arXiv:2311.02182v3 [quant-ph] 8 Aug 2025

Noise-robust proofs of quantum network nonlocality

Sadra Boreiri†, Bora Ulu†, Nicolas Brunner, and Pavel Sekatski

Department of Applied Physics, University of Geneva, Switzerland

Quantum networks allow for novel forms of quantum nonlocality. By exploiting the combination of entangled states and entangled measurements, strong nonlocal correlations can be generated across the entire network. So far, all proofs of this effect are essentially restricted to the idealized case of pure entangled states and projective local measurements. Here we present noise-robust proofs of network quantum nonlocality, for a class of quantum distributions on the triangle network that are based on entangled states and entangled measurements. The key ingredient is a result of approximate rigidity for local distributions that satisfy the so-called "parity token counting" property with high probability. Our methods can be applied to any type of noise. As illustrative examples, we consider quantum distributions obtained with imperfect sources and obtain a noise robustness up to $\sim 80\%$ for dephasing noise and up to $\sim 0.5\%$ for white noise. Additionally, we prove that all distributions in the vicinity of some ideal quantum distribution are nonlocal, with a bound on the total-variation distance $\sim 0.25\%$. Our work opens interesting perspectives towards the practical implementation of quantum network nonlocality.

1 Introduction

A growing interest has recently been devoted to the question of quantum nonlocality in networks, see e.g. [1] for a review. A general framework has been developed to investigate nonlocal correlations in networks featuring independent sources [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13]. A central motivation comes from the fact that the network setting allows for novel forms of nonlocality compared to the more standard Bell test, where a single source distributes a physical resource (e.g. an entangled quantum state) to the parties.

Notably, it is possible to observe quantum nonlocality with parties performing a fixed measurement [4, 3]; see also [14] for a recent experiment. That is, in each round of the experiment, each party provides an output, but they receive no input. This in contrast to the standard Bell test where the use of measurement inputs is essential.

Another key aspect of the network setting is the possibility to exploit joint entangled measurements (where one or more eigenstates are non-separable) [15]. Similarly, as in quantum teleportation, the combination of entangled states and entangled joint measurements allows for strong correlations to be distributed across the entire network. Notably, this leads to novel forms of quantum nonlocal correlations that are genuine to the network structure [16], in the sense that they crucially rely on the use of entangled joint measurements [17, 18]. This is possible for a simple triangle network, as in Fig. 1.

While these works represent significant progress in our understanding of quantum nonlocality, they suffer from one major limita-

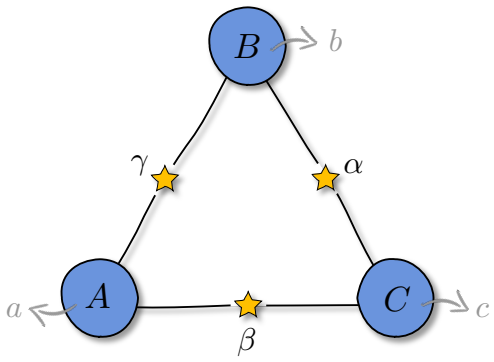


Figure 1: The triangle network features three distant parties, connected pairwise by three independent sources distributing physical systems.

† These authors contributed equally to this work

tion. Namely, these proofs of genuine quantum network nonlocality are derived in an ideal noiseless setting, where the sources distribute ideal (pure) quantum states and nodes perform ideal (projective) quantum measurements. Formally, these proofs are based on the concept of "token counting" rigidity [16, 19, 20], which so far only applies to the ideal (noiseless) case. Hence, these nonlocality proofs no longer apply when any realistic noise level is considered. This represents a major hurdle towards any practical implementation or application of these ideas, as any experiment features a certain amount of noise originating from unavoidable technical imperfections. Moreover, numerical analysis based on neural networks indicates noise-robust nonlocality [21].

One can draw a parallel with the situation of nonlocality in the mid-sixties. Indeed, John Bell's initial proof of nonlocality applied only to the idealized case of a pure entangled state and projective local measurements [22]. A key step was then made by Clauser, Horne, Shimony, and Holt who derived a noise-robust proof of quantum nonlocality via their famous CHSH Bell inequality [23], which enabled the first Bell experiments [24, 25]. This is still today the most commonly used test of nonlocality [26, 27, 28], and the basis of many applications within the device-independent approach (see e.g. [29]).

In this work we address the question of noise-robust proofs of quantum network nonlocality. We derive methods for this problem and apply them to several examples of quantum nonlocal distributions that are based on the concept of "parity token counting" (PTC) [20]. Our main contribution is a result of approximate rigidity for local distributions that satisfy the PTC property with a high (but non-unit) probability. Importantly, this method can be applied to any type of noise. We present illustrative examples, considering quantum distributions obtained from imperfect sources, featuring either dephasing or white noise. Additionally, we show that all distributions in the vicinity of some ideal quantum distribution are nonlocal, with a bound on the total-variation distance, which can be viewed as a nonlinear Bell inequality for the triangle network. Our work opens interesting perspectives towards the practical implementation of Bell nonlocality in quantum networks.

2 Summary of results

To start, we first provide a summary of the main contributions of this work.

1. We present a general family of quantum distributions on the triangle network (without inputs) and prove their nonlocality in the ideal (noiseless) case, see Section 5. These distributions include as a special case the distributions of Renou et al. [16], from now on referred to as RGB4. The nonlocality proof is based on the "parity token counting" (PTC) rigidity property [20]. In words, it states that on the triangle network with binary outcomes, any local distribution for which the sum of the outputs is always odd (1 or 3) has a unique corresponding local model.

2. We derive noise-robust nonlocality proofs for these quantum distributions, see Section 5.2. Our methods can be applied to arbitrary type of noise. As an illustration, we demonstrate the nonlocality of quantum distributions where up to $\sim 80\%$ dephasing noise, or $\sim 0.55\%$ white noise, is added at each source. We also investigated other noise models, such as no-click and photon loss; see Appendix E. Moreover, in Section 5.3 we characterize the distributions in the vicinity of a given nonlocal distribution. For a specific quantum distribution, we can show that any distribution in a $\sim 0.25\%$ total-variation distance ball around it is nonlocal. This can be viewed as a Bell inequality, in the sense that it guarantees that a full-measure region of the correlation set (the ball around the quantum distribution) is nonlocal.

3. The key ingredient which allows us to prove noise-robust nonlocality, is a result on the approximate PTC rigidity, discussed in Section 4. It gives a quantitative characterization of all local models leading to distributions for which the PTC condition holds approximately. That is, the distributions for which the sum of the outputs is odd with high probability. With the help of this result, we also derive an analytical outer relaxation of the set of binary output correlations compatible with a triangle-local model.

Below we start with Section 3 where we introduce more formally the context, the notations, and review the concept of parity token counting (PTC) distributions.

3 Background

To start we recall some background notions and fix the notations. Consider a triangle network as in Fig. 1, where three independent sources (labeled α, β, and γ) distribute physical systems to three distant parties (nodes A, B, and C). Each party thus receives two physical subsystems (from two different sources), and produces a measurement outcome, denoted by the classical variables a, b, and c (of finite cardinality d). The experiment results in the joint probability distribution $P(a,b,c)$.

We aim to understand which distributions are compatible with the triangle causal network depicted in Fig. 1. The set of possible distributions depends on the underlying physical theory. We are interested here in distributions that can be realized in classical physics (where the sources produce correlated classical variables) and in quantum theory (where the sources can produce entangled states). Our focus will be on quantum distributions that cannot be realized classically, hence demonstrating network nonlocality.

First, a distribution $P_L(a,b,c)$ is said to be triangle-local if it admits a classical model, i.e. it can be decomposed as

$$P_L(a,b,c) = \mathbb{E}\Big(P_A(a|\beta,\gamma)P_B(b|\gamma,\alpha)P_C(c|\alpha,\beta)\Big), \tag{1}$$

where $\mathbb{E}$ is the expected value over the independently distributed classical variables α,β,γ [30]. Furthermore, the conditional probabilities $P_X(x|\xi,\xi')$ can be considered deterministic without loss of generality, and represented by the response functions $x(\xi,\xi')$. A distribution that does not admit a decomposition of the form (1) is termed triangle-nonlocal. For brevity, we refer to triangle-(non)local correlations simply as (non)local throughout the text.

Second, we say that a distribution $P_Q(a,b,c)$ is triangle-quantum if it can be written as

$$P_Q(a,b,c) = \mathrm{tr}\Big(\rho^{\alpha}_{B_LC_R}\otimes\rho^{\beta}_{C_LA_R}\otimes\rho^{\gamma}_{A_LB_R} \times E^{a}_{A_LA_R}\otimes E^{b}_{B_LB_R}\otimes E^{c}_{C_LC_R}\Big) \tag{2}$$

where ρ^{ξ} denote the quantum states distributed by the sources (density operators), and $\{E^{x}_{YY'}\}$ denote the measurements performed at each party and are given by positive operator valued measures (POVMs). Note that one should pay attention to the order of subsystems when computing (2).

For a given output cardinality d we denote the set of all triangle-local distributions $\mathcal{L}^d_\triangle$ and the set of all triangle-quantum distributions $\mathcal{Q}^d_\triangle$. In general, characterizing these sets is a challenging problem. Outer approximations can be derived [31, 6, 9, 8], but typically only provide loose bounds.

For $d \geq 3$, it is known that $\mathcal{L}^d_\triangle \subset \mathcal{Q}^d_\triangle$, i.e. there exist quantum distributions that are nonlocal (see [20] and references therein). For $d = 2$, this is still an open question.

Known proofs of quantum nonlocality for the triangle network are of two types. The first are distributions that can be viewed as a clever embedding of a standard Bell test (e.g. CHSH) on the triangle network without inputs [4]. The idea is to distribute effective inputs via two classical sources. Each party will broadcast the received inputs via their output. The third source distributes the entangled state as in the usual Bell test. See also [32, 33, 34, 20] for more examples. For this class of distributions, noise-robust methods have been recently developed [35], leading to a first experiment [14]. Due to their connection with standard Bell tests, these distributions can be implemented without the need for any entangled measurements. Hence, the nonlocality of these distributions originates solely from the use of an entangled state.

In contrast, there exist quantum distributions on the triangle network based on the judicious combination of entangled states and measurements, as first shown by Renou and colleagues [16]. It was recently shown that this distribution requires in fact the use of entangled measurements [18], hence demonstrating genuine network quantum nonlocality [17]. So far, these nonlocality proofs are essentially restricted to an ideal noiseless scenario; see e.g. [16, 19, 36, 20, 37]. That is, they consider a setting where the shared states ρ^{ξ} are pure and the local measurements $\{E^{x}_{YY'}\}$ are projective. However, these proofs no longer apply when any realistic amount of noise is

included[1]. On the other hand, numerical analysis based on machine learning suggests that the non-locality of these quantum distributions is in fact robust to noise [21], up to $\approx 10\%$ white noise at the sources for some instances of the RGB4 distribution. Deriving noise-robust analytical proofs is therefore an important open problem, which we address in this work.

3.1 PTC distributions and rigidity

A key class of distributions for our work is termed parity token counting (PTC). These are local distributions on the triangle network, obtained via the following model. Each source possesses a single token that it sends to either one of its connected parties with some given probability. For instance, the source α sends its token to Bob with probability p_α and to Charlie with probability $1 - p_\alpha$. Then each party outputs the parity of the total number of received tokens. Therefore, consider that we define the binary random variable $t_\alpha \in \{0, 1\}$ as the number of tokens α sends to Bob, $t_\beta \in \{0, 1\}$ as the number of tokens β sends to Charlie, and $t_\gamma \in \{0, 1\}$ as the number of tokens γ sends to Alice. Then for example, Alice receives t_γ from the source γ and $t_\beta \oplus 1$ from the source β and her response function would be $a(\beta, \gamma) = t_\gamma \oplus t_\beta \oplus 1$, where $\oplus$ denotes the sum modulo two.

The distribution $P(a, b, c)$ resulting from the above PTC model has binary outputs ($d = 2$). By construction, any such distribution satisfies

$$a \oplus b \oplus c = 1, \tag{3}$$

which we refer to as the PTC condition.

Let us now explore the reverse link. Starting from a distribution that satisfies the PTC condition, what can we infer about the underlying classical model? Interestingly, PTC distributions demonstrate a form of rigidity [20]; note that this property of rigidity was initially discussed for token counting models [19]. Rigidity means that a PTC distribution can only be obtained via a PTC model (up to irrelevant relabellings). That is, for any local model leading to a PTC distribution, there exists for each source a token function

$$\begin{aligned} &T_\xi : \xi \to t_\xi \in \{0, 1\} \qquad \text{such that} \\ &x(\xi, \xi') = t_\xi \oplus t_{\xi'} \oplus 1, \end{aligned} \tag{4}$$

for the sources $\xi, \xi' \in \{\alpha, \beta, \gamma\}$ and $x \in \{a, b, c\}$ representing the party connected to the sources ξ, ξ'. In addition, if the probability distribution is such that $E_a, E_b, E_c \neq 0$ for $E_x = \Pr(x = 0) - \Pr(x = 1)$, the token functions can be chosen to fulfill

$$p_\alpha(t) := \Pr(t_\alpha = t) = \frac{1}{2}\left(1 + (-1)^t \sqrt{\left|\frac{E_b E_c}{E_a}\right|}\right) \tag{5}$$

and similar equalities of the other sources. This means that there is essentially only a unique local model over the triangle which can simulate such a distribution. In [20] one can find a simple characterization of the local PTC set $\mathcal{L}^2_{\triangle|PTC}$, containing all triangle-local distributions that satisfy the PTC condition of Eq. (3). A graphical representation of $\mathcal{L}^2_{\triangle|PTC}$ is given in Fig. 2, we find that it covers only about 12.5% of the volume[2] of $\mathcal{P}^2_{PTC}$ – the set of binary probability distributions $P(a, b, c)$ fulfilling Eq. (3).

Remarkably, the PTC rigidity, albeit formulated differently, is also true for any quantum PTC distribution [18]. In particular, this implies that in the slice of the probability set where the parity condition in Eq. (3) holds (ideal case without noise), there is no separation between the quantum and local sets, i.e. $\mathcal{Q}^2_{\triangle|PTC} = \mathcal{L}^2_{\triangle|PTC}$.

4 Approximate PTC rigidity for local distributions

Let us now consider a local distribution $P(a, b, c)$ for which the parity condition only holds approximately, i.e.

$$\Pr(a \oplus b \oplus c = 1) \geq 1 - \varepsilon. \tag{6}$$

Intuitively, one expects the underlying local model to verify some approximate form of rigidity. This is what we now establish.

[1] Note the semi-analytical methods of inflation [6] can detect a specific instance of the RGB4 distribution, but the resulting noise-robustness is around 10^{-7} for white noise at the sources, so irrelevant from any practical perspective.

[2] We have computed the volume of the set by uniformly sampling 10^6 points P from the PTC slice $\mathcal{P}^2_{PTC}$ and verifying for each point if it admits a local model or not. For symmetric distributions $E_A = E_B = E_C$ we find that the set $\mathcal{L}^2_{\triangle|PTC}$ is characterized by a simple inequality $P(1, 1, 1) \geq \frac{1}{4}$.

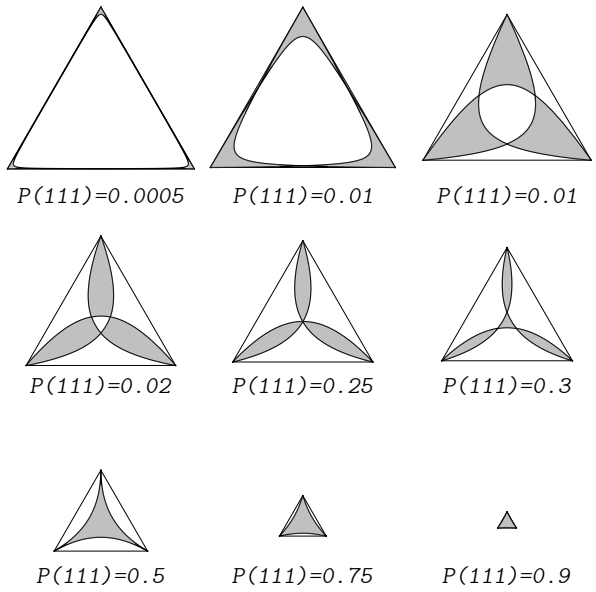


Figure 2: A layer-by-layer representation of the set $\mathcal{L}^2_{\triangle|PTC} = \mathcal{Q}^2_{\triangle|PTC}$, containing all triangle-local (or equivalently quantum) distributions $P(a,b,c)$ with binary outcomes, satisfying the deterministic PTC constraint $\Pr(a \oplus b \oplus c = 1) = 1$. Each layer depicts the set for a given value of $P(a = b = c = 1)$. The distributions inside the set can be represented by the vector $\boldsymbol{v} = \big(P(1,0,0), P(0,1,0), P(0,0,1)\big)$, with the remaining nonzero probability $P(1,1,1)$ given by normalization. Each plot in the figure is a slice of the set for a fixed value of $P(1,1,1)$, i.e. it is a cut of the set with the plane define by the equation $\boldsymbol{v} \cdot (1,1,1) = 1 - P(1,1,1)$. In each layer, the outer triangle delimits the set of valid probabilities. A 3D representation of the set $\mathcal{L}^2_{\triangle|PTC}$ in printable format can be obtained from the authors upon request.

Since the response functions are deterministic[3], there is a set

$$\Lambda = \{\boldsymbol{\xi} | a(\beta,\gamma) \oplus b(\alpha,\gamma) \oplus c(\alpha,\beta) = 1\} \tag{7}$$

containing all the values $\boldsymbol{\xi} = (\alpha,\beta,\gamma)$ for which the parity condition in Eq. (3) is true. By assumption we have $\Pr(\boldsymbol{\xi} \in \Lambda) \geq 1 - \varepsilon$.

Let us now take any value γ_* for which $\Pr(\boldsymbol{\xi} \in \Lambda | \gamma = \gamma_*) \geq 1 - \varepsilon$. Such a value must exist, since this probability cannot be strictly lower than its average on all γ. For the chosen value of γ, we define the token function to be

$$t_{\gamma_*} = 1. \tag{8}$$

Next, take $G_* = \{(\alpha,\beta) | (\alpha,\beta,\gamma_*) \in \Lambda\}$ to be the set of all pairs (α,β) for which the parity condition holds when $\gamma = \gamma_*$. On this set we now define the token functions t_α and t_β as

$$\begin{aligned} t_\alpha &= b(\alpha,\gamma_*) \\ t_\beta &= a(\beta,\gamma_*). \end{aligned} \tag{9}$$

Let us check that for $\gamma = \gamma_*$ and any values α and β in G_*, the PTC rigidity of Eq.(4) holds. By construction we have $t_\beta \oplus t_{\gamma_*} \oplus 1 = t_\beta = a(\beta,\gamma_*)$ and $t_\alpha \oplus t_{\gamma_*} \oplus 1 = b(\alpha,\gamma_*)$, while

$$t_\beta \oplus t_\alpha \oplus 1 = a(\beta,\gamma_*) \oplus b(\alpha,\gamma_*) \oplus 1 = c(\alpha,\beta) \tag{10}$$

follows from the parity condition at $(\alpha,\beta,\gamma_*) \in \Lambda$. In addition, we have a lower bound on the probability that (α,β)s are sampled in G_*

$$\Pr((\alpha,\beta) \in G_*) = \Pr(\boldsymbol{\xi} \in \Lambda | \gamma = \gamma_*) \geq 1 - \varepsilon. \tag{11}$$

Finally, consider the set

$$\Omega = \{\boldsymbol{\xi} | (\alpha,\beta) \in G_* \text{ and } \boldsymbol{\xi} \in \Lambda\}. \tag{12}$$

which has the probability

$$\begin{aligned} \Pr(\boldsymbol{\xi} \in \Omega) &\geq 1 - \Pr(\boldsymbol{\xi} \notin \Lambda) - \Pr((\alpha,\beta) \notin G_*) \\ &\geq 1 - 2\varepsilon. \end{aligned} \tag{13}$$

For any point (α,β,γ) from this set, we can define the prolongation of the token function t_γ as

$$\begin{aligned} t_\gamma^{(\alpha,\beta)} &= 1 \oplus a(\beta,\gamma) \oplus a(\beta,\gamma_*) \\ &= 1 \oplus b(\alpha,\gamma) \oplus b(\alpha,\gamma_*), \end{aligned} \tag{14}$$

where the equality of the two definitions follows from the fact that both points (α,β,γ) and (α,β,γ_*) are inside Λ by construction. It is also easy to see that this value t_γ satisfies the PTC rigidity, i.e. $t_\gamma^{(\alpha,\beta)} \oplus t_\beta \oplus 1 = a(\beta,\gamma)$ and the other equations, for the variables value (α,β,γ) (see Appendix A). The problem however is that the prolongations of t_γ must not necessarily agree for different choices of (α,β). Hence this strategy can not be used to define the token function t_γ on the whole set Ω. Nevertheless, we show in

[3] Note that any measurement that is not described by a deterministic response function can be decomposed as an additional local source of randomness followed by a deterministic response function. In this case, the randomness source can be merged with one of the sources present in the network, and all of the following arguments can be applied at the level of the merged sources.

Appendix A, that there exists a subset $\Upsilon \subset \Omega$ such that

$$\Pr(\boldsymbol{\xi} \in \Upsilon) \geq 1 - \mathrm{f}_\varepsilon \geq 1 - 3\varepsilon \quad \text{with}$$
$$1 - \mathrm{f}_\varepsilon = \begin{cases} \frac{1}{2} - \varepsilon + \frac{1-4\varepsilon}{2\sqrt{1-2\varepsilon}} & \varepsilon \leq \frac{1}{4} \\ \frac{1-2\varepsilon}{2} & \varepsilon > \frac{1}{4} \end{cases} \tag{15}$$

on which the token function t_γ can be prolonged consistently. The key idea is to use the fact that the definitions in Eq. (14) must agree for any two points in Ω with the same α or β. This demonstrates the following result holds.

Result 1. (Approximate PTC rigidity) *For any triangle-local model leading to a distribution $P(a,b,c) \in \mathcal{L}^2_\triangle$ that fulfills the approximate parity condition $Pr(a \oplus b \oplus c = 1) \geq 1 - \varepsilon$, there exists a subset Υ of local variable values $\boldsymbol{\xi} = (\alpha, \beta, \gamma)$ with $Pr(\boldsymbol{\xi} \in \Upsilon) \geq 1 - \mathrm{f}_\varepsilon$ (see Eq. 15), on which one can define the token functions $T_\alpha, T_\beta, T_\gamma$ that satisfy the PTC rigidity $a(\beta,\gamma) = T_\beta \oplus T_\gamma \oplus 1$ etc.*

For any triangle-local model, the set Υ gives rise to the subnormalized distribution $P(a,b,c;\boldsymbol{\xi} \in \Upsilon)$ representing the probability of the event "$\xi = (\alpha,\beta,\gamma) \in \Upsilon$ and $a(\beta,\gamma) = a, b(\alpha,\gamma) = b, c(\alpha,\beta) = c$" for the output values a, b, c. These probabilities satisfy $\sum_{a,b,c} P(a,b,c;\boldsymbol{\xi} \in \Upsilon) \geq 1 - \mathrm{f}_\varepsilon$. For simplicity let us assume that this bound is tight (otherwise one can repeat the following discussion with a second parameter $\mathrm{f}' \leq \mathrm{f}_\varepsilon$ and reach the same conclusions). It follows that

$$P(a,b,c;\boldsymbol{\xi} \in \Upsilon) + \mathrm{f}_\varepsilon P'(a,b,c) = P(a,b,c), \tag{16}$$

where $P'(a,b,c)$ is some[4] (non necessarily local) distribution.

Now, let us extend the token functions T_ξ to all the values of the local variables that do not appear in Υ, this can be done by e.g. assigning $t_\xi = 0$ to all such values. With the token functions defined everywhere, we can now define the corresponding extended probability distribution as follows

$$\widetilde{P}(a,b,c) \tag{17}$$
$$= \mathbb{E}\Big(P_{ptc}(a|t_\beta, t_\gamma) P_{ptc}(b|t_\alpha, t_\gamma) P_{ptc}(c|t_\alpha, t_\beta)\Big)$$

[4]Note that we know the values $P'(a,b,c)$ on the outcomes with wrong parity and will exploit this fact later.

where the outputs are obtained with the PTC response functions

$$P_{ptc}(x|t_\xi, t_{\xi'}) = \begin{cases} 1 & x = t_\xi \oplus t_{\xi'} \oplus 1 \\ 0 & x \neq t_\xi \oplus t_{\xi'} \oplus 1 \end{cases}. \tag{18}$$

By construction, this distribution is triangle-local and PTC. Furthermore, we know that $P(a,b,c)$ and $\widetilde{P}(a,b,c)$ agree on the subset $\boldsymbol{\xi} \in \Upsilon$, hence the identity

$$P(a,b,c;\boldsymbol{\xi} \in \Upsilon) + \mathrm{f}_\varepsilon \widetilde{P}'(a,b,c) = \widetilde{P}(a,b,c) \tag{19}$$

holds for some PTC (non-necessarily local) distributions $\widetilde{P}'$. Combining Eqs. (16,19) we conclude that the observed local distribution must satisfy $P(a,b,c) = \widetilde{P}(a,b,c) - \mathrm{f}_\varepsilon \widetilde{P}'(a,b,c) + \mathrm{f}_\varepsilon P'(a,b,c)$, which can be nicely expressed in terms of total variation distance

$$\delta(P, \widetilde{P}) = \frac{1}{2} \sum_{a,b,c} |P(a,b,c) - \widetilde{P}(a,b,c)| \leq \mathrm{f}_\varepsilon. \tag{20}$$

We have thus shown that any local distribution $P(a,b,c)$ satisfying the approximate parity condition of Eq. (6), must be close to the local PTC set

$$\delta(P, \mathcal{L}^2_{\triangle|PTC}) = \min_{\widetilde{P} \in \mathcal{L}^2_{\triangle|PTC}} \delta(P, \widetilde{P}) \leq \mathrm{f}_\varepsilon \tag{21}$$

In turn, the approximate parity condition is equivalent to a bound on the total variation distance between P and the PTC slice of the probability set $\delta(P, \mathcal{P}^2_{PTC}) = \min_{Q \in \mathcal{P}^2_{PTC}} \delta(P,Q) \leq \varepsilon$. Therefore, we have shown the following result.

Result 2. (Relaxation of the local set $\mathcal{L}^2_\triangle$) *Any triangle-local distribution $P(a,b,c) \in \mathcal{L}^2_\triangle$ that is close to the PTC slice of the probability set $\delta(P, \mathcal{P}^2_{PTC}) \leq \varepsilon$ is also close to the local (also triangle-local) PTC set $\delta(P, \mathcal{L}^2_{\triangle|PTC}) \leq \mathrm{f}_\varepsilon$ (see Eq. 15).*

In words, the approximate rigidity allows us to define a cone of distributions $\widehat{\mathcal{L}}^2_\triangle$ that are accessible when going away from the PTC slice, as illustrated in Fig. 4. Loosely speaking – it contains triangle-local distributions at a distance ε to the PTC slice that must also be close enough to the local PTC set such that we can not move more than a distance ε *inside* the slice. Any distribution outside of the cone is guaranteed to be

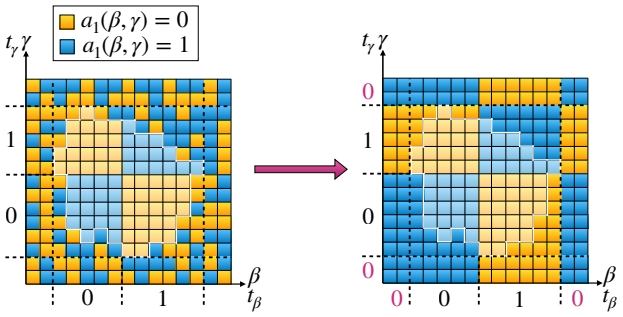


Figure 3: A graphical representation of the construction used in the proof of approximate PTC rigidity. The coordinates along the axes represent different values of the local variables β and γ, here discretized to 15 possible values each. The color of each square gives the value of Alice's outcome $a(\beta,\gamma)=0$ (yellow) and $a(\beta,\gamma)=1$ (blue). On the left, the lighter region depicts the set Υ on which the binary token functions t_β and t_γ are defined (given below the axes) and satisfy $a(\beta,\gamma)=t_\beta\oplus t_\gamma\oplus 1$. On the right, the token functions are extended on the whole set of local variables by setting them to 0 outside of Υ. The values obtained by applying the PTC response function $a(\beta,\gamma)=t_\beta\oplus t_\gamma\oplus 1$ (right) are not guaranteed to match the observed values $a(\beta,\gamma)$ (left) outside of Υ.

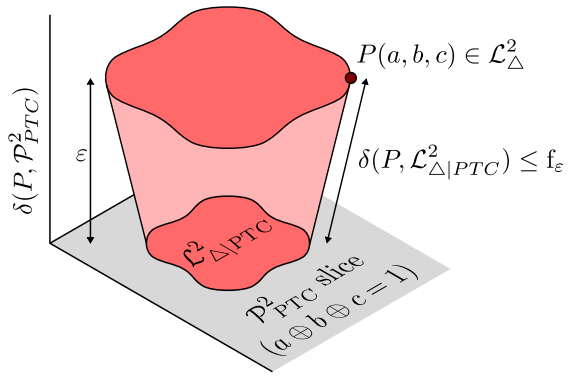


Figure 4: A graphical representation of Result 2. Consider any triangle-local distribution $P(a,b,c)\in\mathcal{L}^2_\triangle$ which is at a distance $\varepsilon=\delta(P,\mathcal{P}^2_{PTC})$ to the PTC slice containing all probability distributions satisfying $\Pr(a\oplus b\oplus c\neq 1)=0$. Result 2 implies that $P(a,b,c)$ must also be "close enough" $\delta(P,\mathcal{L}^2_{\triangle|PTC})\leq \mathrm{f}_\varepsilon$ to the local PTC set $\mathcal{L}^2_{\triangle|PTC}$ of which we have full characterization. Here, δ is the total variation distance, and the function $\mathrm{f}_\varepsilon\leq 3\varepsilon$ is defined in Eq. (15).

nonlocal, i.e. we have established an outer approximation of the local set $\mathcal{L}^2_\triangle\subset\widehat{\mathcal{L}}^2_\triangle$. It is an open question whether all quantum distributions are also contained in this cone.

To illustrate Result 2 we discuss the nonlocality of the noisy W distributions (see Appendix B).

5 Noise-robust nonlocality in $\mathcal{Q}^4_\triangle$

We now consider a particular family of four-outcome quantum distributions on the triangle. Those are obtained from the following quantum models. Each source distributes a two-qubit state

$$|\psi\rangle=\lambda_0|01\rangle+\lambda_1|10\rangle\,, \tag{22}$$

with real and positive $\lambda_0^2+\lambda_1^2=1$, to the neighboring parties. Each party performs the same projective measurement $\{\Pi_{x_1,x_2}=|\phi_{x_1,x_2}\rangle\langle\phi_{x_1,x_2}|\}$ on the two qubits they receive from the neighboring sources. The POVM elements are given by projectors on the states:

$$\begin{aligned}|\phi_{1,0}\rangle&=u|0,1\rangle+v|1,0\rangle\\ |\phi_{1,1}\rangle&=v|0,1\rangle-u|1,0\rangle\\ |\phi_{0,0}\rangle&=w|0,0\rangle+z|1,1\rangle\\ |\phi_{0,1}\rangle&=z|0,0\rangle-w|1,1\rangle\end{aligned} \tag{23}$$

with real positive u,v,w,z satisfying $u^2+v^2=w^2+z^2=1$. Here, the element $|\phi_{x_1,x_2}\rangle$ corresponds to the 2-bit output x_1,x_2 for the party $x\in\{a,b,c\}$. The resulting family of quantum distributions $P(a_1,a_2,b_1,b_2,c_1,c_2)$ is termed "tilted Bell state measurements" (TBSM) distribution, since the local measurements can be viewed as tilted Bell state measurements. Elements of such a distribution are given by

$$\begin{aligned}P(1,i,1,j,1,k)&=(\lambda_0^3u_iu_ju_k+\lambda_1^3v_iv_jv_k)^2\\ P(1,i,0,j,0,k)&=(\lambda_0\lambda_1^2u_iw_jz_k+\lambda_1\lambda_0^2v_iw_kz_j)^2\ \circlearrowleft\end{aligned} \tag{24}$$

where $u_0=u$, $v_0=v$, $w_0=w$, $z_0=z$, $u_1=v$, $v_1=-u$, $w_1=z$ and $z_1=-w$. Here $\circlearrowleft$ means that the equation is valid up to cyclic permutations of the parties.

By setting $w=1-z=0$ one recovers the RGB4 family of distributions [16], while merging the outcomes $(0,0)$ and $(0,1)$ leads to the three-output family of quantum distributions introduced in [20].

As a warm-up, we now present a proof of nonlocality of the distributions in Eq. (24) for a subset of parameter values (λ_0,u,w). It has the advantage of streamlining the proof techniques used in [16] and subsequently [20], while relying on the same basic idea. This will be helpful when we move to the noisy case.

5.1 Nonlocality of the noiseless TBSM distributions

The nonlocality proof proceeds by contradiction, so let us assume that the distribution is local. It is easy to see that when the second output is discarded, the resulting binary distributions

$$P(a_1, b_1, c_1) = \sum_{a_2,b_2,c_2} P(a_1, a_2, b_1, b_2, c_1, c_2) \tag{25}$$

satisfy the parity condition $a_1 \oplus b_1 \oplus c_1 = 1$. By rigidity, it follows that for any underlying local model there exist the binary token values $t_\alpha, t_\beta, t_\gamma$ that explain the distribution

$$P(\boldsymbol{x}_1) = \mathbb{E}\Big(P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t})\Big), \tag{26}$$

where $\mathbb{E}$ is the expected value over $\boldsymbol{t}$ and we introduced the compact notation $\boldsymbol{x}_i = (a_i, b_i, c_i)$, $\boldsymbol{t} = (t_\alpha, t_\beta, t_\gamma)$ and $P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) = P_{ptc}(a_1|t_\beta, t_\gamma)P_{ptc}(b_1|t_\alpha, t_\gamma)P_{ptc}(c_1|t_\alpha, t_\beta)$, with the response functions of Eq. (18). Furthermore, if $E_{a_1}, E_{b_1}, E_{c_1} \neq 0$ the token distribution $p(\boldsymbol{t}) = p_\alpha(t_\alpha)p_\beta(t_\beta)p_\gamma(t_\gamma)$ is known exactly via Eq. (5).

This conclusion is, of course, independent of whether we discard the outcomes $\boldsymbol{x}_2$ or not. Hence, for any local model resulting in the TBSM distribution, we can define the random variable $\boldsymbol{t}$, indicating the behavior of the tokens. Using the chain rule, we have $P(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{t}} p(\boldsymbol{t})P(\boldsymbol{x}_2|\boldsymbol{t})P(\boldsymbol{x}_1|\boldsymbol{t}, \boldsymbol{x}_2)$. From the rigidity, we can infer that $\boldsymbol{x}_1$ is solely a function of $\boldsymbol{t}$, leading to

$$P(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{t}} p(\boldsymbol{t})P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t})P(\boldsymbol{x}_2|\boldsymbol{t}). \tag{27}$$

Here, all the distributions $P(\boldsymbol{x}_2|\boldsymbol{t}) \in \mathcal{L}^2_\triangle$ must be triangle-local. Furthermore, these distributions are not independent, since the output of a party (e.g. a_2) can not be influenced by the source that is not connected to it (that is by α and the token value t_α). This observation implies the following constraints on the marginal distributions $P(a_2|\boldsymbol{t}) = \sum_{b_2,c_2} P(\boldsymbol{x}_2|\boldsymbol{t})$

$$P(a_2|t_\alpha = 0, t_\beta, t_\gamma) = P(a_2|t_\alpha = 1, t_\beta, t_\gamma) \quad \circlearrowleft \tag{28}$$

and similar constraints for all permutations of parties, as indicated by the symbol $\circlearrowleft$. The above discussion can be summarized as follows.

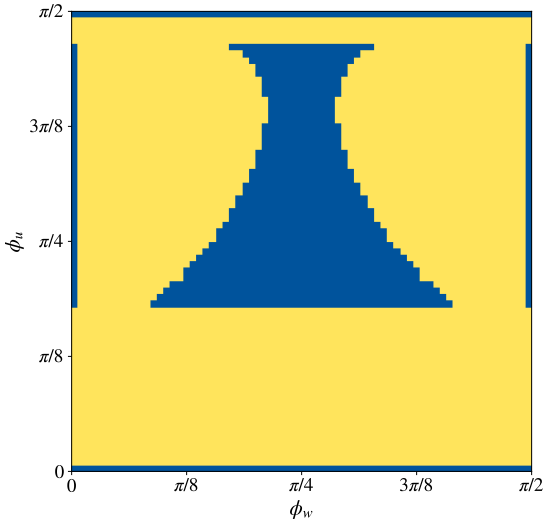


Figure 5: **Nonlocality of noiseless TBSM distributions.** Nonlocal region depicted in yellow. Here distributions are parametrized via angles: $u = \cos(\phi_u)$, $w = \cos(\phi_w)$. Parameter: $\lambda_0^2 = 0.22$.

Result 3. (Nonlocality of TBSM distributions) *For any triangle-local distribution $P(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{L}^4_\triangle$ satisfying $a_1 \oplus b_1 \oplus c_1 = 1$, there exist eight probability distributions $P(\boldsymbol{x}_2|\boldsymbol{t})$ such that*

$$\begin{aligned} &(3.i)\ P(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{t}} p(\boldsymbol{t})P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t})P(\boldsymbol{x}_2|\boldsymbol{t}) \\ &(3.ii)\ P(a_2|t_\alpha = 0, t_\beta, t_\gamma) = P(a_2|t_\alpha = 1, t_\beta, t_\gamma) \quad \circlearrowleft \\ &(3.iii)\ P(\boldsymbol{x}_2|\boldsymbol{t}) \in \mathcal{L}^2_\triangle \qquad \forall \boldsymbol{t} \end{aligned} \tag{29}$$

where P_{PTC} is the parity token counting response function and the token distribution $p(\boldsymbol{t}) = p_\alpha(t_\alpha)p_\beta(t_\beta)p_\gamma(t_\gamma)$ is uniquely determined from $P(\boldsymbol{x}_1)$ via Eq. (5) (except degenerate cases where it is not unique).

If we were able to show that these conditions cannot be satisfied together, we could conclude that the distribution $P(\boldsymbol{x}_1, \boldsymbol{x}_2)$ is nonlocal. Yet, the locality condition $(3.iii)$ here is very challenging to formalize since we do not have a good characterization of the local set $\mathcal{L}^2_\triangle$. A simple solution is to relax it and keep the other two conditions. This relaxation gives rise to a simple linear program (LP) that must be feasible for any triangle-local $P(\boldsymbol{x}_1, \boldsymbol{x}_2)$ with $a_1 \oplus b_1 \oplus c_1 = 1$, that we give in Appendix C.

In Fig. 5, we present the results of the LP for $\lambda_0^2 = 0.22$. One notes that Result 3 can be used to prove the nonlocality of the noisy distribution if the noise does not perturb the parity of the

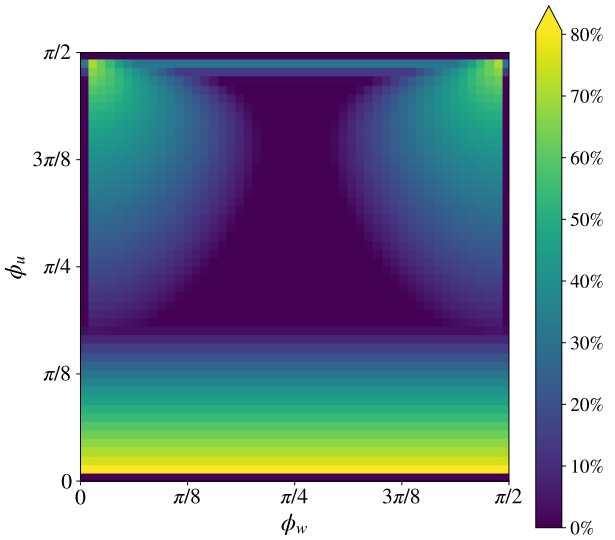


Figure 6: **Dephasing Noise Robustness.** The amount of local dephasing noise d in Eq. (30) below which the TBSM distribution is guaranteed to be nonlocal. Parameters: $u = \cos(\phi_u)$, $w = \cos(\phi_w)$ and $\lambda_0^2 = 0.22$.

$\boldsymbol{x}_1$ outcomes. For example, the local dephasing noise given by the channel $\mathcal{E}_d : |\psi\rangle\langle\psi| \mapsto \varrho$ with

$$\varrho = (1-d)\,|\psi\rangle\langle\psi| + d\left(\lambda_0^2\,|01\rangle\langle 01| + \lambda_1^2\,|10\rangle\langle 10|\right) \tag{30}$$

is of this type. In Appendix C we give the LP which finds the minimal amount of dephasing noise such that the distribution becomes triangle-local. Its results are presented in Fig. 6 for $\lambda_0^2 = 0.22$.

Nevertheless, this still only applies to a slice of the quantum set $\mathcal{Q}_\triangle^4$. In the next Section, we derive an approach that singles out the nonlocality of a full-measure subset of triangle quantum distributions. Interestingly, we observe that the largest robustness to white noise, as well as for total-variation distance, is obtained for the same parameters $\lambda_0^2 = 0.22, \phi_u = 0.42, \phi_w = 0$, with $u = \cos(\phi_u)$ and $w = \cos(\phi_w)$.

5.2 Nonlocality of the noisy TBSM distributions

Our main idea now is pretty simple – adjust the LP-based nonlocality proof of the previous Section to the setting of approximate PTC rigidity discussed in Section 4.

Indeed, let us consider any triangle-local distribution

$$\begin{aligned} &P(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ &= \mathbb{E}(P_A(a_1, a_2|\beta,\gamma)P_B(b_1, b_2|\alpha,\gamma)P_C(\dots)) \end{aligned}$$

such that $\Pr(a_1 \oplus b_1 \oplus c_1 = 1) \geq 1 - \varepsilon$. By approximate PTC rigidity, we know that there exists a local PTC distribution $\widetilde{P}(\boldsymbol{x}_1)$ such that $\delta(P(\boldsymbol{x}_1), \widetilde{P}(\boldsymbol{x}_1)) \leq \mathrm{f}_\varepsilon$. This distribution was constructed by applying the token functions T_ξ to the sources, and replacing the detectors with the parity token counting response functions (see Fig. 3). This construction does not require us to discard the outcomes $\boldsymbol{x}_2$, and we can define a full distribution

$$\begin{aligned} &\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) \\ &= \mathbb{E}(P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t})P_A(a_2|\beta,\gamma)P_B(b_2|\alpha,\gamma)\dots). \end{aligned} \tag{31}$$

The distribution $\widetilde{P}$ is triangle-local, by construction it agrees with the initial distribution on outcomes $\boldsymbol{x}_2$ ($\widetilde{P}(\boldsymbol{x}_2) = P(\boldsymbol{x}_2)$), and is PTC local for the outcomes $\boldsymbol{x}_1$. Furthermore, we know that the two distributions match $P(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\xi} \in \Upsilon) = \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2; \boldsymbol{\xi} \in \Upsilon)$ on the subset Υ of local variable values with $\Pr(\boldsymbol{\xi} \in \Upsilon) \geq 1 - \mathrm{f}_\varepsilon$, which is equivalent to a bound on their total variation distance

$$\delta(P(\boldsymbol{x}_1, \boldsymbol{x}_2), \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq \mathrm{f}_\varepsilon. \tag{32}$$

We have thus reached the following conclusion.

Result 4. (Nonlocality of noisy TBSM distribution) *For any triangle-local distribution $P(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{L}_\triangle^4$ satisfying $Pr(a_1 \oplus b_1 \oplus c_1 = 1) \geq 1 - \varepsilon$, there exists a triangle-local distribution $\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ such that*

$$\begin{aligned} &(4.i)\ \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0 \quad \text{for} \quad a_1 \oplus b_1 \oplus c_1 \neq 1 \\ &(4.ii)\ \delta(P(\boldsymbol{x}_1, \boldsymbol{x}_2), \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq \mathrm{f}_\varepsilon. \\ &(4.iii)\ \sum_{\boldsymbol{x}_1} \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} P(\boldsymbol{x}_1, \boldsymbol{x}_2) \end{aligned} \tag{33}$$

In turn, the distribution $\widetilde{P}$ satisfies the premises of Result 3. We could thus verify the nonlocality of P by showing that no distribution $\widetilde{P}$ compatible with both Result 3 and 4 exists. A complication here is that the relation (5) between the distribution $\widetilde{P}(\boldsymbol{x}_1)$ and the underlying token distribution $p(\boldsymbol{t})$ is nonlinear. It is thus easier to introduce the token distributions $p_\alpha, p_\beta, p_\gamma$ as

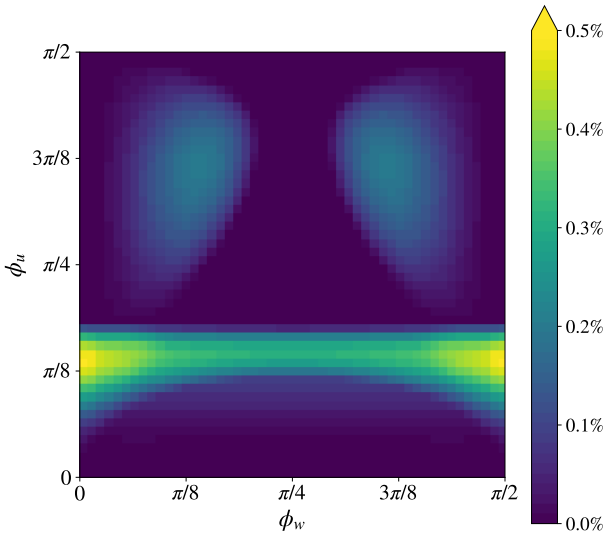


Figure 7: **White Noise Robustness.** Maximal amount of white noise below which we can prove the nonlocality of TBSM distributions. The largest amount of noise that can be tolerated is here $\omega = 0.5\%$; note that this value can be improved slightly to $\omega = 0.544\%$ (see Appendix D.1). Parameters: $u = \cos(\phi_u)$, $w = \cos(\phi_w)$ and $\lambda_0^2 = 0.22$.

additional variables rather than computing them from $\widetilde{P}$. That way we can combine results 3 and 4 into the following corollary.

Corollary of Result 4. *For any triangle-local distribution $P(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{L}^4_{\triangle}$ satisfying $Pr(a_1 \oplus b_1 \oplus c_1 = 1) \geq 1 - \varepsilon$, there exist probability distributions $P(\boldsymbol{x}_2|\boldsymbol{t})$, $p_\xi(t_\xi)$ and $\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ such that $(4.ii), (4.iii), (3.ii), (3.iii)$ and*

$$(*)\ \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{t}} p(\boldsymbol{t}) P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) P(\boldsymbol{x}_2|\boldsymbol{t}) \quad (34)$$

for $p(\boldsymbol{t}) = p_\alpha(t_\alpha) p_\beta(t_\beta) p_\gamma(t_\gamma)$.

As in the ideal case to verify this feasibility problem we start by forgetting the condition $(3.iii)$ requiring the distributions $P(\boldsymbol{x}_2|\boldsymbol{t})$ to be local. Note also that the condition $(4.ii)$ on the total variation distance $\delta(P, \widetilde{P}) \leq \mathrm{f}_\varepsilon$ is equivalent to the existence of two probability distributions $P'(\boldsymbol{x}_1, \boldsymbol{x}_2)$ and $\widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2)$ fulfilling $P + \mathrm{f}_\varepsilon P' = \widetilde{P} + \mathrm{f}_\varepsilon \widetilde{P}'$ (see Lemma 1 in Appendix D). By introducing these probabilities as variables we can rewire the conditions $(4.iii)$ and $(*)$ as

$$(4.iii')\ \sum_{\boldsymbol{x}_1} \widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} P'(\boldsymbol{x}_1, \boldsymbol{x}_2) \quad (35)$$

$$(*')\ P(\boldsymbol{x}_1, \boldsymbol{x}_2) + \mathrm{f}_\varepsilon (P'(\boldsymbol{x}_1, \boldsymbol{x}_2) - \widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2)) = \sum_{\boldsymbol{t}} p(\boldsymbol{t}) P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) P(\boldsymbol{x}_2|\boldsymbol{t}) \quad (36)$$

and get rid of $(4.ii)$ and the variable $\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)$. Finally, it is also possible to directly bound the values of the token probabilities $p_\xi(t_\xi)$ from $(4.ii)$. One easily sees that the total variation distance implies $|\widetilde{E}_{x_1} - E_{x_1}| \leq 2\mathrm{f}_\varepsilon$. In fact, a slightly better bound can be obtained by using the fact that $\widetilde{P}$ is in the PTC slice

$$|\widetilde{E}_{x_1} - E^\Lambda_{x_1}| \leq 2\mathrm{f}_\varepsilon - \varepsilon, \quad (37)$$

where $E^\Lambda_{x_1} = \Pr(x_1 = 0, a_1 \oplus b_1 \oplus c_1 = 1) - \Pr(x_1 = 1, a_1 \oplus b_1 \oplus c_1 = 1)$. Hence, if all the correlators for the observed distribution satisfy $|E^\Lambda_{x_1}| > (2\mathrm{f}_\varepsilon - \varepsilon)$ one can use Eq. (5) to obtain

$$\frac{1}{2}\left(1 + \sqrt{\frac{(|E^\Lambda_{b_1}| - (2\mathrm{f}_\varepsilon - \varepsilon))(|E^\Lambda_{c_1}| - (2\mathrm{f}_\varepsilon - \varepsilon))}{|E^\Lambda_{a_1}| + (2\mathrm{f}_\varepsilon - \varepsilon)}}\right) \leq p_\alpha(0) \leq \frac{1}{2}\left(1 + \sqrt{\frac{(|E^\Lambda_{b_1}| + (2\mathrm{f}_\varepsilon - \varepsilon))(|E^\Lambda_{c_1}| + (2\mathrm{f}_\varepsilon - \varepsilon))}{|E^\Lambda_{a_1}| - (2\mathrm{f}_\varepsilon - \varepsilon)}}\right), \quad (38)$$

up to cyclic permutations.

In Appendix D we show how using this bound the feasibility problem in Eq. (34) can be relaxed to an LP, or a tighter collection of LPs for a grid of values $p_\xi(0)$. To illustrate this approach we now consider the family of noisy TBSM distributions, where local white noise is added to the pure state in Eq. (22) leading to the density operator

$$\rho = (1 - \omega) |\psi\rangle\langle\psi| + \omega \frac{\mathbb{1}}{2} \otimes \frac{\mathbb{1}}{2}. \quad (39)$$

The resulting distribution satisfies $\Pr(a_1 \oplus b_1 \oplus c_1 = 1) = \frac{1}{2}(1 + (1 - \omega)^3)$. In Fig. 7 we plot the maximal amount of white noise ω for which we can prove the nonlocality of the distribution for $\lambda_0^2 = 0.22$ and different measurement parameters. The maximal value that we find is about $\omega = 0.55\%$ at $\phi_u = 0.42$.

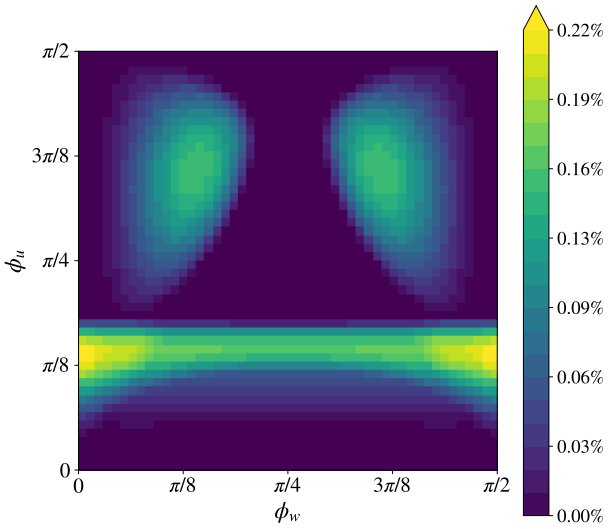


Figure 8: **Total Variation.** The size of the total variation distance ball $\mathcal{B}_\varepsilon(\bar{P})$ in Eq. (40) around the ideal TBSM distributions, for which we prove nonlocality. The maximal value in the plot is $\varepsilon = 0.22\%$, which can be improved to $\varepsilon = 0.24\%$ (see Appendix D.1). Parameters: $u = \cos(\phi_u)$, $w = \cos(\phi_w)$ and $\lambda_0^2 = 0.22$.

In Appendix E we also discuss the nonlocality of the distributions for white noise at the measurements, measurements with a no-click probability, as well as photon-number-resolving detectors with limited efficiency in the case of the single-photon entanglement implementation [36].

5.3 Nonlocality of regions of the quantum set.

So far we argued how the nonlocality of a given distribution $P(\boldsymbol{x}_1, \boldsymbol{x}_2)$ can be demonstrated. However, to better understand the separation between the quantum and the local sets it is more insightful to derive expressions that single out whole regions of the quantum sets that are guaranteed to be nonlocal, similarly to the usual Bell tests. In the case of nonlocality without inputs, considered here, such an expression must be nonlinear. We now derive such "Bell tests" in terms of total variation distance balls around a target quantum distribution.

Given a target quantum distribution $\bar{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{Q}_\triangle^4$ define the total variation distance ball as the following subset of the correlation set

$$\mathcal{B}_\varepsilon(\bar{P}) = \{P \in \mathcal{P}_\triangle^4 | \delta(P, \bar{P}) \leq \varepsilon\}. \tag{40}$$

We will now show that for well-chosen target distributions $\bar{P}$ and small enough ε the whole set $\mathcal{B}_\varepsilon(\bar{P})$ is nonlocal.

Concretely, consider $\bar{P}$ such that $a_1 \oplus b_1 \oplus c_1 = 1$. Any distribution $P \in \mathcal{B}_\varepsilon(\bar{P})$ must satisfy

$$\Pr(a_1 \oplus b_1 \oplus c_1 = 1) \geq 1 - \varepsilon. \tag{41}$$

In addition, if the ball contains some triangle-local distribution P by Result 4 there must exist a local PTC distribution $\widetilde{P}$ with $\delta(\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2), P(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq \mathrm{f}_\varepsilon$ by $(4.ii)$ and $\delta(\widetilde{P}(\boldsymbol{x}_2), P(\boldsymbol{x}_2)) = 0$ by $(4.iii)$. By the triangle inequality, $\widetilde{P}$ must also satisfy

$$\begin{aligned} \delta(\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2), \bar{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)) &\leq \delta(\widetilde{P}, P) + \delta(P, \bar{P}) \\ &\leq \mathrm{f}_\varepsilon + \varepsilon \\ \delta(\widetilde{P}(\boldsymbol{x}_2), \bar{P}(\boldsymbol{x}_2)) &\leq 0 + \delta(P, \bar{P}) \leq \varepsilon, \end{aligned}$$

leading to the following conclusion.

Result 5. (Nonlocality of a ball $B_\varepsilon(\bar{P})$) *If the ball $B_\varepsilon(\bar{P}) \subset \mathcal{P}_\triangle^4$ contains triangle-local distributions, there exists a triangle-local distribution $\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)$ such that*

$$\begin{aligned} &(5.i)\ \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) = 0 \quad \text{for} \quad a_1 \oplus b_1 \oplus c_1 \neq 1 \\ &(5.ii)\ \delta(\bar{P}(\boldsymbol{x}_1, \boldsymbol{x}_2), \widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq \mathrm{f}_\varepsilon + \varepsilon. \\ &(5.iii)\ \delta(\bar{P}(\boldsymbol{x}_2), \widetilde{P}(\boldsymbol{x}_2)) \leq \varepsilon. \end{aligned} \tag{42}$$

Naturally, this can be combined with Result 3 in order to show that no local distribution $\widetilde{P}$ fulfilling $(5.i) - (5.iii)$ can exist. The argument for this is fairly similar to the corollary of Result 4. In Appendix F we present an LP relaxation of this feasibility problem. In Fig. 8 we plot the size of the ball ε around the TBSM distributions with $\lambda_0^2 = 0.22$ which we prove to be nonlocal. The largest value we find is $\varepsilon = 0.24\%$ and it is achieved for an RGB4 distribution ($w = 0$).

6 Discussion

We have derived noise-robust proofs of quantum nonlocality in the triangle network without inputs. Our key ingredient is a result of approximate rigidity for parity token (PTC) local distributions. These methods can be applied to an arbitrary noisy distribution. In particular, we illustrate the relevance of the method by obtaining bounds on the admissible noise for a general family of quantum distributions that we introduced, generalizing the RGB4 distributions [16]. We characterized noise-robustness in terms of limited visibility for the source states, considering

dephasing and white noise. Alternatively, we can also characterize distributions in the vicinity of a target distribution, via a bound on the total-variation distance.

Our work opens different perspectives. First, it would be interesting to see if our methods can be applied to larger networks and distributions satisfying PTC. This may lead to proofs of network nonlocality that are more robust to noise compared to the present results.

Another direction is to see how to strengthen the noise robustness for the nonlocality proof in our Result 4. A promising option is to revisit Result 3, and strengthen the last condition (3.iii). We used the standard implementation of the inflation technique, but stronger versions could be obtained by exploiting nonlinear constraints. While we could prove robustness to white noise up to 0.55%, there is still potential for significant improvements; indeed numerical methods [21] suggest robustness of up to $\sim 10\%$. Any progress in this direction would be relevant, in particular from the perspective of experimental demonstrations. Concerning dephasing noise, we demonstrated strong robustness, up to $\sim 80\%$. We conjecture that there exist quantum distributions that can tolerate any amount of dephasing noise arbitrarily close to one.

It would also be interesting to see if the property of approximate PTC rigidity can be extended from local models to quantum ones. If not, one can hope to find a quantum distribution violating the approximate PTC rigidity, indicating the existence of quantum nonlocal correlations within the triangle network with binary outputs and no inputs.

Code availability.— Our implementation of the method for the triangle network can be found at https://github.com/SadraBoreiri/Robust-Network-Nonlocality

Acknowledgements.— We thank Wolfgang Dür for discussions at the early stage of this project, and Victor Gitton for useful comments on the first version of the paper. We acknowledge financial support from the Swiss National Science Foundation (project 2000021 and NCCR SwissMAP) and by the Swiss Secretariat for Education, Research and Innovation (SERI) under contract number UeM019-3.

References

[1] Armin Tavakoli, Alejandro Pozas-Kerstjens, Ming-Xing Luo, and Marc-Olivier Renou. "Bell nonlocality in networks". Reports on Progress in Physics **85**, 056001 (2022).

[2] C. Branciard, N. Gisin, and S. Pironio. "Characterizing the nonlocal correlations created via entanglement swapping". Phys. Rev. Lett. **104**, 170401 (2010).

[3] Cyril Branciard, Denis Rosset, Nicolas Gisin, and Stefano Pironio. "Bilocal versus nonbilocal correlations in entanglement-swapping experiments". Physical Review A**85** (2012).

[4] Tobias Fritz. "Beyond Bell's theorem: correlation scenarios". New Journal of Physics **14**, 103001 (2012).

[5] Denis Rosset, Cyril Branciard, Tomer Jack Barnea, Gilles Pütz, Nicolas Brunner, and Nicolas Gisin. "Nonlinear bell inequalities tailored for quantum networks". Phys. Rev. Lett. **116**, 010403 (2016).

[6] Elie Wolfe, Robert W Spekkens, and Tobias Fritz. "The inflation technique for causal inference with latent variables". Journal of Causal Inference **7**, 20170020 (2019).

[7] Marc-Olivier Renou, Yuyi Wang, Sadra Boreiri, Salman Beigi, Nicolas Gisin, and Nicolas Brunner. "Limits on correlations in networks for quantum and no-signaling resources". Phys. Rev. Lett. **123**, 070403 (2019).

[8] Johan Åberg, Ranieri Nery, Cristhiano Duarte, and Rafael Chaves. "Semidefinite tests for quantum network topologies". Phys. Rev. Lett. **125**, 110505 (2020).

[9] Elie Wolfe, Alejandro Pozas-Kerstjens, Matan Grinberg, Denis Rosset, Antonio Acín, and Miguel Navascues. "Quantum inflation: A general approach to quantum causal compatibility". Physical Review X**11** (2021).

[10] Patricia Contreras-Tejada, Carlos Palazuelos, and Julio I. de Vicente. "Genuine multipartite nonlocality is intrinsic to quantum networks" (2020). arXiv:2004.01722.

[11] Nicolas Gisin, Jean-Daniel Bancal, Yu Cai, Patrick Remy, Armin Tavakoli, Emmanuel Zambrini Cruzeiro, Sandu Popescu, and Nicolas Brunner. "Constraints on nonlocality in networks from no-signaling and independence". Nature Communications **11**, 1–6 (2020).

[12] Ya-Li Mao, Hu Chen, Bixiang Guo, Shiting Liu, Zheng-Da Li, Ming-Xing Luo, and Jingyun Fan. "Certifying network topologies and nonlocalities of triangle quantum networks". Phys. Rev. Lett. **132**, 240801 (2024).

[13] Ming-Xing Luo. "Fully device-independent model on quantum networks". Phys. Rev. Res. **4**, 013203 (2022).

[14] Emanuele Polino, Davide Poderini, Giovanni Rodari, Iris Agresti, Alessia Suprano, Gonzalo Carvacho, Elie Wolfe, Askery Canabarro, George Moreno, Giorgio Milani, Robert W. Spekkens, Rafael Chaves, and Fabio Sciarrino. "Experimental nonclassicality in a causal network without assuming freedom of choice". Nature Communications**14** (2023).

[15] Nicolas Gisin. "Entanglement 25 years after quantum teleportation: testing joint measurements in quantum networks". Entropy **21**, 325 (2019).

[16] Marc-Olivier Renou, Elisa Bäumer, Sadra Boreiri, Nicolas Brunner, Nicolas Gisin, and Salman Beigi. "Genuine quantum nonlocality in the triangle network". Physical Review Letters**123** (2019).

[17] Ivan Šupić, Jean-Daniel Bancal, Yu Cai, and Nicolas Brunner. "Genuine network quantum nonlocality and self-testing". Phys. Rev. A **105**, 022206 (2022).

[18] Pavel Sekatski, Sadra Boreiri, and Nicolas Brunner. "Partial self-testing and randomness certification in the triangle network". Phys. Rev. Lett. **131**, 100201 (2023).

[19] Marc-Olivier Renou and Salman Beigi. "Network nonlocality via rigidity of token counting and color matching". Physical Review A **105**, 022408 (2022).

[20] Sadra Boreiri, Antoine Girardin, Bora Ulu, Patryk Lipka-Bartosik, Nicolas Brunner, and Pavel Sekatski. "Towards a minimal example of quantum nonlocality without inputs". Phys. Rev. A **107**, 062413 (2023).

[21] Tamás Kriváchy, Yu Cai, Daniel Cavalcanti, Arash Tavakoli, Nicolas Gisin, and Nicolas Brunner. "A neural network oracle for quantum nonlocality problems in networks". npj Quantum Information **6**, 1–7 (2020).

[22] J. S. Bell. "On the Einstein Podolsky Rosen paradox". Physics Physique Fizika **1**, 195–200 (1964).

[23] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. "Proposed experiment to test local hidden-variable theories". Physical review letters **23**, 880 (1969).

[24] Alain Aspect, Philippe Grangier, and Gérard Roger. "Experimental tests of realistic local theories via bell's theorem". Phys. Rev. Lett. **47**, 460–463 (1981).

[25] Alain Aspect, Jean Dalibard, and Gérard Roger. "Experimental test of bell's inequalities using time-varying analyzers". Phys. Rev. Lett. **49**, 1804–1807 (1982).

[26] B. Hensen, N. Kalb, M. S. Blok, A. E. Dréau, A. Reiserer, R. F. L. Vermeulen, R. N. Schouten, M. Markham, D. J. Twitchen, K. Goodenough, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson. "Loophole-free bell test using electron spins in diamond: second experiment and additional analysis". Scientific Reports**6** (2016).

[27] Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. "Strong loophole-free test of local realism". Physical review letters **115**, 250402 (2015).

[28] Marissa Giustina, Marijn A. M. Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, Waldimar Amaya, Valerio Pruneri, Morgan W. Mitchell, Jörn Beyer, Thomas Gerrits, Adriana E. Lita, Lynden K. Shalm, Sae Woo Nam, Thomas Scheidl, Rupert Ursin, Bernhard Wittmann,

and Anton Zeilinger. “Significant-loophole-free test of bell’s theorem with entangled photons”. Phys. Rev. Lett. **115**, 250401 (2015).

[29] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. “Bell nonlocality”. Rev. Mod. Phys. **86**, 419–478 (2014).

[30] Denis Rosset, Nicolas Gisin, and Elie Wolfe. “Universal bound on the cardinality of local hidden variables in networks”. Quantum Information & Computation **18**, 910–926 (2018).

[31] Rafael Chaves, Christian Majenz, and David Gross. “Information–theoretic implications of quantum causal structures”. Nature Communications**6** (2015).

[32] Thomas C. Fraser and Elie Wolfe. “Causal compatibility inequalities admitting quantum violations in the triangle structure”. Phys. Rev. A **98**, 022113 (2018).

[33] Mirjam Weilenmann and Roger Colbeck. “Non-shannon inequalities in the entropy vector approach to causal structures”. Quantum **2**, 57 (2018).

[34] Ivan Šupić, Jean-Daniel Bancal, and Nicolas Brunner. “Quantum nonlocality in networks can be demonstrated with an arbitrarily small level of independence between the sources”. Phys. Rev. Lett. **125**, 240403 (2020).

[35] Rafael Chaves, George Moreno, Emanuele Polino, Davide Poderini, Iris Agresti, Alessia Suprano, Mariana R. Barros, Gonzalo Carvacho, Elie Wolfe, Askery Canabarro, Robert W. Spekkens, and Fabio Sciarrino. “Causal networks and freedom of choice in bell’s theorem”. PRX Quantum **2**, 040323 (2021).

[36] Paolo Abiuso, Tamás Kriváchy, Emanuel-Cristian Boghiu, Marc-Olivier Renou, Alejandro Pozas-Kerstjens, and Antonio Acín. “Single-photon nonlocality in quantum networks”. Phys. Rev. Research **4**, L012041 (2022).

[37] Alejandro Pozas-Kerstjens, Nicolas Gisin, and Marc-Olivier Renou. “Proofs of network quantum nonlocality in continuous families of distributions”. Phys. Rev. Lett. **130**, 090201 (2023).

[38] J. A. Nelder and R. Mead. “A Simplex Method for Function Minimization”. The Computer Journal **7**, 308–313 (1965).

[39] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. Nature Methods **17**, 261–272 (2020).

[40] MOSEK ApS. “Mosek optimizer api for python 10.1.16”. (2019). url: `https://docs.mosek.com/latest/pythonapi/index.html`.

A Approximate PTC rigidity

In a triangle-local model, consider variables α, β, γ sampled from the sets A, B, Γ respectively, according to some probability distributions. Without loss of generality, we can assume that the model includes deterministic response functions $a(\beta,\gamma), b(\alpha,\gamma), c(\alpha,\beta) \in \{0,1\}$[5].

Approximate PTC rigidity states that, for any triangle-local model leading to a binary output distribution $P(a,b,c)$ satisfying the approximate parity condition $\Pr(a\oplus b\oplus c=1) \geq 1-\varepsilon$, there exists a subset $\Upsilon \subset A\times B\times \Gamma$ of local variables with high probability $Pr(\boldsymbol{\xi} \in \Upsilon) \geq 1-O(\varepsilon)$, within which it is feasible to define the token functions $T_\alpha, T_\beta, T_\gamma$ satisfying the PTC rigidity $a(\beta,\gamma) = T_\beta \oplus T_\gamma \oplus 1$ etc. We prove this here, in two main steps. In the first step, we characterize the subset Υ and show the consistent assignment of the token functions inside this subset. In the second step, we prove that this subset happens with high probability, specifically $Pr(\Upsilon) \geq 1-3\varepsilon$. Moreover, in Section A.1 we prove a slightly more involved and tighter lower bound, showing that $Pr(\Upsilon) \geq 1-\mathrm{f}_\varepsilon = 1-\frac{5}{2}\varepsilon+O(\varepsilon^2)$.

Step 1: Defining a subset Υ of local variables with consistent assignment of token functions

Here our objective is to specify $\Upsilon \subset A\times B\times\Gamma$ a subset of local variables on which we can consistently assign the token functions $T_\alpha, T_\beta, T_\gamma$ satisfying the PTC condition

To do so, first let Λ be the subset of $A\times B\times\Gamma$ in which the outputs satisfy the PTC condition, i.e.

$$a(\beta,\gamma)\oplus b(\alpha,\gamma)\oplus c(\alpha,\beta) = 1 \quad \forall\, \boldsymbol{\xi} = (\alpha,\beta,\gamma) \in \Lambda \tag{43}$$

$$\Pr(\Lambda) = \Pr(\boldsymbol{\xi}\in\Lambda) = 1-\varepsilon. \tag{44}$$

For each value γ define the set

$$G_\gamma = \{(\alpha,\beta)|(\alpha,\beta,\gamma)\in\Lambda\} \subset A\times B \tag{45}$$

as the slice of Λ for a fixed γ. The probabilities of these sets $\Pr(G_\gamma) = \Pr((\alpha,\beta)\in G_\gamma)$ satisfy the following equality

$$\mathbb{E}(\Pr(G_\gamma)) = \Pr((\alpha,\beta,\gamma)\in\Lambda) = \Pr(\Lambda) = 1-\varepsilon, \tag{46}$$

where the expected value on the left-hand side is taken with respect to γ. Let us now identify the value γ_* for which $\Pr(G_\gamma)$ is maximal, i.e.

$$\gamma_* = \mathrm{argmax}_\gamma \Pr(G_\gamma) \tag{47}$$

$$1-\varepsilon_* = \max_\gamma \Pr(G_\gamma). \tag{48}$$

By construction we have $\Pr(G_\gamma) \leq 1-\varepsilon_*$ for all γ, and $\varepsilon_* \leq \varepsilon$ since the maximum is larger than the average. To shorten the notation we dub $G_* = G_{\gamma_*}$.

On the set G_* define the token functions

$$T_\alpha := b(\alpha,\gamma_*) \qquad T_\beta := a(\beta,\gamma_*) \tag{49}$$

as well as the value $T_{\gamma_*} = 1$. It is easy to see that on the set $G_*\times\{\gamma_*\} \subset \Lambda$ these functions satisfy the PTC rigidity conditions given by

$$T_\beta \oplus T_{\gamma_*} \oplus 1 = a(\beta,\gamma_*) \tag{50}$$

$$T_\alpha \oplus T_{\gamma_*} \oplus 1 = b(\alpha,\gamma_*) \tag{51}$$

$$T_\beta \oplus T_\alpha \oplus 1 = a(\beta,\gamma_*)\oplus b(\alpha,\gamma_*)\oplus 1 = c(\alpha,\beta), \tag{52}$$

[5]For any indeterministic local response function at the parties we can simply move the indeterminacy to the sources α, β, γ.

where we used $a(\beta,\gamma_*)\oplus b(\alpha,\gamma_*)\oplus c(\alpha,\beta)=1$ in the last line.

Now consider a different value γ and the corresponding set G_γ. Furthermore, define the set

$$S_\gamma = G_\gamma \cap G_* = \{(\alpha,\beta)|(\alpha,\beta)\in G_*, (\alpha,\beta,\gamma)\in\Lambda\}. \tag{53}$$

It is a subset of both G_* and G_γ, hence each $(\alpha,\beta)\in S_\gamma$, both (α,β,γ) and (α,β,γ_*) are inside Λ. We denote

$$1-\delta_\gamma = \Pr(S_\gamma), \tag{54}$$

which must satisfy $\delta_\gamma \geq \varepsilon_*$ because $S_\gamma \subset G_*$.

Remaining in the slice γ, we can define the following binary functions

$$t^z_\alpha := b(\alpha,\gamma)\oplus z\oplus 1 \tag{55}$$
$$t^z_\beta := a(\beta,\gamma)\oplus z\oplus 1 \tag{56}$$
$$t^z_\gamma := z. \tag{57}$$

For both choices $z=0$ and $z=1$, these functions satisfy the PTC conditions on the set $S_\gamma\times\{\gamma\}$

$$t^z_\alpha\oplus t^z_\gamma\oplus 1 = b(\alpha,\gamma)\oplus z\oplus 1\oplus z\oplus 1 = b(\alpha,\gamma) \tag{58}$$
$$t^z_\beta\oplus t^z_\gamma\oplus 1 = a(\beta,\gamma)\oplus z\oplus 1\oplus z\oplus 1 = a(\beta,\gamma) \tag{59}$$
$$t^\gamma_\alpha\oplus t^\gamma_\beta\oplus 1 = b(\alpha,\gamma)\oplus a(\beta,\gamma)\oplus 1 = c(\alpha,\beta), \tag{60}$$

since $(\alpha,\beta,\gamma)\in\Lambda$. However, depending on the value of z, the functions t^z_α and t^z_β might not agree with the token functions T_α and T_β defined previously. More precisely, let us now define the

$$A^z_\gamma = \{\alpha|T_\alpha\oplus t^z_\alpha = 0\} \tag{61}$$
$$B^z_\gamma = \{\alpha|T_\beta\oplus t^z_\beta = 0\}, \tag{62}$$

such that A^0_γ with A^1_γ and B^0_γ with B^1_γ are disjoint, and $S_\gamma\subset(A^0_\gamma\cup A^1_\gamma)\times(B^0_\gamma\cup B^1_\gamma)$. In fact for any point $(\alpha,\beta)\in S_\gamma$ it is true that

$$T_\alpha\oplus t^z_\alpha\oplus T_\beta\oplus t^z_\beta = (T_\alpha\oplus T_\beta)\oplus(t^z_\alpha\oplus t^z_\beta) = (c(\alpha,\beta)\oplus 1)\oplus(c(\alpha,\beta)\oplus 1) = 0, \tag{63}$$

because it is both in G_* and G_γ. Hence, any such point is either in $A^0_\gamma\times B^0_\gamma$ or $A^1_\gamma\times B^1_\gamma$, and we conclude that

$$S_\gamma\subset(A^0_\gamma\times B^0_\gamma)\cup(A^1_\gamma\times B^1_\gamma). \tag{64}$$

This allows us to split $S_\gamma = S^0_\gamma\cup S^1_\gamma$ in two disjoint sets with $S^z_\gamma\subset A^z_\gamma\times B^z_\gamma$. Note that for the choice $t_\gamma = z$ the token functions T_α and T_γ are PTC on $S^z_\gamma\times\{\gamma\}$. Our goal is of course to pick the value of z for which $\Pr(S^z_\gamma)$ is maximized. Let us define this probability

$$\hat{P}_\gamma = \max_z \Pr(S^z_\gamma) \leq \Pr(S_\gamma) = 1-\delta_\gamma. \tag{65}$$

We will now lower-bound this probability. First, note that there is a trivial bound given by $\hat{P}_\gamma\geq\frac{1-\delta_\gamma}{2}$. Furthermore, we have

$$\hat{P}_\gamma = \Pr(S_\gamma) - \min_z \Pr(S^z_\gamma) \geq 1-\delta_\gamma - \min_z \Pr(A^z_\gamma\times B^z_\gamma). \tag{66}$$

To this note that for $z' = \mathrm{argmin}_z\Pr(S^z_\gamma)$ we have $\Pr(S^{z'}_\gamma)\leq\Pr(A^{z'}_\gamma\times B^{z'}_\gamma)$, and $\Pr(S^{z'}_\gamma)\leq\Pr(S^{z'\oplus 1}_\gamma)\leq\Pr(A^{z'\oplus 1}_\gamma\times B^{z'\oplus 1}_\gamma)$ (since $S^z_\gamma\subset A^z_\gamma\times B^z_\gamma$). In addition, we know that

$$1-\delta_\gamma = P_\gamma = \Pr(S^0_\gamma)+\Pr(S^1_\gamma)\leq\Pr(A^0_\gamma\times B^0_\gamma)+\Pr(A^1_\gamma\times B^1_\gamma). \tag{67}$$

Denoting $p_z = \Pr(A^z_\gamma)$ and $q_z = \Pr(A^z_\gamma)$ with $p_0+p_1, q_0+q_1 \leq 1$, we want to upper bound $\min_z \Pr(A^z_\gamma \times B^z_\gamma)$ subject to the constraint in Eq. (67). This is equivalent to solving the following min-max problem

$$\begin{aligned} \max_{p_0,p_1,q_0,q_1} \quad & \min\{p_0q_0, p_1q_1\} \\ \text{such that} \quad & p_0+p_1, q_0+q_1 \leq 1 \\ & p_0, p_1, q_0, q_1 \geq 0 \\ & p_0q_0 + p_1q_1 \geq 1-\delta_\gamma. \end{aligned} \tag{68}$$

In Section A.2 we solve this optimization, proving the following bound

$$\min_z \Pr(A^z_\gamma \times B^z_\gamma) \leq \left(\frac{1-\sqrt{1-2\delta_\gamma}}{2}\right)^2 \quad \Longrightarrow \quad \hat{P}_\gamma \geq 1-\delta_\gamma - \left(\frac{1-\sqrt{1-2\delta_\gamma}}{2}\right)^2. \tag{69}$$

Recalling that $\hat{P}_\gamma$ must also satisfy $\hat{P}_\gamma \geq \frac{1-\delta_\gamma}{2}$ we obtain the desired lower bound

$$\max_z \Pr(S^z_\gamma) \geq f(\delta_\gamma) := \max\left\{1-\delta_\gamma - \left(\frac{1-\sqrt{1-2\delta_\gamma}}{2}\right)^2, \frac{1-\delta_\gamma}{2}\right\} = \begin{cases} 1-\delta_\gamma - \left(\frac{1-\sqrt{1-2\delta_\gamma}}{2}\right)^2 & \delta_\gamma \leq \frac{1}{2} \\ \frac{1-\delta_\gamma}{2} & \delta_\gamma > \frac{1}{2} \end{cases} \tag{70}$$

Let us now define the set $\hat{S}_\gamma = S^{\hat{z}(\gamma)}_\gamma$ with $\hat{z}(\gamma) = \text{argmax}_z \Pr(S^z_\gamma)$ to be the branch for which the probability is maximal, and prolong the token function

$$T_\gamma = \hat{z}(\gamma) \tag{71}$$

on the whole set (if $\hat{S}_\gamma$ is empty one can assign any value to T_γ). By taking the union of all the γ-slices we can now define the set

$$\Upsilon = \bigcup_\gamma \hat{S}_\gamma \times \{\gamma\} = \{(\alpha,\beta,\gamma) | (\alpha,\beta) \in \hat{S}_\gamma\} \tag{72}$$

on which we have shown a consistent assignment of the token functions $T_\alpha, T_\beta, T_\gamma$ fulfilling the PTC condition (for $\gamma = \gamma_*$ we, of course, have $\hat{S}_{\gamma^*} = G_*$).

Step 2: proving $\Pr(\Upsilon) \geq 1-3\varepsilon$.

Finally, we have to lower bound the probability $\Pr(\Upsilon)$. To do so we first consider the set

$$\Omega = \bigcup_\gamma S_\gamma \times \{\gamma\} = \{(\alpha,\beta,\gamma) | (\alpha,\beta) \in S_\gamma\} = \{(\alpha,\beta,\gamma) | (\alpha,\beta) \in G_*, (\alpha,\beta,\gamma) \in \Lambda\} \tag{73}$$

for which the probability is easy to bound

$$\Pr(\Omega) = 1 - \Pr(\boldsymbol{\xi} \notin \Lambda \text{ or } (\alpha,\beta) \notin G_*) \geq 1 - \Pr(\boldsymbol{\xi} \notin \Lambda) - \Pr((\alpha,\beta) \notin G_*) \geq 1-\varepsilon-\varepsilon_*. \tag{74}$$

The probabilities $\Pr(\Omega)$ and $\Pr(\Upsilon)$ can also be expressed as expected values over γ

$$\Pr(\Omega) = \mathbb{E}(\Pr(S_\gamma)) = \mathbb{E}(1-\delta_\gamma) \geq 1-\varepsilon-\varepsilon_* \tag{75}$$

$$\Pr(\Upsilon) = \mathbb{E}(\Pr(\hat{S}_\gamma)) \geq \mathbb{E}(f(\delta_\gamma)). \tag{76}$$

The last step is to minimize the rhs of the last inequality under the constraint of Eq. (75), that is

$$\min \quad \mathbb{E}(f(\delta_\gamma)) \tag{77}$$

$$\text{such that} \quad \mathbb{E}(\delta_\gamma) \leq \varepsilon + \varepsilon_* \tag{78}$$

$$\delta_\gamma \geq \varepsilon_*, \tag{79}$$

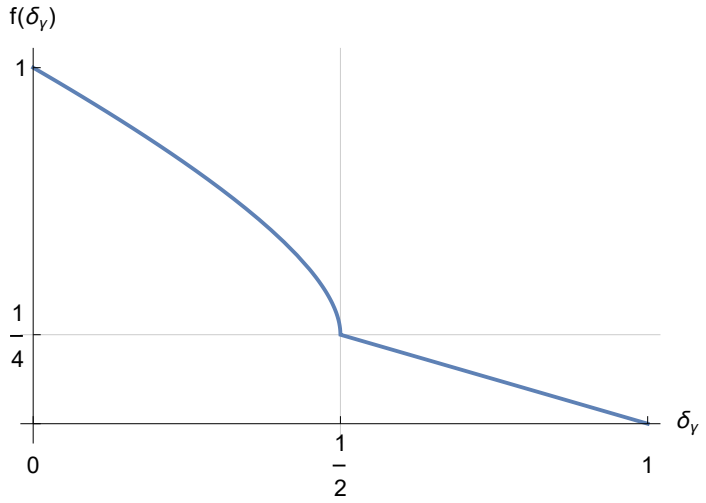


Figure 9: The function $f(\delta_\gamma)$ evaluated in the interval $[0,1]$. The concavity of the function can be observed in the $[0,1/2]$ interval.

where the last line comes from $\Pr(S_\gamma) \leq \Pr(G_\gamma) \leq \Pr(G_*) = 1-\varepsilon_*$. The function $f(\delta_\gamma)$ is depicted in Fig. 9. We can relax this minimization to a trivial one if we find a convex decreasing lower bound on the goal function $f(\delta_\gamma) \geq f_{\text{con}}(\delta_\gamma)$ on the interval $\delta_\gamma \in [\varepsilon_*, 1]$, since such a function must satisfy

$$\mathbb{E}(f(\delta_\gamma)) \geq \mathbb{E}(f_{\text{con}}(\delta_\gamma)) \geq f_{\text{con}}(\mathbb{E}(\delta_\gamma)) \geq f_{\text{con}}(\varepsilon+\varepsilon_*). \tag{80}$$

Note that the function $f(\delta_\gamma)$ is concave on the interval $[0,1/2]$, which follows from the negativity of its second derivative

$$f''(\delta_\gamma) = -\frac{1}{2(1-2\delta_\gamma)^{3/2}} < 0. \tag{81}$$

It follows that a convex lower bound on the whole interval $\delta_\gamma \in [0,1]$ is provided by a linear function

$$f(\delta_\gamma) \geq f_{\text{con}}(\delta_\gamma) = 1 - \frac{3}{2}\delta_\gamma, \tag{82}$$

which results in the following bound

$$\Pr(\Upsilon) \geq \mathbb{E}(f(\delta_\gamma)) \geq 1 - \frac{3}{2}(\varepsilon+\varepsilon_*) \geq 1 - 3\varepsilon, \tag{83}$$

by $\varepsilon_* \leq \varepsilon$. In the next Section, we derive a tighter upper bound which uses the fact that $\delta_\gamma \geq \varepsilon_*$.

A.1 Tighter bound on $\Pr(\Upsilon)$

A tight convex lower bound on $f(\delta_\gamma)$ depicted in Fig. 9 on the interval $[\varepsilon_*, 1]$ is given by the piecewise linear function which connects the points $\{\varepsilon_*, f_* = f(\varepsilon_*)\}$ with $\{1/2, 1/4\}$ for $\delta_\gamma \in [\varepsilon_*, 1/2]$ and the points $\{1/2, 1/4\}$ with $\{1, 0\}$ for $\delta_\gamma \in [1/2, 1]$. Formally, this lower bound is given by

$$f^{\varepsilon_*}_{\text{con}}(\delta_\gamma) = \begin{cases} \frac{2f_*-\varepsilon_*}{2-4\varepsilon_*} - \frac{4f_*-1}{2-4\varepsilon_*}\delta_\gamma & \delta_\gamma \in [\varepsilon_*, 1/2] \\ \frac{1-\delta_\gamma}{2} & \delta_\gamma \in (1/2, 1] \end{cases} \tag{84}$$

This implies a tight lower bound

$$\begin{aligned} \min \quad & \mathbb{E}(f(\delta_\gamma)) \geq f^{\varepsilon_*}_{\text{con}}(\varepsilon+\varepsilon_*) && (85) \\ \text{such that} \quad & \mathbb{E}(\delta_\gamma) \leq \varepsilon+\varepsilon_* && (86) \\ & \delta_\gamma \geq \varepsilon_*, && (87) \end{aligned}$$

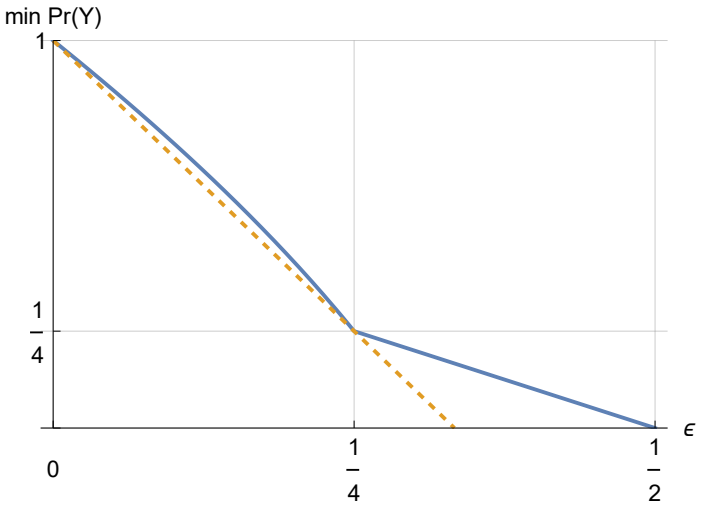


Figure 10: The lower bounds f_ε (full line) and $1-3\varepsilon$ (dashed line) on the probability $\Pr(\Upsilon)$ of the PTC set as functions of the probability $\varepsilon = \Pr(a \oplus b \oplus c \neq 1)$ to observe the wrong parity of the outcomes.

and $\Pr(\Upsilon) \geq f^{\varepsilon_*}_{\mathrm{con}}(\varepsilon+\varepsilon_*)$ for any scenario with $\Pr(G_*) = 1-\varepsilon_*$. But since ε_* is not known, we need to take the worst-case lower bound

$$\Pr(\Upsilon) \geq 1-\mathrm{f}_\varepsilon := \min_{\varepsilon_* \in [0,\varepsilon]} f^{\varepsilon_*}_{\mathrm{con}}(\varepsilon+\varepsilon_*) \tag{88}$$

Finally, let us show that

$$1-\mathrm{f}_\varepsilon = f^{\varepsilon}_{\mathrm{con}}(2\varepsilon) = \begin{cases} \frac{1}{2} - \varepsilon + \frac{1-4\varepsilon}{2\sqrt{1-2\varepsilon}} & \varepsilon \leq \frac{1}{4} \\ \frac{1-2\varepsilon}{2} & \varepsilon > \frac{1}{4} \end{cases} \tag{89}$$

This follows from the fact that the function $f^{\varepsilon_*}_{\mathrm{con}}(\varepsilon+\varepsilon_*)$ is decreasing in ε_* hence the minimum is reached by setting $\varepsilon_* = \varepsilon$. Indeed, its derivative

$$\frac{\partial}{\partial \varepsilon_*} f^{\varepsilon_*}_{\mathrm{con}}(\varepsilon+\varepsilon_*) = \begin{cases} -\frac{\left(1+\sqrt{1-2\varepsilon_*}\right)(1-2\varepsilon_*)-2\varepsilon}{2(1-2\varepsilon_*)^{3/2}} & \varepsilon+\varepsilon_* \leq \frac{1}{2} \\ -\frac{1}{2} & \varepsilon+\varepsilon_* > \frac{1}{2} \end{cases} \tag{90}$$

is negative. This concludes the proof of the lower bound

$$\Pr(\Upsilon) \geq 1-\mathrm{f}_\varepsilon \geq 1-3\varepsilon, \tag{91}$$

illustrated in Fig. 10. As a final remark, consider the first order expansion $1-\mathrm{f}_\varepsilon = 1-\frac{5}{2}\varepsilon + O(\varepsilon^2)$ to see that for very small ε the two lower bounds are quite different.

A.2 Solution of the optimization in Eq. (68)

Here we solve the optimization problem

$$\begin{aligned} \max_{p_0,p_1,q_0,q_1} \quad & \min\{p_0q_0, p_1q_1\} \\ \text{such that} \quad & p_0+p_1, q_0+q_1 \leq 1 \\ & p_0,p_1,q_0,q_1 \geq 0 \\ & p_0q_0 + p_1q_1 \geq 1-\delta_\gamma. \end{aligned} \tag{92}$$

Without loss of generality consider the case $p_0q_0 \leq p_1q_1$, and rewrite the problem as

$$\begin{aligned} \max_{p_0,p_1,q_0,q_1} \quad & p_0q_0 \\ \text{such that} \quad & p_0q_0 \leq p_1q_1 \\ & p_0q_0 \geq 1-\delta_\gamma - p_1q_1 \\ & p_0 \leq 1-p_1 \\ & q_0 \leq 1-q_1 \\ & p_0,p_1,q_0,q_1 \geq 0. \end{aligned} \tag{93}$$

The variables p_1 and q_1 now do not appear in the goal function. Furthermore, the constraints on p_0 and q_0 are easier to satisfy with larger values of p_1, q_1 and p_1q_1. Hence the maximum is attained when these auxiliary variables take the maximal values $p_1 = 1-p_0$ and $q_1 = 1-p_0$. This allows us to rewrite the maximization in terms of two variables $p = p_0$ and $q = q_0$ as follows

$$\begin{aligned} \max_{p,q\in[0,1]} \quad & p\,q \\ \text{such that} \quad & p\,q \leq (1-p)(1-q) \\ & p\,q \geq 1-\delta_\gamma-(1-p)(1-q) \end{aligned} \quad \text{which can be rewritten as} \quad \begin{aligned} \max_{p,q\in[0,1]} \quad & p\,q \\ \text{such that} \quad & p+q \leq 1 \\ & 2\,p\,q \geq p+q-\delta_\gamma \end{aligned} . \tag{94}$$

Let us introduce the variables $g = \sqrt{p\,q}$ and $a = \frac{p+q}{2}$ which are the geometric and the algebraic means of p, q. With their help, we write the following optimization and they are equivalent as knowing a, g gives p, q as well with $p, q = a \pm \sqrt{a^2-g^2}$

$$\begin{aligned} \max_{g,a} \quad & g^2 \\ \text{such that} \quad & a \leq \frac{1}{2} \\ & g^2 \geq a - \frac{1}{2}\delta_\gamma, \\ & a \geq g. \end{aligned} \tag{95}$$

The final constraint arises because the algebraic and geometric means satisfy the inequality $a \geq g$, and guarantees that p and q are well defined, which can be saturated $a = g$ by the choice $p = q$. Note that for a fixed free variable a, the choice $g = a$ is the one that maximizes the goal function and relaxes the constraints the most. We can thus set $a = g$ without loss of generality in order to obtain

$$\begin{aligned} \max_{g\in[0,\frac{1}{2}]} \quad & g^2 \\ \text{such that} \quad & g^2 - g + \frac{1}{2}\delta_\gamma \geq 0. \end{aligned} \tag{96}$$

Here $g^2 - g + \frac{1}{2}\delta_\gamma$ is a decreasing function of $g \in [0, 1/2]$. It admits a zero at $g = \frac{1}{2}\left(1-\sqrt{1-2\delta_\gamma}\right)$ inside the interval, and is thus negative for larger values of g. We conclude that this is precisely the value maximizing g and g^2 under the constraint. Hence, we have shown that

$$\begin{aligned} \max_{p_0,p_1,q_0,q_1} \quad & \min\{p_0q_0, p_1q_1\} \qquad = \left(\frac{1-\sqrt{1-2\,\delta_\gamma}}{2}\right)^2 \\ \text{such that} \quad & p_0+p_1, q_0+q_1 \leq 1 \\ & p_0,p_1,q_0,q_1 \geq 0 \\ & p_0q_0 + p_1q_1 \geq 1-\delta_\gamma. \end{aligned} \tag{97}$$

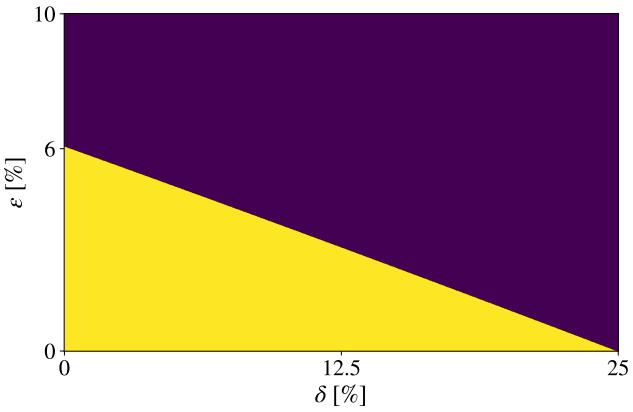


Figure 11: In yellow we show the parameter region (ε, δ) where the noisy W-distributions $P_W^{(\varepsilon,\delta)}(a,b,c)$ in Eq. (98) appear to be nonlocal.

B The noisy W-distribution

In this Section, we illustrate Result 2 and discuss the nonlocality of noisy W-distributions. Such distributions $P_W^{(\varepsilon,\delta)}$ are given by

$$P_W^{(\varepsilon,\delta)}(100) = \frac{1-\varepsilon-\delta}{3} \circlearrowleft, \qquad P_W^{(\varepsilon,\delta)}(111) = \delta \tag{98}$$

and the other terms are left unspecified. As mentioned in the main text the $\circlearrowleft$ symbol means the equation is valid up to cyclic permutations of the parties. It satisfies the PTC condition with probability $1-\varepsilon$. Hence, by our Result 2 we know that there must exist a local PTC distribution $P(a,b,c)$ such that $\delta(P_W^{(\varepsilon,\delta)}, P) \leq \mathrm{f}_\varepsilon$. With straightforward algebra, this condition can be rewritten as

$$\Delta^{(\varepsilon,\delta)}(P) = \sum_{(a,b,c)=(100),(010),(001)} \left| \frac{1-\varepsilon-\delta}{3} - P(a,b,c) \right| + |\delta - P(111)| - 2\mathrm{f}_\varepsilon \leq 0. \tag{99}$$

Now, the set of the PTC distributions can be parameterized with the three "token probabilities" $p_\xi(0) = q_\xi \in [0,1]$ leading to

$$P(111) = q_\alpha q_\beta q_\gamma + (1-q_\alpha)(1-q_\beta)(1-q_\gamma) \tag{100}$$

$$P(100) = (1-q_\alpha) q_\beta q_\gamma + q_\alpha (1-q_\beta)(1-q_\gamma) \qquad \circlearrowleft . \tag{101}$$

Therefore the implication of result 2 is equivalent to

$$\Delta_{min}^{(\varepsilon,\delta)} = \min_{q_\alpha, q_\beta, q_\gamma} \Delta^{(\varepsilon,\delta)}(q_\alpha q_\beta q_\gamma) \leq 0, \tag{102}$$

since there must exist a combination of $q_\alpha, q_\beta, q_\gamma$ for which $\Delta^{(\varepsilon,\delta)}(P) \leq 0$ holds. To prove that a distribution $P_W^{(\varepsilon,\delta)}$ is not triangle-local it is thus sufficient to show that $\Delta_{min}^{(\varepsilon,\delta)} > 0$.

To perform the minimization in Eq. (102), we used the Nelder-Mead or simplex search method [38] via the Python library SciPy [39]. The results of the minimization are reported in Fig. 11. Note that this is an empirical minimization algorithm, which does not provide a convergence guarantee. Note that the goal function $\Delta^{(\varepsilon,\delta)}(q_\alpha, q_\beta, q_\gamma)$ is a smooth and well-behaved function of three variables on a compact interval; therefore, it is plausible that the solver has found its global minimum. Nevertheless, it should be stressed that this only proves the nonlocality of a given distribution under the assumption that the global minimum was indeed found. In the future, it is desirable to perform the same minimization with more computational resources and the guarantee of convergence. This can e.g. be done with the help of the branch-and-bound algorithm, given that the goal function is Lipschitz-continuous on a compact three-parameter subset.

C Linear Program for TBSM when PTC is satisfied exactly

Definition 1. *For a binary tuple of outputs* $\boldsymbol{x} = (a, b, c)$*, and a binary tuple of tokens* $\boldsymbol{t} = (t_\alpha, t_\beta, t_\gamma)$*, we define the PTC response function as follows:*

$$P_{PTC}(\boldsymbol{x}|\boldsymbol{t}) = \begin{cases} 1 & a = t_\beta \oplus t_\gamma \oplus 1,\ b = t_\gamma \oplus t_\alpha \oplus 1\ , c = t_\alpha \oplus t_\beta \oplus 1 \\ 0 & o.w. \end{cases}.$$

This function is, indeed, an indicator function of whether the outputs $\boldsymbol{x}$ and tokens $\boldsymbol{t}$ are consistent. Note that clearly for any $\boldsymbol{t}$ there is only one consistent $\boldsymbol{x}$, and more interestingly for any PTC tuple $\boldsymbol{x} = (a, b, c)$, i.e. $a \oplus b \oplus c = 1$, there are exactly two consistent tokens $\boldsymbol{t}$, which are flipped of each other. For example for $\boldsymbol{x} = (1, 1, 1)$ we find $\boldsymbol{t} = (1, 1, 1)$, or $\boldsymbol{t} = (0, 0, 0)$.

Noiseless case: Nonlocality of $P(\boldsymbol{x}_1\boldsymbol{x}_2)$ can be checked by the infeasibility of the following linear program with 8 probability distributions (64 variables) $P(\boldsymbol{x}_2|\boldsymbol{t})$, where we use the compact notation $\boldsymbol{x}_i = (a_i, b_i, c_i)$, $\boldsymbol{t} = (t_\alpha, t_\beta, t_\gamma)$

$$\begin{aligned} &(i)\ P(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{t}} p(\boldsymbol{t}) P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) P(\boldsymbol{x}_2|\boldsymbol{t}) \\ &(ii)\ \sum_{b_2,c_2} P(a_2, b_2, c_2|0, t_\beta, t_\gamma) - \sum_{b_2,c_2} P(a_2, b_2, c_2|1, t_\beta, t_\gamma) = 0 \quad \circlearrowleft \end{aligned} \tag{103}$$

Note that as mentioned before, for any PTC tuple of $\boldsymbol{x}_1$, there are exactly two consistent $\boldsymbol{t}$; therefore the sum in (i), is only over two terms.

Dephasing noise: the local dephasing noise given by the channel $\mathcal{E}_d : |\psi\rangle\langle\psi| \mapsto \varrho$ with

$$\varrho = (1 - d)\,|\psi\rangle\langle\psi| + d\left(\lambda_0^2\,|01\rangle\langle 01| + \lambda_1^2\,|10\rangle\langle 10|\right) \tag{104}$$

where $|\psi\rangle = \lambda_0\,|01\rangle + \lambda_1\,|10\rangle$. Note that the resulting distribution P in the presence of the dephasing noise is still a PTC distribution and can be written as

$$P = (1 - d)^3 P_Q + (1 - (1 - d)^3) P_C$$

Where P_Q is the quantum distribution obtained without the presence of dephasing noise($d = 0$), and P_C is the classical distribution($d = 1$), therefore we have, $P = P_Q + (P_C - P_Q)e$ for $e := (1 - (1 - d)^3)$ and using (103) finding the minimum amount of dephasing noise d for which the distribution becomes local can be written as

$$\text{Minimize} \quad e \tag{105}$$

$$\text{s.t.} \quad (i)\ P_Q(\boldsymbol{x}_1\boldsymbol{x}_2) = \sum_{\boldsymbol{t}} p(\boldsymbol{t}) P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) P(\boldsymbol{x}_2|\boldsymbol{t}) + (P_Q(\boldsymbol{x}_1\boldsymbol{x}_2) - P_C(\boldsymbol{x}_1\boldsymbol{x}_2))e \tag{106}$$

$$(ii)\ \sum_{b_2,c_2} P(a_2, b_2, c_2|0, t_\beta, t_\gamma) - \sum_{b_2,c_2} P(a_2, b_2, c_2|1, t_\beta, t_\gamma) = 0 \quad \circlearrowleft \tag{107}$$

Solving this LP, we find the optimum value e^*, which translates to the optimum dephasing noise $d^* = 1 + \sqrt[3]{e^* - 1}$. We used the MOSEK Optimizer API for Python to solve the Linear Program [40].

D Linear Program for noisy TBSM distributions

Lemma 1. *Let $p(x)$ and $q(x)$ be probability distributions defined on the discrete probability space $(\Omega, \mathcal{F})$. The following statements are equivalent:*

(i) There exist probability distributions p' and q' on $(\Omega, \mathcal{F})$ such that $p(x) = q(x) + \varepsilon p'(x) - \varepsilon q'(x)$
(ii) $\delta_{TV}(p, q) \leq \varepsilon$

Proof. $(i) \Rightarrow (ii)$:

$$\delta_{TV}(p,q) = \frac{1}{2}\sum_x |p(x) - q(x)| = \frac{\varepsilon}{2}\sum_x |p'(x) - q'(x)| = \varepsilon\ \delta_{TV}(p', q') \leq \varepsilon$$

$(ii) \Rightarrow (i)$: let's define $A \subset \Omega$ as the subset within which $p(x) \geq q(x)$, then (ii) implies that:

$$\sum_{x\in A} \left((p(x) - q(x))\right) + \sum_{x\in \bar{A}} \left((q(x) - p(x))\right) \leq 2\varepsilon$$

let's say $\sum_{x\in A} \left((p(x)-q(x))\right)+\sum_{x\in \bar{A}} \left((q(x)-p(x))\right) = 2\varepsilon'$ for $\varepsilon' \leq \varepsilon$. Moreover, we have $\sum_{x\in A} \left((p(x)-q(x))\right) - \sum_{x\in \bar{A}} \left((q(x) - p(x))\right) = \sum_{x\in\Omega} \left((p(x) - q(x))\right) = 0$, therefore

$$\sum_{x\in A} \left((p(x) - q(x))\right) = \sum_{x\in \bar{A}} \left((q(x) - p(x))\right) = \varepsilon' \tag{108}$$

Let's define $p'(x)$ and $q'(x)$ as follows:

$$p'(x) = \begin{cases} \frac{p(x)-q(x)}{\varepsilon} + \frac{\varepsilon-\varepsilon'}{\varepsilon} f(x) & x \in A \\ \frac{\varepsilon-\varepsilon'}{\varepsilon} f(x) & x \in \bar{A} \end{cases}$$
$$q'(x) = \begin{cases} \frac{\varepsilon-\varepsilon'}{\varepsilon} f(x) & x \in A \\ \frac{q(x)-p(x)}{\varepsilon} + \frac{\varepsilon-\varepsilon'}{\varepsilon} f(x) & x \in \bar{A} \end{cases}$$

for any valid probability distribution $f(x)$ on the same space. It is easy to see that $p'(x)$ and $q'(x)$ are valid probability distributions satisfying (i). It is also interesting to note that, in the case of equality $\delta_{TV}(p, q) = \varepsilon$, $p'(x)$ and $q'(x)$ respectively have a non-zero support only on A and $\bar{A}$ □

Note that equality $\delta_{TV}(p, q) = \varepsilon'$ is achieved iff ε' is the maximum ε satisfying (i) .

LP for an observed distribution with $\Pr(a_1 \oplus b_1 \oplus c_1 = 1) \geq 1 - \varepsilon$: Nonlocality of an observed distribution $P(\boldsymbol{x}_1\boldsymbol{x}_2)$ can be checked by the infeasibility of the following quadratic constrained program with 6 probability distributions $P(\boldsymbol{x}_2, \boldsymbol{t}), P'(\boldsymbol{x}_1, \boldsymbol{x}_2), \widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2), p_\alpha(t), p_\beta(t), p_\gamma(t)$.

$$(*')\ P(\boldsymbol{x}_1, \boldsymbol{x}_2) = \underbrace{\sum_{\boldsymbol{t}} P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) P(\boldsymbol{x}_2, \boldsymbol{t})}_{\widetilde{P}(\boldsymbol{x}_1\boldsymbol{x}_2)} - \mathrm{f}_\varepsilon P'(\boldsymbol{x}_1, \boldsymbol{x}_2) + \mathrm{f}_\varepsilon \widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2)$$
$$(4.iii')\ \sum_{\boldsymbol{x}_1} \widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2) = \sum_{\boldsymbol{x}_1} P'(\boldsymbol{x}_1, \boldsymbol{x}_2) \tag{109}$$
$$(ii)\ p_\alpha(1) \sum_{b_2,c_2} P(a_2, b_2, c_2, 0, t_\beta, t_\gamma) - p_\alpha(0) \sum_{b_2,c_2} P(a_2, b_2, c_2, 1, t_\beta, t_\gamma) = 0 \quad \circlearrowleft \tag{110}$$

Note that $P(t_\alpha, t_\beta, t_\gamma) = p_\alpha(t_\alpha)p_\beta(t_\beta)p_\gamma(t_\gamma)$ is implied by the last constraint. Furthermore, this constraint is coming from the network structure, which implies $P(a_2|0, t_\beta, t_\gamma) = P(a_2|1, t_\beta, t_\gamma)$. Moreover, note that here $p_\alpha(t), p_\beta(t), p_\gamma(t)$ correspond to the PTC distribution $\widetilde{P}$ and not the observed

distribution P, therefore we do not know their values and we consider them as additional variables. Note that the quadratic constraints in (ii) are not convex, i.e. if we write these constraints in the form of $x^T Q x$, the matrix Q is not positive semidefinite.

Note that although we do not know the distributions $p_\alpha(t), p_\beta(t), p_\gamma(t)$ but we can bound them:

First notice that $p_\alpha(t_\alpha = 0) = \frac{1}{2}(1 + \sqrt{|\frac{\widetilde{E}_{b_1}\widetilde{E}_{c_1}}{\widetilde{E}_{a_1}}|})$. Where, $\widetilde{E}_{x_1} = \widetilde{P}(x_1 = 0) - \widetilde{P}(x_1 = 1)$. It is straightforward to see that the total variation distance leads to the conclusion that $|\widetilde{E}_{x_1} - E_{x_1}| \leq 2\mathrm{f}_\varepsilon$. In fact, a slightly better bound can be obtained by using the fact that $\widetilde{P}$ is in the PTC slice, see Section D.2 below,

$$|\widetilde{E}_{x_1} - E^\Lambda_{x_1}| \leq (2\mathrm{f}_\varepsilon - \varepsilon), \tag{111}$$

where $E^\Lambda_{x_1} = P(x_1 = 0, a_1 \oplus b_1 \oplus c_1 = 1) - P(x_1 = 1, a_1 \oplus b_1 \oplus c_1 = 1)$ is computed from the PTC part of the observed distribution P. Hence, if all the correlators for the observed distribution satisfy $|E^\Lambda_{x_1}| \geq (2\mathrm{f}_\varepsilon - \varepsilon)$ one can use Eq. (5) to obtain

$$\begin{aligned} l_\alpha := \frac{1}{2}\left(1 + \sqrt{\frac{(|E^\Lambda_{b_1}| - (2\mathrm{f}_\varepsilon - \varepsilon))(|E^\Lambda_{c_1}| - (2\mathrm{f}_\varepsilon - \varepsilon))}{|E^\Lambda_{a_1}| + (2\mathrm{f}_\varepsilon - \varepsilon)}}\right) &\leq p_\alpha(0) \\ &\leq \frac{1}{2}\left(1 + \sqrt{\frac{(|E^\Lambda_{b_1}| + (2\mathrm{f}_\varepsilon - \varepsilon))(|E^\Lambda_{c_1}| + (2\mathrm{f}_\varepsilon - \varepsilon))}{|E^\Lambda_{a_1}| - (2\mathrm{f}_\varepsilon - \varepsilon)}}\right) := u_\alpha \end{aligned} \tag{112}$$

Therefore (ii) can be written as:

$$l_\alpha \sum_{b_2,c_2} P(a_2, b_2, c_2, 0, t_\beta, t_\gamma) - (1 - l_\alpha) \sum_{b_2,c_2} P(a_2, b_2, c_2, 1, t_\beta, t_\gamma) \leq 0 \tag{113}$$

$$u_\alpha \sum_{b_2,c_2} P(a_2, b_2, c_2, 0, t_\beta, t_\gamma) - (1 - u_\alpha) \sum_{b_2,c_2} P(a_2, b_2, c_2, 1, t_\beta, t_\gamma) \geq 0. \tag{114}$$

As an illustration, we can obtain bounds on the white noise robustness of the RGB4 distributions, see Fig. 12.

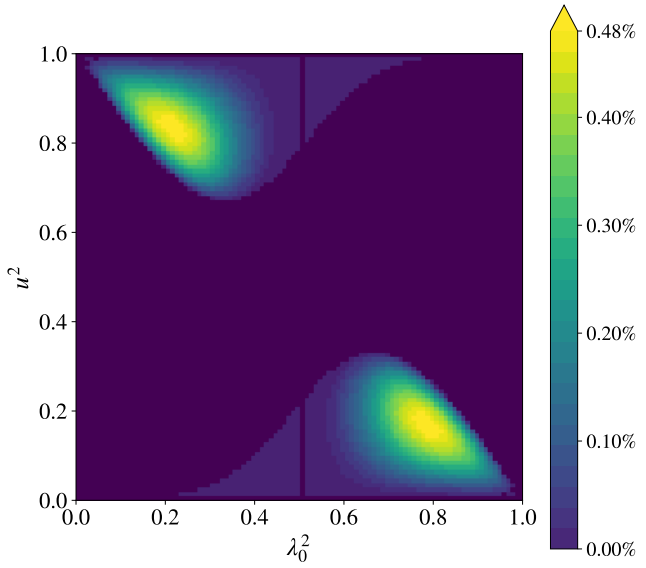


Figure 12: **Noise robustness of the RGB4 distributions.** Maximal amount of white noise below which we can prove the nonlocality of RGB4 distributions (TBSM distributions with $\phi_w = 0$).

D.1 Collection of LPs (the grid technique for $p_\xi(t_\xi)$)

Finally, note that the LP has to be satisfied for some values $p_\xi(t_\xi)$ of the token probabilities in the intervals $[l_\alpha, u_\alpha]$. In practice, it is convenient to divide each interval into smaller intervals $[l_\alpha + k\frac{u_\alpha - l_\alpha}{M}, l_\alpha + (k+1)\frac{u_\alpha - l_\alpha}{M}]$, and verify that no solution can be found within any of the M^3 possible combinations of intervals. This gives a better result as compared to the harsher relaxation in Eqs. (113,114). In particular, for the figures 7,8,12,13, and 14 we used $M = 16$, while to obtain the best bounds $\omega = 0.544\%, \omega' = 0.504\%$ white noise and $\varepsilon = 0.24\%$ total-variation distance we went to $M = 128$.

Furthermore, to see how good is this relaxation of the $p_\xi(t_\xi)$ variable to a finite set of intervals, we also check the feasibility of the LP for the honest values of $p_\xi(0) = q_\xi = \lambda_0^2$, i.e. the values which appear in the quantum strategy and are thus feasible by construction. This gives an upper bound on the maximum amount of noise for which the original feasibility problem could guarantee that the distribution is nonlocal.

D.2 Development of a tighter bound on $\widetilde{E}_{x_1}$

Note that we have:

$$P(\boldsymbol{x}_1, \boldsymbol{\xi} \in \Lambda) = P(\boldsymbol{x}_1; \boldsymbol{\xi} \in \Upsilon) + (\mathrm{f}_\varepsilon - \varepsilon)P''(\boldsymbol{x}_1)$$
$$\widetilde{P}(\boldsymbol{x}_1) = P(\boldsymbol{x}_1; \boldsymbol{\xi} \in \Upsilon) + \mathrm{f}_\varepsilon \widetilde{P'}(\boldsymbol{x}_1)$$

Which can be rewritten as :

$$\widetilde{P}(\boldsymbol{x}_1) = P(\boldsymbol{x}_1, \boldsymbol{\xi} \in \Lambda) + \mathrm{f}_\varepsilon \widetilde{P}'(\boldsymbol{x}_1) - (\mathrm{f}_\varepsilon - \varepsilon)P''(\boldsymbol{x}_1)$$

Therefore with $E^\Lambda_{x_1} = P(x_1 = 0, a_1 \oplus b_1 \oplus c_1 = 1) - P(x_1 = 1, a_1 \oplus b_1 \oplus c_1 = 1)$ we have

$$\widetilde{E}_{x_1} = E^\Lambda_{x_1} + \mathrm{f}_\varepsilon \widetilde{E}'_{x_1} - (\mathrm{f}_\varepsilon - \varepsilon)E''_{x_1}$$

Therefore,

$$\widetilde{E}_{x_1} \in [E^\Lambda_{x_1} - (2\mathrm{f}_\varepsilon - \varepsilon), E^\Lambda_{x_1} + (2\mathrm{f}_\varepsilon - \varepsilon)].$$

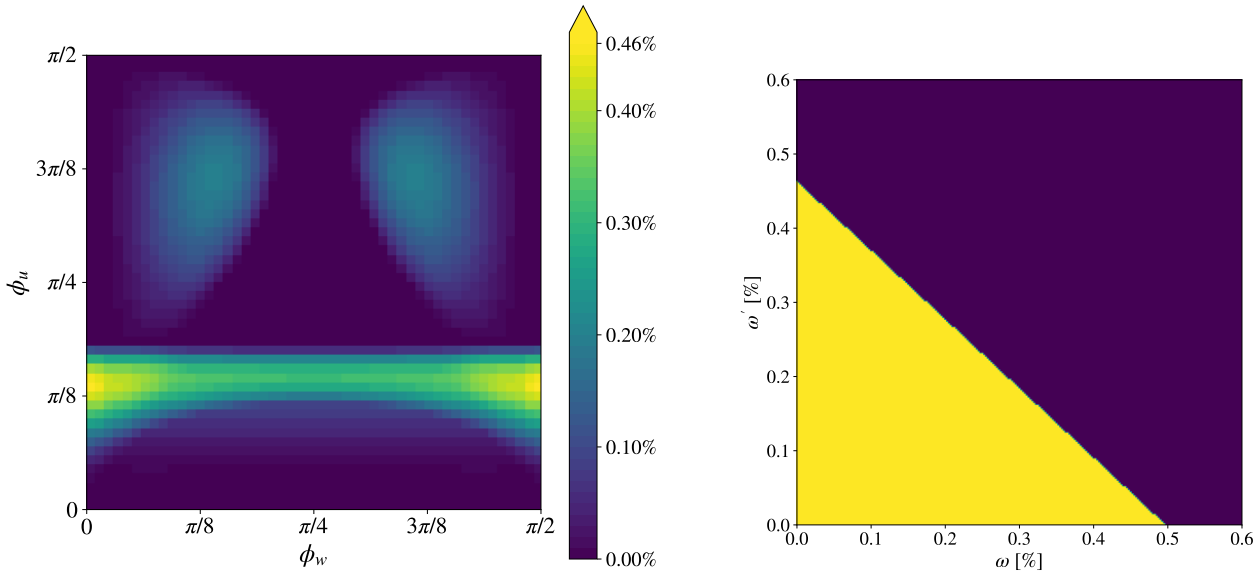
E Other noise models

White noise at the measurement. Alternatively to the white noise at the source, discussed in the main text, one can consider white noise affecting the measurements. In this case the POVM elements describing the measurement are replaced by

$$E_{x_1,x_2} = (1-\omega')\,|\phi_{x_1,x_2}\rangle\langle\phi_{x_1,x_2}| + \omega'\frac{\mathbb{1}}{2}\otimes\frac{\mathbb{1}}{2}. \tag{115}$$

Note that here for the resulting distribution we have $\Pr(a_1 \oplus b_1 \oplus c_1 = 1) = \frac{1}{2}\left(1+(1-\omega')^3\right)$. In Fig. 13(a), we depict the robustness to measurement white noise of the TBSM distributions with $\lambda_0^2 = 0.22$. The maximal amount of noise for which we can prove that nonlocality of the distribution is given by $\omega' = 0.504\%$ for the case $\phi_u = 0.42$ and $\phi_w = 0$.

Considering the combined effect of white noise at the sources and the parties we get, $\Pr(a_1 \oplus b_1 \oplus c_1 = 1) = \frac{1}{2}\left(1+(1-\omega')^3(1-\omega)^3\right)$. In Fig. 13(b), we depict the nonlocality of the TBSM distribution with $\phi_u = 0.42$ and $\phi_w = 0$ as a function of both the noise parameters.



(a) **White noise at the measurement** (b) **Noisy states and measurements**

Figure 13: **(a)** Maximal amount of white noise at the level of measurements per party below which we can prove the nonlocality of the TBSM distribution. Parameters: $u = \cos(\phi_u)$, $w = \cos(\phi_w)$ and $\lambda_0^2 = 0.22$. **(b)** Noise robustness of the TBSM distribution when noise is added to both states (ω) and measurements (ω'). The distribution is nonlocal in the yellow region, for parameters $\lambda_0^2 = 0.22, \phi_u = 0.42, \phi_w = 0$.

No-click noise. Another natural imperfection to consider is the possibility that the measurement does not produce an outcome with probability p. This is described by five-outcome POVM given by

$$E_{x_1,x_2} = (1-p)\,|\phi_{x_1,x_2}\rangle\langle\phi_{x_1,x_2}| \qquad \forall x_1, x_2 = 0, 1 \tag{116}$$

$$E_{\emptyset} = p\,\mathbb{1} \tag{117}$$

where $E_{\emptyset}$ is the element associated with the no-click event. The probability distribution obtained with such measurements has five outcomes for each party and is not readily analyzed with the results presented in this paper. However, a trick commonly used for device-independent quantum key distribution is to relabel the no-click outcome to one of the four "good" outcomes restoring output cardinality.

We find that the optimal strategy (for noise robustness) is for each party to relabel the no-click outcome to $\varnothing \to (1,0)$, which corresponds to the POVM

$$E_{1,0} = (1-p)\,|\phi_{1,0}\rangle\langle\phi_{1,0}| + p\,\mathbb{1} \tag{118}$$
$$E_{x_1,x_2} = (1-p)\,|\phi_{x_1,x_2}\rangle\langle\phi_{x_1,x_2}| \qquad \forall (x_1,x_2) \neq (1,0). \tag{119}$$

For this strategy, we prove the nonlocality of the resulting distribution for up to a no-click probability of $p = 0.7\%$, see Fig. 14(a).

Note that if e.g. the sources produce two photons entangled in polarization $|0\rangle = |1_h\rangle$ and $|1\rangle = |1_v\rangle$, a limited photon transmission (or detector efficiency) η would correspond to the no-click noise with $(1-p) = \eta^2$.

Single photon entanglement implementation with loss. For $\phi_w = 0$ the TBSM distributions (RGB4) can be realised with single-photon entanglement (vacuum state $|0\rangle$ and single photon $|1\rangle$), linear optics and photon number resolving detectors [36]. In this case the dominant noise is the limited efficiency of the detectors, which can be modeled as a photon loss channel with transmission η. Formally, we have the channel

$$\mathcal{E}(\rho) = K_0 \rho K_0^\dagger + K_1 \rho K_1^\dagger \tag{120}$$
$$K_0 = |0\rangle\langle 0| + \sqrt{\eta}\,|1\rangle\langle 1| \tag{121}$$
$$K_1 = \sqrt{1-\eta}\,|0\rangle\langle 1| \tag{122}$$

which is applied on all of the six modes. Alternatively, the noise can be absorbed in the measurements by considering the POVM elements

$$E_{x_1,x_2} = \mathcal{E}^* \otimes \mathcal{E}^*(\Pi_{x_1,x_2}) = \sum_{i,j=0}^{1} K_i^\dagger \otimes K_j^\dagger\, \Pi_{x_1,x_2}\, K_i \otimes K_j. \tag{123}$$

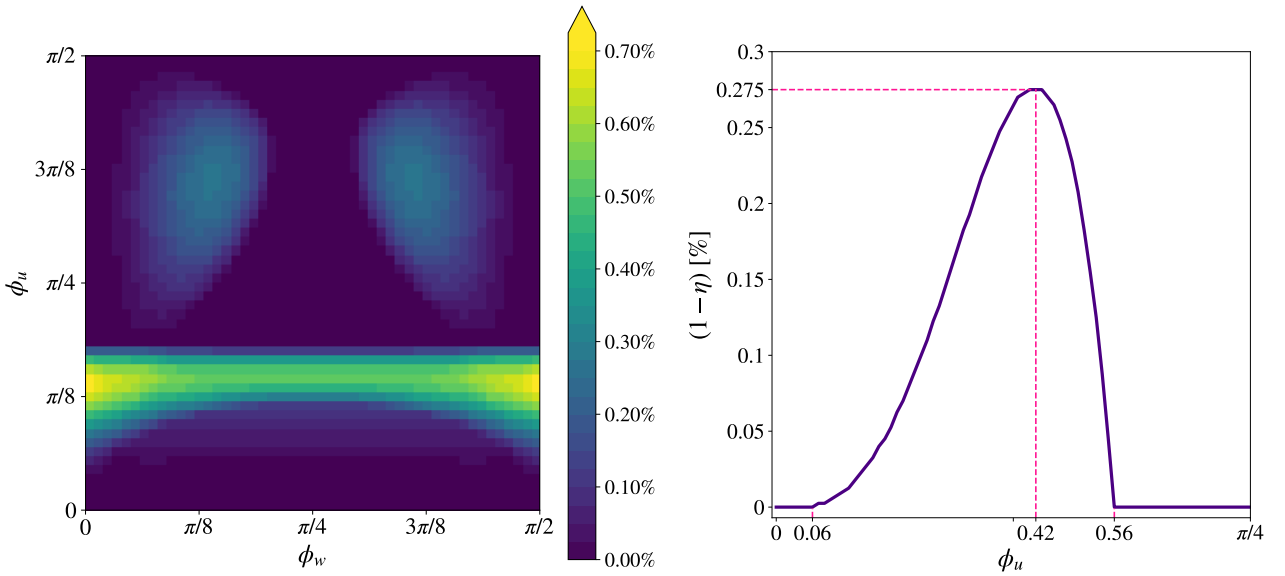
Recall that for $\phi_w = 0$ the original projectors for $x_1 = 1$ read $\Pi_{00} = |00\rangle\langle 00|$ and $\Pi_{01} = |11\rangle\langle 11|$, while the lossy POVM elements read

$$E_{0,0} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1-\eta & 0 & 0 \\ 0 & 0 & 1-\eta & 0 \\ 0 & 0 & 0 & (1-\eta)^2 \end{pmatrix} \qquad E_{0,1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \eta^2 \end{pmatrix} \tag{124}$$

$$E_{1,0} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \eta u^2 & \eta uv & 0 \\ 0 & \eta uv & \eta v^2 & 0 \\ 0 & 0 & 0 & (1-\eta)\eta \end{pmatrix} \qquad E_{1,1} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & \eta v^2 & -\eta uv & 0 \\ 0 & -\eta uv & \eta u^2 & 0 \\ 0 & 0 & 0 & (1-\eta)\eta \end{pmatrix} \tag{125}$$

in the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$.

The nonlocality of the resulting distributions can be detected up to a transmission loss of $1-\eta = 0.275\%$, see Fig. 14 (b).



(a) **No-click scenario** (b) **Single-photon entanglement scenario with loss**

Figure 14: **(a)** Maximal no-click probability at each party below which we can prove the nonlocality of the distribution. Parameters: $u = \cos(\phi_u)$, $w = \cos(\phi_w)$ and $\lambda_0^2 = 0.22$. **(b)** Maximal amount of photon loss at measurements below which we can prove the nonlocality of the RGB4 distribution (TBSM with $\phi_w = 0$) for the single-photon entanglement implementation. Here η denotes the detection efficiency, i.e. the probability that an emitted photon is detected. Parameters: $u = \cos(\phi_u)$, $\phi_w = 0$ and $\lambda_0^2 = 0.22$

F Linear Program for the total variation distance

Given a target quantum distribution $\bar{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) \in \mathcal{Q}_\triangle^4$ define the total variation distance ball as the following subset of the correlation set

$$\mathcal{B}_\varepsilon(\bar{P}) = \{P \in \mathcal{P}_\triangle^4 | \delta(P, \bar{P}) \leq \varepsilon\}. \tag{126}$$

We will now demonstrate that for well-chosen target distributions $\bar{P}$ and small enough ε the entire set $\mathcal{B}_\varepsilon(\bar{P})$ is nonlocal. Concretely, consider $\bar{P}$ such that $a_1 \oplus b_1 \oplus c_1 = 1$. Any distribution $P \in \mathcal{B}_\varepsilon(\bar{P})$ must satisfy

$$\Pr(a_1 \oplus b_1 \oplus c_1 = 1) \geq 1 - \varepsilon. \tag{127}$$

We will prove that all distributions in $\mathcal{B}_\varepsilon(\bar{P})$ are nonlocal by contradiction. Let us assume that the ball contains some triangle-local distribution P, then from Result 3 there must exist a local PTC distribution $\widetilde{P}$ with $\delta(\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2), P(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq \mathrm{f}_\varepsilon$ and $\delta(\widetilde{P}(\boldsymbol{x}_2), P(\boldsymbol{x}_2)) = 0$. By triangle inequality, $\widetilde{P}$ must also satisfy

$$\delta(\widetilde{P}(\boldsymbol{x}_1, \boldsymbol{x}_2), \bar{P}(\boldsymbol{x}_1, \boldsymbol{x}_2)) \leq \delta(\widetilde{P}, P) + \delta(P, \bar{P}) \leq \varepsilon + \mathrm{f}_\varepsilon \tag{128}$$

$$\delta(\widetilde{P}(\boldsymbol{x}_2), \bar{P}(\boldsymbol{x}_2)) \leq 0 + \delta(P, \bar{P}) \leq \varepsilon. \tag{129}$$

Therefore, from Eq. (128) and Lemma 1

$$\bar{P}(\boldsymbol{x}_1, \boldsymbol{x}_2) = \underbrace{\sum_{\boldsymbol{t}} P_{PTC}(\boldsymbol{x}_1|\boldsymbol{t}) P(\boldsymbol{x}_2, \boldsymbol{t})}_{\widetilde{P}(\boldsymbol{x}_1\boldsymbol{x}_2)} - (\varepsilon + \mathrm{f}_\varepsilon)\bar{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2) + (\varepsilon + \mathrm{f}_\varepsilon)\widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2). \tag{130}$$

Taking the marginal we get

$$\bar{P}(\boldsymbol{x}_2) = \widetilde{P}(\boldsymbol{x}_2) - (\varepsilon + \mathrm{f}_\varepsilon)\bar{P}'(\boldsymbol{x}_2) + (\varepsilon + \mathrm{f}_\varepsilon)\widetilde{P}'(\boldsymbol{x}_2). \tag{131}$$

Now from this equation and Eq. (129) we get

$$\delta(\widetilde{P}'(\boldsymbol{x}_2), \bar{P}'(\boldsymbol{x}_2)) = \frac{1}{\varepsilon + \mathrm{f}_\varepsilon}\delta(\widetilde{P}(\boldsymbol{x}_2), \bar{P}(\boldsymbol{x}_2)) \leq \frac{\varepsilon}{\varepsilon + \mathrm{f}_\varepsilon} := \alpha. \tag{132}$$

Note that as $2.5\varepsilon \leq \mathrm{f}_\varepsilon \leq 3\varepsilon$, we have $1/4 \leq \alpha \leq 2/7$, moreover for small values of ε, we have $\mathrm{f}_\varepsilon \approx 2.5\varepsilon$ and $\alpha \approx 2/7$. Again by Lemma 1, we have

$$(4.iii') \sum_{\boldsymbol{x}_1} \widetilde{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2) - \sum_{\boldsymbol{x}_1} \bar{P}'(\boldsymbol{x}_1, \boldsymbol{x}_2) + \alpha\widetilde{P}''(\boldsymbol{x}_2) - \alpha\bar{P}''(\boldsymbol{x}_2) = 0. \tag{133}$$

Moreover, like before, we use the network constraint that $P(a_2|0, t_\beta, t_\gamma) = P(a_2|1, t_\beta, t_\gamma)$ written as the following

$$(ii)\ p_\alpha(0) \sum_{b_2,c_2} P(a_2, b_2, c_2, 0, t_\beta, t_\gamma) - p_\alpha(1) \sum_{b_2,c_2} P(a_2, b_2, c_2, 1, t_\beta, t_\gamma) = 0 \quad \circlearrowleft. \tag{134}$$

Note that here $p_\alpha(t), p_\beta(t), p_\gamma(t)$ correspond to the PTC distribution $\widetilde{P}$ which is not observed, but we can bound $p_\alpha(t), p_\beta(t), p_\gamma(t)$:

Notice that $p_\alpha(t_\alpha = 0) = \frac{1}{2}(1 + \sqrt{|\frac{\widetilde{E}_{b_1}\widetilde{E}_{c_1}}{\widetilde{E}_{a_1}}|})$. Where, $\widetilde{E}_{x_1} = \widetilde{P}(x_1 = 0) - \widetilde{P}(x_1 = 1)$. It is straightforward to see that the total variation distance $\delta(\widetilde{P}, \bar{P}) \leq \mathrm{f}_\varepsilon + \varepsilon$, leads to the conclusion that $|\widetilde{E}_{x_1} - \bar{E}_{x_1}| \leq 2(\mathrm{f}_\varepsilon + \varepsilon)$. Where $\bar{E}_{x_1} = \bar{P}(x_1 = 0) - \bar{P}(x_1 = 1)$ is computed from the target PTC distribution $\bar{P}$. Hence, if all the correlators for the target distribution satisfy $|\bar{E}_{x_1}| \geq 2(\mathrm{f}_\varepsilon + \varepsilon)$ one can use Eq. (5) to obtain

$$\begin{aligned} l_\alpha := \frac{1}{2}\left(1 + \sqrt{\frac{(|\bar{E}_{b_1}| - 2(\mathrm{f}_\varepsilon + \varepsilon))(|\bar{E}_{c_1}| - 2(\mathrm{f}_\varepsilon + \varepsilon))}{|\bar{E}_{a_1}| + 2(\mathrm{f}_\varepsilon + \varepsilon)}}\right) \leq\ & p_\alpha(0) \\ p_\alpha(0) \leq \frac{1}{2}\left(1 + \sqrt{\frac{(|\bar{E}_{b_1}| + 2(\mathrm{f}_\varepsilon + \varepsilon))(|\bar{E}_{c_1}| + 2(\mathrm{f}_\varepsilon + \varepsilon))}{|\bar{E}_{a_1}| - 2(\mathrm{f}_\varepsilon + \varepsilon)}}\right) :=\ & u_\alpha \quad \circlearrowleft. \end{aligned} \tag{135}$$

Therefore (ii) can be written as:

$$\begin{aligned} l_\alpha \sum_{b_2,c_2} P(a_2, b_2, c_2, 0, t_\beta, t_\gamma) - (1 - l_\alpha) \sum_{b_2,c_2} P(a_2, b_2, c_2, 1, t_\beta, t_\gamma) \leq 0 \\ u_\alpha \sum_{b_2,c_2} P(a_2, b_2, c_2, 0, t_\beta, t_\gamma) - (1 - u_\alpha) \sum_{b_2,c_2} P(a_2, b_2, c_2, 1, t_\beta, t_\gamma) \geq 0. \end{aligned} \tag{136}$$

Combining Equations 130, 133, and 136, yields a set of linear constraints that should be satisfied if there exists any local distribution inside the ball $\mathcal{B}_\varepsilon(\bar{P})$. Therefore, the infeasibility of this LP proves the non-locality of all of the points in $\mathcal{B}_\varepsilon(\bar{P})$. (see Fig. 8). Again, to find tighter bounds, we use the grid technique.