Optimality of semiquantum nonlocality in the presence of high inconclusive rates

Charles Ci Wen Lim

Optimality of semi-quantum nonlocality
in the presence of high inconclusive rates

Charles Ci Wen Lim

Quantum Information Science Group, Computational Sciences and Engineering Division,
Oak Ridge National Laboratory, Oak Ridge, TN 37831-6418, USA.

Quantum nonlocality is a counterintuitive phenomenon that lies beyond the purview of causal influences. Recently, Bell inequalities have been generalized to the case of quantum inputs, leading to a powerful family of semi-quantum Bell inequalities that are capable of detecting any entangled state. Here, we focus on a different problem and investigate how the local-indistinguishability of quantum inputs and postselection may affect the requirements to detect semi-quantum nonlocality. To this end, we consider a semi-quantum nonlocal game based on locally-indistinguishable qubit inputs, and derive its postselected local and quantum bounds by using a novel connection to the local-distinguishability of quantum states. Interestingly, we find that the postselected local bound is independent of the measurement efficiency, and the achievable postselected Bell violation increases with decreasing measurement efficiency.

It is known that in quantum physics, there exist experiments in which correlations from measurements on entangled systems are at odds with our causal world views. These correlations may be verified by using a family of statistical tests called Bell inequalities [1, 2], which are linear constraints on the set of correlations that are compatible with the principle of local causes [3]. In other words, if the correlations violate a Bell inequality, then the underlying physics must be nonlocal in nature. Remarkably, apart from their foundational significance, Bell inequalities have also found practical applications in quantum cryptography and quantum state estimation [4–9]. For these reasons, quantum nonlocality is one of the most widely studied topics in quantum information science.

Recently, a new paradigm called *semi-quantum nonlocality* has emerged [10], where observers use quantum inputs—instead of classical inputs—to specify their desired measurement settings. Interestingly, by doing so, all entangled states are "nonlocal", in that for any entangled state there is always a semi-quantum Bell inequality with which violation can be achieved. This feature suggests that certain semi-quantum Bell inequalities are strong entanglement witnesses and thus could provide an unprecedented level of confidence in detecting entanglement using untrusted measurement devices. For instance, see Ref. [11] for a generic procedure that converts entanglement witnesses into measurement-device-independent entanglement witnesses, and Ref. [12] for the corresponding proof-of-principle experiment.

On a more general level, semi-quantum nonlocality also admits the possibility of working with *locally-indistinguishable* quantum inputs, a notion that is central to local quantum state discrimination [13–15] and quantum data hiding [16, 17]. For our purposes, we define such quantum inputs as quantum states that are indistinguishable at the level of local operations and shared randomness (LOSR) [10], but distinguishable at the level of local quantum measurements assisted with

shared entanglement (henceforth referred to as *quantum strategies*). In particular, our theoretical contribution recognizes that semi-quantum Bell inequalities using locally-indistinguishable quantum inputs may acquire the following two interesting properties: (1) the ability to safely perform postselection and (2) the ability to achieve higher Bell violations with decreasing measurement efficiencies [18].

The first property is based on the observation that postselection strategies due to the detection loophole [19–21] are local filtering processes assisted with shared randomness. Thus by the above definition, this means that it is impossible for LOSR models to produce postselected correlations that are semi-quantum nonlocal—even if arbitrarily low measurement efficiencies are allowed. The second property is due to the fact that the violation of a semi-quantum Bell inequality is directly related to the local-distinguishability of the quantum input states. This connection implies that with a suitable choice of quantum inputs, it is possible to devise a semi-quantum Bell inequality whose optimal violation is achieved only if the measurement efficiencies fall below a certain threshold; analogous to the optimal discrimination of nonorthogonal quantum states whereby inconclusive measurement elements are necessary [22].

To illustrate the above properties, we analyze a semi-quantum Bell experiment inspired by the Clauser-Horne-Shimony-Holt (CHSH) Bell experiment [23], and derive its postselected local bound and postselected maximum quantum bound for a given measurement efficiency. To start with, let us first clarify the meaning of using quantum states to choose the measurements. While it is clear what it means by classically choosing a measurement setting (e.g., turning a knob), in the case of quantum inputs, the notion of choosing a measurement is somewhat less obvious. To sharpen this notion, we propose to think in terms of programmable quantum measurement (PQM) devices [24, 25]. More specifically, a PQM device is a measurement device that accepts

two quantum inputs, namely a quantum target system and a quantum program system, and then performs a measurement (determined by the state of the program system) on the target system. That is, one uses the state of the quantum program system to choose the desired measurement. Thus we may view measurement devices in the semi-quantum nonlocality framework as *untrusted* PQM devices whose measurements are purportedly determined by *trusted* quantum input systems, i.e., see Fig. (1).

Semi-quantum CHSH inequality. We consider a semi-quantum Bell experiment involving two distant observers, called Alice and Bob, who each have a trusted local source of randomness, a trusted qubit preparation device, and an untrusted PQM device. Note that the measurement-independence condition [26] is thus assumed. In each run of the experiment, Alice generates two random bits $\bar{x} = x_1 x_2$ and prepares a program qubit using the following encoding scheme: $|\bar{x}\rangle\langle\bar{x}| = H^{x_1}|x_2\rangle\langle x_2|H^{x_1}$, where $\{|x_2\rangle\}_{x_2=0,1}$ is the computational basis and H is the Hadamard matrix. Then, she sends the prepared qubit to her PQM device for measurement and receives an outcome $a \in \{0, 1, \varnothing\}$, where all inconclusive outcomes are assigned to $\varnothing$. Likewise for Bob, we write $\bar{y} = y_1 y_2$ and b to denote his measurement choice and measurement outcome, respectively. Furthermore, in what follows, we will refer to Alice's and Bob's qubit input systems as A and B, respectively, and their corresponding quantum target systems as A′ and B′.

In the LOSR framework, the untrusted measurements are modeled by a classical distribution $\{\Pr[\lambda]\}_\lambda$ and a corresponding set of conditional local positive-operator valued measure (POVM) operators, $\{Q_a^\lambda\}_a$, $\{R_b^\lambda\}_b$, acting on systems A and B, respectively. Here, the classical variable λ is a diagonal quantum state living in the Hilbert space of A′ ⊗ B′, and thus captures all the classical randomness that is pre-shared between the two measurement devices. For a given pair of measurement choices, $\omega_{\bar{x}} := |\bar{x}\rangle\langle\bar{x}|$ and $\tau_{\bar{y}} := |\bar{y}\rangle\langle\bar{y}|$, the conditional probability of observing outcomes a and b is given as

$$\Pr[a, b|\bar{x}, \bar{y}] = \sum_\lambda \Pr[\lambda] \mathrm{Tr}\left[Q_a^\lambda \omega_{\bar{x}}\right] \mathrm{Tr}\left[R_b^\lambda \tau_{\bar{y}}\right], \quad (1)$$

which is synonymous to the locality condition assumed in standard Bell inequalities. Also, we write $\{M_{a,b}\}_{a,b}$ to denote the effective two-qubit measurement acting on the qubit inputs, i.e., $M_{a,b} = \sum_\lambda \Pr[\lambda] Q_a^\lambda \otimes R_b^\lambda$. Note that if $M_{a,b}$ is not separable for some a, b, then by definition the joint target state must be entangled, i.e., see Fig. (1). Accordingly, a violation of Eq. (1) implies that the local PQM devices must share entanglement.

Following standard arguments [19, 20], we suppose $\sum_{a\neq\varnothing} \Pr[a|\bar{x}] = \gamma$, $\sum_{b\neq\varnothing} \Pr[b|\bar{y}] = \gamma$, and $\sum_{a,b\neq\varnothing} \Pr[a, b|\bar{x}, \bar{y}] = \gamma^2$ for all measurements choices, where $\gamma \in (0, 1]$ is the measurement efficiency. With that, our
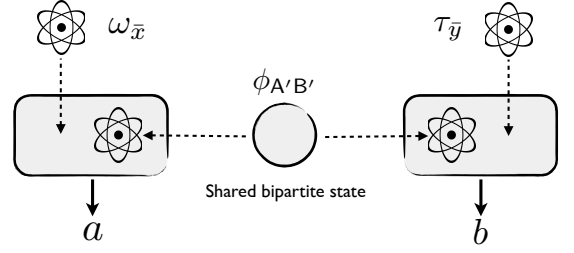

FIG. 1. **Operational interpretation.** Alice's and Bob's measurement choices are encoded into trusted qubit systems and then sent to their respective untrusted PQM devices. The PQM devices share a bipartite state (denoted by $\phi_{A'B'}$) which may or may not be entangled. To test for entanglement, Alice and Bob compute Eq. (2): if the inequality is violated, they conclude $\phi_{A'B'}$ is entangled, otherwise, the experiment is not conclusive. It is useful to mention that like standard Bell experiments, the PQM devices and the source device are all part of the test.

postselected inequality reads

$$S(\gamma) = \frac{1}{4} \sum_{\bar{x}, \bar{y}} (-1)^{f(\bar{x}, \bar{y})} \frac{C(\bar{x}, \bar{y})}{\gamma^2} \leq \beta(\gamma|\mathrm{LOSR}), \quad (2)$$

where $f(\bar{x}, \bar{y}) := x_1 \wedge y_1 \oplus x_2 \oplus y_2$ is a balanced boolean function, and $C(\bar{x}, \bar{y}) := \Pr[a = b|\bar{x}, \bar{y}] - \Pr[a \neq b|\bar{x}, \bar{y}]$ for $a, b \neq \varnothing$ is the conditional correlation function. Our goal is to derive the postselected local bound, $\beta(\gamma|\mathrm{LOSR})$, and to see how it scales with the measurement efficiency, γ.

For pedagogical reasons, we first discuss what happens when the inputs are classical. In this picture, our inequality can be seen as a symmetric extension of the CHSH inequality. To see this connection, we note that the first bit of each party, x_1, y_1, determines his or her measurement setting, and the second bit, x_2, y_2, determines if he or she should flip the measurement outcome. Indeed, it can be easily verified that Eq. (2) is an average of four CHSH inequalities conditioned on x_2 and y_2, therefore the local bound of our inequality assuming classical inputs is 2. However, despite these similarities, there is a subtle difference between the CHSH inequality and Eq. (2) with respect to classical local models. That is, a classical local model that outputs fixed correlated outcomes independently of the inputs would give a CHSH value of 2, whereas with Eq. (2) the Bell value is zero. This example illustrates that the additional randomness injected via x_2 and y_2 plays an interesting role in constraining the efficacy of certain classical local models.

Moving on to quantum inputs, the measurement basis is now determined by the basis in which the program qubit is prepared, and the bit flip value is given by the eigenvector of that basis. Notice that this encoding scheme is inspired by the celebrated quantum conjugate coding scheme used in quantum cryptography [27, 28]. The main advantage of this scheme is that the probability of learning each bit is upper bounded by $(1 + 1/\sqrt{2})/2 \approx 0.853$ [29]. Based on these observations,

we thus expect correlations generated by LOSR models to be weakly correlated with Alice's and Bob's measurement choices.

Recall that we want to derive the postselected local bound and the postselected maximum quantum bound for Eq. (2). As mentioned earlier, the former is denoted by $\beta(\gamma|\mathrm{LOSR})$ and is defined as the maximization of $S(\gamma)$ over all LOSR measurements for a fixed measurement efficiency γ. At this point, it is useful to mention that all postselection strategies conceivable by LOSR models are automatically accounted for in the maximization. That is, any postselection strategy employed by the underlying LOSR model must be captured by the local *filtering* POVMs $Q_0^\lambda + Q_1^\lambda$ and $R_0^\lambda + R_1^\lambda$, which are also optimized as part of the maximization together with the distribution $\{\Pr[\lambda]\}_\lambda$. Moving on, the postselected maximum quantum bound is denoted by $\beta(\gamma)$ and is defined as the maximization of $S(\gamma)$ over the set of quantum strategies, $\{\phi_{\mathsf{A'B'}}, \{Q_a\}_a, \{R_b\}_b\}$.

Connection to quantum state discrimination. The above maximization problems can be solved by using a connection to the local-distinguishability of quantum inputs. To illustrate this connection, we first note that the proposed semi-quantum CHSH experiment is equivalent to a guessing game in which the *untrusted* local measurement devices have to guess the bit value $f(\bar{x}, \bar{y})$ when given quantum inputs $\omega_{\bar{x}} \otimes \tau_{\bar{y}}$. More precisely, the devices win the game if they output $a \oplus b = f(\bar{x}, \bar{y})$ whenever $a, b \neq \varnothing$, i.e., the game is counted only for jointly conclusive events. The conditional guessing probability can be written in terms of Eq. (2),

$$G(\gamma) := \frac{\Pr[a \oplus b = f(\bar{x}, \bar{y})]}{\gamma^2} = \frac{1}{2} + \frac{S(\gamma)}{8}, \qquad (3)$$

where $S(\gamma)/8$ can be seen as the distinguishing advantage. Then, it can be easily verified that

$$\Pr[a \oplus b = f(\bar{x}, \bar{y})] = \frac{\mathrm{Tr}\left[\rho_0 \Pi_{a \oplus b = 0} + \rho_1 \Pi_{a \oplus b = 1}\right]}{2}, \quad (4)$$

where we used

$$\rho_0 = \frac{1}{8} \sum_{\substack{\bar{x}, \bar{y} \\ \mathrm{s.t.} f(\bar{x}, \bar{y}) = 0}} \omega_{\bar{x}} \otimes \tau_{\bar{y}}, \quad \rho_1 = \frac{1}{8} \sum_{\substack{\bar{x}, \bar{y} \\ \mathrm{s.t.} f(\bar{x}, \bar{y}) = 1}} \omega_{\bar{x}} \otimes \tau_{\bar{y}},$$

and the measurement assignments $\Pi_{a \oplus b = 0} = M_{0,0} + M_{1,1}$, $\Pi_{a \oplus b = 1} = M_{0,1} + M_{0,1}$ and $\Pi_\varnothing = \mathbb{1} - \Pi_{a \oplus b = 0} - \Pi_{a \oplus b = 1}$.

We may interpret Eq. (4) as follows. In each run of the experiment, the measurement devices are given a product quantum state $\omega_{\bar{x}} \otimes \tau_{\bar{y}}$ randomly chosen from one of the two sets of states, $\{\omega_{\bar{x}} \otimes \tau_{\bar{y}} : f(\bar{x}, \bar{y}) = 0\}$ and $\{\omega_{\bar{x}} \otimes \tau_{\bar{y}} : f(\bar{x}, \bar{y}) = 1\}$, and the devices have to guess which set the given state is drawn from. In other words, the local devices have to collectively guess the global identity $f(\bar{x}, \bar{y})$ of $\omega_{\bar{x}} \otimes \tau_{\bar{y}}$ using whatever resources they are given with. Indeed, the figure of merit in this case is exactly given by Eq. (3), which is the conditional guessing probability uniformly averaged over all product states. Using the Born's rule and the linearity of the trace operator, this guessing game can be simplified to the local-distinguishability of two non-orthogonal mixed states ρ_0 and ρ_1 assuming a fixed conclusive rate of γ^2. Therefore, the maximization Eq. (2) is equivalent to the maximization of Eq. (4) (up to the constant normalization factor of $1/\gamma^2$).

The advantage of local-distinguishability games is that they can be analytically solved through semidefinite programming [15], a form of convex optimization that maximizes a linear function over the intersection of a semidefinite cone and an affine plane [30]. For brevity, we present only the primal programs and defer the corresponding dual programs and optimal solutions to the Supplementary Material. The primal program for computing the maximum quantum guessing probability assuming a fixed $\gamma^2 \in (0, 1]$ is given by

$$\mathtt{maximize}: \frac{1}{2}\mathrm{Tr}\left[\rho_0 \Pi_{a \oplus b = 0} + \rho_1 \Pi_{a \oplus b = 1}\right]$$
$$\mathtt{subject\ to}: \Pi_{a \oplus b = 0} + \Pi_{a \oplus b = 1} + \Pi_\varnothing = \mathbb{1}_{\mathsf{A} \otimes \mathsf{B}},$$
$$\mathrm{Tr}\left[(\omega_{\bar{x}} \otimes \tau_{\bar{y}} \Pi_\varnothing\right] = 1 - \gamma^2, \quad \forall\, \bar{x}, \bar{y}$$
$$\Pi_i \succeq 0, \quad i = 0, 1, \varnothing,$$

and the optimal values are found to be

$$\max G(\gamma) = \begin{cases} \frac{1}{2}\left(1 + \frac{1}{\gamma^2 2\sqrt{2}}\right) & \text{if} \quad \gamma > \frac{1}{\sqrt{2}} \\ \frac{1}{2} + \frac{1}{2\sqrt{2}} & \text{if} \quad \gamma \leq \frac{1}{\sqrt{2}} \end{cases}. \quad (5)$$

Here, an important remark is in order. These optimal values are obtained over the whole set of two-qubit POVMs acting on $\mathsf{A} \otimes \mathsf{B}$, which is larger than the set of quantum strategies, i.e., $M_{a,b} = \mathrm{Tr}_{\mathsf{A'B'}}\left[\phi_{\mathsf{A'B'}}(Q_a \otimes R_b)\right]$. Thus strictly speaking, Eq. (5) is an upper bound on the maximum quantum bound, i.e., $\beta(\gamma) \leq 2\sqrt{2}$ for $0 < \gamma \leq 1/\sqrt{2}$ and $\beta(\gamma) \leq \sqrt{2}/\gamma^2$ for $1/\sqrt{2} < \gamma \leq 1$. However, as we will see later, this upper bound is tight for $\gamma \in (0, 1/2]$.

Similarly, the maximization for LOSR measurements is based on a circuitous method, which nevertheless also leads to a tight upper bound on $G(\gamma|\mathrm{LOSR})$. More precisely, we optimize over all measurements compatible with the positive partial transpose (PPT) condition [31] instead of LOSR measurements. The reason is that PPT measurements admit a much simpler characterization and can be formulated as linear constraints in the semidefinite programs, i.e., we only need to add $\Pi_i^{T_\mathsf{B}} \succeq 0$, for $i = 0, 1, \varnothing$, where T_B means the partial transpose with respect to Bob's measurements. Moreover, we use the fact that PPT and separable measurements are equivalent at the level of two-qubit positive operators [31]. Therefore, the optimal bound for PPT measurements is an upper bound on that of LOSR measurements, i.e., $\beta(\gamma|\mathrm{LOSR}) \leq \beta(\gamma|\mathrm{Sep}) = \beta(\gamma|\mathrm{PPT})$. The optimal value for PPT models is found to be independent of γ,

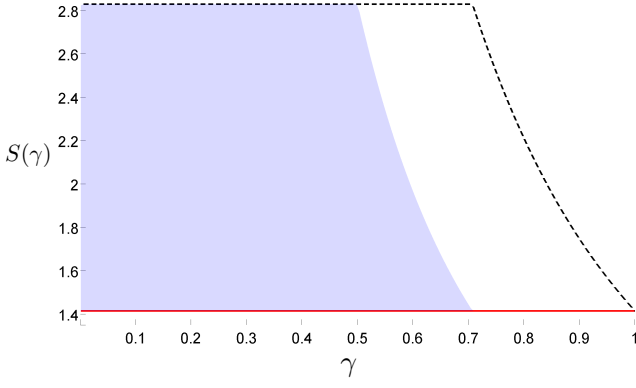$$\max G(\cdot|\mathrm{PPT}) = \frac{1}{2} + \frac{1}{4\sqrt{2}}. \qquad (6)$$

FIG. 2. **Quantum violation vs efficiency.** The vertical axis is the postselected Bell value $S(\gamma)$ and the horizontal axis is the measurement efficiency, $\gamma \in (0,1]$. The (black) dashed line is given by the maximum quantum bound, Eq. (5), which is obtained using general two-qubit measurements. The (red) solid line is the postselected local bound, Eq. (6). The (blue) shaded area is due to the pretty good quantum strategy.

Interestingly, it turns out that the optimal measurements are given by LOSR measurements. To show this, suppose the qubit inputs are given by $\omega_{0x_2} = (\mathbb{1}_A + (-1)^{x_2}\mathbb{X})/2$, $\omega_{1x_2} = (\mathbb{1}_A + (-1)^{x_2}\mathbb{Y})/2$, and $\tau_{y_1 y_2} = (\mathbb{1}_B + (-1)^{y_2}(\mathbb{X} + (-1)^{y_1}\mathbb{Y})/\sqrt{2})/2$, where $\mathbb{X}$ and $\mathbb{Y}$ are Pauli matrices [32]. Then, it can be verified that the joint input states are jointly diagonal in the standard Bell basis:

$$\rho_0 = \begin{pmatrix} \alpha^+ & 0 & 0 & 0 \\ 0 & \alpha^- & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}, \quad \rho_1 = \begin{pmatrix} \alpha^- & 0 & 0 & 0 \\ 0 & \alpha^+ & 0 & 0 \\ 0 & 0 & \frac{1}{4} & 0 \\ 0 & 0 & 0 & \frac{1}{4} \end{pmatrix}, \quad (7)$$

where the eigenvalues are $\alpha^\pm := (1 \pm 1/\sqrt{2})/4$, and the corresponding eigenvectors are ordered as: $|\Psi^+\rangle$, $|\Psi^-\rangle$, $|\Phi^+\rangle$ and $|\Phi^-\rangle$ [33]. For example, we have $\langle\Psi^+|\rho_0|\Psi^+\rangle = \langle\Psi^-|\rho_1|\Psi^-\rangle = \alpha^+$. A simple LOSR measurement that achieves Eq. (6) is one that uses only local measurements, i.e., no shared randomness is needed. More specifically, the strategy is $Q_a = \gamma(\mathbb{1}_A + (-1)^a\mathbb{X})/2$, $R_b = \gamma(\mathbb{1}_B + (-1)^b\mathbb{X})/2$ for $a, b = 0, 1$, and $Q_\varnothing = (1-\gamma)\mathbb{1}_A$, $R_\varnothing = (1-\gamma)\mathbb{1}_B$ for the inconclusive outcomes. That is, each measurement device with probability γ measures in the $\mathbb{X}$ basis, and with probability $1 - \gamma$ outputs $\varnothing$ without measurement. Another strategy is to measure in the $\mathbb{Y}$ basis instead of $\mathbb{X}$, or to use a combination of these two strategies assisted with shared randomness.

A pretty good quantum strategy. As mentioned above, the optimal solutions to Eq. (5) are given in terms of two-qubit POVMs and thus do not provide a clear exposition on the optimal quantum strategy (i.e., the optimal entangled bipartite state and local PQMs) needed to achieve the maximum quantum bound. To this end, we

provide an explicit quantum strategy that reaches the upper bound in the region of $0 < \gamma < 1/2$, i.e., see the shaded area in Fig. (2). Again, we refer to the aforementioned encoding scheme, i.e., Eq. (7). The optimal joint target system is a two-qubit maximally entangled state, $\phi_{A'B'} = |\Psi^+\rangle\langle\Psi^+|$, and the optimal PQMs are

$$Q_0 = \gamma_1 |\Psi^+\rangle\langle\Psi^+| + \gamma_2(|00\rangle\langle00| + |11\rangle\langle11|)$$
$$Q_1 = \gamma_1 |\Psi^-\rangle\langle\Psi^-| + \gamma_2(|00\rangle\langle00| + |11\rangle\langle11|)$$
$$Q_\varnothing = \mathbb{1}_{A\otimes B} - Q_0 - Q_1,$$

and likewise $R_i = Q_i$ for $i = 0, 1, \varnothing$, where $\gamma_1 = \min\{2\gamma, 1\}$ and $\gamma_2 = \max\{\gamma - 1/2, 0\}$. Note that the PQMs are inefficient Bell-state measurements (BSMs), i.e., they can only discriminate between $|\Psi^+\rangle$ and $|\Psi^-\rangle$. The Bell values using these states and measurements are $S(\gamma) = 2\sqrt{2}$ for $0 < \gamma \leq 1/2$ and $S(\gamma) = 1/(\gamma^2\sqrt{2})$ for $1/2 < \gamma < 1/\sqrt{2}$. We remark that this quantum strategy is however sub-optimal when it comes to detecting weakly entangled states. For instance, in the case of two-qubit Werner states [34], $\phi_{A'B'} = F|\Psi^-\rangle\langle\Psi^-| + (1 - F)\mathbb{1}_{A'\otimes B'}/4$, it can be shown that violation is obtained only for $F > 1/2$; note that these Werner states are separable for $F \leq 1/3$. On the other hand, we have upper bounds on the achievable Bell violations for $F > 1/3$, which suggest that Eq. (2) might be able to detect all entangled two-qubit Werner states; see Supplementary Material.

Discussion. A way to interpret our result is to examine the optimality conditions for discriminating ρ_0 and ρ_1. To begin with, we remind that these mixed states share the same support and can be simultaneously diagonalized in the Bell basis. The first point implies that unambiguous state discrimination [22] is not possible, thus the best measurement scheme, for our purpose, is probabilistic minimum-error state discrimination [35, 36]. From the optimality conditions of this scheme, it can be easily verified that the maximum success probability for which ρ_0 and ρ_1 are optimally discriminated is $1/2$, which is indeed the value given in Eq. (5). This also explains the trend seen in Fig. (2) wherein higher Bell violations are achieved with higher inconclusive rates/lower measurement efficiencies.

From the second point, it is clear that the optimal measurement that discriminates between ρ_0 and ρ_1 must consists of entangled POVMs: the positive and negative eigenspaces of $\rho_0 - \rho_1$ are maximally entangled subspaces. This means that no LOSR measurement (or more generally, separable measurement) can coherently access these entangled eigenspaces. Crucially, this limitation also holds in the presence of inconclusive outcomes, i.e., entanglement cannot be created using local operations and classical communication (local filtering with shared randomness in our case).

Conclusion. In the above, we have provided a semi-quantum Bell experiment that safely allows for postselection and is defined by a loss-independent local bound

that is violated only in the region of imperfect measurement efficiencies. On the conceptual level, our result suggests that semi-quantum nonlocality is much more powerful than previously recognized. For instance, Eq. (2) does not require the so-called *fair-sampling condition* [21, 23], which is typically assumed in standard Bell experiments involving postselection to ensure that the conclusive/detected events are representative of the underlying quantum system. Most interestingly, Fig. (2) shows that in order to (optimally) violate Eq. (2), it is necessary to use highly inefficient measurements, which up to the best of our knowledge, is the first time that such a trend has been found. Furthermore, the maximal quantum violation $2\sqrt{2}$ can be achieved for a continuum of measurement efficiencies, i.e., $\gamma \in (0, 1/2)$, unlike standard Bell inequalities which can only reach their maximum violations in the limit of perfect measurement efficiency.

Finally, we remark that on the practical side, our inequality provides a semi-device-independent method for testing entanglement in *detected quantum systems*. That is, as mentioned above, the inequality allows one to restrict the analysis to detected events without assuming the fair-sampling condition. For example, this application could be useful for entanglement-based experiments suffering from high detection losses, e.g., those based on practical entangled photon-pair sources [37].

[1] Bell, J. S. *Physics* **1**, 195–200 (1964)
[2] Brunner. N., Cavalcanti. D., Pironio. S., Scarani. V. & Wehner. S. *Rev. Mod. Phys.* **86**, 419 (2014)
[3] That is, events happening in one space-time region are independent of events in space-like separated regions.
[4] Acin, A. *et al. Phys. Rev. Lett.* **98**, 230501 (2007)
[5] Pironio. S. *et al. Nature* **464**, 1021–1024 (2010)
[6] Lim. C. C. W., Portmann. C., Tomamichel. M., Renner. R. & Gisin. N. *Phys. Rev. X.* **3**, 031006 (2013)
[7] Reichardt. B. W., Unger. F. & Vazirani. U. *Nature* **496**, 456–460 (2013)
[8] Vazirani. U. & Vidick. T. *Phys. Rev. Lett.* **113**, 140501 (2014)
[9] Yang. T. H., Vertesi. T., Bancal. J.-D., Scarani. V. & Navascues. M. *Phys. Rev. Lett.* **113**, 040401 (2014)
[10] Buscemi. F. *Phys. Rev. Lett.* **108**, 200401 (2012)
[11] Branciard. C., Rosset. D., Liang. Y.-C. & Gisin. N. *Phys. Rev. Lett.* **110**, 060405 (2013)
[12] Xu. P. *et al. Phys. Rev. Lett.* **112**, 140506 (2014)
[13] Peres. A. & Wootters. W. K. *Phys. Rev. Lett.* **66**, 1119 (1991)
[14] Bennett. C. H. *et al. Phys. Rev. A.* **59**, 1070 (1999)
[15] Cosentino. A. *Phys. Rev. A.* **87**, 012321 (2013)
[16] Terhal. B. M., DiVincenzo. D. P. & Leung. D. W. *Phys. Rev. Lett.* **86**, 5807 (2001)
[17] DiVincenzo. D. P., Leung. D. W. & Terhal. B. M. *IEEE Trans. Inf Theory* Vol. **48**. No. 3, 580–599 (2002)
[18] The measurement efficiency of a measurement device is defined as the probability of obtaining a conclusive outcome whenever the device receives an input.
[19] Eberhard. P. H. *Phys. Rev. A.* **47**, R747(R) (1993)
[20] Branciard. C. *Phys. Rev. A.* **83**, 032123 (2011)
[21] Berry. D. W., Jeong, H., Stobińska, M. & Ralph. T. *Phys. Rev. A.* **81**, 012109 (2010)
[22] Barnett. S. M. & Croke. S. *Adv. Opt. Photon.* **1**, 238–278 (2009)
[23] Clauser, J. F., Horne, M. A., Shimony, A. & Holt, R. A. *Phys. Rev. Lett.* **23**, 880–884 (1969)
[24] Nielsen. M. A. & Chuang. I. L. *Phys. Rev. Lett.* **79**, 321 (1997)
[25] Fiurasek. J., Dusek. M. & Flip. R. *Phys. Rev. Lett.* **89**, 190401 (2002)
[26] Barrett. J. & Gisin. N. *Phys. Rev. Lett.* **106**, 100406 (2011)
[27] Wiesner. S. Conjugate coding. *Sigact News* **15**(1), 7888 (1983)
[28] Bennett, C. H. & Brassard, G. in *Proc. IEEE Int. Conf. on Computers, Systems and Signal Processing* 175–179 (IEEE Computer Society, 1984)
[29] Ambainis. A., Nayak. A., Ta-Shma. A & Vazirani. U. *Proceedings of the 31st Annual ACM Symposium on Theory of Computing (STOC' 99)*, 376–383 (1999)
[30] Vandenberghe. L. & Boyd. *SIAM Rev.* **38**, 49–95 (1996)
[31] Horodecki. M., Horodecki. P. & Horodecki. R. *Phys. Lett. A.* **223**, 1–2 (1996)
[32] More precisely, the eigenvectors of $\mathbb{X}$ are $\{(|0\rangle \pm |1\rangle)/\sqrt{2}\}$ and the ones of $\mathbb{Y}$ are $\{(|0\rangle \pm i|1\rangle)/\sqrt{2}\}$.
[33] The standard Bell basis is defined as $|\Phi^{\pm}\rangle := (|00\rangle \pm |11\rangle)/\sqrt{2}$ and $|\Psi^{\pm}\rangle := (|01\rangle \pm |10\rangle)/\sqrt{2}$.
[34] Werner. R. F. *Phys. Rev. A.* **40**, 4277 (1989)
[35] Bagan. E., Munoz-Tapia. R., Olivares-Renteria. G. A. & Bergou. J. A. *Phys. Rev. A.* **86**, 040303(R) (2012)
[36] Herzog. U. *Phys. Rev. A.* **86**, 032314 (2012)
[37] Caprara Vivoli. V *et al. Phys. Rev. A.* **91**, 012107 (2015)