,\mathbf{a},\mathbf{b}) = Tr[U^{ideal}_{\mathbf{a}}\rho U^{ideal}_{\mathbf{a}})^\dagger P^M_x\otimes P^P_{y,\mathbf{b}}]$.

Remark III.3 *In spite of the notation $P^{ideal}(x,y|\rho,\mathbf{a},\mathbf{b})$ used here and in the next section, this is the probability to get the outcomes x and y when measuring the factorized observable*

$$\mathbf{a}'(\phi, u_0, v_0)\cdot\boldsymbol{\sigma}\otimes\mathbf{b}\cdot\boldsymbol{\sigma}\,,$$

and not $\mathbf{a}(\phi)\cdot\boldsymbol{\sigma}\otimes\mathbf{b}\cdot\boldsymbol{\sigma}$, according to Remark III.1.

For any choice of $\mathbf{a}, \mathbf{b}$, the difference between the detection probabilities $P^{real}(x,y|\rho,\mathbf{a},\mathbf{b})$ and the ones related to product observables $P^{ideal}(x,y|\rho,\mathbf{a},\mathbf{b})$ can be bounded by

$$\begin{aligned}&|P^{real}(x,y|\rho,\mathbf{a},\mathbf{b}) - P^{ideal}(x,y|\rho,\mathbf{a},\mathbf{b})|\\ &\qquad\le 2\sqrt{\|R_{\mathbf{a}}R^\dagger_{\mathbf{a}}\|}+\|R_{\mathbf{a}}R^\dagger_{\mathbf{a}}\| \le 2\sqrt{e}+e,\end{aligned} \tag{35}$$

where e is given by (32).

A similar bound can be derived for the difference between the CHSH parameter (4) associated to the detection probabilities (15)

$$S^{real} = \mathrm{Tr}[\rho \sum_{a,b} c_{ab} (U^{real}_{\mathbf{a}})^\dagger (I \otimes U^\dagger_{\mathbf{b}})\sigma_3 \otimes \sigma_3 (I \otimes U_{\mathbf{b}}) U^{real}_{\mathbf{a}}]$$

and an ideal one associated to measurements of product observables on the same state ρ

$$S^{ideal} = \mathrm{Tr}[\rho \sum_{\mathbf{a},\mathbf{b}} c_{\mathbf{ab}} (U^{ideal}_{\mathbf{a}})^\dagger (I \otimes U^\dagger_{\mathbf{b}})\sigma_3 \otimes \sigma_3 (I \otimes U_{\mathbf{b}}) U^{ideal}_{\mathbf{a}}],$$

Where $c_{\mathbf{a}_0,\mathbf{b}_0} = c_{\mathbf{a}_1,\mathbf{b}_0} = c_{\mathbf{a}_0,\mathbf{b}_1} = 1$ and $c_{\mathbf{a}_1,\mathbf{b}_1} = -1$.

As discussed above, for any choice of the vector $\mathbf{a}$ associated to the phase shift ϕ the difference operator $R_{\mathbf{a}} = U^{real}_{\mathbf{a}} - U^{ideal}_{\mathbf{a}}$ can be written as

$$R_{\mathbf{a}} = e^{i\phi/2} \left(f_1(\phi) R_1 + f_2(\phi) R_2 \right) \tag{36}$$

where $f_1(\phi) = \cos(\phi/2)$, $f_2(\phi) = \sin(\phi/2)$ and the two operators R_1, R_2 do not depend on ϕ and are given by

$$R_1 = \begin{pmatrix} R_1^H & 0 \\ 0 & R_1^V \end{pmatrix}, \quad R_2 = \begin{pmatrix} R_2^H & 0 \\ 0 & R_2^V \end{pmatrix}$$

where

$$\begin{aligned} R_1^H &= (\cos(\alpha_1^H + \alpha_2^H) - \cos(u+v)) I_{2\times 2} + i(\sin(\alpha_1^H + \alpha_2^H) \\ &\qquad - \sin(u+v))\sigma_x \\ R_2^H &= i(\cos(\alpha_1^H - \alpha_2^H) - \cos(x-y))\sigma_z + i(\sin(\alpha_1^H - \alpha_2^H) \\ &\qquad - \sin(u-v))\sigma_y \\ R_1^V &= (\cos(\alpha_1^V + \alpha_2^V) - \cos(u+v)) I_{2\times 2} + i(\sin(\alpha_1^V + \alpha_2^V) \\ &\qquad - \sin(u+v))\sigma_x \\ R_2^V &= i(\cos(\alpha_1^V - \alpha_2^V) - \cos(u-v))\sigma_z + i(\sin(\alpha_1^V - \alpha_2^V) \\ &\qquad - \sin(u-v))\sigma_y \end{aligned}$$

with $u = (\alpha_1^H + \alpha_1^V)/2$ and $v = (\alpha_2^H + \alpha_2^V)/2$. In the case the two beam splitters have similar features, i.e. $\alpha_1^H \sim \alpha_2^H$ and $\alpha_1^V \sim \beta_2^V$ then $R_2 = 0$. Since $R_1 R_2^\dagger + R_1^\dagger R_2 = 0$ we have $\|R_{\mathbf{a}}\|^2_{HS} = f_1^2(\phi)\|R_1\|^2_{HS} + f_2^2(\phi)\|R_2\|^2_{HS}$. Moreover

$$R_1 R_1^\dagger = R_1^\dagger R_1 = 4 \sin^2\left(\frac{\alpha_1^H + \alpha_2^H - \alpha_1^V - \alpha_2^V}{4} \right) I_{4\times 4},$$

$$R_2 R_2^\dagger = R_2^\dagger R_2 = 4 \sin^2\left(\frac{(\alpha_1^H - \alpha_2^H) - (\alpha_1^V - \alpha_2^V)}{4} \right) I_{4\times 4},$$

hence

$$\|R_1\| = 2 \left| \sin\left(\frac{\alpha_1^H + \alpha_2^H - \alpha_1^V - \alpha_2^V}{4} \right) \right|$$

$$\|R_2\| = 2 \left| \sin\left(\frac{(\alpha_1^H - \alpha_2^H) - (\alpha_1^V - \alpha_2^V)}{4} \right) \right| \tag{37}$$

By expanding $U^{real}_{\mathbf{a}} = U^{ideal}_{\mathbf{a}} + R_{\mathbf{a}}$ we get

$$\begin{aligned} S^{real} &= S^{ideal} + \mathrm{Tr}[\rho \sum_{a,b} c_{ab} (R_{\mathbf{a}})^\dagger \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} U^{ideal}_{\mathbf{a}}] \\ &\quad + \mathrm{Tr}[\rho \sum_{a,b} c_{ab} (U^{ideal}_{\mathbf{a}})^\dagger \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} R_{\mathbf{a}}] \\ &\quad + \mathrm{Tr}[\rho \sum_{a,b} c_{ab} R^\dagger_{\mathbf{a}} \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} R_{\mathbf{a}}] \end{aligned}$$

Where $R_{\mathbf{a}}$ is given by (36). In particular:

$$\begin{aligned} &\mathrm{Tr}[\rho \sum_{a,b} c_{ab} (R_{\mathbf{a}})^\dagger \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} U^{ideal}_{\mathbf{a}}] \\ &\quad = \mathrm{Tr}[\rho R_1^\dagger \sum_{a,b} c_{ab} e^{-i\phi/2} f_1(\phi(\mathbf{a}))\sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} U^{ideal}_{\mathbf{a}}] \\ &\quad + \mathrm{Tr}[\rho R_2^\dagger \sum_{a,b} c_{ab} e^{-i\phi/2} f_2(\phi(\mathbf{a}))\sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} U^{ideal}_{\mathbf{a}}] \end{aligned}$$

Each term on the right hand side has the form $\mathrm{Tr}[\rho R_i^\dagger O_i]$, with $i = 1, 2$ and

$$\begin{aligned} O_i &= \sum_{a,b} c_{ab} e^{-i\phi/2} f_i(\phi(\mathbf{a}))\sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} U^{ideal}_{\mathbf{a}} \\ &= A_0 B_0 + A_0 B_1 + A_1 B_0 - A_1 B_1 \end{aligned}$$

with $B_j = I \otimes \mathbf{b}_j \cdot \boldsymbol{\sigma}$ and $A_j = e^{-i\phi/2} f_i(\phi(\mathbf{a}_j))(\boldsymbol{\sigma} \otimes I) U^{ideal}_{\mathbf{a}}$, $j = 0, 1$. By explicit computation we have:

$$\begin{aligned} \|O_i O_i^\dagger\| &\leq 2c_0^2 + 2c_1^2 + 2|c_0^2 - c_1^2| + 4|c_0 c_1| \\ &\leq \max_{|c_0|,|c_1| \in [0,1]} (2c_0^2 + 2c_1^2 + 2|c_0^2 - c_1^2| + 4 c_0 c_1) \\ &= 8 \end{aligned}$$

with $c_j := f_i(\phi(\mathbf{a}_j))$. Hence, by Von Neumann's trace inequality

$$|\mathrm{Tr}[\rho R_i^\dagger O_i]| \leq \|R_i^\dagger O\| \, \mathrm{Tr}[\rho] \leq 2\sqrt{2}\sqrt{\|R_i^\dagger R_i\|}.$$

and

$$\begin{aligned} &\mathrm{Tr}[\rho \sum_{a,b} c_{ab} (R_{\mathbf{a}})^\dagger \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} U^{ideal}_{\mathbf{a}}] \\ &\qquad \leq 2\sqrt{2}\left(\sqrt{\|R_1^\dagger R_1\|} + \sqrt{\|R_2^\dagger R_2\|}\right). \end{aligned}$$

The same bound holds for the term $\mathrm{Tr}[\rho \sum_{a,b} c_{ab} (U^{ideal}_{\mathbf{a}})^\dagger \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} R_{\mathbf{a}}]$. Similarly

$$\mathrm{Tr}[\rho \sum_{a,b} c_{ab} R^\dagger_{\mathbf{a}} \sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma} R_{\mathbf{a}}] = \sum_{i,j=1,2} \mathrm{Tr}[R_j \rho R_i^\dagger O_{ij}] \tag{38}$$

with

$$O_{ij} = \sum_{a,b} c_{ab} f_i(\phi(\mathbf{a})) f_j(\phi(\mathbf{a})\sigma_3 \otimes \mathbf{b}\cdot\boldsymbol{\sigma}.$$

Since $\|O_{ij}O_{ij}^\dagger\| \leq 2c_0^2 + 2c_1^2 + 2|c_0^2 - c_1^2|$, with $c_k = f_i(\phi(\mathbf{a}_k))f_j(\phi(\mathbf{a}_k))$, $k = 0, 1$, we have $\|O_{ij}\| \leq 2$ if $i = j$ and $\|O_{ij}\| \leq 1$ if $i \neq j$, hence the right hand of (38) is bounded by $2(\|R_1R_1^\dagger\| + \|R_2R_2^\dagger\| + \|R_1\|\|R_2\|)$. Eventually we have:

$$\begin{aligned} |S^{real} - S^{ideal}| &\leq 4\sqrt{2}(\|R_1\| + \|R_2\|) \\ &\quad + 2(\|R_1\|^2 + \|R_2\|^2 + \|R_1\|\|R_2\|) \end{aligned} \quad (39)$$

where $\|R_1\|$ and $\|R_2\|$ are given by (37).

C. Generally lossy beam splitters

Let us consider now the general case where the two BS in the MZI providing the rotation of momentum qubit present different values of t_H, t_V, r_H, r_V and condition (17) is not satisfied. Hence, the approximations adopted in the previous section, in particular (18) and (19), are no longer feasible. The detection probabilities are still given by (15), with $U^{real}_{\mathbf{a},\mathbf{b}} = (I \otimes \mathbf{b} \cdot \boldsymbol{\sigma}) U^{real}_{\mathbf{a}(\phi)}$, with $U^{real}_{\mathbf{a}(\phi)}$ described by (21) and (22), but in the realistic case both BS present losses, which depend explicitly on the polarization:

$$t_{H,k}^2 + r_{H,k}^2 \neq t_{V,k}^2 + r_{V,k}^2, \quad k = 1, 2. \quad (40)$$

In fact, condition (40) does not allow to get rid of the denominator in (15) and obtain formula (18). In the following we are going to estimate an upper bound for the difference between the detection probabilities (15) and the simplified ones

$$\mathrm{Tr}[\tilde{U}^{real}_{\mathbf{a},\mathbf{b}} \rho (\tilde{U}^{real}_{\mathbf{a},\mathbf{b}})^\dagger \, P^M_x \otimes P^P_y], \quad (41)$$

with :

$$\tilde{U}^{real}_{\mathbf{a},\mathbf{b}} = (I \otimes U_{\mathbf{b}}) \tilde{U}^{BS}_1 (V(\phi) \otimes I) \tilde{U}^{BS}_2. \quad (42)$$

where the unitary operators $\tilde{U}^{BS}_1$ and $\tilde{U}^{BS}_2$ are defined as

$$\tilde{U}^{BS}_k = \begin{pmatrix} \tilde{t}_{H,k} & i\tilde{r}_{H,k} & 0 & 0 \\ i\tilde{r}_{H,k} & \tilde{t}_{H,k} & 0 & 0 \\ 0 & 0 & \tilde{t}_{V,k} & i\tilde{r}_{V,k} \\ 0 & 0 & i\tilde{r}_{V,k} & \tilde{t}_{V,k} \end{pmatrix}, \quad (43)$$

where $\tilde{t}_{H,k} = \frac{t_{H,k}}{\sqrt{t_{H,k}^2 + r_{H,k}^2}}$, $\tilde{r}_{H,k} = \frac{r_{H,k}}{\sqrt{t_{H,k}^2 + r_{H,k}^2}}$, $\tilde{t}_{V,k} = \frac{t_{V,k}}{\sqrt{t_{V,k}^2 + r_{V,k}^2}}$, $\tilde{r}_{V,k} = \frac{r_{V,k}}{\sqrt{t_{V,k}^2 + r_{V,k}^2}}$, $k = 1, 2$.

Remark III.4 *The above formula may include the contribution of the pair of realistic mirrors in the preparation stage of Figure 1. Each mirror is here permitted to be lossy (with losses depending on the polarization channel) but we assume that the two mirrors have very similar physical characteristics. The difference with the ideal case discussed in Remark II.1, is just that the matrix representing the couple of mirrors has a further numerical factor* $\eta \in (0, 1)$ *in front of the unitary matrix* V_{mir} *in (7). When passing to the description in terms of* 4×4 *complex matrices as in (11) to take the two polarization into account, the factor may be different for the two* 2×2 *matrices on the principal diagonal and we may have two coefficients* $\eta_H, \eta_V \in (0, 1)$*. The net final effect of these two further factors is just to rescale the coefficients appearing in (22) with* $k = 2$ *to*

$$V_2^{BS} = \begin{pmatrix} \eta_H t_{H,2} & i\eta_H r_{H,2} & 0 & 0 \\ i\eta_H r_{H,2} & \eta_H t_{H,2} & 0 & 0 \\ 0 & 0 & \eta_V t_{V,2} & i\eta_V r_{V,2} \\ 0 & 0 & i\eta_V r_{V,2} & \eta_V t_{V,2} \end{pmatrix}, \quad (44)$$

which is, then, used in the expression (21) of $U^{real}_{\mathbf{a}(\phi)}$*. Notice that, if the mirrors act differently on the two polarization channels, we can pass from the situation of equally lossy polarization described in Section III B to the generic situation described in this section, and* viceversa*, depending on the value of* η_H *and* η_V*.*

Let $\tilde{e}_{\mathbf{a},\mathbf{b}}$ be defined as the difference between the detection probabilities (15) and the simplified ones (18)

$$\begin{aligned} \tilde{e}_{\mathbf{a},\mathbf{b}} := \Bigg| & \frac{\mathrm{Tr}[U^{real}_{\mathbf{a},\mathbf{b}} \rho (U^{real}_{\mathbf{a},\mathbf{b}})^\dagger \, P^M_x \otimes P^P_y]}{\mathrm{Tr}[U^{real}_{\mathbf{a},\mathbf{b}} \rho (U^{real}_{\mathbf{a},\mathbf{b}})^\dagger]} \\ & - \mathrm{Tr}[\tilde{U}^{real}_{\mathbf{a},\mathbf{b}} \rho (\tilde{U}^{real}_{\mathbf{a},\mathbf{b}})^\dagger \, P^M_x \otimes P^P_y] \Bigg| \end{aligned} \quad (45)$$

and let $\tilde{e}$ be the supremum over all possible choices of unit vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^3$

$$\tilde{e} := \sup_{\mathbf{a},\mathbf{b}} \tilde{e}_{\mathbf{a},\mathbf{b}} \, . \quad (46)$$

The coefficient $\tilde{e}$ in (46) depends also on the density matrix ρ. In order to take into account this, it is convenient to introduce the decomposition

$$\rho = \alpha P_H \rho_H P_H + \beta P_V \rho_V P_V + P_V p P_H + P_H p^\dagger P_V \, , \quad (47)$$

where P_H resp. P_V are the two orthogonal projections operators onto the subspaces of $\mathcal{H}_M \otimes \mathcal{H}_P$ spanned by the vectors $\{|0H\rangle, |1H\rangle\}$ resp. $\{|0V\rangle, |1V\rangle\}$. The operators $\rho_H : \mathbb{C}^2_H \to \mathbb{C}^2_H$ and $\rho_V : \mathbb{C}^2_V \to \mathbb{C}^2_V$ are 2×2 density matrices and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$, whereas $p : \mathbb{C}^2_H \to \mathbb{C}^2_V$ is a linear operator.
By introducing the two constants c_H and c_V defined as

$$\begin{aligned} c_H &= \sqrt{(t_{H,1}^2 + r_{H,1}^2)(t_{H,2}^2 + r_{H,2}^2)} \\ c_V &= \sqrt{(t_{V,1}^2 + r_{V,1}^2)(t_{V,2}^2 + r_{V,2}^2)} \, , \end{aligned} \quad (48)$$

we have:

$$\tilde{e} \leq \left| \frac{\alpha\beta(c_H^2 - c_V^2)}{\alpha c_H^2 + \beta c_V^2} \right| + \left| \frac{\sqrt{\alpha\beta}(c_H c_V - \alpha c_V^2 - \beta c_H^2)}{\alpha c_H^2 + \beta c_V^2} \right| \quad (49)$$

The details of the derivation are postponed to appendix.

The values of the constants α, β can be estimated by taking into account the generation stage (I) (see fig. 1).

Indeed, the state ρ of the photon entering the preparation stage (II) is the result of the action of a collimator and a polarization filter selecting the vertical component, of a beam splitter, with transmission and reflection coefficients $t_{V,0}, r_{V,0}$, and a polarization rotator located along the reflected path and converting the vertical polarization into the horizontal one. If we assume that the two mirrors in the generation stage are lossy, but essentially identical to each other, the net effect of such non-ideal mirrors is just to change $t_{V,0}, r_{V,0}$ with a common factor $\eta \in (0,1)$. According to this procedure, the generation stage produces a state ρ of the form (47) with coefficients α, β given by:

$$\alpha = t_{V,0}^2/(t_{V,0}^2 + r_{V,0}^2), \quad \beta = r_{V,0}^2/(t_{V,0}^2 + r_{V,0}^2). \tag{50}$$

Taking into account the analysis above and the one in the previous section, the difference between the detection probabilities, $P^{real}(x, y|\rho, \mathbf{a}, \mathbf{b})$, and those associated to product observables can be estimated by combining the found bounds as

$$|P^{real}(x, y|\rho, \mathbf{a}, \mathbf{b}) - P^{ideal}(x, y|\rho, \mathbf{a}, \mathbf{b})| \leq 2\sqrt{e} + e + \tilde{e} \tag{51}$$

with e given by (32) and $\tilde{e}$ bounded by (49). Analogously, the difference between the CHSH parameter (4) associated to the detection probabilities (15) and an ideal one associated to measurements of product observables on the same state ρ can be estimated as

$$|S^{real} - S^{ideal}| \leq 4\sqrt{2}(\|R_1\| + \|R_2\|) + 2(\|R_1\|^2 + \|R_2\|^2 + \|R_1\|\|R_2\|) + 16\tilde{e} \tag{52}$$

where $\|R_1\|$ and $\|R_2\|$ are given by (37).

Sharper bounds can be obtained numerically as shown in [19].

Remark III.5 *The bounds (51) and (52) have a precise operative meaning. The experimental setting represented in Fig. 1 should ideally allow the measurements of factorized observables* $\mathbf{a} \cdot \boldsymbol{\sigma} \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$*, where* $\mathbf{a}$ *and* $\mathbf{b}$ *are initially chosen in the Bloch sphere. Furthermore, from this ideal perspective, the state of the photon exiting the generation stage in Fig. 1 should be the Bell state* $|\Psi_+\rangle = \frac{1}{\sqrt{2}}(|0H\rangle + i|1V\rangle)$*. However, the presence of non-idealities in the optical components inevitably modifies the factorized form of the actually measured observables on the one hand as well as the state of the photon on the other hand. We can nevertheless extract from the experimental data the values of measurement of generic factorized observables* $\mathbf{a}' \cdot \boldsymbol{\sigma} \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$ *up to a certain error depending on the given technical specifications of the components of the setup. As discussed in Remark III.1, we can in particular choose* $\mathbf{a}' = \mathbf{a}'(\phi, u_0, v_0) \equiv \mathbf{a}_0$ *in such a way that the error attains its minimal value. In this way the experimental data can be interpreted as a measurement of* $\mathbf{a}_0 \cdot \boldsymbol{\sigma} \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$*, affected by the said minimal error, which is estimated in (51) where* P^{ideal} *is ascribed to* $\mathbf{a}_0 \cdot \boldsymbol{\sigma} \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$*. Referring to a set of four similar observables, the same argument leads to the estimate (52) for the value of the CHSH parameter* S*.*

IV. SEMI-DEVICE-INDEPENDENT ENTROPY CERTIFICATION OF A QRNG BASED ON BELL INEQUALITY VIOLATION BY SPE STATES

Random numbers are a fundamental resource in several practical applications, ranging from Monte Carlo simulations to cryptography. In the latter case, the unpredictability of the sequence of random bits is an important issue, since it affects the security of the associated criptographic protocols. For this reason, it is important to certify randomness, i.e. to prove that the random numbers are uniformly distributed, uncorrelated and unpredictable. The first two features can be rather easily checked by running suitable statistical tests for the distribution of the output string of random bits. On the other hand, the proof of unpredictability is a rather challenging problem that cannot be tackled by properly tailored statistical tests. In principle a QRNG, whose entropy source is a quantum process, is more secure than a generic true random number generator since quantum physics is intrinsically probabilistic; most quantum phenomena are unpredictable and the theory allows to compute only the statistical distribution of the possible outcomes. In addition, from a practical point of view the detailed modelling of the underlying entropy source as well as the estimate of the min-entropy of the output string is robust and independent of additional classical noise sources [26]. In this framework, a rather challenging class of QRNG recently proposed consists in the so-called *device independent* QRNG, which in principle allow a certification of the quality and the security of the random numbers they produce *independently of any detailed model of the devices.* The main figure of merit characterizing the randomness as well as the security of the output string in the criptographic applications is the *guessing probability* p_g, defined as the probability of the most probable digit and the corresponding *min-entropy* H_∞, given by $H_\infty = -\log_2 p_g$ (see [24] for an operational meaning of this quantity). Hence, when a random number generator is used, it is important to provide a certification that the output string contains a certain amount of min-entropy, in such way that the application of a randomness extractor allows to obtain a sequence of almost-uniform random bits [27].

Let's come back to the framework described in section II, where measurements of two commuting observables $\mathbf{a} \cdot \boldsymbol{\sigma} \otimes I$ and $I \otimes \mathbf{b} \cdot \boldsymbol{\sigma}$ are performed on a quantum system with two independent degrees of freedom and associated Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ with $\mathcal{H}_A = \mathcal{H}_B = \mathbb{C}^2$. In the case where the state ρ of the system is a pure state, i.e. $\rho = |\psi\rangle\langle\psi|$, with ψ unit vector in $\mathcal{H}_A \otimes \mathcal{H}_B$, the amount of "quantum randomness" contained in the measurement outcomes is quantified by the guessing probability $G(\psi, \mathbf{a}, \mathbf{b})$ defined as

$$G(\psi, \mathbf{a}, \mathbf{b}) := \max_{x,y} P(x, y|\psi, \mathbf{a}, \mathbf{b}) \tag{53}$$

where $P(x, y|\psi, \mathbf{a}, \mathbf{b})$ are given by (2). An upper bound for $G(\psi, \mathbf{a}, \mathbf{b})$ lower than 1 certificates quantum randomness in the distribution P. In the general case of a mixed state ρ, the *quantum guessing probability* is defined as

$$G(\rho, \mathbf{a}, \mathbf{b}) = \sup_{\{(\mu(\lambda), \psi_\lambda)_{\lambda\in\Lambda}\}} \int_\Lambda G(\psi_\lambda, \mathbf{a}, \mathbf{b}) d\mu(\lambda) \ , \tag{54}$$

where the sup is taken over all possible decomposition $\{(\mu(\lambda), \psi_\lambda)_{\lambda\in\Lambda}\}$ of ρ into a convex superposition of pure states [28]:

$$\rho = \int_\Lambda |\psi_\lambda\rangle\langle\psi_\lambda| d\mu(\lambda) \ . \tag{55}$$

The idea behind expression (54) is to separate the quantum randomness of every pure state ψ_λ from its *classical weight* $\mu(\lambda)$, considering the exact value of the parameter λ as a piece of information in principle accessible to a potential eavesdropper. As the notation itself says, the quantum guessing probability (54) has an explicit dependence on the quantum state ρ and on the couple of measured observables $\mathbf{a}\cdot\boldsymbol{\sigma}\otimes I$ and $I\otimes\mathbf{b}\cdot\boldsymbol{\sigma}$. A further step towards a more robust certification of the quantum randomness leads to the *realization-independent quantum guessing probability* $G(P_{\rho,\mathbf{a},\mathbf{b}})$ associated to a given distribution

$$P_{\rho,\mathbf{a},\mathbf{b}} := \{P(x, y|\rho, \mathbf{a}, \mathbf{b}) \mid x, y \in \{-1, 1\}\}$$

of the form (2), which is defined by

$$G(P_{\rho,\mathbf{a},\mathbf{b}}) := \sup_{\{\tilde\rho,\tilde{\mathbf{a}},\tilde{\mathbf{b}}\}} G(\tilde\rho, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}), \tag{56}$$

where the supremum is evaluated over all possible triples $\{\tilde\rho, \tilde{\mathbf{a}}, \tilde{\mathbf{b}}\}$ of states $\tilde\rho$ and local observables $\tilde{\mathbf{a}}\cdot\boldsymbol{\sigma}\otimes I$ and $I\otimes\tilde{\mathbf{b}}\cdot\boldsymbol{\sigma}$ compatible with the distribution $P_{\rho,\mathbf{a},\mathbf{b}}$, i.e.:

$$P(x, y|\rho, \mathbf{a}, \mathbf{b}) = Tr(\tilde\rho P_x^{\tilde{\mathbf{a}}_i} \otimes P_y^{\tilde{\mathbf{b}}_j}) \quad \text{for all } x, y \in \{-1, 1\}.$$

This quantity actually provides a quantification of the amount of secure quantum randomness present in the distribution $P_{\rho,\mathbf{a},\mathbf{b}}$ independently of any accessible (classical) side information and any particular description of the system. In fact $G(P_{\rho,\mathbf{a},\mathbf{b}})$ depends only on the observed distribution of outcomes and gives a rather conservative bound on the probability of the most probable outcome in any (even the worst-case) scenario under the only assumption that the quantum distribution P is obtained from the measurement of observables in product form. Correspondingly, the min-entropy $H_\infty := -\log_2 G(P_{\rho,\mathbf{a},\mathbf{b}})$ expresses this guessing probability in bits. In our specific model with the above definition of $G(P_{\rho,\mathbf{a},\mathbf{b}})$, H_∞ varies in $[0, \alpha]$ where α is about 1.2 (see the analysis in [20] and the curve (a) in fig.2 therein, in particular). However, the upper bound (here $\alpha > 1$) depends on the considered system and on the precise definition of $G(P_{\rho,\mathbf{a},\mathbf{b}})$. In any cases, the more quantum randomness the analysed probability distribution contains, the more its entropy is positive and far from 0. As a consequence, lower bounds for H_∞ – i.e., upper bounds for $G(P_{\rho,\mathbf{a},\mathbf{b}})$ – may be of crucial interest in applications.

To this regard, in [22], the authors considered a Bell test (readapted to a quantum contextuality test in our case) with a measured CHSH parameter S obtained by the measured probability distributions $P_{\rho,\mathbf{a},\mathbf{b}}$ of four pairs observables $\mathbf{a}_i\cdot\boldsymbol{\sigma}\otimes I$ and $I\otimes\mathbf{b}_j\cdot\boldsymbol{\sigma}$, $(i, j = 1, 2)$, on the state ρ. They proved that the realization independent quantum guessing probability $G(P_{\rho,\mathbf{a},\mathbf{b}})$ of each of the four measured distributions $((\mathbf{a}, \mathbf{b}) \in \{(\mathbf{a}_i, \mathbf{b}_j) \mid i, j = 1, 2\})$ is bounded by

$$G(P_{\rho,\mathbf{a},\mathbf{b}}) \le \frac{1}{2} + \frac{1}{2}\sqrt{2 - \frac{S^2}{4}} \ . \tag{57}$$

Note that the right-hand side gives rise to a meaningful (i.e. < 1) upper bound for $G(P_{\rho,\mathbf{a},\mathbf{b}})$ only if $2 < S \le 2\sqrt{2}$, i.e. when S stays between the CHSH classical threshold and the Tsirelson limit.

In view of the discussion around the estimates (51) and (52), in the case of a test of quantum contextuality on SPE photons with realistic devices, inequality (57) has to be modified in the following way

$$G(P^{real}_{\rho,\mathbf{a}_i,\mathbf{b}_j}) \le \frac{1}{2} + \frac{1}{2}\sqrt{2 - \frac{(S^{real} - e_s))^2}{4}} + e_p \tag{58}$$

with $e_s = 4\sqrt{2}(\|R_1\| + \|R_2\|) + 2(\|R_1\|^2 + \|R_2\|^2 + \|R_1\|\|R_2\|) + 16\tilde e$ and $e_p = 2\sqrt{e} + e + \tilde e$, with e given by (32). This yields an equivalent lower bound for the associated min-entropy of the considered measured probability distributions with the following form

$$H_\infty \ge -\log_2\left(\frac{1}{2} + \frac{1}{2}\sqrt{2 - \frac{(S^{real} - e_s)^2}{4}} + e_p\right) \tag{59}$$

for the min entropy of the probability distribution computed out of the raw data of (nominal) measurements of $\mathbf{a}_i\cdot\boldsymbol{\sigma}\otimes\mathbf{b}_j\cdot\boldsymbol{\sigma}$ with the apparatus of Fig. 1, one of the four choices to obtain S^{real} to insert in right-hand side. These estimates are *semi-device independent* as, e.g., the right-hand side depends only on the technical features of beam splitters and mirrors of the preparation stage (see Fig. 1) embodied in e_s and e_p. The beam splitters are here considered as quite realistic: they can be different from each other, lossy, and acting differently on states with different polarization, though they are assumed not to change the polarization of the photons they handle. The estimate (59) is robust under classical side information.

In the practical implementation of the protocol described in [19], the estimate of S^{real} from the experimental data requires the fair sampling assumption. In

[22], the measured probabilities $P_{\rho,\mathbf{a},\mathbf{b}}$ is evaluated by randomly choosing the value of $(\mathbf{a},\mathbf{b})$ for each round of the experiment. This is possible also within our semi-device independent protocol, but is not strictly necessary. A predetermined sequence of $(\mathbf{a},\mathbf{b})$ can be used as long as this knowledge cannot be exploited by an adversary to maliciously modify the outcomes of the experiment [21].

V. SPE WITH REALISTIC DETECTORS

In this section, we analyse quantitatively realistic detectors and provide a technique for the construction of confidence intervals for the quantum probabilities (15) taking into account a finite number of experimental data. In particular, we develop a Markov model for the memory effects present in the sequence of measurement outcomes due to detector non idealities such as dead time, afterpulsing and dark counts rate (DCR). Finally, we propose an unbiased estimator for the quantum transition probabilities out of the collection of experimental data. This analysis plays an important role when the features of the light source do not allow a strict control of photon arrival times, for example when an attenuated classical source is used as in [17]. As anticipated in the introduction, the following analysis relies upon the fair sampling assumption. Specifically, in the application to the semi device independent QRNG protocol described in section III C, the provider of the detectors is assumed to be trusted.

As explained in [17], in an experiment of Bell inequality violation by SPE photons it is possible to associate to the photons of the incoming beam a sequence of independent identically distributed random variables $\{\xi_n\}_{n\geq 1}$ with four possible outcomes $i = 1,2,3,4$ associated to the four final channels of the measuring apparatus, i.e. the four detectors, and the corresponding probabilities p_i. We assume that the parameters characterizing the state preparation and the measurement stage are stable during the acquisition time, in such a way that the process $\{\xi_n\}_{n\geq 1}$ is stationary.

In the ideal case, i.e. if we neglect afterpulsing and dead time and we assume a detection efficiency equal to 100%, the sequence of measurement outcomes allows to estimate the theoretical probabilities p_i in term of the corresponding empirical frequencies N_i/N, where N_i, $i = 1,2,3,4$ is the number of counts on the i-th detector and $N = \sum_i N_i$ is the total number of detected photons. However, if we take into account the presence of detector non-idealities, such as dead time and afterpulsing, then we can no longer assume the independence of the sequence of measurement outcomes and memory effects arise. Indeed, when a photon is detected, the SPAD remains blind for a time interval of length T_d where it is unable to detect additional photons that may reach its sensitive volume. On the other hand, there is a non-negligible probability of afterpulsing, i.e., of a subsequent readout caused by a secondary event produced in the SPAD instead of a further incoming photon.

In the following, we shall assume that the four detectors have equal efficiency $\eta \in (0,1]$ and shall take it into account by replacing the intensity λ of the photon beam with an effective intensity $\lambda_e = \eta\lambda$. The probability of afterpulsing will be denoted by p_a and assumed to be of order 10^{-2} or less. The symbol T_d will denote the detector dead time, while $N(T_d)$ will be the number of photons reaching the detectors during the dead time. Let us call $\mathbb{P}(N(T_d) \geq n)$ the probability that a number of photons greater or equal to n reaches the detector photons during the dead time. We shall assume that $\mathbb{P}(N(T_d) = 1) = \epsilon$, with ϵ of order 10^{-2} or less, and that $\mathbb{P}(N(T_d) > 1) = o(\epsilon)$. This condition is fulfilled, e.g., if the photon source is an attenuated laser, yielding a Poissonian distribution of the arrival times. In this case we have $\mathbb{P}(N(T_d) = 1) = \lambda_e T_d e^{-\lambda_e T_d} \sim \lambda_e T_d$ and $\mathbb{P}(N(T_d) > 1) = 1 - e^{-\lambda_e T_d} - \lambda_e T_d e^{-\lambda_e T_d}$. If the expected value $\lambda_e T_d$ of $N(T_d)$ is of order 10^{-2} we have $\mathbb{P}(N(T_d) = 1) \sim \lambda_e T_d$ and $\mathbb{P}(N(T_d) > 1) \sim (\lambda_e T_d)^2/2 = o(\lambda_e T_d)$. Eventually, we shall assume that the time of afterpulsing T_a is of the order of the dead time, i.e. $T_a \sim CT_d$ with $C = O(1)$ and correspondingly $\mathbb{P}(N(T_a) = 1) = O(\epsilon)$ and $\mathbb{P}(N(T_a) > 1) = o(\epsilon)$.

Under these approximations it is possible to develop a Markov model for the correlations among the subsequent readouts of the detectors caused by non-idealities.

Let $\{\eta_n\}_{n\geq 1}$ be the sequence of random variables with 4 possible outcomes $i = 1,2,3,4$ associated to the subsequent readouts of the four detectors. Each realization of the sequence $\{\eta_n\}_{n\geq 1}$ actually gives the temporal sequence of outcomes of the measurements. In other words, with the statement $\eta_n = i$ we mean that the n-th data is collected by the i-th detector. In the ideal case, if dead time, DCR and afterpulsing are neglected, the two sequences $\{\xi_n\}_{n\geq 1}$ and $\{\eta_n\}_{n\geq 1}$ will have the same distribution. In the realistic case, the distribution of $\{\eta_n\}_{n\geq 1}$ is actually affected by the non-idealities of the measuring apparatus in the way we are going to describe. Let us consider first the case where DCR gives a negligible contribution and focus on the first detected photon. Since in this case neither of the four detectors is in dead time nor in afterpulsing caused by previous detections, the first variable η_1 has the same distribution of ξ_1, i.e. $\mathbb{P}(\eta_1 = i) = p_i$, $i = 1,...,4$. Let us consider now the second detection, whose statistics is described by the random variable η_2. The correlations between η_1 and η_2 are described by the set of conditional probabilities $\mathbb{P}(\eta_2 = j|\eta_1 = i)$, with $i,j = 1,...,4$. Taking into account afterpulsing, whose occurrence is denoted with the symbol AFP (while AFP^c denotes the complementary event), we have

$$\begin{aligned}\mathbb{P}(\eta_2 = j|\eta_1 = i) = \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP)p_a \\ + \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c)(1-p_a).\end{aligned}$$

We have implicitly assumed that all detectors have the same probability of afterpulsing $\mathbb{P}(AFP|\eta_1 = i) = p_a$, $i = 1,\ldots,4$. Concerning the first term, denoting by τ the interarrival time between the first and the second photon, we have:

$$\begin{aligned} &\mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP)\\ &= \mathbb{P}(\eta_2 = j|\eta_1 = i\cap AFP\cap \tau < T_a)\mathbb{P}(\tau < T_a|\eta_1 = i\cap AFP)\\ &+\mathbb{P}(\eta_2 = j|\eta_1 = i\cap AFP\cap \tau > T_a)\mathbb{P}(\tau > T_a|\eta_1 = i\cap AFP)\\ &= \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP \cap \tau < T_a)\mathbb{P}(N(T_a) > 0)\\ &\qquad + \delta_{ij}\mathbb{P}(N(T_a) = 0) \end{aligned}$$

Now, observing that this term has to be multiplied by $p_a \sim 10^{-2}$, in the case where $\mathbb{P}(N(T_a) > 0) = O(\epsilon)$, we can neglect the first term obtaining:

$$\mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP) \sim \delta_{ij}\mathbb{P}(N(T_a) = 0).$$

Let us consider now the probability $\mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c)$ that, given the result of the first measurement is i and no afterpulsing occurs, the result of the second measurement is j. Denoting with $N(T_d)$ the number of photons reaching the detectors during the dead time, we have

$$\begin{aligned} &\mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c)\\ &= \sum_{k=0} \mathbb{P}(\eta_2 = j \cap N(T_d) = k|\eta_1 = i \cap AFP^c)\\ &= \sum_{k=0} \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c \cap N(T_d) = k)\mathbb{P}(N(T_d) = k)\\ &= \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c \cap N(T_d) = 0)\mathbb{P}(N(T_d) = 0)\\ &+ \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c \cap N(T_d) = 1)\mathbb{P}(N(T_d) = 1)\\ &\qquad + o(\epsilon)\\ &= p_j(1-\epsilon+o(\epsilon)) + q_{ij}\epsilon + o(\epsilon) \end{aligned}$$

where, if $i = j$,

$$q_{ij} = \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c \cap N(T_d) = 1) = p_j^2,$$

while, if $i \neq j$,

$$q_{ij} = \mathbb{P}(\eta_2 = j|\eta_1 = i \cap AFP^c \cap N(T_d) = 1) = p_j + p_i p_j$$

Under the assumption that ϵ is so small than we can neglect all the terms of order $o(\epsilon)$, the conditional probabilities satisfy the Markov property

$$\begin{aligned} &\mathbb{P}(\eta_{n+1} = i_{n+1}|\eta_1 = i_1, \ldots, \eta_n = i_n)\\ &\quad= \mathbb{P}(\eta_{n+1} = i_{n+1}|\eta_n = i_n) = \mathbb{P}(\eta_2 = i_{n+1}|\eta_1 = i_n), \end{aligned} \tag{60}$$

since, in fact, the left hand side of (60) is equal to the right hand side plus additional terms which account for the cases where $N(T_d) \geq 2$ and these have a negligible probability. In summary, according to our approximations, the sequence of random variables $\{\eta_n\}_{n\geq 1}$ is a stationary Markov chain with transition probabilities $\mathbb{P}(\eta_{n+1} = j|\eta_n = i) = P_{ij}$ given (up to term of order $o(\epsilon)$ by:

$$P_{ij} = p_a\delta_{ij} + (1-p_a)\left((1-\epsilon)p_j + \epsilon q_{ij}\right)$$

Equivalently, the stochastic matrix P is equal to

$$P = p_a I_{4\times 4} + (1-p_a)\left((1-\epsilon)\tilde{P} + \epsilon Q\right),$$

with

$$\tilde{P} = \begin{pmatrix} p_1 & p_2 & p_3 & p_4\\ p_1 & p_2 & p_3 & p_4\\ p_1 & p_2 & p_3 & p_4\\ p_1 & p_2 & p_3 & p_4 \end{pmatrix}$$

$$Q = \begin{pmatrix} p_1^2 & p_2(1+p_1) & p_3(1+p_1) & p_4(1+p_1)\\ p_1(1+p_2) & p_2^2 & p_3(1+p_2) & p_4(1+p_2)\\ p_1(1+p_3) & p_2(1+p_3) & p_3^2 & p_4(1+p_3)\\ p_1(1+p_4) & p_2(1+p_4) & p_3(1+p_4) & p_4^2 \end{pmatrix}$$

If $p_i > 0$ for all $i = 1,2,3,4$, then the Markov chain is irreducible and by the ergodic theorem the empirical frequencies converge to the unique invariant distribution $(f_i)_{i=1,\ldots 4}$. More precisely, by denoting $N_n^i := \sum_{k=1}^n 1_{\eta_k = i}$, we have

$$\mathbb{P}\left(\lim_{n\to\infty}\frac{N_n^i}{n} - f_i = 0\right) = 1.$$

where (f_1, f_2, f_3, f_4) can be computed as the left eigenvector of the matrix P with eigenvalue 1:

$$f_i = \frac{p_i}{1+\epsilon p_i}\left(\sum_{j=1}^4 \frac{p_j}{1+\epsilon p_j}\right)^{-1} \sim p_i + \epsilon\, p_i(\sum_j p_j^2 - p_i)$$

The latter relation can be easily inverted up to terms of order $o(\epsilon)$ yielding:

$$p_i \sim f_i\left(1 + \epsilon\left(f_i + \sum_{j=1}^4 f_j^2\right)\right)$$

This formula provides a rough estimator of the theoretical probabilities p_i in terms of the empirical frequencies f_i, $i = 1, ..., 4$.

More precise and unbiased estimators for the parameters p_i, $i = 1, ..., 4$, as well as the corresponding confidence intervals can be obtained via the maximum likelihood principle. Given a realization of the Markov chain described above, i.e. a sequence of outcomes $\{x_i\}_{i=1,\ldots,n}$, with $x_i = 1,2,3,4$, its probability is given by

$$p_{x_1}\prod_{i=1}^{n-1} P_{x_i x_{i+1}},$$

and the corresponding log-likelihood is equal to

$$l(P) := \log(p_{x_1}\prod_{i=1}^{n-1} P_{x_i x_{i+1}}) \tag{61}$$

$$= \log(p_{x_1}) + \sum_{i,j=1,2,3,4} N_{ij}\log(P_{ij}), \tag{62}$$

where N_{ij} is the number of transitions from i to j. The estimated values of the parameters p_i as well as the corresponding confidence interval can be obtained by maximization of (61) under the four constraints $\sum_j P_{ij} = 1$, $i = 1, ..., 4$ (see [29, 30] for the underlying theory). A practical implementation of this technique is presented in [19], where the the statistical programming language R has been used.

Eventually, if DCR is not negligible then the model described above can be easily updated by replacing the distribution of the initial random variables ξ_n

$$\mathbb{P}(\xi_n = i) = p_i, \qquad i = 1, 2, 3, 4,$$

with the corrected values

$$\mathbb{P}(\xi_n = i) = \tilde{p}_i = (1 - p_{DCR})p_i + \frac{p_{DCR}}{4}, \qquad i = 1, 2, 3, 4,$$

where p_{DCR} is the total fraction of detected photons due to dark counts.

For an application of this technique to an experiment of CHSH violation by SPE photons coming from an attenuated laser source see [19].

VI. CONCLUSIONS

This work shows how the entropy of a realistic QRNG based on momentum-polarization single-photon entangled states can be certified despite the non-idealities of the employed optical and electronic devices. Our analysis starts by taking into account the polarization-dependent responses of beam splitters and mirrors to estimate, first, the real conditional probabilities $P^{real}_{\rho,\mathbf{a},\mathbf{b}}$ due to the optical setup, and, from that, the real CHSH correlation function S^{real}. Then, an upper bound e_P is carefully evaluated between the real conditional probabilities and the ideal ones. Consequently, an upper bound e_S is estimated for the distance $|S^{real} - S^{ideal}|$ between the real and the ideal CHSH functions used for proving the entanglement. According to device independent protocols where the violation of the CHSH inequality ensures a level of min-entropy, S^{real} together with e_P and e_S concur to finally define a modified expression for the certified lower value of the min-entropy. The resulting certified QRNG falls into the class of the *semi-device independent* ones, as a modeling is needed due to some feature of the preparation stage (beam splitters and mirrors optical responses), but not needed for others (momentum and polarization angles). Moreover, the fact that single photons are measured with realistic detectors having non-unitary efficiency and affected by dead time, afterpulsing and dark counts, is here considered by means of a Markovian model used to take memory effects into account by correcting for the measured probabilities and allowing for the construction of the corresponding confidence intervals tackling the issue of finite statistics. The model implicitly relies on the fair sampling assumption and on the hypothesis that the preparation and measurement parameters of the system are stable during the acquisition time. Finally, the present QRNG protocol is robust under classical side information (see [21]). We plan to generalize it to the case of quantum side information in a future paper along the lines of, e.g., [31, 32].

Our analysis demonstrates that single-photon entanglement is not simply interesting from a fundamental point of view [16], but it can be a practical resource in quantum information, thanks in particular to the fact that attenuated classical light sources can be used [17]. An example of application of entropy certification of a QRNG based on entangled single photons from a weak laser beam is demonstrated by the experiment reported in [19]. Nevertheless, our analysis of optical non-idealities can also be applied to experimental tests of quantum contextuality of single-particle entangled states generated from heralded single photons [8, 33–35], where the photon spin or polarization is necessarily one of the two degrees of freedom involved. Indeed, in these cases as well, beam splitters and mirrors - the optical source of an unwanted coupling between the different degrees of freedom - are needed to build a Mach-Zehnder interferometer serving as a gate for the other employed qubit, be it linear [8, 33] or angular [34, 35] momentum. In this sense, our results could be exploited to enable the certification of the entropy of a QRNG based on single-photon entanglement involving degrees of freedom others than momentum and polarization. And even more generally, they could potentially be of interest to tests of quantum contextuality of single-particle entanglement exploiting particles other than photons, e.g. neutrons [36, 37] or atoms [38].

ACKNOWLEDGMENTS

We acknowledge helpful discussions with C. Agostinelli on the code implementation of the Markovian model. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 820405. NL was supported by a fellowship of Q@TN within the PAT(AI) grant.

Appendix: Proof of equation (33)

Let us compute the minimum Hilbert-Schmidt distance between the operator $\tilde{U}^{real}_{\mathbf{a}}$ given by (28) (denoted below by $\tilde{U}^{real}$ for shortness) and a unitary operator of the form $A \otimes I$, with $A = e^{i\zeta}e^{i\alpha\,\hat{m}\cdot\boldsymbol{\sigma}}$. By explicit computation we have:

$$\begin{aligned}\|\tilde{U}^{real} - (A\otimes I)\|^2_{HS} &= \mathrm{Tr}[(|\tilde{U}^{real} - (A\otimes I))(|\tilde{U}^{real} - (A\otimes I)^{\dagger})] \\ &= 8 - 4\cos\zeta(\cos\alpha(\cos\theta_H + \cos\theta_V) \\ &\quad + \sin\alpha\,\hat{m}\cdot(\sin\theta_H\,\hat{n}_h + \sin\theta_V\,\hat{n}_V)) \qquad (A.1)\end{aligned}$$

Clearly, the minimum of the distance $\|\tilde{U}^{real} - (A\otimes I)\|_2$ is attained for those values of $\zeta, \alpha, \hat{m}$ that maximize the function $\cos(\zeta)\, f(\alpha, \hat{m})$, with $f(\alpha, \hat{m}) := \cos\alpha(\cos\theta + \cos\theta') +$

$\sin\alpha\,\hat{m}\cdot(\sin\theta\,\hat{n}+\sin\theta'\,\hat{n}')$. Since $\cos(\zeta)\in[-1,+1]$ the problem is reduced to maximize the absolute value of $f(\alpha,\hat{m})$, for $\alpha\in[0,2\pi]$ and $\hat{m}\in\mathbb{R}^3$, $\|\hat{m}\|=1$. By direct computation we obtain:

$$\max_{\alpha\in[0,2\pi],\|\hat{m}\|=1}|f(\alpha,\hat{m})| = (2+2\cos^2\phi/2\cos(\alpha+\beta-\alpha'-\beta') + 2\sin^2\phi/2\cos(\alpha-\beta-\alpha'+\beta'))^{1/2}$$

Hence, by direct computation

$$\max_{\phi/2\in[0,2\pi]}\ \min_{\alpha\in[0,2\pi],\|\hat{m}\|=1}\|\tilde{U}^{real}-(A\otimes I)\|_2^2 = 8\Big(1-\min\Big\{\left|\cos\left(\frac{\alpha_1^H+\alpha_2^H-\alpha_1^V-\alpha_2^V}{2}\right)\right|, \left|\cos\left(\frac{\alpha_1^H-\alpha_2^H-\alpha_1^V+\alpha_2^V}{2}\right)\right|\Big\}\Big)$$

and this final result coincides with (32).

More generally, we can consider the case where the unitary operator $A\otimes I$, with $A=e^{i\zeta}U_M$, $\zeta\in[0,2\pi]$ and $U_M\in SU(2)$, is replaced by the general unitary operator in the product form $A\otimes U_P$, $U_P\in SU(2)$. We shall use the notation $U_M=e^{i\alpha\,\hat{m}\cdot\boldsymbol{\sigma}}$ and $U_P=e^{i\beta\,\hat{k}\cdot\boldsymbol{\sigma}}$, $\alpha,\beta\in[0,2\pi]$, $\hat{m},\hat{k}\in\mathbb{R}^3$, $\|m\|=\|k\|=1$. In particular, if U_P has the form

$$U_P=\begin{pmatrix}u_{HH} & u_{HV}\\ u_{VH} & u_{VV}\end{pmatrix} = \begin{pmatrix}\cos\beta+i\sin\beta\,k_z & \sin\beta(ik_x+k_y)\\ \sin\beta(ik_x-k_y) & \cos\beta-i\sin\beta\,k_z\end{pmatrix}$$

the matrix associated to the tensor product $A\otimes U_P$ assumes the following block form

$$\begin{pmatrix}u_{HH}A & u_{HV}A\\ u_{VH}A & u_{VV}A\end{pmatrix}$$

Analogously as before, we look for the minimum of the Hilbert-Schmidt distance between the opearator $\tilde{U}^{real}=\begin{pmatrix}U(\theta,\hat{n}) & 0\\ 0 & U(\theta',\hat{n}')\end{pmatrix}$ and $A\otimes U_P$. By explicit computation, using the identity

$$|u_{HH}|^2+|u_{VV}|^2+|u_{VH}|^2+|u_{HV}|^2 = 2\cos^2\beta+2\sin^2\beta\,k_z^2+2\sin^2\beta\,(k_x^2+k_y^2)=2$$

we get

$$\|\tilde{U}^{real}-(A\otimes U_P)\|_2^2 = \mathrm{Tr}[(|\tilde{U}^{real}-(A\otimes U_P))(|\tilde{U}^{real}-(A\otimes U_P)^\dagger)] = \mathrm{Tr}[4I_{2\times2}-(u_{HH}U(\theta,\hat{n})^\dagger A+h.c.)-(u_{VV}U(\theta',\hat{n}')^\dagger A+h.c)].$$

Hence, we have to find the values of $\zeta,\alpha,\beta\in[0,2\pi]$ and $\hat{m},\hat{k}\in\mathbb{R}^3$, $\|\hat{m}\|=\|\hat{k}\|=1$, maximizing the function

$$\begin{aligned} g(\zeta,\alpha,\beta,\hat{m},\hat{k}) &:= \mathrm{Tr}[(u_{HH}U(\theta_H,\hat{n}_H)^\dagger A+h.c.) + (u_{VV}U(\theta_V,\hat{n}_V)^\dagger A+h.c\\ &= \cos\zeta\cos\beta\big(\cos\alpha(\cos\theta_H+\cos\theta_V)+\sin\alpha\,\hat{m}\cdot(\sin\theta_H\,\hat{n}_H + \sin\theta_V\,\hat{n}_V)\big)\\ &\quad -\sin\zeta\sin\beta k_z\big(\cos\alpha(\cos\theta_H-\cos\theta_V) + \sin\alpha\,\hat{m}\cdot(\sin\theta_H\,\hat{n}_H-\sin\theta_V\,\hat{n}_V)\big)\end{aligned}$$

By explicit computation we get:

$$\max_{\zeta,\alpha,\beta,\hat{m},\hat{k}} g(\zeta,\alpha,\beta,\hat{m},\hat{k}) = \max_{\alpha,\hat{m}}|f(\alpha,\hat{m})|$$

which yields again the same result in (32).

Appendix: Proof of inequality (49)

It is convenient to introduce the following notation for later use. Let V_H,V_V and be the matrices defined as

$$V_H:=\begin{pmatrix}t_{H,1} & ir_{H,1}\\ ir_{H,1} & t_{H,1}\end{pmatrix}V(\phi)\begin{pmatrix}t_{H,2} & ir_{H,2}\\ ir_{H,2} & t_{H,2}\end{pmatrix},$$
$$V_V:=\begin{pmatrix}t_{V,1} & ir_{V,1}\\ ir_{V,1} & t_{V,1}\end{pmatrix}V(\phi)\begin{pmatrix}t_{V,2} & ir_{V,2}\\ ir_{V,2} & t_{V,2}\end{pmatrix}$$

with $V(\phi)$ given by (7). Analogously, let U_H,U_V be the matrices defined as

$$U_H:=\frac{V_H}{c_H},\quad U_V:=\frac{V_V}{c_V}$$

where c_H and c_V are the two positive constant given by (48). We actually have:

$$U_H:=\begin{pmatrix}\tilde{t}_{H,1} & i\tilde{r}_{H,1}\\ i\tilde{r}_{H,1} & \tilde{t}_{H,1}\end{pmatrix}V(\phi)\begin{pmatrix}\tilde{t}_{H,2} & i\tilde{r}_{H,2}\\ i\tilde{r}_{H,2} & \tilde{t}_{H,2}\end{pmatrix}$$
$$U_V:=\begin{pmatrix}\tilde{t}_{V,1} & i\tilde{r}_{V,1}\\ i\tilde{r}_{V,1} & \tilde{t}_{V,1}\end{pmatrix}V(\phi)\begin{pmatrix}\tilde{t}_{V,2} & i\tilde{r}_{V,2}\\ i\tilde{r}_{V,2} & \tilde{t}_{V,2}\end{pmatrix}$$

In particular the operators $\tilde{U}^{real}_{\mathbf{a},\mathbf{b}}=(I\otimes U_{\mathbf{b}})\otimes U^{real}_{\mathbf{a}}$ and $U^{real}_{\mathbf{a},\mathbf{b}}$, defined respectively in (21) and (42), can be represented as

$$\begin{aligned}\tilde{U}^{real}_{\mathbf{a},\mathbf{b}} &= (I\otimes U_{\mathbf{b}})(P_HU_HP_H+P_VU_VP_V),\\ U^{real}_{\mathbf{a},\mathbf{b}} &= (I\otimes U_{\mathbf{b}})(c_HP_HU_HP_H+c_VP_VU_VP_V)\end{aligned}\tag{A.1}$$

where P_H resp. P_V are the projections operators on the subspaces of $\mathcal{H}_M\otimes\mathcal{H}_P$ spanned by the vectors $\{|0H\rangle,|1H\rangle\}$ resp. $\{|0V\rangle,|1V\rangle\}$.

By introducing the notation $P_y^{P,\mathbf{b}}:=U_{\mathbf{b}}^{+}P_y^PU_{\mathbf{b}}$, and using (A.1), we get that the difference parameter $\tilde{e}_{\mathbf{a},\mathbf{b}}$ defined in (45) can be estimated as:

$$\begin{aligned}\tilde{e}_{\mathbf{a},\mathbf{b}} = \Big|\,\mathrm{Tr}\Big[\Big(\Big(\frac{c_H^2}{D}-1\Big)P_HU_HP_H\rho P_HU_H^\dagger P_H &+ \Big(\frac{c_Vc_H}{D}-1\Big)\Big(P_VU_VP_V\rho P_HU_H^\dagger P_H+h.c.\Big)\\ &+\Big(\frac{c_V^2}{D}-1\Big)P_VU_VP_V\rho P_VU_V^\dagger P_V\Big)P_x^M\otimes P_y^{P,\mathbf{b}}\Big]\Big|\end{aligned}$$

where $D = c_H^2\,\mathrm{Tr}[P_H \rho P_H] + c_V^2\,\mathrm{Tr}[P_V \rho P_V]$.

Let $\mathcal{R}_{\mathbf{a}}$ be the operator defined as

$$\begin{aligned}\mathcal{R}_{\mathbf{a}} := \Big(&\left(\frac{c_H^2}{D} - 1\right) P_H U_H P_H \rho P_H U_H^\dagger P_H \\ &+ \left(\frac{c_V c_H}{D} - 1\right) \left(P_V U_V P_V \rho P_H U_H^\dagger P_H + h.c.\right) \\ &\qquad + \left(\frac{c_V^2}{D} - 1\right) P_V U_V P_V \rho P_V U_V^\dagger P_V \Big)\end{aligned}$$

in such a way that, e.g., $\tilde{e}_{\mathbf{a},\mathbf{b}} \leq \|\mathcal{R}_{\mathbf{a}}\|$ for any choice of $\mathbf{a}, \mathbf{b}$.

By adopting the decomposition (47) for the density matrix ρ:

$$\rho = \alpha P_H \rho_H P_H + \beta P_V \rho_V P_V + P_V p P_H + P_H p^\dagger P_V \ ,$$

where $\rho_H : \mathbb{C}_H^2 \to \mathbb{C}_H^2$ and $\rho_V : \mathbb{C}_V^2 \to \mathbb{C}_V^2$ are 2×2 density matrices and $\alpha, \beta \geq 0$ with $\alpha + \beta = 1$, whereas $p : \mathbb{C}_H^2 \to \mathbb{C}_V^2$. The positivity of the density matrix ρ yields

$$|\langle x, py\rangle|^2 \leq \langle x, \alpha\rho_H x\rangle\langle y, \beta\rho_V y\rangle \quad \forall x, y \in \mathbb{C}^2 \equiv \mathbb{C}_H^2 \equiv \mathbb{C}_V^2.$$

In particular p vanishes when either $\alpha = 0$ or $\beta = 0$. Therefore, it is safe to rename p in the previous decomposition as $\alpha\beta p$. The above decomposition now reads

$$\rho = \alpha P_H \rho_H P_H + \beta P_V \rho_V P_V + \sqrt{\alpha\beta} P_V p P_H + \sqrt{\alpha\beta} P_H p^\dagger P_V,$$

where $\alpha, \beta, \rho_H, \rho_V$ are as above and $p : \mathbb{C}_H^2 \to \mathbb{C}_V^2$ is an arbitrary operator satisfying

$$|\langle x, py\rangle|^2 \leq \langle x, \rho_H x\rangle\langle y, \rho_V y\rangle \quad \forall x, y \in \mathbb{C}^2 \equiv \mathbb{C}_H^2 \equiv \mathbb{C}_V^2 \ . \tag{A.2}$$

By exploiting this decomposition, the operator $\mathcal{R}_{\mathbf{a}}$ can be rephrased to

$$\mathcal{R}_{\mathbf{a}} = g_1(\alpha, \beta) Q_1 + g_2(\alpha, \beta) Q_2$$

where

$$g_1(\alpha, \beta) := \frac{\sqrt{\alpha\beta}(c_H c_V - \alpha c_V^2 - \beta c_H^2)}{\alpha c_H^2 + \beta c_V^2} \ ,$$

$$Q_1 := \begin{pmatrix} 0 & U_H p^\dagger U_V^\dagger \\ U_V p U_H^\dagger & 0 \end{pmatrix} \ ,$$

and

$$g_2(\alpha, \beta) := \frac{\alpha\beta(c_H^2 - c_V^2)}{\alpha c_H^2 + \beta c_V^2}$$

$$Q_2 = \begin{pmatrix} U_H \rho_H U_H^\dagger & 0 \\ 0 & -U_V \rho_V U_V^\dagger \end{pmatrix} \ .$$

As a consequence

$$||g_1 Q_1||^2 = g_1^2 ||Q^\dagger Q|| \ .$$

The last norm can be estimated observing that

$$Q_1^\dagger Q_1 := \begin{pmatrix} U_H p^\dagger p U_V^\dagger & 0 \\ 0 & U_V p p^\dagger U_H^\dagger \end{pmatrix}$$

hence

$$||Q_1^\dagger Q_1|| = \max\{||U_H p^\dagger p U_V^\dagger||, ||U_V p p^\dagger U_H^\dagger||\}$$

$$= \max\{||p^\dagger p||, ||p p^\dagger||\} = \max\{||p||^2, ||p^\dagger||^2\} = ||p||^2 \ .$$

From (A.2), with $y \in \mathbb{C}^2$ with $||y|| = 1$

$$||py||^4 = |\langle py, py\rangle|^2 \leq \langle py, \rho_H py\rangle\langle y, \rho_V y\rangle = ||\sqrt{\rho_H} py||^2 ||\sqrt{\rho_V} y||^2$$

hence

$$||py||^4 \leq ||\sqrt{\rho_H}||^2 ||py||^2 ||\sqrt{\rho_V} y||^2,$$

so that

$$||p||^2 \leq ||\sqrt{\rho_H}||^2 ||\sqrt{\rho_V}||^2 = ||\rho_H||\,||\rho_V|| \ ,$$

and eventually

$$||g_1(\alpha, \beta) Q_1|| \leq |g_1(\alpha, \beta)| \sqrt{||\rho_H||\,||\rho_V||}$$

The operator Q_2 is in block form and its eigenvalues can be computed as the eigenvalues of the two blocks, which are given by density matrices (with eigenvalues bounded by 1). This allows to conclude that for any choice of ρ the operator norm of this term is bounded by 1. Analogously, a similar estimate can be obtained for the term $|\sqrt{||\rho_H||\,||\rho_V||}|$. This gives

$$\|\mathcal{R}_{\mathbf{a}}\| \leq |g_1(\alpha, \beta)| + |g_2(\alpha, \beta)| \tag{A.3}$$

where now the constants $g_i(\alpha, \beta)$, $i = 1, 2$, do not depend explicitly on ϕ, or equivalently on $\mathbf{a}$, and we eventually obtain (49).

[1] A. Einstein, B. Podolsky, and N. Rosen, Can quantum-mechanical description of physical reality be considered complete?, Physical Review **47**, 777 (1935).

[2] J. Bell, On the einstein podolski rosen paradox, Physics **1**, 195 (1964).

[3] A. K. Ekert, Quantum cryptography based on bell's theorem, Phys. Rev. Lett. **67**, 661 (1991).

[4] C. H. Bennett and S. J. Wiesner, Communication via

one- and two-particle operators on einstein-podolsky-rosen states, Phys. Rev. Lett. **69**, 2881 (1992).

[5] A. Ekert, R. Jozsa, and P. Marcer, Quantum algorithms: Entanglement-enhanced information processing [and discussion], Philosophical Transactions: Mathematical, Physical and Engineering Sciences **356**, 1769 (1998).

[6] C. Macchiavello, On the role of entanglement in quantum information, Physica A: Statistical Mechanics and its Applications **338**, 68 (2004), proceedings of the conference A Nonlinear World: the Real World, 2nd International Conference on Frontier Science.

[7] M. A. Nielsen and I. Chuang, Quantum computation and quantum information (2002).

[8] B. R. Gadway, E. J. Galvez, and F. De Zela, Bell-inequality violations with single photons entangled in momentum and polarization, Journal of Physics B: Atomic, Molecular and Optical Physics **42**, 015503 (2009).

[9] A. Vallés, V. D'Ambrosio, M. Hendrych, M. Mičuda, L. Marrucci, F. Sciarrino, and J. P. Torres, Generation of tunable entanglement and violation of a bell-like inequality between different degrees of freedom of a single photon, Physical Review A **90**, 052326 (2014).

[10] Y. Hasegawa, R. Loidl, G. Badurek, M. Baron, and H. Rauch, Violation of bell-type inequality in single-neutron interferometry: quantum contextuality, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **529**, 182 (2004).

[11] Y. Hasegawa, K. Durstberger-Rennhofer, S. Sponar, and H. Rauch, Kochen–specker theorem studied with neutron interferometer, Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment **634**, S21 (2011).

[12] C. Simon, M. Żukowski, H. Weinfurter, and A. Zeilinger, Feasible kochen-specker experiment with single particles, Physical Review Letters **85**, 1783 (2000).

[13] Y.-F. Huang, C.-F. Li, Y.-S. Zhang, J.-W. Pan, and G.-C. Guo, Experimental test of the kochen-specker theorem with single photons, Physical Review letters **90**, 250401 (2003).

[14] P. Saha and D. Sarkar, Robustness measure of hybrid intra-particle entanglement, discord, and classical correlation with initial werner state, Quantum Information Processing **15**, 791 (2016).

[15] S. Adhikari, D. Home, A. S. Majumdar, A. K. Pan, A. Shenoy H., and R. Srikanth, Toward secure communication using intra-particle entanglement, Quantum Information Processing **14**, 1451 (2015).

[16] S. Azzini, S. Mazzucchi, V. Moretti, D. Pastorello, and L. Pavesi, Single-particle entanglement, Advanced Quantum Technologies **3**, 2000014 (2020).

[17] M. Pasini, N. Leone, S. Mazzucchi, V. Moretti, D. Pastorello, and L. Pavesi, Bell-inequality violation by entangled single-photon states generated from a laser, an led, or a halogen lamp, Physical Review A **102**, 063708 (2020).

[18] J.-Å. Larsson, Bell's inequality and detector inefficiency, Physical Review A **57**, 3304 (1998).

[19] N. Leone, S. Azzini, S. Mazzucchi, V. Moretti, and L. Pavesi, Certified quantum random numbers based on single-photon entanglement (2021), arXiv:2104.04452 [quant-ph].

[20] 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, and et al., Random numbers certified by bell's theorem, Nature **464**, 1021 (2010).

[21] S. Pironio and S. Massar, Security of practical private randomness generation, Physical Review A **87**, 012336 (2013).

[22] A. Acín, S. Massar, and S. Pironio, Randomness versus nonlocality and entanglement, Physical review letters **108**, 100402 (2012).

[23] A. Acín and L. Masanes, Certified randomness in quantum physics, Nature **540**, 213 (2016).

[24] R. Konig, R. Renner, and C. Schaffner, The operational meaning of min-and max-entropy, IEEE Transactions on Information theory **55**, 4337 (2009).

[25] J. Silman, S. Pironio, and S. Massar, Device-independent randomness generation in the presence of weak cross-talk, Physical review letters **110**, 100504 (2013).

[26] D. Frauchiger, R. Renner, and M. Troyer, True randomness from realistic quantum devices, arXiv preprint arXiv:1311.4547 (2013).

[27] N. Nisan and A. Ta-Shma, Extracting randomness: A survey and new constructions, Journal of Computer and System Sciences **58**, 148 (1999).

[28] Identity (55) is meant in *weak sense*, where μ is any probability measure over a parameter set Λ and the vectors ψ_λ are not mutually orthogonal in general.

[29] P. Billingsley, Statistical methods in markov chains, The Annals of Mathematical Statistics , 12 (1961).

[30] E. Ammicht and H. Wenzelburger, Maximum-likelihood estimators for the transition probabilities of a reversible markov chain, IFAC Proceedings Volumes **15**, 1113 (1982).

[31] R. Arnon-Friedman, F. Dupuis, O. Fawzi, R. Renner, and T. Vidick, Practical device-independent quantum cryptography via entropy accumulation, Nature communications **9**, 1 (2018).

[32] Y. Zhang, H. Fu, and E. Knill, Efficient randomness certification by quantum probability estimation, Physical review research **2**, 013016 (2020).

[33] M. Michler, H. Weinfurter, and M. Żukowski, Experiments towards falsification of noncontextual hidden variable theories, Phys. Rev. Lett. **90**, 250401 (2000).

[34] L. Chen and W. She, Single-photon spin-orbit entanglement violating a Bell-like inequality, J. Opt. Soc. Am. B **27**, A7 (2010).

[35] E. Karimi, J. Leach, S. Slussarenko, B. Piccirillo, L. Marrucci, L. Chen, W. She, S. Franke-Arnold, M. J. Padgett, and E. Santamato, Spin-orbit hybrid entanglement of photons and quantum contextuality, Phys. Rev. A **82**, 022115 (2010).

[36] H. Geppert, T. Denkmayr, S. Sponar, H. Lemmel, and Y. Hasegawa, Improvement of the polarized neutron interferometer setup demonstrating violation of a Bell-like inequality, Nuclear Instruments and Methods in Physics Research A **763**, 417 (2014).

[37] J. Shen, S. J. Kuhn, R. M. Dalgliesh, V. O. de Haan, N. Geerits, A. A. M. Irfan, F. Li, S. Lu, S. R. Parnell, J. Plomp, A. A. van Well, A. Washington, D. V. Baxter, G. Ortiz, W. M. Snow, and R. Pynn, Unveiling contextual realities by microscopically entangling a neutron, Nature Communications **11**, 930 (2020).

[38] F. Jeske, T. Stöferle, and M. DeKieviet, Massive spin-momentum entanglement measured in an atomic beam

spin echo experiment, Eur. Phys. J. D **82**, 25 (2011).