



## Coincidence postselection for genuine multipartite nonlocality: Causal diagrams and threshold efficiencies

Valentin Gebhart  and Augusto Smerzi 

*QSTAR, INO-CNR and LENS, Largo Enrico Fermi 2, 50125 Firenze, Italy*



(Received 28 July 2022; accepted 9 November 2022; published 2 December 2022)

Genuine multipartite nonlocality (GMN), the strongest form of multipartite nonlocality that describes fully collective nonlocal correlations among all experimental parties, can be observed when different distant parties each locally measure a particle from a shared entangled many-particle state. For the demonstration of GMN, the experimentally observed statistics are typically postselected: Events for which some parties do not detect a particle must be discarded. This coincidence postselection generally leads to the detection loophole that invalidates a proper nonlocality demonstration. In this work, we address how to close the detection loophole for a coincidence detection in demonstrations of nonlocality and GMN. We first show that if the number of detected particles is conserved, i.e., using ideal and noiseless experimental devices, one can employ causal diagrams and the no-signaling principle to prove that a coincidence postselection cannot create any detection loophole. Furthermore, for realistic experimental devices with finite detection efficiencies, we show how a general Bell inequality can be sharpened such that its new version is still valid after a postselection of the measurement data. In this case, there are threshold detection efficiencies that, if surpassed in the experiment, lead to the possibility of demonstrating nonlocality and GMN without opening the detection loophole. Our results imply that genuine  $N$ -partite nonlocality can be generated from  $N$  independent particle sources even when allowing for nonideal detectors.

DOI: [10.1103/PhysRevA.106.062202](https://doi.org/10.1103/PhysRevA.106.062202)

### I. INTRODUCTION

Bell nonlocality [1,2] is one of the most intriguing aspects of quantum systems and plays a central role in modern research of foundational physics and the development of quantum-enhanced technologies [3], such as quantum key distribution and quantum random number generators. For a proper experimental demonstration of nonlocality, it is essential to exclude any local-realist explanation of the observed measurement results that appear to violate a Bell inequality, including any possible “loopholes” that the explanation could potentially utilize. Two main loopholes in Bell experiments are (i) the locality loophole, if the different parts of experimental configuration are not separated distantly enough to exploit the principles of special relativity [4], and (ii) the detection loophole, if the measured statistics must be postselected due to a nonideal detection efficiency or particle losses [5,6], because of the possibility that the postselection generates fake nonlocal correlation via the selection bias [7].

The most common way to address the detection loophole is to assume fair sampling [8–11], i.e., to assume that the postselected statistics is a fair sample of the statistics that would have been observed using ideal experimental tools. However, the fair sampling assumption does not necessarily hold in real experiments: The detection loophole has been exploited to create false demonstrations of nonlocality [12–15], corrupting the security of quantum technological applications [16,17]. Therefore, for an unambiguous demonstration of nonlocality, the detection loophole has to be closed. To do so, one can

include the nondetection events in the statistics, i.e., one does not discard any measurement data [6,18–20], such that there is no effect due to postselection. The second approach is to postselect data but, at the same time, to sharpen the Bell inequality accordingly [21–23]. Both of these approaches yield a (minimal) threshold detection efficiency of the experimental apparatus that must be achieved, and, in this way, the detection loophole (and the locality loophole) was eventually closed in recent experiments [24–29]. The precise values of the threshold efficiencies depends on the Bell inequality in question and has been subject to a long line of research [19,21–23,30–36]. However, to our knowledge, there is no analysis of how to close the detection loophole in demonstrations of genuine multipartite nonlocality (GMN) [37–39]. GMN is the strongest form of multipartite nonlocality that requires that the correlations cannot be explained by nonlocal correlations shared only by some groups of the experimental parties, and constitutes the quantum resource for different quantum technologies [40–46]. Furthermore, in most studies, threshold efficiencies were derived for setups where, in the ideal noiseless limit, each party receives a single particle. These results are not applicable to Bell scenarios in which the particles’ destinations are prepared in a superposition [20], such as, the proposal by Yurke and Stoler (YS) to generate nonlocality from independent particle sources [47,48].

In this work, we consider a general  $N$ -partite Bell scenario with a coincidence postselection, i.e., a postselection of events for which each of the  $N$  parties detects a single particle. This postselection may be necessary due to nonideal detectors and

particle losses, or a random distribution of particles among the parties, or both. We first examine an ideal experimental apparatus where the number of detected particles is conserved. In this case, we use causal diagrams and  $d$ -separation rules [7], together with the no-signaling principle, to show that a coincidence postselection is valid for demonstrations of GMN, extending the results of Refs. [49,50]. Second, we analyze general Bell inequalities (testing for nonlocality or GMN) if noisy experimental devices are employed, in which case causal diagrams cannot prove a valid postselection anymore. Instead, we derive sharpened Bell inequalities that must be used to close the detection loophole when postselecting the measurement results [22,23]. The sharpened inequalities yield threshold detection efficiencies that, if surpassed in experiments, enable a demonstration of multipartite nonlocality or GMN. Our results can be used to demonstrate GMN also in setups where the particles are randomly distributed among the parties [20,47,48], showing that one can create genuine  $N$ -partite nonlocality from  $N$  independent particle sources even for nonideal detectors.

## II. COINCIDENCE POSTSELECTION WITH IDEAL DETECTORS: CAUSAL DIAGRAMS

Here, we consider a Bell scenario with ideal detectors and no particle losses, and in which a constant number  $N_T$  of particles is shared among  $N$  parties. Thus, the number of detected particles of each party is completely determined by the number of detected particles of the remaining parties. We can then employ causal inference and  $d$ -separation rules<sup>1</sup> [7], together with the no-signaling principle,<sup>2</sup> to show that a coincidence postselection, i.e., a postselection of events in which each of the parties receives a single particle, is valid for demonstrations of nonlocality and GMN. We will focus on the case that  $N_T = N$  particles are distributed, and note that the analysis also holds for  $N_T > N$  if each party should receive a fixed number of particles. The analysis can also be applied for  $N_T < N$  but, in this case, GMN cannot be observed because not all parties receive a particle.

### A. Bipartite nonlocality

For simplicity, we first analyze a bipartite Bell scenario, consisting of two parties, Alice and Bob, who share two parts of a quantum system and each perform local measurements on their subsystem. Alice (Bob) can choose different

<sup>1</sup>The  $d$ -separation rules dictate how to infer the statistical dependence between two nodes of a causal diagram, also if some of the others variables are conditioned on. In general, any path of causal arrows that connects two nodes of the diagram can lead to a dependence. The  $d$ -separation rules say that (i) a path is blocked if there is a collider (a node where the path's arrows collide) along the path, (ii) a path is blocked if along it there is a noncollider that is conditioned on, and (iii; selection bias) a path is open if along it there is a collider that is conditioned on.

<sup>2</sup>The no-signaling principle states that the measurement-setting choice of any party cannot influence the results of any other spacelike-separated party, even if all (hidden) variables of the system were known [39,51,52]. For a formal definition, see Eq. (6).

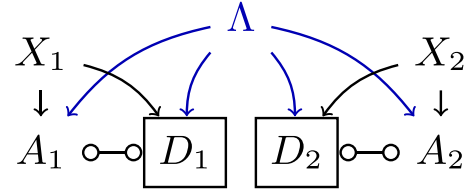


FIG. 1. Causal diagram of a local hidden variable (LHV) model in a bipartite Bell scenario where the  $k$ th party ( $k = 1, 2$ ) can choose a measurement setting  $x_k$  and observes the outcomes  $a_k$  and  $d_k$ , where  $d_k$  is the number of detected particles. The statistics are then postselected on coincidence events, i.e., on events for which  $d_1 = d_2 = 1$ , as indicated by the boxes. Generally, one can allow for any causal influence between  $A_k$  and  $D_k$  as indicated as the lines with circular endings.

measurement settings, labeled by the variable  $X_1$  ( $X_2$ ), and observes an outcome denoted as a random variable  $A_1$  ( $A_2$ ). Furthermore, we indicate the number of detected particles at Alice's and Bob's measurement station as the variables  $D_1$  and  $D_2$ , respectively. To derive a Bell inequality, one assumes that the observed correlations can be described by a local hidden variable (LHV) model [1,2],

$$p_{a_1, a_2, d_1, d_2 | x_1, x_2} = \int d\lambda p_\lambda p_{a_1, d_1 | x_1, \lambda} p_{a_2, d_2 | x_2, \lambda}, \quad (1)$$

where  $\Lambda$  is a LHV and each probability  $p_\lambda$ ,  $p_{a_1, d_1 | x_1}$ , and  $p_{a_2, d_2 | x_2}$  sums to 1. Note that we indicate the possible values of a random variable  $X$  as lower-case letters  $x$  and write  $p_x$  for the probability  $P(X = x)$ . The causal diagram of this LHV model is shown in Fig. 1: The LHV  $\Lambda$  can influence all measurement outcomes, while the local setting  $X_k$  ( $k = 1, 2$ ) can only influence the outcomes  $A_k$  and  $D_k$ . Furthermore, we make no restriction on possible causal influences between  $A_k$  and  $D_k$ , which we indicate as a bidirected arrow with circular ends; this includes influences of the form  $A_k \rightarrow D_k$  or  $D_k \rightarrow A_k$ , and a hidden common cause between  $A_k$  and  $D_k$  (which can be included in  $\Lambda$ ).

Now consider a coincidence postselection of the statistics, i.e., a postselection of events where each party detects a single particle. We denote the detection of a single particle as  $D_k = d_k$ , while any other number of detected particles (e.g.,  $D_k = 0$  and  $D_k = 2$ ) will be grouped to the outcome  $D_k = d_k$ . In Fig. 1, we indicate the coincidence postselection, i.e., the conditioning on  $D_k = d_k$ , as boxes around the variables  $D_k$ . If the postselected statistics,  $p_{a_1, a_2 | d_1, d_2, x_1, x_2}$ , can be written as a LHV model similar to Eq. (1), they also fulfill the Bell inequality and, thus, the postselection does not open any detection or postselection loophole. Using  $p_{a_1, a_2 | d_1, d_2, x_1, x_2} = \int d\lambda p_\lambda |_{d_1, d_2, x_1, x_2} p_{a_1, a_2 | d_1, d_2, x_1, x_2, \lambda}$ , we see that  $p_{a_1, a_2 | d_1, d_2, x_1, x_2}$  can be described as a LHV model if the conditions

$$p_\lambda |_{d_1, d_2, x_1, x_2} = p_\lambda |_{d_1, d_2}, \quad (2)$$

$$p_{a_1, a_2 | d_1, d_2, x_1, x_2, \lambda} = p_{a_1 | d_1, d_2, x_1, \lambda} p_{a_2 | d_1, d_2, x_2, \lambda} \quad (3)$$

are satisfied [11,49].

Equation (3) can be inferred directly from Fig. 1 using the  $d$ -separation rules<sup>1</sup>: Any path that connects  $(A_1, X_1)$  to  $(A_2, X_2)$  passes through  $\Lambda$  and is blocked because  $\Lambda$  is a

noncollider that is conditioned on. To show Eq. (2), we must use the fact that a constant number of particles is distributed among the parties and that we employ ideal noiseless (number-resolving) detectors. In this case, the value of  $D_1$  can be inferred by the value of  $D_2$ , and the other way around. Now, to show the independence of  $X_1$  and  $\Lambda$  when conditioning on  $D_1$  and  $D_2$ , we must consider the two possible paths  $X_1 \rightarrow D_1 \leftarrow \Lambda$  and  $X_1 \rightarrow A_1 \rightarrow D_1 \leftarrow \Lambda$  that both appear open as  $D_1$  is a collider that is conditioned on. However, if there was a nonvanishing influence from  $X_1$  to  $D_1$  (along any path), since  $D_2$  is completely determined by  $D_1$ , there would also be a nonvanishing influence from  $X_1$  to  $D_2$ , in conflict with the no-signaling principle.

**B. (Genuine) multipartite nonlocality**

To test for general nonlocality in the  $N$ -partite Bell scenario, one must extend Eq. (1) to  $N$  parties. One again writes the correlations as  $p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda} = \int d\lambda p_\lambda p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda}$  with the factorization

$$p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda} = \prod_{k=1}^N p_{a_k,d_k|x_k,\lambda}, \tag{4}$$

where the  $k$ th party can choose the measurement setting  $x_k$  and observes the outcomes  $a_k$  and  $d_k$ , and we have used the notation  $\mathbf{a} = (a_1, \dots, a_N)$  and similarly for  $\mathbf{d}$  and  $\mathbf{x}$ . However, for  $N > 2$ , one can also test for a stronger form of nonlocality called genuine  $N$ -partite nonlocality. Here, one allows for nonlocal correlations shared among subgroups of the  $N$  parties. Thus, one only requires that  $p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda}$  must factorize in at least two factors, yielding a hybrid local-nonlocal hidden variable (HLNHV) model [37–39]. For instance, one possible factorization is given by

$$p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda} = p_{a_k,d_k|x_k,\lambda} p_{\mathbf{a}\setminus a_k,\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k,\lambda} \tag{5}$$

for some  $1 \leq k \leq N$ , where we introduced the notation  $\mathbf{a}\setminus a_k = (a_1, \dots, a_{k-1}, a_{k+1}, \dots, a_N)$  and similarly for  $\mathbf{d}$  and  $\mathbf{x}$ . In the HLNHV model, we furthermore assume that all nonlocal correlations fulfill the no-signaling conditions<sup>3</sup> [39,51,52], e.g.,

$$p_{\mathbf{a}\setminus a_k,\mathbf{d}\setminus d_k|\mathbf{x},\lambda} = p_{\mathbf{a}\setminus a_k,\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k,\lambda}, \tag{6}$$

for any  $k$ . This ensures that the measurement setting  $X_k$  of the  $k$ th party has no influence on the measurement outcomes of the other parties, even if conditioned on the hidden variable  $\Lambda$ .

We now focus on a coincidence postselection of a HLNHV model and note that the  $N$ -partite LHV model, Eq. (4), can be discussed in completely analogy to Sec. II A. In the case of

<sup>3</sup>We note that, in the literature, the no-signaling principle is sometimes also defined in an operational form, e.g.,  $p_{a|xy} = p_{a|x}$ , i.e., excluding the hidden variable  $\Lambda$ . This definition is used, e.g., in the discussion of Popescu-Rohrlich boxes [3,53]. In the context of GMN, the no-signaling principle is usually defined as  $p_{a|xy\lambda} = p_{a|x\lambda}$ , i.e., that any party’s measurement choice cannot influence the outcome of a second party even if the hidden variable  $\Lambda$  was known [39,51,52], which implies the operational no-signaling principle.

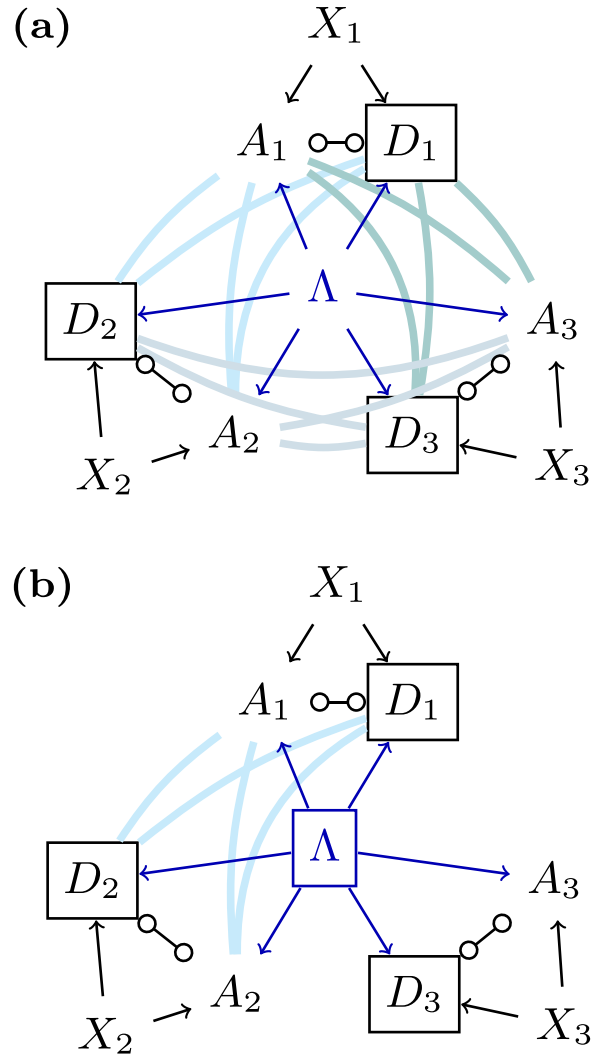


FIG. 2. Causal diagrams of a hybrid local-nonlocal hidden variable (HLNHV) model in a three-partite Bell scenario. (a) Besides the commonly shared LHV  $\Lambda$ , any pair of parties can share nonlocal quantum correlations (indicated as the light blue lines) that are subject to the no-signaling fine-tuning conditions. (b) For a specific value of the LHV  $\Lambda$  (i.e., conditioning on  $\Lambda$  as indicated by the box), one of the three parties factorizes with the other two parties, while the latter can still share nonlocal quantum correlations.

three parties, we sketch the causal diagram of the HLNHV model in Fig. 2. We indicate the nonlocal correlations that can be shared between any two of the three parties as light blue lines. These correlations are subject to the no-signaling conditions, Eq. (6), representing fine-tuning conditions for the causal diagram [54,55]. Furthermore, if we condition on a specific value  $\lambda$ , one of the three parties factorizes with the other two; see, e.g., Fig. 2(b). Again, the postselection of the events for which  $D_k = d_k$  is indicated as boxes around the variables  $D_k$ . Similar to the conditions Eqs. (2) and (3) for nonlocality in the bipartite case, there are conditions on the postselected statistics  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  that, if fulfilled, validate the postselection for a GMN demonstration [50]. The first condition is that if  $p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda}$  factorizes in a specific way for a given  $\lambda$ , e.g., into two groups of  $k$  and  $N - k$  parties as Eq. (5)

for  $k = 1$ , then the probabilities  $p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda}$  must factorize in the same way. This can be shown directly with the  $d$ -separation rules: For instance, in the case of Fig. 2(b), every possible path that connects  $A_3$  and  $X_3$  to Alice's and Bob's settings and outcomes passes through  $\Lambda$  and is thus blocked because  $\Lambda$  is a noncollider that is conditioned on.

Second, we have to show the condition  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d}}$ , similar to Eq. (2). Here, as in Sec. II A, we must use that a constant number of particles is distributed among the parties, and that we have ideal number-resolving detectors. Thus, the number of detected particles of the  $k$ th measurement station is determined by the number of detected particles at the other stations, i.e., it can be written as a function  $d_k = f_k(\mathbf{d}\setminus d_k)$ . For instance, for three particles distributed among three parties, we have  $d_1 = 3 - d_2 - d_3$ .

To prove that, e.g.,  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d},\mathbf{x}\setminus x_1}$ , we first calculate that

$$p_{\lambda,\mathbf{d},\mathbf{x}} = p_{d_1|\mathbf{d}\setminus d_1,\mathbf{x},\lambda} p_{\mathbf{d}\setminus d_1|\mathbf{x},\lambda} p_{\mathbf{x},\lambda} \quad (7)$$

$$= \delta_{d_1,f_1(\mathbf{d}\setminus d_1)} p_{\mathbf{d}\setminus d_1|\mathbf{x}\setminus x_1,\lambda} p_{\mathbf{x},\lambda}, \quad (8)$$

where we have used that  $d_1 = f_1(\mathbf{d}\setminus d_1)$  and the no-signaling principle Eq. (6). Thus, we obtain

$$p_{\lambda|\mathbf{d},\mathbf{x}} = \frac{p_{\lambda,\mathbf{d},\mathbf{x}}}{p_{\mathbf{d},\mathbf{x}}} \quad (9)$$

$$= \frac{\delta_{d_1,f_1(\mathbf{d}\setminus d_1)} p_{\mathbf{d}\setminus d_1|\mathbf{x}\setminus x_1,\lambda} p_{\mathbf{x},\lambda}}{\int d\lambda \delta_{d_1,f_1(\mathbf{d}\setminus d_1)} p_{\mathbf{d}\setminus d_1|\mathbf{x}\setminus x_1,\lambda} p_{\mathbf{x},\lambda}} \quad (10)$$

$$= p_{\lambda|\mathbf{d},\mathbf{x}\setminus x_1}, \quad (11)$$

where, in the last line, we have used the free-choice assumption  $p_{\mathbf{x},\lambda} = p_{\mathbf{x}} p_{\lambda}$  to see that the dependence on the setting  $x_1$  cancels. Similarly, one can remove the influence of any  $X_k$  on  $\Lambda$  when conditioning on  $\mathbf{D} = \mathbf{d}$ , and we obtain  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d}}$ .

We note that  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d}}$  can also be shown using the  $d$ -separation rules, together with the no-signaling conditions and that the number of detected particles is conserved. However, these additional conditions on the causal diagram have the effect that using the  $d$ -separation rules does not provide a faster way to demonstrate  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d}}$  than a direct application of the conditions. For instance, to prove that  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d},\mathbf{x}\setminus x_1}$ , we must check all possible paths that connect  $X_1$  and  $\Lambda$  when conditioning on  $\mathbf{D} = \mathbf{d}$  and  $\mathbf{X}\setminus X_1 = \mathbf{x}\setminus x_1$ . As can be seen in Fig. 2(b), all such paths that are potentially open begin as  $X_1 \rightarrow D_1 \dots$ ,  $X_1 \rightarrow A_1 \rightarrow D_1 \dots$ ,  $X_1 \rightarrow A_1 \rightarrow A_k \dots$ , or  $X_1 \rightarrow A_1 \rightarrow D_k \dots$  for  $k \neq 1$ . The last two possibilities are blocked due to the no-signaling principle, Eq. (6). The first two possibilities instead appear open because of the possible influence from  $X_1$  to  $D_1$ . To show that there can be no such influence, we must employ that the number of detected particles is conserved, yielding that

$$p_{d_1|\mathbf{x},\mathbf{d}\setminus d_1} = \delta_{d_1,f_1(\mathbf{d}\setminus d_1)}. \quad (12)$$

Since the right-hand side is independent of  $X_1$ ,  $X_1$  has no influence on  $D_1$  if conditioned on  $\mathbf{D}\setminus D_1 = \mathbf{d}\setminus D_1$  and  $\mathbf{X}\setminus X_1 = \mathbf{x}\setminus x_1$ . We thus observe that  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d},\mathbf{x}\setminus x_1}$ . Similarly, one can show the independence of the other settings  $X_k$  for  $k > 1$ , and one obtains  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda|\mathbf{d}}$ .

To conclude this section, we want to compare our result to Refs. [49,50]. Reference [49] shows that a postselection that can be equivalently decided when excluding any of the

$N$  parties is valid for the demonstration of  $N$ -partite nonlocality. This can be applied to a coincidence postselection in an ideal scenario as considered here: Any party can be excluded in the postselection decision because its number of detected particles is determined by the number of the detected particles of the remaining parties. However, Ref. [49] cannot be applied to demonstrations of GMN. In Ref. [50], it was shown that, in an  $N$ -partite scenario, a collective postselection that can be equivalently decided when excluding any half of the parties is valid to demonstrate genuine  $N$ -partite nonlocality. The coincidence postselection in the noiseless case considered here admits an additional structure such that it is valid for demonstrating genuine  $N$ -partite nonlocality even if all but one parties have to be included in the postselection: The coincidence postselection can be written by conditioning on  $N$  local variables  $D_k$ , which is not possible for a general collective postselection.

### III. COINCIDENCE POSTSELECTION WITH INEFFICIENT DETECTORS: THRESHOLD INEFFICIENCIES

We now consider the  $N$ -partite Bell scenario case with nonideal detectors and transmission losses. In this case, the number of detected particles is not conserved and we cannot use the reasoning of the previous section. In particular, we cannot follow that the parties measurement settings have no influence on the postselection. However, one can limit the strength of this influence using measurable quantities. As causal diagrams make no statement about the strength of the causal influences, they are not useful here. Instead, to close the detection loophole, one can either include nondetection events in the statistics [6,18–20], i.e., one does not discard any results, or one can postselect on coincidence events but must sharpen the corresponding Bell inequalities accordingly [21–23]. Both approaches lead to threshold efficiencies that must be surpassed in the experiment to demonstrate nonlocality [19,21–23,30–36]. In the following, we derive threshold efficiencies that also apply for experiments in which the particles are randomly distributed [20,47,48], and for Bell inequalities that demonstrate GMN. Typically, even if we assume perfect detectors and no losses in the setup, undesired events occur with high probability. For instance, in the ideal three-partite YS setup [48], the desired events occur only with a probability of  $p = 1/4$  and the remaining events show no multipartite correlations (because one of the parties receives no particle). Thus, when including all events, Bell inequalities that test for GMN are not violated, even in an ideal setup. We therefore take the second approach of sharpening the Bell inequality and postselecting the desired events.

A general Bell inequality in the  $N$ -partite scenario can be written as

$$\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{x}} \leq I, \quad (13)$$

where  $c_{\mathbf{a},\mathbf{x}}, I \in \mathbb{R}$ , and the  $k$ th party can choose from  $M_k$  different measurement settings  $x_k \in \{1, \dots, M_k\}$  and observes the outcome  $a_k$  from a finite set of possible outcomes. Note that this form includes Bell inequalities that, if violated, demonstrate multipartite nonlocality [56] and GMN [37].

As in Sec. II, we consider a coincidence postselection, i.e., the  $k$ th party additionally has the variable  $D_k$ , the number of detected particles, where  $D_k = d_k$  denotes the detection of a single particle (or, more generally, the desired number of particles). We thus want to postselect the events for which  $\mathbf{D} = \mathbf{d}$ , so we are left with the probabilities  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$ . Since the probability of observing  $\mathbf{d}$  may depend on the measurement settings  $\mathbf{x}$ , the distribution of the LHV  $\Lambda$  of each summand of inequality (13) generally depends on  $\mathbf{x}$  as well. Thus, the Bell inequality is generally not valid for the postselected statistics  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  without further assuming fair sampling [8–11] (see Appendix A for further details and explanations).

We now follow the approach by Larsson [22] to sharpen the multipartite Bell inequalities using a measurable detection efficiency. In particular, for perfect detectors and no transmission losses, and if the number of distributed particles is constant, the sharpened Bell inequality should converge to the initial Bell inequality (13). In this case, due to continuity, there is some threshold detection efficiency above which the sharpened Bell inequality can be violated by quantum mechanics (assuming there are quantum states that violate the initial Bell inequality). Similarly to Ref. [22], we sharpen the Bell inequality using the minimal conditional detection efficiency

$$\eta_c = \min_{k,\mathbf{x}} p_{d_k|\mathbf{d}\setminus d_k,\mathbf{x}} = \min_{k,\mathbf{x}} \frac{p_{\mathbf{d}|\mathbf{x}}}{p_{\mathbf{d}\setminus d_k|\mathbf{x}}}. \quad (14)$$

The efficiency  $\eta_c$  corresponds to the minimal probability of the detection of a single particle by the  $k$ th party, given that all other parties detect a single particle and given the measurement settings  $\mathbf{x}$ , minimized over the party  $k$  and all possible settings  $\mathbf{x}$ .

We emphasize why it is crucial to use the conditional detection efficiency  $\eta_c$  to sharpen the Bell inequality if we want to obtain a useful result for setups with a random distribution of particles per party [20,47,48]. This is because, in the limit of perfect detectors, we find that  $\eta_c = 1$ , while a detection efficiency such as  $\eta_{\text{coincidence}} = \min_{\mathbf{x}} p_{\mathbf{d}|\mathbf{x}}$ , that in the standard scenario (one particle per party) yields  $\eta_{\text{coincidence}} = 1$ , would have a lower value (e.g.,  $\eta_{\text{coincidence}} = 1/4$  in the ideal YS scenario with  $N = 3$  [48]). Thus, even in the ideal YS setup, quantum mechanics could not violate the sharpened Bell inequality if the threshold efficiency  $\eta_{\text{coincidence}}^*$  is larger than  $1/4$ .

The threshold conditional efficiency  $\eta_c^*$  depends on the Bell inequality of interest, e.g., on the number of parties  $N$  and on the number of measurement settings  $M_k$ . Furthermore, the method of how to sharpen the Bell inequality differs if one assumes an underlying LHV model (multipartite nonlocality) or an underlying HLNHV model (GMN). In the case of an underlying LHV model, one can directly generalize the proof of Ref. [22] and find the sharpened Bell inequality

$$\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x}} \leq C + (I - C) \left[ 1 - \frac{1 - \eta_c}{\eta_c} \left( \sum_k M_k - N \right) \right], \quad (15)$$

where we defined  $C = \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}|$ . Inequality (15) is demonstrated in Appendix B and reduces to the results of

TABLE I. Threshold conditional efficiency  $\eta_c^*$  for the demonstration of nonlocality for  $N = 2$  (CHSH) and  $N = 3$  (Mermin), and for the demonstration of GMN for  $N = 3$  (Svetlichny). In the standard setup where each party receives one particle (and assuming noiseless transmission), one has a threshold detection efficiency  $\eta_{\text{det}}^* = \eta_c^*$ . In the Yurke-Stoler (YS) setup [47,48], assuming perfect transmission and  $\eta_{1|2} = 2\eta_{\text{det}}(1 - \eta_{\text{det}})$ , one finds the  $\eta_{\text{det}}^*$  listed in the right column.

Bell inequality	$\eta_c^*$	$\eta_{\text{det}}^*$ in YS setup
CHSH [8]	0.83 [18,21,22]	0.91 [20]
Mermin [56]	0.75 [33]	0.9
Svetlichny [37]	0.967	0.989

Ref. [22] for the corresponding Bell inequalities. We note that, for  $\eta_c = 1$ , we recover the original Bell inequality (13).

In the case of an underlying HLNHV model, the technique of Ref. [22] cannot be applied, but one can still demonstrate the sharpened Bell inequality (see Appendix C for a detailed derivation)

$$\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x}} \leq I + 4CN \frac{(1 - \eta_c)}{\eta_c}, \quad (16)$$

which, again, for  $\eta_c = 1$ , reduces to the original Bell inequality (13). We note that, in inequality (16), one can slightly optimize the sharpened Bell inequality by using an optimized  $C_{\text{opt}}$  instead of  $C^4$ ; see Appendix C.

Using the maximal value  $I_Q$  predicted by quantum mechanics that can be reached for the left-hand sides of inequalities (15) and (16), one obtains a threshold conditional efficiency  $\eta_c^*$ . For experiments with  $\eta_c > \eta_c^*$ , one can thus potentially demonstrate (genuine) multipartite nonlocality while closing the detection loophole. We emphasize that our results are derived in a general setting. For specific Bell inequalities, there may be more specialized approaches that yield smaller  $\eta_c^*$ ; see, e.g., Ref. [33] for the Mermin inequality that we discuss below. In this context, we note that the main objective of this work is to find some threshold conditional efficiency  $\eta_c^* < 1$  for Bell inequalities that certify GMN, such that the results can be applied to setups with a random distribution of particles among the parties [48].

We finally note that one often has an inequality of the form

$$\sum_{\mathbf{x}} \tilde{c}_{\mathbf{x}} \left\langle \prod_{k=1}^N A_k \right\rangle_{\mathbf{x}} \leq I, \quad (17)$$

where  $\langle \prod_{k=1}^N A_k \rangle_{\mathbf{x}} = \sum_{\mathbf{a}} a_1 a_2 \cdots a_N p_{\mathbf{a}|\mathbf{x}}$ , for instance, the CHSH inequality [8] for  $N = 2$ , and the Mermin inequality [56] and Svetlichny inequality [37] for  $N = 3$ . In this case, one has  $c_{\mathbf{a},\mathbf{x}} = a_1 a_2 \cdots a_N \tilde{c}_{\mathbf{x}}$ . Usually, the results are binary,  $a_k \in \{-1, 1\}$ , and thus  $\max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}| = |\tilde{c}_{\mathbf{x}}|$ .

<sup>4</sup>The optimized  $C_{\text{opt}}$  is defined as  $C_{\text{opt}} = \min_{\mathbf{y}} \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}| D(\mathbf{x}, \mathbf{y}) / N$ , where  $D$  is a discrete distance defined as  $D(\mathbf{x}, \mathbf{y}) = \sum_k \delta_{x_k, y_k}$ . Since  $D(\mathbf{x}, \mathbf{y}) \leq N$ , we have  $C_{\text{opt}} \leq C$ .

### A. Application to standard Bell scenarios

We now discuss our results for different Bell experiments with  $N = 2$  and  $N = 3$  parties, as summarized in the second column of Table I. In the bipartite case, we consider the CHSH inequality [8], and for  $N = 3$  we consider the Mermin inequality [56] for three-partite nonlocality and the Svetlichny inequality [37] for genuine three-partite nonlocality. For the CHSH inequality, we have  $C = 4$ ,  $I = 2$ , and  $I_Q = 2\sqrt{2}$ . We obtain  $\eta_c^* = 2(\sqrt{2} - 1) \approx 0.83$ , similarly to Ref. [22]. We note that we could also use the sharpened Bell inequality (16) instead, yielding  $\eta_c^* = 8/(7 + \sqrt{2}) \approx 0.95$ . Thus, inequality (15) yields a smaller  $\eta_c^*$  than inequality (16). For the Mermin inequality, we have  $C = 4$ ,  $I = 2$ , and  $I_Q = 4$ , such that we find  $\eta_c^* = 3/4$ , similarly to Ref. [23]. Finally, for the Svetlichny inequality, we must use inequality (16), and with  $C_{\text{opt}} = 4$ ,  $I = 4$ , and  $I_Q = 4\sqrt{2}$  we find  $\eta_c^* = 12/(11 + \sqrt{2}) \approx 0.967$ .

In the standard Bell scenario, one particle is sent to each party. If one party does not detect its particle, this might be due to a nonideal detection efficiency  $\eta_{\text{det}}$ , or due to a loss in the transmission of the particles described by the transmission efficiency  $\eta_{\text{tra}}$ . Assuming that  $\eta_{\text{det}}$  and  $\eta_{\text{tra}}$  are the same for any party, one has  $\eta_c = (\eta_{\text{det}}\eta_{\text{tra}})^N / (\eta_{\text{det}}\eta_{\text{tra}})^{N-1} = \eta_{\text{det}}\eta_{\text{tra}}$ . In the case of  $\eta_{\text{tra}} = 1$ , we have thus  $\eta_{\text{det}}^* = \eta_c^*$  and we recover the results of Refs. [18,21,22] for  $N = 2$  and of Refs. [23,33] for  $N = 3$ . Note that, using a precertification of the presence of the particle in their respective measurement stations [57,58], one can use  $\eta_{\text{tra}} = 1$  even for noisy transmissions. For the  $N$ -partite Mermin inequality for  $N$  odd ( $C = 2^{(N-1)}$ ,  $I = 2^{(N-1)/2}$ , and  $I_Q = 2^{(N-1)}$ ), we find  $\eta_{\text{det}}^* = N/(N+1)$ , which is larger than the (optimal)  $\eta_{\text{det}}^* = N/(2N-2)$  found in Ref. [33] for  $N > 3$ . This is because, in the derivation of  $\eta_{\text{det}}^*$  of Ref. [33], further structure [in the form of Greenberger-Horne-Zeilinger (GHZ) correlations] is used, while, in our derivation, we specified no further information about the observed correlations or the Bell inequality.

### B. Application to the Yurke-Stoler scenario

Finally, we consider the  $N$ -partite YS setup [48] that distributes  $N$  independent particles among the  $N$  parties. The parties are arranged in a circular configuration, where each two neighboring parties share a single-particle source between them. In the ideal noiseless case, each particle ends up at either of the parties with a probability  $p = 1/2$ . Therefore, the probability of each party detecting a single particle is  $p = 2/2^N$ , and the quantum state corresponding to these events is a  $N$ -partite GHZ state that displays genuine  $N$ -partite nonlocality [50,59,60]. However, for the remaining events that occur with a probability of  $p = 1 - 2/2^N$ , at least one party does not receive a particle, such that these events show no GMN, and we must employ a coincidence postselection to violate the Bell inequalities.

Since  $N$  particles are shared among the  $N$  parties, if one party detects two particles, a second party does not receive a particle. We thus have  $\eta_c = 1$  in the noiseless case, and, using the coincidence postselection and the sharpened Bell inequality (16), we can demonstrate GMN using the appropriate Bell inequality [37,59,60]. This recovers the results of Sec. II. If we assume a constant transmission inefficiency  $\eta_{\text{tra}}$ ,

and that all detectors have the same detection efficiency  $\eta_{\text{det}}$  to detect an incoming particle and a probability  $\eta_{1|2}$  to detect a single particle if two particles arrive, one calculates that (see Appendix D)

$$\eta_c = \frac{2\eta_{\text{det}}\eta_{\text{tra}}}{2 + (N-1)[\eta_{\text{tra}}\eta_{1|2}/\eta_{\text{det}} + 2(1 - \eta_{\text{tra}})]}. \quad (18)$$

Note that in the ideal case, i.e.,  $\eta_{\text{det}} = \eta_{\text{tra}} = (1 - \eta_{1|2}) = 1$ , we have  $\eta_c = 1$ .

For the case of  $N = 2$  [ $\eta_c^* = 2(\sqrt{2} - 1)$ ], if we assume  $\eta_{\text{tra}} = 1$ , and that the particles are detected independently, i.e.,  $\eta_{1|2} = 2\eta_{\text{det}}(1 - \eta_{\text{det}})$ , we find that  $\eta_{\text{det}}^* = 4/(3 + \sqrt{2}) \approx 0.906$ , in accordance with Ref. [20]. For an experiment that only employs on-off detectors, i.e., detectors that cannot differentiate between one and two particles, then even in the noiseless case ( $\eta_{\text{det}} = \eta_{\text{tra}} = \eta_{1|2} = 1$ ), we obtain  $\eta_c = 2/3 < \eta_c^*$ . We thus find that if no number-resolving detectors are available, the detection loophole cannot be closed even in the noiseless bipartite scenario, and fair sampling must be assumed to demonstrate nonlocality [11].

Finally, for the three-partite case with  $\eta_{\text{tra}} = 1$  and  $\eta_{1|2} = 2\eta_{\text{det}}(1 - \eta_{\text{det}})$ , we obtain  $\eta_c = \eta_{\text{det}}/(3 - 2\eta_{\text{det}})$ . Thus, for the demonstration of three-partite nonlocality ( $\eta_c^* = 3/4$ ), we find  $\eta_{\text{det}}^* = 9/10$ . For the demonstration of genuine three-partite nonlocality [ $\eta_c^* = 12/(11 + \sqrt{2})$ ], we find  $\eta_{\text{det}}^* = 36/(35 + \sqrt{2}) \approx 0.989$ . We note that, using GMN Bell inequalities for  $N > 3$  [59,60], one obtains  $\eta_{\text{det}}^* < 1$  for any  $N > 3$ . Thus, we observe that genuine  $N$ -partite nonlocality can be created from  $N$  independent particle sources, even if nonideal detectors are used.

## IV. CONCLUSIONS

We have considered a coincidence postselection in Bell experiments, i.e., a postselection of measurement results for which each measurement party detects a single particle. For this postselection, we have shown how to close the detection loophole that is created due to the selection bias [7]. If the number of detected particles is constant (requiring an ideal noiseless experimental apparatus), we have shown how to use causal diagrams and  $d$ -separation rules, together with the no-signaling principle, to validate a coincidence postselection for the demonstration of nonlocality and genuine multipartite nonlocality (GMN). In a realistic experiment with nonideal detection efficiencies, we have shown how to sharpen the Bell inequalities for both nonlocality and GMN such that they are still valid for the postselected statistics. This results in threshold detection efficiencies that, if reached in experiments, enable a demonstration of nonlocality and GMN while closing the detection loophole. Finally, we have applied our results to the  $N$ -partite Yurke-Stoler (YS) setup [48] to demonstrate that genuine  $N$ -partite nonlocality can be created from  $N$  independent particle sources, even if nonideal detectors are employed.

## ACKNOWLEDGMENT

This work was supported by the European Commission through the H2020 QuantERA ERA-NET Cofund in Quantum Technologies project ‘‘MENTA’’.

### APPENDIX A: HIDDEN VARIABLE MODELS OF POSTSELECTED STATISTICS

Here, we discuss why, generally, the postselected statistics  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  do not fulfill the Bell inequality (13). The Bell inequality  $\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{x}} \leq I$  is proven by assuming that the probabilities  $p_{\mathbf{a}|\mathbf{x}}$  can be written as a hidden variable model  $p_{\mathbf{a}|\mathbf{x}} = \int d\lambda p_{\lambda} p_{\mathbf{a}|\mathbf{x},\lambda}$ , where  $p_{\mathbf{a}|\mathbf{x},\lambda}$  must factorize as Eq. (4) for a LHV model or as Eq. (5) for a HLNHV model. We can always write

$$p_{\mathbf{a}|\mathbf{d},\mathbf{x}} = \int d\lambda p_{\lambda|\mathbf{d},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda}. \quad (\text{A1})$$

The probabilities  $p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda}$  again factorize in the desired way: For instance, if in the HLNHV model for a specific value  $\lambda$ , we have that  $p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda} = p_{a_k, d_k|x_k, \lambda} p_{\mathbf{a} \setminus a_k, \mathbf{d} \setminus d_k | \mathbf{x} \setminus x_k, \lambda}$ , Eq. (5), one shows that

$$p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda} = \frac{p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda}}{p_{\mathbf{d}|\mathbf{x},\lambda}} \quad (\text{A2})$$

$$= \frac{p_{a_k, d_k|x_k, \lambda} p_{\mathbf{a} \setminus a_k, \mathbf{d} \setminus d_k | \mathbf{x} \setminus x_k, \lambda}}{p_{d_k|x_k, \lambda} p_{\mathbf{d} \setminus d_k | \mathbf{x} \setminus x_k, \lambda}} \quad (\text{A3})$$

$$= p_{a_k|d_k, x_k, \lambda} p_{\mathbf{a} \setminus a_k | \mathbf{d} \setminus d_k, \mathbf{x} \setminus x_k, \lambda}. \quad (\text{A4})$$

However, the distribution  $p_{\lambda|\mathbf{d},\mathbf{x}}$  of the hidden variable  $\Lambda$  generally depends on the setting  $\mathbf{x}$ , such that we cannot write  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}} = \int d\lambda q_{\lambda} p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda}$  for some distribution  $q_{\lambda}$  and, thus, we cannot prove the Bell inequality.

In the following sections, we will define a fixed distribution  $q_{\lambda}$  (the distribution  $q_{\lambda}^{(\text{MN})}$  for LHV models and the distribution  $q_{\lambda}^{(\text{GMN})}$  for HLNHV models) such that we can bound the difference between  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  and

$$p_{\mathbf{a}|\mathbf{d},\mathbf{x},q} = \int d\lambda q_{\lambda} p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda}, \quad (\text{A5})$$

for any measurement setting  $\mathbf{x}$ , using the experimentally measurable  $\eta_c$ , Eq. (14). We then find new Bell inequalities for the postselected statistics  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  by using the fact that the probabilities  $p_{\mathbf{a}|\mathbf{d},\mathbf{x},q}$ , being written in a setting-independent distribution  $q_{\lambda}$ , fulfill the original Bell inequality,

$$\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},q} \leq I. \quad (\text{A6})$$

We finally want to note that, in this context, one can easily see the effect of the fair sampling assumption  $p_{\mathbf{d}|\mathbf{x},\lambda} = p_{\mathbf{d}|\lambda}$  [9–11]. First note that the fair sampling assumption also implies that  $p_{\mathbf{d}|\mathbf{x}} = \int d\lambda p_{\lambda|\mathbf{x}} p_{\mathbf{d}|\mathbf{x},\lambda} = p_{\mathbf{d}}$ , where we have used the free will assumption  $p_{\mathbf{x},\lambda} = p_{\mathbf{x}} p_{\lambda}$ . Therefore, one finds that

$$p_{\lambda|\mathbf{d},\mathbf{x}} = \frac{p_{\lambda,\mathbf{d},\mathbf{x}}}{p_{\mathbf{d},\mathbf{x}}} = \frac{p_{\mathbf{d}|\mathbf{x},\lambda} p_{\mathbf{x},\lambda}}{p_{\mathbf{d}} p_{\mathbf{x}}} = \frac{p_{\mathbf{d}|\lambda} p_{\lambda}}{p_{\mathbf{d}}} = p_{\lambda|\mathbf{d}}. \quad (\text{A7})$$

Thus, the distributions  $p_{\lambda|\mathbf{d},\mathbf{x}}$  are independent of  $\mathbf{x}$  and the postselected statistics  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  fulfill the original Bell inequality.

### APPENDIX B: SHARPENED BELL INEQUALITIES FOR MULTIPARTITE NONLOCALITY

In this Appendix, we derive the sharpened Bell inequality (15) that can be used to demonstrate multipartite nonlocality. We thus consider an underlying LHV model, such that  $p_{\mathbf{a},\mathbf{d}|\mathbf{x},\lambda}$  factorizes as in Eq. (4). In the following, we generalize the approach of Larsson [22] to a general multipartite Bell scenario with  $N$  parties,  $M_k$  settings for the  $k$ th party and a finite number of possible outcomes for each party. In contrast to Ref. [22], we do not assume a deterministic LHV model. We note that, for a LHV model, this restriction can be made without loss of generality [3,61]. For the HLNHV model discussed in the next section, this restriction is generally not valid [54].

We first define the LHV distribution

$$q_{\lambda}^{(\text{MN})} = \frac{p_{\lambda} p_{\lambda}^{\text{prod}}}{p^{\text{prod}}}, \quad (\text{B1})$$

where we defined  $p_{\lambda}^{\text{prod}} = \prod_{k=1}^N \prod_{x_k=1}^{M_k} p_{d_k|x_k, \lambda}$  and  $p^{\text{prod}} = \int d\lambda p_{\lambda} p_{\lambda}^{\text{prod}}$ , and  $p_{\lambda}$  is the initial LHV distribution. Furthermore, as in Ref. [22], we define

$$\delta = \min_{\mathbf{x}} \frac{p^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x}}}. \quad (\text{B2})$$

Now, after introducing the notation  $p_{\lambda}^{\text{prod}|\mathbf{x}} = p_{\lambda}^{\text{prod}} / p_{\mathbf{d}|\mathbf{x},\lambda}$  and  $C = \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}|$ , and using the triangle inequality, we can calculate

$$\sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x}} = \delta \sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}} + \sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} (p_{\mathbf{a}|\mathbf{d},\mathbf{x}} - \delta p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}}) \quad (\text{B3})$$

$$\leq \delta \sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}} + \left| \sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} (p_{\mathbf{a}|\mathbf{d},\mathbf{x}} - \delta p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}}) \right| \quad (\text{B4})$$

$$\leq \delta I + \left| \sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}} \left( \frac{p^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x}}} - \delta \right) + \sum_{\mathbf{a},\mathbf{x}} c_{\mathbf{a},\mathbf{x}} \int d\lambda p_{\lambda} p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda} p_{\mathbf{d}|\mathbf{x},\lambda} \frac{1 - p_{\lambda}^{\text{prod}|\mathbf{x}}}{p_{\mathbf{d}|\mathbf{x}}} \right| \quad (\text{B5})$$

$$\leq \delta I + \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}| \left( \frac{p^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x}}} - \delta \right) \sum_{\mathbf{a}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}} + \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}| \int d\lambda p_{\lambda} p_{\mathbf{d}|\mathbf{x},\lambda} \frac{1 - p_{\lambda}^{\text{prod}|\mathbf{x}}}{p_{\mathbf{d}|\mathbf{x}}} \sum_{\mathbf{a}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda} \quad (\text{B6})$$

$$= \delta I + C \left( \frac{p^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x}}} - \delta \right) + C \left( 1 - \frac{p^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x}}} \right) \quad (\text{B7})$$

$$= C + (I - C)\delta. \quad (\text{B8})$$

In the third line, we have used that  $p_{\lambda|\mathbf{d},\mathbf{x}} = p_{\lambda} p_{\mathbf{d}|\mathbf{x},\lambda} / p_{\mathbf{d}|\mathbf{x}}$ . In the fourth line, we have used that  $\delta \leq p^{\text{prod}} / p_{\mathbf{d}|\mathbf{x}}$  and that  $|c_{\mathbf{a},\mathbf{x}}| \leq \max_{\mathbf{a}} |c_{\mathbf{a},\mathbf{x}}|$ . In the fifth line, we have used that  $\sum_{\mathbf{a}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}} = 1$  and  $\sum_{\mathbf{a}} p_{\mathbf{a}|\mathbf{d},\mathbf{x},\lambda} = 1$ .

Finally, we have to find an upper bound for  $\delta$  using the experimentally measurable  $\eta_c = \min_{k,\mathbf{x}} p_{\mathbf{d}|\mathbf{d}\setminus d_k,\mathbf{x}}$ , Eq. (14). We first derive some useful relations. Using the LHV factorization, Eq. (4), we calculate

$$p_{\lambda}^{\text{prod}\setminus\mathbf{x}} = \frac{p_{\lambda}^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x},\lambda}} = \prod_{k=1}^N \prod_{\substack{y_k=1 \\ y_k \neq x_k}}^{M_k} p_{d_k|y_k,\lambda} \geq \sum_{k=1}^N \sum_{\substack{y_k=1 \\ y_k \neq x_k}}^{M_k} p_{d_k|y_k,\lambda} - \left( \sum_k M_k - N \right) + 1, \quad (\text{B9})$$

where, in the last step, we have used that for  $p_i \in [0, 1]$  one has  $\prod_{i=1}^L p_i \geq \sum_{i=1}^L p_i - L + 1$ , which can be proven by induction over  $L$ : For  $L = 2$ , we have that  $p_1 + p_2 - 1 = p_1(1 - p_2) + p_2 - 1 + p_1 p_2 \leq 1 - p_2 + p_2 - 1 + p_1 p_2$ . Then, assuming that  $\prod_{i=1}^L p_i \geq \sum_{i=1}^L p_i - L + 1$  holds for some  $L$ , we find that

$$\prod_{i=1}^{L+1} p_i = p_{L+1} \prod_{i=1}^L p_i \geq p_{L+1} + \prod_{i=1}^L p_i - 1 \geq p_{L+1} + \sum_{i=1}^L p_i - L + 1 - 1 = \sum_{i=1}^{L+1} p_i - (L + 1) + 1. \quad (\text{B10})$$

Next, we calculate that, for any  $k$  and  $y_k$ ,

$$\frac{\int d\lambda p_{\lambda} p_{\mathbf{d}|\mathbf{x},\lambda} p_{d_k|y_k,\lambda}}{p_{\mathbf{d}|\mathbf{x}}} = \frac{\int d\lambda p_{\lambda} p_{\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k,\lambda} p_{d_k|x_k,\lambda} p_{d_k|y_k,\lambda}}{p_{\mathbf{d}|\mathbf{x}}} \quad (\text{B11})$$

$$\geq \frac{\int d\lambda p_{\lambda} p_{\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k,\lambda} (p_{d_k|x_k,\lambda} + p_{d_k|y_k,\lambda} - 1)}{p_{\mathbf{d}|\mathbf{x}}} \quad (\text{B12})$$

$$= 1 + \frac{p_{\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k}}{p_{\mathbf{d}|\mathbf{x}}} \frac{\int d\lambda p_{\lambda} p_{\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k,\lambda} (p_{d_k|y_k,\lambda} - 1)}{p_{\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k}} \quad (\text{B13})$$

$$\geq 1 + \frac{1}{\eta_c} (\eta_c - 1) \quad (\text{B14})$$

$$= 2 - \frac{1}{\eta_c}, \quad (\text{B15})$$

where, in the second line, we have used again that  $p_1 p_2 \geq p_1 + p_2 - 1$  for  $p_i \in [0, 1]$ , and, in the fourth line, we have used that  $p_{d_k|y_k,\lambda} \leq 1$  and that  $\eta_c \leq p_{\mathbf{d}|\mathbf{d}\setminus d_k,\mathbf{x}} = p_{\mathbf{d}|\mathbf{x}} / p_{\mathbf{d}\setminus d_k|\mathbf{x}\setminus x_k}$  for any  $k$  and  $\mathbf{x}$ .

Finally, we combine the results of Eqs. (B9) and (B11)–(B15) to derive that

$$\frac{p^{\text{prod}}}{p_{\mathbf{d}|\mathbf{x}}} = \frac{\int d\lambda p_{\lambda} p_{\mathbf{d}|\mathbf{x},\lambda} p_{\lambda}^{\text{prod}\setminus\mathbf{x}}}{p_{\mathbf{d}|\mathbf{x}}} \quad (\text{B16})$$

$$\geq \frac{\sum_{k=1}^N \sum_{y_k \neq x_k} \int d\lambda p_{\lambda} p_{\mathbf{d}|\mathbf{x},\lambda} p_{d_k|y_k,\lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \left( \sum_k M_k - N \right) + 1 \quad (\text{B17})$$

$$\geq \left( \sum_k M_k - N \right) \left( 2 - \frac{1}{\eta_c} \right) - \left( \sum_k M_k - N \right) + 1 \quad (\text{B18})$$

$$= 1 - \frac{1 - \eta_c}{\eta_c} \left( \sum_k M_k - N \right). \quad (\text{B19})$$

We therefore find that  $\delta \geq 1 - \frac{1 - \eta_c}{\eta_c} (\sum_k M_k - N)$  and, inserting this bound into Eq. (B8), we obtain the sharpened Bell inequality for LHV models, inequality (15).

### APPENDIX C: SHARPENED BELL INEQUALITIES FOR GENUINE MULTIPARTITE NONLOCALITY

Here, we derive the sharpened Bell inequality (16) that can be used for demonstrations of GMN. We thus consider an underlying HLNHV model, Eq. (5), such that we cannot use the LHV factorization structure of the coincidence detection probability, i.e., we cannot assume that  $p_{\mathbf{d}|\mathbf{x}} = \int d\lambda p_{\lambda} \prod_k p_{d_k|x_k,\lambda}$ . Therefore, we cannot use the previously defined HV distribution  $q_{\lambda}^{(\text{MN})}$  to approximate the  $p_{\mathbf{a}|\mathbf{d},\mathbf{x}}$  with  $p_{\mathbf{a}|\mathbf{d},\mathbf{x},q^{(\text{MN})}}$  using the conditional detection efficiency  $\eta_c$ . In particular, the derivation of Appendix B breaks down at Eqs. (B9) and (B11)–(B15). Instead, here we use the hidden variable distribution defined as

$$q_{\lambda}^{(\text{GMN})} = p_{\lambda|\mathbf{d},\mathbf{y}} = \frac{1}{p_{\mathbf{d}|\mathbf{y}}} p_{\lambda} p_{\mathbf{d}|\mathbf{y},\lambda}, \quad (\text{C1})$$



where  $\mathbf{y}$  is an arbitrary fixed measurement setting. Since  $q_\lambda^{(\text{GMN})}$  is independent on the measurement settings, the Bell inequality  $\sum_{\mathbf{a}, \mathbf{x}} c_{\mathbf{a}, \mathbf{x}} p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, q^{(\text{GMN})}} \leq I$  holds; see Eq. (A6).

To sharpen the Bell inequality, we first compute that, for any measurement setting  $\mathbf{x}$ ,

$$\int d\lambda p_\lambda \left| \frac{p_{\mathbf{d}|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \frac{p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}} \right| = \frac{1}{p_{\mathbf{d}|\mathbf{x}}} \int d\lambda p_\lambda \left| p_{\mathbf{d}|\mathbf{x}, \lambda} - p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} \frac{p_{\mathbf{d}|\mathbf{x}}}{p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}} \right| \quad (\text{C2})$$

$$\leq \frac{1}{p_{\mathbf{d}|\mathbf{x}}} \int d\lambda p_\lambda \left[ |p_{\mathbf{d}|\mathbf{x}, \lambda} - p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}| + p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} \left( 1 - \frac{p_{\mathbf{d}|\mathbf{x}}}{p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}} \right) \right] \quad (\text{C3})$$

$$\leq \frac{1}{p_{\mathbf{d}|\mathbf{x}}} \int d\lambda p_\lambda [p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} - p_{\mathbf{d}|\mathbf{x}, \lambda} + p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} (1 - \eta_c)] \quad (\text{C4})$$

$$\leq \frac{1}{p_{\mathbf{d}|\mathbf{x}}} [p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} - p_{\mathbf{d}|\mathbf{x}} + p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} (1 - \eta_c)] \quad (\text{C5})$$

$$\leq \frac{2(1 - \eta_c)}{\eta_c}. \quad (\text{C6})$$

In the second line we have used that, for  $a, b, c \in [0, 1]$ , it holds that  $|a - bc| = |a - b + b(1 - c)| \leq |a - b| + b(1 - c)$ . In the third line, we have used that  $p_{\mathbf{d}|\mathbf{x}, \lambda} = p_{d_k|\mathbf{d}, \mathbf{x}, \lambda} p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} < p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} = p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}$  [note the use of the no-signaling principle, Eq. (6)] and that  $\eta_c \leq p_{\mathbf{d}|\mathbf{x}}/p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}$ , which we again used twice in the last line.

It follows that for any two measurement settings  $\mathbf{x}$  and  $\tilde{\mathbf{x}}$  that differ only in the  $k$ th entry, one has

$$\int d\lambda p_\lambda \left| \frac{p_{\mathbf{d}|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \frac{p_{\mathbf{d}|\tilde{\mathbf{x}}, \lambda}}{p_{\mathbf{d}|\tilde{\mathbf{x}}}} \right| \leq \int d\lambda p_\lambda \left[ \left| \frac{p_{\mathbf{d}|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \frac{p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda}} \right| + \left| \frac{p_{\mathbf{d}|\tilde{\mathbf{x}}, \lambda}}{p_{\mathbf{d}|\tilde{\mathbf{x}}}} - \frac{p_{\mathbf{d}|\mathbf{d}_k|\tilde{\mathbf{x}}, \lambda}}{p_{\mathbf{d}|\mathbf{d}_k|\tilde{\mathbf{x}}, \lambda}} \right| \right] \leq \frac{4(1 - \eta_c)}{\eta_c}, \quad (\text{C7})$$

where we have used that  $p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} = p_{\mathbf{d}|\mathbf{d}_k|\tilde{\mathbf{x}}, \lambda}$  and  $p_{\mathbf{d}|\mathbf{d}_k|\mathbf{x}, \lambda} = p_{\mathbf{d}|\mathbf{d}_k|\tilde{\mathbf{x}}, \lambda}$  due to the no-signaling principle and  $\tilde{\mathbf{x}} \setminus \tilde{x}_k = \mathbf{x} \setminus x_k$ . Next, we find that for any measurement setting  $\mathbf{x}$ ,

$$\int d\lambda p_\lambda \left| \frac{p_{\mathbf{d}|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \frac{p_{\mathbf{d}|\mathbf{y}, \lambda}}{p_{\mathbf{d}|\mathbf{y}}} \right| \leq \int d\lambda p_\lambda \left[ \left| \frac{p_{\mathbf{d}|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \frac{p_{\mathbf{d}|\mathbf{x}_{D(\mathbf{x}, \mathbf{y})-1}, \lambda}}{p_{\mathbf{d}|\mathbf{x}_{D(\mathbf{x}, \mathbf{y})-1}}} \right| + \dots + \left| \frac{p_{\mathbf{d}|\mathbf{x}_1, \lambda}}{p_{\mathbf{d}|\mathbf{x}_1}} - \frac{p_{\mathbf{d}|\mathbf{y}, \lambda}}{p_{\mathbf{d}|\mathbf{y}}} \right| \right] \leq \frac{4D(\mathbf{x}, \mathbf{y})(1 - \eta_c)}{\eta_c}, \quad (\text{C8})$$

where  $D$  is a discrete distance defined as  $D(\mathbf{x}, \mathbf{y}) = \sum_k \delta_{x_k, y_k}$ , and we have used the sequence  $\mathbf{x}_i$  [ $i = 0, \dots, D(\mathbf{x}, \mathbf{y}) - 1$ ] starting from  $\mathbf{x}_0 = \mathbf{y}$ , where  $\mathbf{x}_i$  is obtained from  $\mathbf{x}_{i-1}$  by changing the  $k_i$ th component from  $y_{k_i}$  to  $x_{k_i}$ , where  $k_i$  is the  $i$ th entry where  $\mathbf{x}$  and  $\mathbf{y}$  differ. Thus, any  $\mathbf{x}_i$  and  $\mathbf{x}_{i-1}$  only differ in only one entry, which also holds for  $\mathbf{x}_{D(\mathbf{x}, \mathbf{y})-1}$  and  $\mathbf{x}$ , and for  $\mathbf{x}_1$  and  $\mathbf{y}$ .

We now finally obtain

$$\sum_{\mathbf{a}, \mathbf{x}} c_{\mathbf{a}, \mathbf{x}} p_{\mathbf{a}|\mathbf{d}, \mathbf{x}} = \sum_{\mathbf{a}, \mathbf{x}} c_{\mathbf{a}, \mathbf{x}} p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, q^{(\text{GMN})}} + \sum_{\mathbf{a}, \mathbf{x}} c_{\mathbf{a}, \mathbf{x}} (p_{\mathbf{a}|\mathbf{d}, \mathbf{x}} - p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, q^{(\text{GMN})}}) \quad (\text{C9})$$

$$\leq \sum_{\mathbf{a}, \mathbf{x}} c_{\mathbf{a}, \mathbf{x}} p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, q^{(\text{GMN})}} + \left| \sum_{\mathbf{a}, \mathbf{x}} c_{\mathbf{a}, \mathbf{x}} (p_{\mathbf{a}|\mathbf{d}, \mathbf{x}} - p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, q^{(\text{GMN})}}) \right| \quad (\text{C10})$$

$$\leq I + \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a}, \mathbf{x}}| \int d\lambda p_\lambda \left| \frac{p_{\mathbf{d}|\mathbf{x}, \lambda}}{p_{\mathbf{d}|\mathbf{x}}} - \frac{p_{\mathbf{d}|\mathbf{y}, \lambda}}{p_{\mathbf{d}|\mathbf{y}}} \right| \sum_{\mathbf{a}} p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, \lambda} \quad (\text{C11})$$

$$\leq I + 4 \frac{1 - \eta_c}{\eta_c} \sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a}, \mathbf{x}}| D(\mathbf{x}, \mathbf{y}). \quad (\text{C12})$$

In the third line, we have again used that  $|c_{\mathbf{a}, \mathbf{x}}| \leq \max_{\mathbf{a}} |c_{\mathbf{a}, \mathbf{x}}|$ , and in the last line, we have used Eq. (C8) and that  $\sum_{\mathbf{a}} p_{\mathbf{a}|\mathbf{d}, \mathbf{x}, \lambda} = 1$ . This bound can be optimized by a minimization of  $\sum_{\mathbf{x}} \max_{\mathbf{a}} |c_{\mathbf{a}, \mathbf{x}}| D(\mathbf{x}, \mathbf{y})$  over  $\mathbf{y}$ . Finally, if we use that  $D(\mathbf{x}, \mathbf{y}) \leq N$ , we obtain the sharpened Bell inequality for GMN, Eq. (16).

#### APPENDIX D: DETECTION PROBABILITIES IN THE YURKE-STOLER SETUP

In this section, we derive the detection probabilities and the conditional detection efficiency in the  $N$ -partite Yurke-Stoler (YS) setup [48]. First, to compute the probabilities of different particle distributions at the measurement parties, we consider the simplified version of the YS setup in which each the  $k$ th party only measures the number of incoming particles  $D_k$ . In this setup, there are  $N$  single-particle sources

$S_k$  that are arranged in a circular configuration. The particle created at  $S_k$  is sent in an equal superposition to the  $(k - 1)$ th and the  $k$ th measurement party (the particle created at  $S_1$  is divided between the first and the  $N$ th party), e.g., using a beam splitter if the particles are photons. The  $k$ th party then measures the number of particles, labeled as  $D_k$ . This setup is sketched in Fig. 3. We note that to generate nonlocality, the  $k$ th party must also measure a second observable  $A_k$ ; see Fig. 2.

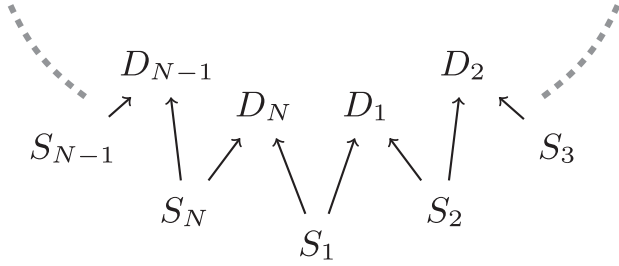


FIG. 3. Sketch of the particle distribution in the  $N$ -partite Yurke-Stoler (YS) scenario [48]. Each of  $N$  single-particle sources  $S_k$  distributes its particle with equal probability (in an equal superposition) to the  $(k-1)$ th and  $k$ th measurement parties ( $S_1$  is distributed to the first and  $N$ th party). The  $k$ th party measures the number of incoming particles  $D_k$ , while for a nonlocality experiment, the  $k$ th party should also measure a second observable  $A_k$ ; see Fig. 2.

If the particle sent from the  $k$ th source is detected by the  $(k-1)$ th party, we indicate this configuration as  $S_k = l$  (left), and  $S_k = r$  (right) otherwise. For a coincidence detection, i.e.,  $D_k = 1$  for all  $k$ , we have the two possible configurations  $\mathbf{S} = (l, l, \dots, l)$  and  $\mathbf{S} = (r, r, \dots, r)$ . Since there are  $2^N$  configurations in total that all have the same probability, the probability for a coincidence in the ideal setup with no losses is

$$p_{\text{coin}} = \frac{2}{2^N}. \quad (\text{D1})$$

To calculate  $\eta_c$ , Eq. (14), we must also compute  $p_{\mathbf{d} \setminus d_k | \mathbf{x}}$ , where we note that in the YS setup, we have  $p_{\mathbf{d} \setminus d_k | \mathbf{x}} = p_{\mathbf{d} \setminus d_k}$  so we discard the measurement settings in the following. The possible configurations that can contribute to this event are a configuration where all parties receive a particle and a configuration where the  $k$ th party receives no particle and one of the other  $(k-1)$  parties receives two (but only detects one). Therefore, we must calculate the probability  $p_{m^2, \bar{k}}$  that the  $m$ th party receives two particles and the  $k$ th party receives none, while the remaining parties each receive one particle. If, e.g., the first party receives two particles,  $m = 1$ , we must have had a configuration with  $S_1 = r$  and  $S_2 = l$ ; see Fig. 3.

For the  $k$ th party receiving no photon, we must have  $S_k = l$  and  $S_{k+1} = r$ . Finally, for the remaining parties to detect a single photon, we must have that  $S_n = l$  for  $n = 3, \dots, k-1$ , and  $S_n = r$  for  $n = k+2, \dots, N$ . A similar reasoning holds also for  $m \neq 1$ . Thus, there is only one configuration that contributes to the probability  $p_{m^2, \bar{k}}$  and, in the ideal noiseless case, we have

$$p_{m^2, \bar{k}} = \frac{1}{2^N}. \quad (\text{D2})$$

Finally, we can compute  $p_{\mathbf{d}}$  and  $p_{\mathbf{d} \setminus d_k}$  when including a finite transmission efficiency  $\eta_{\text{tra}}$ , a finite single-particle detection efficiency  $\eta_{\text{det}}$ , and a probability  $\eta_{1|2}$  for the detection of only one particle if two particles are received. Since a coincidence detection can only occur if every party receives and detects a single particle, we have

$$p_{\mathbf{d}} = p_{\text{coin}} \eta_{\text{det}}^N \eta_{\text{tra}}^N. \quad (\text{D3})$$

Next, to observe  $\mathbf{d} \setminus d_k$  could first have a single particle per party ( $p = p_{\text{coin}}$ ), and all  $N-1$  parties detect their particle with probability  $p = \eta_{\text{det}}^{N-1} \eta_{\text{tra}}^{N-1}$ . Second, one could have that the  $k$ th party receives no particle and the  $m$ th ( $m \neq k$ ) party receives two while the remaining parties receive and detect a single particle ( $p = p_{m^2, \bar{k}} \eta_{\text{det}}^{N-2} \eta_{\text{tra}}^{N-2}$ ). Furthermore, the  $m$ th party must only detect a single particle, which may happen because one particle is lost in transmission and the other one is detected [ $p = 2\eta_{\text{det}} \eta_{\text{tra}} (1 - \eta_{\text{tra}})$ ], or both particles are received but only one is detected ( $p = \eta_{\text{tra}}^2 \eta_{1|2}$ ). Since there are  $N-1$  possibilities for  $m$  (that are all equally probable), we obtain

$$p_{\mathbf{d} \setminus d_k} = p_{\text{coin}} \eta_{\text{det}}^{N-1} \eta_{\text{tra}}^{N-1} + (N-1) p_{m^2, \bar{k}} \eta_{\text{det}}^{N-2} \eta_{\text{tra}}^{N-2} \times [\eta_{\text{tra}}^2 \eta_{1|2} + 2\eta_{\text{det}} \eta_{\text{tra}} (1 - \eta_{\text{tra}})] \quad (\text{D4})$$

After some simplifications, we find Eq. (18) of the main text,

$$\eta_c = \frac{p_{\mathbf{d}}}{p_{\mathbf{d} \setminus d_k}} = \frac{2\eta_{\text{det}} \eta_{\text{tra}}}{2 + (N-1)[\eta_{\text{tra}} \eta_{1|2} / \eta_{\text{det}} + 2(1 - \eta_{\text{tra}})]}. \quad (\text{D5})$$

- 
- [1] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Phys. Phys. Fiz.* **1**, 195 (1964).
- [2] J. S. Bell, in *Speakable and Unsayable in Quantum Mechanics: Collected Papers on Quantum Philosophy*, 2nd ed. (Cambridge University Press, Cambridge, 2004), pp. 52–62.
- [3] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
- [4] A. Aspect, Proposed experiment to test the nonseparability of quantum mechanics, *Phys. Rev. D* **14**, 1944 (1976).
- [5] P. M. Pearle, Hidden-variable example based upon data rejection, *Phys. Rev. D* **2**, 1418 (1970).
- [6] J. F. Clauser and M. A. Horne, Experimental consequences of objective local theories, *Phys. Rev. D* **10**, 526 (1974).
- [7] J. Pearl, *Causality: Models, Reasoning, and Inference* (Cambridge University Press, Cambridge, 2009).
- [8] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed Experiment to Test Local Hidden-Variable Theories, *Phys. Rev. Lett.* **23**, 880 (1969).
- [9] D. W. Berry, H. Jeong, M. Stobińska, and T. C. Ralph, Fair-sampling assumption is not necessary for testing local realism, *Phys. Rev. A* **81**, 012109 (2010).
- [10] D. Orsucci, J.-D. Bancal, N. Sangouard, and P. Sekatski, How post-selection affects device-independent claims under the fair sampling assumption, *Quantum* **4**, 238 (2020).
- [11] V. Gebhart and A. Smerzi, Extending the fair sampling assumption using causal diagrams, [arXiv:2207.09348](https://arxiv.org/abs/2207.09348).
- [12] D. S. Tasca, S. P. Walborn, F. Toscano, and P. H. Souto Ribeiro, Observation of tunable Popescu-Rohrlich correlations through postselection of a Gaussian state, *Phys. Rev. A* **80**, 030101 (2009).

- [13] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov, and C. Kurtsiefer, Experimentally Faking the Violation of Bell's Inequalities, *Phys. Rev. Lett.* **107**, 170404 (2011).
- [14] E. Pomarico, B. Sanguinetti, P. Sekatski, H. Zbinden, and N. Gisin, Experimental amplification of an entangled photon: what if the detection loophole is ignored?, *New J. Phys.* **13**, 063031 (2011).
- [15] J. Romero, D. Giovannini, D. S. Tasca, S. M. Barnett, and M. J. Padgett, Tailored two-photon correlation and fair-sampling: A cautionary tale, *New J. Phys.* **15**, 083047 (2013).
- [16] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Hacking commercial quantum cryptography systems by tailored bright illumination, *Nat. Photon.* **4**, 686 (2010).
- [17] J. Jogenfors, A. M. Elhassan, J. Ahrens, M. Bourennane, and J. Åke Larsson, Hacking the Bell test using classical light in energy-time entanglement-based quantum key distribution, *Sci. Adv.* **1**, e1500793 (2015).
- [18] N. D. Mermin, The EPR experiment—Thoughts about the “loophole”, *Ann. N.Y. Acad. Sci.* **480**, 422 (1986).
- [19] P. H. Eberhard, Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment, *Phys. Rev. A* **47**, R747 (1993).
- [20] F. Sciarrino, G. Vallone, A. Cabello, and P. Mataloni, Bell experiments with random destination sources, *Phys. Rev. A* **83**, 032112 (2011).
- [21] A. Garg and N. D. Mermin, Detector inefficiencies in the Einstein-Podolsky-Rosen experiment, *Phys. Rev. D* **35**, 3831 (1987).
- [22] J.-A. Larsson, Bell's inequality and detector inefficiency, *Phys. Rev. A* **57**, 3304 (1998).
- [23] J.-A. Larsson, Necessary and sufficient detector-efficiency conditions for the Greenberger-Horne-Zeilinger paradox, *Phys. Rev. A* **57**, R3145 (1998).
- [24] M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, Experimental violation of a Bell's inequality with efficient detection, *Nature (London)* **409**, 791 (2001).
- [25] D. N. Matsukevich, P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, Bell Inequality Violation with Two Remote Atomic Qubits, *Phys. Rev. Lett.* **100**, 150404 (2008).
- [26] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin, and P. G. Kwiat, Detection-Loophole-Free Test of Quantum Nonlocality, and Applications, *Phys. Rev. Lett.* **111**, 130406 (2013).
- [27] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor *et al.*, Strong Loophole-Free Test of Local Realism, *Phys. Rev. Lett.* **115**, 250402 (2015).
- [28] M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin *et al.*, Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons, *Phys. Rev. Lett.* **115**, 250401 (2015).
- [29] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. Vermeulen, R. N. Schouten, C. Abellán *et al.*, Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, *Nature (London)* **526**, 682 (2015).
- [30] S. Massar, Nonlocality, closing the detection loophole, and communication complexity, *Phys. Rev. A* **65**, 032121 (2002).
- [31] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig, Combinatorics and Quantum Nonlocality, *Phys. Rev. Lett.* **91**, 047903 (2003).
- [32] N. Brunner, N. Gisin, V. Scarani, and C. Simon, Detection Loophole in Asymmetric Bell Experiments, *Phys. Rev. Lett.* **98**, 220403 (2007).
- [33] A. Cabello, D. Rodríguez, and I. Villanueva, Necessary and Sufficient Detection Efficiency for the Mermin Inequalities, *Phys. Rev. Lett.* **101**, 120402 (2008).
- [34] T. Vértesi, S. Pironio, and N. Brunner, Closing the Detection Loophole in Bell Experiments Using Qudits, *Phys. Rev. Lett.* **104**, 060401 (2010).
- [35] R. Chaves and J. B. Brask, Feasibility of loophole-free nonlocality tests with a single photon, *Phys. Rev. A* **84**, 062110 (2011).
- [36] N. Miklin, A. Chaturvedi, M. Bourennane, M. Pawłowski, and A. Cabello, Exponentially Decreasing Critical Detection Efficiency for Any Bell Inequality, *Phys. Rev. Lett.* **129**, 230403 (2022).
- [37] G. Svetlichny, Distinguishing three-body from two-body nonseparability by a Bell-type inequality, *Phys. Rev. D* **35**, 3066 (1987).
- [38] J.-D. Bancal, C. Branciard, N. Gisin, and S. Pironio, Quantifying Multipartite Nonlocality, *Phys. Rev. Lett.* **103**, 090503 (2009).
- [39] J.-D. Bancal, J. Barrett, N. Gisin, and S. Pironio, Definitions of multipartite nonlocality, *Phys. Rev. A* **88**, 014102 (2013).
- [40] M. Hillery, V. Bužek, and A. Berthiaume, Quantum secret sharing, *Phys. Rev. A* **59**, 1829 (1999).
- [41] M. Epping, H. Kampermann, C. Macchiavello, and D. Bruß, Multi-partite entanglement can speed up quantum key distribution in networks, *New J. Phys.* **19**, 093012 (2017).
- [42] M. Pivoluska, M. Huber, and M. Malik, Layered quantum key distribution, *Phys. Rev. A* **97**, 032312 (2018).
- [43] J. Ribeiro, G. Murta, and S. Wehner, Fully device-independent conference key agreement, *Phys. Rev. A* **97**, 022307 (2018).
- [44] G. Murta, F. Grasselli, H. Kampermann, and D. Bruß, Quantum conference key agreement: A review, *Adv. Quantum Technol.* **3**, 2000025 (2020).
- [45] T. Holz, H. Kampermann, and D. Bruß, Genuine multipartite Bell inequality for device-independent conference key agreement, *Phys. Rev. Res.* **2**, 023251 (2020).
- [46] M. Proietti, J. Ho, F. Grasselli, P. Barrow, M. Malik, and A. Fedrizzi, Experimental quantum conference key agreement, *Sci. Adv.* **7**, eabe0395 (2021).
- [47] B. Yurke and D. Stoler, Bell's-inequality experiments using independent-particle sources, *Phys. Rev. A* **46**, 2229 (1992).
- [48] B. Yurke and D. Stoler, Einstein-Podolsky-Rosen Effects from Independent Particle Sources, *Phys. Rev. Lett.* **68**, 1251 (1992).
- [49] P. Blasiak, E. Borsuk, and M. Markiewicz, On safe post-

- selection for Bell tests with ideal detectors: Causal diagram approach, *Quantum* **5**, 575 (2021).
- [50] V. Gebhart, L. Pezzè, and A. Smerzi, Genuine Multipartite Nonlocality with Causal-Diagram Postselection, *Phys. Rev. Lett.* **127**, 140401 (2021).
- [51] M. L. Almeida, D. Cavalcanti, V. Scarani, and A. Acín, Multipartite fully nonlocal quantum states, *Phys. Rev. A* **81**, 052111 (2010).
- [52] R. Gallego, L. E. Würflinger, A. Acín, and M. Navascués, Operational Framework for Nonlocality, *Phys. Rev. Lett.* **109**, 070401 (2012).
- [53] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, *Found. Phys.* **24**, 379 (1994).
- [54] C. J. Wood and R. W. Spekkens, The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning, *New J. Phys.* **17**, 033002 (2015).
- [55] J.-M. A. Allen, J. Barrett, D. C. Horsman, C. M. Lee, and R. W. Spekkens, Quantum Common Causes and Quantum Causal Models, *Phys. Rev. X* **7**, 031021 (2017).
- [56] N. D. Mermin, Extreme Quantum Entanglement in a Superposition of Macroscopically Distinct States, *Phys. Rev. Lett.* **65**, 1838 (1990).
- [57] A. Cabello and F. Sciarrino, Loophole-Free Bell Test Based on Local Precertification of Photon's Presence, *Phys. Rev. X* **2**, 021010 (2012).
- [58] E. Meyer-Scott, D. McCloskey, K. Gołos, J. Z. Salvail, K. A. G. Fisher, D. R. Hamel, A. Cabello, K. J. Resch, and T. Jennewein, Certifying the Presence of a Photonic Qubit by Splitting It in Two, *Phys. Rev. Lett.* **116**, 070501 (2016).
- [59] D. Collins, N. Gisin, S. Popescu, D. Roberts, and V. Scarani, Bell-Type Inequalities to Detect True  $n$ -Body Nonseparability, *Phys. Rev. Lett.* **88**, 170405 (2002).
- [60] M. Seevinck and G. Svetlichny, Bell-Type Inequalities for Partial Separability in  $N$ -Particle Systems and Quantum Mechanical Violations, *Phys. Rev. Lett.* **89**, 060401 (2002).
- [61] A. Fine, Hidden Variables, Joint Probability, and the Bell Inequalities, *Phys. Rev. Lett.* **48**, 291 (1982).