Measurement uncertainty from no-signaling and nonlocality

Justyna Łodyga,[1] Waldemar Kłobus,[1] Ravishankar Ramanathan,[2,3,4] Andrzej Grudka,[1] Michał Horodecki,[2,3] and Ryszard Horodecki[2,3]

[1]*Faculty of Physics, Adam Mickiewicz University, 61-614 Poznań, Poland*
[2]*Institute for Theoretical Physics and Astrophysics, University of Gdańsk, 80-952 Gdańsk, Poland*
[3]*National Quantum Information Centre of Gdańsk, 81-824 Sopot, Poland*
[4]*Laboratoire d'Information Quantique, Université Libre de Bruxelles, Belgium*

(Received 22 February 2016; revised manuscript received 25 January 2017; published 25 July 2017)

One of the formulations of Heisenberg uncertainty principle, concerning so-called measurement uncertainty, states that the measurement of one observable modifies the statistics of the other. Here, we derive such a measurement uncertainty principle from two comprehensible assumptions: impossibility of instantaneous messaging at a distance (no-signaling), and violation of Bell inequalities (nonlocality). The uncertainty is established for a pair of observables of one of two spatially separated systems that exhibit nonlocal correlations. To this end, we introduce a gentle form of measurement which acquires partial information about one of the observables. We then bound disturbance of the remaining observables by the amount of information gained from the gentle measurement, minus a correction depending on the degree of nonlocality. The obtained quantitative expression resembles the quantum mechanical formulations, yet it is derived without the quantum formalism and complements the known qualitative effect of disturbance implied by nonlocality and no-signaling.

DOI: 10.1103/PhysRevA.96.012124

I. INTRODUCTION

In recent decades much effort has been put into understanding quantum mechanics "from the outside". Namely, one considers possible constraints for correlations coming solely from the no-signaling principle, and compares them with quantum mechanical constraints. The first observation was already undertaken in the 1990s by Popescu and Rohrlich [1]. They showed that no-signaling constraints are much weaker, and allow for extremely strong correlations that violate the so-called Bell-CHSH inequality [2,3] to the maximal possible extent, i.e., achieving the maximal algebraic value of the Bell quantity.

On the other hand, much work has been done in order to extract features of quantum formalism that are responsible for various nonclassical effects, such as quantum computational speedup, reduction of communication complexity, quantum key distribution, and expansion or amplification of weak randomness. It turns out that to achieve at least some of those effects, one does not need to employ the full quantum formalism, but just refer to its two features: the impossibility of faster-than-light communication (no-signaling) combined with Bell nonlocality. For example, to obtain secure key distribution, one just uses the no-signaling principle in conjunction with the fact that statistics obtained in distant labs violate Bell inequalities, exhibiting in this way Bell nonlocality [4]. However, such a fundamental rule as the Heisenberg uncertainty principle [5], so far treated as a hallmark of quantum mechanics, has not yet been derived from these simple assumptions.

When considering the Heisenberg uncertainty principle, one may think of either of its two facets: the *preparation uncertainty principle*, stating that one cannot prepare a system in a state exhibiting peaked statistics for each of two incompatible observables [6–9], and the *measurement uncertainty principle*, stating that by measuring one observable, one necessarily disturbs the statistics of the other observable [10–13]. Tomamichel and Hänggi [14] obtained the former principle from nonlocality using quantum formalism. However, the preparation uncertainty cannot be determined solely from no-signaling and nonlocality, as it is not exhibited by the Popescu-Rohrlich box [15]. The measurement uncertainty principle, on the other hand, does not meet such restrictions. It has a closely related formulation as an information gain versus disturbance trade-off [16–21] and has become a basis for quantum cryptography [22,23]: a potential eavesdropper by gaining information about the cryptographic key necessarily disturbs the system, which can be noticed by the parties that are to establish the key. The subject of measurement uncertainty principle in the context of nonlocality and no-signaling was touched upon by Oppenheim and Wehner [15] who showed (in a nonquantitative manner) that Bell nonlocality implies that a sharp measurement, i.e., the measurement with complete knowledge about the outcome, must cause disturbance.

In this paper, we derive a *quantitative* measurement uncertainty relation, in the form of a trade-off implied by Bell nonlocality and no-signaling. To this end, we introduce a notion of gentle measurement as well as a quantitative notion of disturbance, both applicable in the operational scenario, where the only objects are statistics of measurements. In particular, we consider a bipartite scenario where Bob, who exhibits nonlocal correlations with Alice (measured by degree of violation of a chosen Bell inequality), performs consecutive measurements of a pair of his observables. As a result, we find that the very act of his first measurement disturbs the statistics of the second measurement (this happens even if the first measurement is *gentle*, i.e., where he does not acquire full knowledge about the result). Additionally, it appears that the magnitude of such disturbance increases not only with information gain but also with the strength of the Bell inequality violation. We subsequently compare our result with its counterpart obtained within the quantum mechanical framework.

In our findings we use traditional monogamy relations to obtain dynamical-type (or better kinematic-type) relations. The former are static, and state that if two systems are nonlocally correlated, the possible information present in a third system must be limited. In contrast, we consider a time ordered scenario where a party measures observables one by one, exactly like in the measurement uncertainty principle.

II. INFORMATION GAIN VIA A GENTLE MEASUREMENT

We start with an initial bipartite system, one system possessed by Alice, the other by Bob. Alice and Bob can sharply measure their observables A_x, $x = 1, \dots, n$ and B_y, $y = 1, \dots, m$, respectively, and obtain corresponding outcomes a and b. In addition, for Bob we introduce a gentle measurement responsible for the partial gain of information of one of his observables. Hereafter, without loss of generality, we choose the fixed observable B_1 to be measured gently. Bob will perform the gentle measurement before he measures another observable, by coupling his measuring apparatus to the system. Equivalently, we can imagine that a third person—Grace—couples some other system to Bob's one, performs some evolution, and takes away her system. This results in an overall tripartite system: on two of them Alice and Bob can still measure their sharp observables, while Grace can measure her single observable that represents gentle measurement of Bob's chosen observable B_1. In terms of no-signaling boxes, Grace has just one input, which we call B_1^g, with corresponding output b_1^g.

Formally, let us denote the statistics of the original bipartite system as $p(a,b|A_x,B_y)$, and the statistics of the tripartite system as $\tilde{p}(a,b,b_1^g|A_x,B_y,B_1^g)$, where B_1^g—the gentle version of B_1—is the only observable available to Grace. The final bipartite statistics is then given by

$$\tilde{p}\big(a,b\big|A_x,B_y,B_1^g\big) = \sum_{b_1^g} \tilde{p}\big(a,b,b_1^g\big|A_x,B_y,B_1^g\big). \quad (1)$$

We shall now require that Grace's observable is indeed a gentle version of Bob's observable, by imposing the following two conditions (for details, see Appendix A).

(1) The act of Grace's measurement will not affect the statistics of the sharp observable B_1, conditioned on any input and output of Alice, i.e.,

$$p(b_1|B_1,a,A_x) = \tilde{p}\big(b_1|B_1,B_1^g,a,A_x\big), \quad \forall a,x. \quad (2)$$

(2) Grace's output b_1^g will be correlated with Bob's output of measurement of B_1 (again conditioned on any of Alice's inputs and outputs) resulting in the following conditional probability distribution:

$$\tilde{p}\big(b_1^g = i\big|b_1 = j,B_1,B_1^g,a,A_x\big) = \begin{cases} \frac{1}{2}+\epsilon & \text{if } i = j, \\ \frac{1}{2}-\epsilon & \text{if } i \neq j, \end{cases} \quad (3)$$

where the parameter $\epsilon \in [0,\frac{1}{2}]$ quantifies the information gain. For $\epsilon = \frac{1}{2}$, complete information about the observable is acquired, i.e., the sharp measurement gives the same output as the gentle measurement, whereas, for $\epsilon = 0$, the outputs of gentle measurement are completely uncorrelated with the outputs of sharp measurement, hence the information gain is zero.

Let us emphasize that we will not restrict in any way what possible changes may happen to the original bipartite box, other than by the above assumptions—which are imposed just by the very definition of gentle measurement. The resulting change will follow solely from no-signaling and nonlocality.

III. DISTURBANCE

Consider first a (not necessarily quantum mechanical) state ρ and a given observable. We want to quantify how much the observable is disturbed by some other action on the state, which changes it into state $\tilde{\rho}$; in our case the action is the gentle measurement of observable B_1. A natural disturbance measure is the statistical distance between the probability distribution $p(b|B_y,a,A_x)$ obtained by measuring the observable $B_y \neq B_1$ on state ρ (i.e., prior to the gentle measurement) and the distribution $\tilde{p}(b|B_y,B_1^g,a,A_x)$ obtained by measuring this observable on state $\tilde{\rho}$ (after the gentle measurement is performed). While deriving the disturbance from nonlocality, however, we shall not show that the disturbance holds for some particular state. Rather, we prove that the disturbance occurs for some of the states produced by Alice. When Alice chooses an observable A_x and obtains an outcome a, a state ρ_{a,A_x} is created at Bob's side. The state changed by gentle measurement is thus given by $\tilde{\rho}_{a,A_x}$. Note that since the gentle measurement is performed on Bob's system, then due to no-signaling we have $p(a|A_x) = \tilde{p}(a|A_x)$. For a given choice of Alice's observable A_x and an outcome a, the disturbance of the observable $B_y \neq B_1$ is defined as

$$D_{a,x}(B_y) = \sum_b \big|p(b|B_y,a,A_x) - \tilde{p}\big(b\big|B_y,B_1^g,a,A_x\big)\big|. \quad (4)$$

In this work we consider the average total disturbance, where we sum over all of Alice's observables and all of Bob's observables apart from B_1 itself, and average over Alice's outcomes

$$\mathcal{D} = \sum_{a,x} p(A_x)p(a|A_x) \sum_{y\neq 1} D_{a,x}(B_y). \quad (5)$$

In Appendix B we argue that the change of nonlocality necessarily causes disturbance, proving that for arbitrary Bell inequality (with moduli of coefficients bounded by 1, without loss of generality), the average total disturbance $\mathcal{D}$ (5) always satisfies

$$n\mathcal{D} \geqslant |\beta(p) - \beta(\tilde{p})|, \quad (6)$$

where n denotes the number of Alice's measurement choices, and $\beta(p),\beta(\tilde{p})$ are the values of the Bell quantity evaluated on initial statistics $p(a,b|A_x,B_y)$ and final statistics $\tilde{p}(a,b|A_x,B_y,B_1^g)$ given by Eq. (1), respectively.

IV. RELEVANCE OF BELL INEQUALITIES FOR OBSERVABLE

It could happen that a chosen Bell inequality does not cover some of the observables. For example, in Bell-CHSH inequality for a scenario where Alice and Bob hold $n = 2$ and $m = 3$ observables, respectively, one of Bob's observables is not included. Therefore, such observable does not cause any disturbance and the inequality (6) becomes trivial.

To quantify the ability of the observable B_1 to disturb the other observable, given a specific Bell inequality, we introduce a new quantity, namely the notion of relevance $w(B_1)$. For simplicity, and due to our convention that the gentle measurement is always performed on a fixed observable B_1, in $w(B_1)$ we remove the argument B_1 and define relevance w as

$$w = \beta^{\max} - \beta_1^{\max}, \tag{7}$$

where $\beta^{\max}$ denotes the maximal value of Bell quantity for no-signaling probabilistic theories and $\beta_1^{\max}$ the maximal value of Bell quantity where the observable B_1 is deterministic. Relevance w (7) measures how far the observable is from being deterministic, i.e., it quantifies its degree of randomness. Therefore, for the increasing value of relevance w, we observe stronger disturbance properties of observable B_1.

For that reason, the relevance w (7) determines the strength of a monogamy relation related to the value β of a chosen Bell inequality

$$\beta + w\langle B_1^g B_1\rangle \leqslant \beta^{\max}, \tag{8}$$

where $\langle B_1^g B_1\rangle$ stands for a correlation function between B_1^g and B_1. In Appendix C we provide a proof for the relation (8), and show that for the CHSH and chain Bell inequality $w = 2$ for any chosen observable, whereas for so-called total function XOR games (a more general class of correlation Bell inequalities with binary outputs) $w \geqslant \min(\beta^{\max} - \beta_{cl}^{\max}, n)$, where $\beta_{cl}^{\max}$ denotes the maximal classical value of the Bell quantity.

V. MEASUREMENT UNCERTAINTY PRINCIPLE

We now present our main result, i.e., the trade-off between information gained in the gentle measurement of observable B_1 and the disturbance caused by it on the remaining observables. Consider arbitrary Bell inequality, and rescale it so that it can be written as $\beta = \sum_{a,b,x,y} c(a,b,A_x,B_y)p(a,b|A_x,B_y)$, where the coefficients are bounded as $|c(a,b,A_x,B_y)| \leqslant 1$. For such a defined Bell inequality β, the trade-off is of the following general form:

$$n\mathcal{D} \geqslant w\mathcal{I} - \mathcal{L}, \tag{9}$$

where n denotes the number of Alice's observables, and relevance w, introduced in Eq. (7), quantifies the indeterminacy of gently measured observable B_1, hence its ability to disturb the other observables. The derived formula (9) combines the three fundamental quantities: disturbance $\mathcal{D}$, information gain $\mathcal{I}$, and the level of nonlocality $\mathcal{L}$. The first one, explicitly defined in Eq. (5), describes the average statistical distance of probability distributions prior and after the act of a gentle measurement. The information $\mathcal{I}$ gained in such a measurement is parametrized by $\epsilon = [0,\frac{1}{2}]$ introduced in Eq. (3), and it is given by $\mathcal{I} = 2\epsilon$. The scaling factor 2 is added for technical reasons, but actually 2ϵ has the interpretation of a correlation function between B_1^g and B_1, cf. Appendix D. Finally, the degree of locality $\mathcal{L} = \beta^{\max} - \beta$ reports on how the nonlocality of the system deviates from the maximal possible nonlocality in general no-signaling theories, and it is quantified by the violation of a chosen Bell inequality. For different experimental settings, different Bell inequalities can be chosen. Intuitively, one can notice that the more information is gained, the more disturbance is introduced into the system. One can also note that whenever the local content $\mathcal{L}$ vanishes (we are at the extreme point of no-signaling correlations), arbitrarily small information gain causes disturbance.

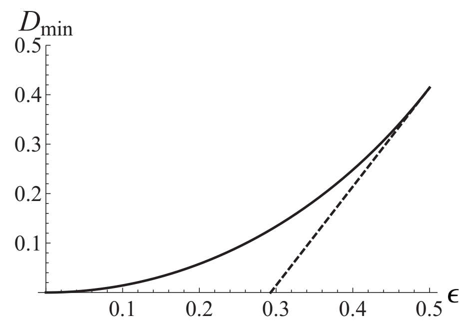

FIG. 1. Lower bound $\mathcal{D}_{\min} = 2\epsilon - \frac{1}{2}(4 - \beta_{\text{CHSH}})$ [RHS of Eq. (11)] on average total disturbance obtained from nonlocality and no-signaling principle for the case of CHSH inequality for $\beta_{\text{CHSH}} = 2\sqrt{2}$ corresponding to maximally nonlocal correlations attainable within the framework of quantum mechanics (dashed line) and $\mathcal{D}_{\min} = \frac{1}{2}(\beta_{\text{CHSH}} - \sqrt{8 - 4(2\epsilon)^2})$ [RHS of Eq. (15)] obtained from quantum predictions (thick line).

VI. EXAMPLES

Two example trade-offs can be obtained from CHSH inequality, and its generalization—chain Bell inequality. The CHSH inequality reads

$$\beta_{\text{CHSH}} = \langle A_1B_1\rangle + \langle A_1B_2\rangle + \langle A_2B_1\rangle - \langle A_2B_2\rangle \leqslant 2, \tag{10}$$

with maximal value $\beta_{\text{CHSH}}^{\max} = 4$. There are just two observables on either side; thus when Bob gently measures B_1, he disturbs the observable B_2, and the trade-off stands as

$$\mathcal{D} = D(B_2) \geqslant 2\epsilon - \tfrac{1}{2}(4 - \beta_{\text{CHSH}}), \tag{11}$$

where we used Eq. (9) with $n = 2$, $w = 2$, $\mathcal{I} = 2\epsilon$, and $\mathcal{L} = 4 - \beta_{\text{CHSH}}$. The right-hand side (RHS) of Eq. (11), i.e., a lower bound $\mathcal{D}_{\min}$ on average total disturbance $\mathcal{D}$, becomes simplified for maximally nonlocal correlations exhibited by the so-called *Popescu-Rohrlich box* with $\beta_{\text{CHSH}} = 4$ to $\mathcal{D}_{\min} = \mathcal{I} = 2\epsilon$. For nonmaximally nonlocal correlations, there is some threshold value of ϵ, denoted ϵ_{th}, for which the lower bound $\mathcal{D}_{\min}$ is nontrivial. For example, at the Tsirelson bound $\beta_{\text{CHSH}} = 2\sqrt{2}$, attained for maximal correlations allowed in quantum regime, $\mathcal{D}_{\min} = 2\epsilon - (2 - \sqrt{2})$ and $\epsilon_{th} = 0.293$, as depicted in Fig. 1 (dashed line).

The chain inequality [25] is given by

$$\beta_{\text{chain}} = \sum_{k=1}^{n-1}(\langle A_kB_k\rangle + \langle A_kB_{k+1}\rangle) + \langle A_nB_n\rangle - \langle A_nB_1\rangle \leqslant 2n - 2 \tag{12}$$

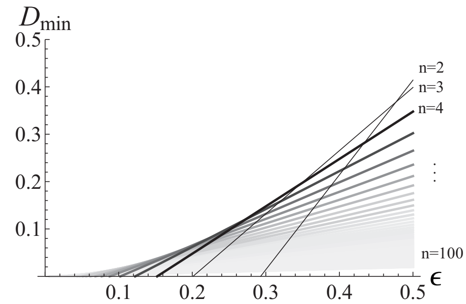

FIG. 2. Lower bound $\mathcal{D}_{\min} = \frac{4}{n}\epsilon - \frac{1}{n}(2n - \beta_{\text{chain}})$ [RHS of Eq. (13)] on average total disturbance obtained from nonlocality and no-signaling principle for the case of chain Bell inequality, where we choose $\beta_{\text{chain}} = 2n\cos(\frac{\pi}{2n})$ [24] with n denoting the number of observables.

and $\beta_{\text{chain}}^{\max} = 2n$. Analogous to the CHSH inequality, for the gentle measurement of B_1, we obtain

$$\mathcal{D} = \sum_{i\neq 1} D(B_i) \geqslant \frac{4}{n}\epsilon - \frac{1}{n}(2n - \beta_{\text{chain}}). \tag{13}$$

The dependence of the lower bound $\mathcal{D}_{\min}$ on average total disturbance [RHS of Eq. (13)] on information gain, as well as on number of observables n, is presented in Fig. 2. Note that the larger the number of observables, the more the threshold $\epsilon_{th}(n)$ moves towards zero. At the same time, the disturbance goes down as $O(\frac{1}{n})$.

In Appendix E we present another example of Bell inequality—generalized chain inequality—in a form of total XOR game for which we provide an optimal quantum strategy. It appears that for some range of parameters the obtained disturbance can be even greater (going down with number of observables n as $O(\frac{1}{n^{1/2+\delta}})$ for small $\delta > 0$) than in the previous two examples.

VII. COMPARISON WITH QUANTUM UNCERTAINTY

We shall now examine how much the uncertainty imposed solely by nonlocality in the no-signaling world is weaker than that implied by nonlocality in the quantum mechanical world. To this end, we use the quantum monogamy relation for the case of CHSH (for derivation see Appendix F),

$$(\beta_{\text{CHSH}})^2 + 4\left|\langle B_1^g B_1\rangle\right|^2 \leqslant 8, \tag{14}$$

which together with Eq. (6) gives the following trade-off:

$$D_q(B_2) \geqslant \tfrac{1}{2}\left(\beta_{\text{CHSH}} - \sqrt{8 - 4(2\epsilon)^2}\right), \tag{15}$$

with $\langle B_1^g B_1\rangle = 2\epsilon$. In Fig. 1 we illustrate this result for $\beta_{\text{CHSH}} = 2\sqrt{2}$ (thick line) and compare with its counterpart in no-signaling world (dashed line). One can notice that the minimal disturbance in the former case is greater than for the latter. Such behavior is expected since no-signaling constraints are in general weaker than quantum mechanical ones [26].

VIII. CONCLUSIONS

In this paper, we have developed a perceptive way of obtaining the measurement uncertainty principle from no-signaling and nonlocality. In particular, we considered a bipartite scenario where one party chooses to measure one of his observables, whereas the second party first performs a gentle measurement of one observable (gaining only partial information about the outcome) and, then, a strong measurement of another observable (where the information gain is maximal). Subsequently, assuming only impossibility of superluminal communication between two parties (i.e., the no-signaling principle) and violation of Bell inequality, we have examined a relation between information gain and disturbance implied by the very act of the gentle measurement. Our results for the case of sharp measurement (i.e., $\epsilon = \frac{1}{2}$) reproduce the extreme case discussed by Oppenheim and Wehner in [15].

Remarkably, while, as we have shown, nonlocality implies measurement uncertainty, the connection between preparation uncertainty and nonlocality is quite opposite: it has been shown [15] that preparation uncertainty excludes too strong nonlocality (cf. [27]).

Our results indicate that for general probabilistic theories obeying the no-signaling principle, the disturbance implied by statistics that can be observed in labs (i.e., the statistics predicted by quantum mechanics) is trivial until information gain reaches some threshold value of ϵ_{th}. This threshold can be shifted towards zero by considering more observables (as in the case of chain Bell inequality).

Moreover, our trade-off has the following cryptographic interpretation. Alice prepares a bipartite system and sends one subsystem to Bob. If the latter subsystem is intercepted and measured by an eavesdropper, then, at the end, Alice and Bob share a disturbed box. For this reason, our results can have potential applications in cryptography based on sending states as in BB84 protocol rather than by performing measurements on shared entangled states of unknown origin.

An open question would be to obtain the ultimate envelope describing the trade-off, i.e., to find the largest possible disturbance for a given information gain. In our work, we have found a Bell inequality that leads to disturbance partially greater than for the usual chain inequality; however, we only observed it to happen for a large number of observables n. Therefore, there still remains an open question of how to obtain the optimal Bell inequality implying the largest possible disturbance for a given information gain ϵ, irrespective of the value of n (for the whole range of n). So far our best bound for such an envelope is the one given by chain inequalities.

ACKNOWLEDGMENTS

We would like to thank P. Kurzyński, D. Lasecki, and A. Wójcik for helpful discussions. This work was supported by ERC Advanced Grant QOLAPS and by John Templeton Foundation through the Grant ID #56033. The opinions expressed in this publication are those of the authors and do not necessarily reflect the views of the John Templeton Foundation. R.R. acknowledges support from the research project "Causality in quantum theory: foundations and applications" of the Fondation Wiener-Anspach and from the Interuniversity Attraction Poles 5 program of the Belgian Science Policy Office under Grant No. IAP P7-35 photonics@be. We also thank A. Smith (University of Manchester) for his kind assistance in language editing of the manuscript.

APPENDIX A: GENTLE MEASUREMENT

In this section we provide a more detailed description of the gentle measurement of observable B_1 performed by Bob. Let us explicitly state the assumptions that the gentle measurement should satisfy. These assumptions are natural and, in particular, are satisfied by a quantum gentle measurement, as we shall see later. Suppose first that we do not measure the gentle observable, but only the sharp one. The probability distribution of the outcome is denoted by $p(b_1|B_1,a,A_x)$. Let us also consider a situation where the observable B_1 is first measured gently (denoted as B_1^g) and then sharply. Since, *a priori*, the statistics of the latter sharp measurement might be disturbed by the preceding gentle measurement, we will for a while denote its outcome by b_1'. The corresponding resulting probability distribution we denote by $\tilde{p}(b_1',b_1^g|B_1,B_1^g,a,A_x)$. We will now make two assumptions. First, we assume that the marginal probability of outcome b_1' is the same as that of b_1, i.e.,

$$\tilde{p}\big(b_1' = j\big|B_1,B_1^g,a,A_x\big) = p(b_1 = j|B_1,a,A_x), \tag{A1}$$

where $\tilde{p}(b_1'|B_1,B_1^g,a,A_x) = \sum_{b_1^g} \tilde{p}(b_1',b_1^g|B_1,B_1^g,a,A_x)$, for any state of the system (recall that various states of Bob's system are prepared by different choices of Alice's observable and by different outcomes of her measurements). Second, we assume that the conditional probability distribution computed from the above-mentioned joint probability distribution is given by

$$\tilde{p}\big(b_1^g = i\big|b_1' = j,B_1,B_1^g,a,A_x\big) = \begin{cases} \frac{1}{2}+\epsilon & \text{if } i = j, \\ \frac{1}{2}-\epsilon & \text{if } i \neq j, \end{cases} \tag{A2}$$

which is almost like Eq. (3) of the main text. The only difference is that instead of b_1 as in Eq. (3), we have b_1'. However, our first assumption implies, in particular, that joint probability distribution of b_1' with Alice's outcomes is the same as that of b_1. Thus, for all our purposes, the two random variables are indistinguishable. Hence we can drop the prime in the above conditions, obtaining Eq. (3).

We will now show that quantum measurements satisfy the above assumptions. To this end, consider a sharp measurement of B_1 described by projection operators,

$$\hat{P}_0 = |0\rangle\langle 0|, \tag{A3}$$

$$\hat{P}_1 = |1\rangle\langle 1|, \tag{A4}$$

performed on an arbitrary qubit state

$$|\Psi\rangle = \beta|0\rangle + \sqrt{1-\beta^2}|1\rangle, \tag{A5}$$

with $\beta \in \mathbb{R}, 0 \leqslant \beta \leqslant 1$, which leads to the following marginal probability distributions for outcomes $b_1 \in \{0,1\}$:

$$p(b_1 = 0|B_1) = \beta^2, \tag{A6}$$

$$p(b_1 = 1|B_1) = 1-\beta^2. \tag{A7}$$

The gentle measurement for B_1 is described by Kraus operators,

$$\hat{E}_0 = \sqrt{\tfrac{1}{2}+\epsilon}|0\rangle\langle 0| + \sqrt{\tfrac{1}{2}-\epsilon}|1\rangle\langle 1|, \tag{A8}$$

$$\hat{E}_1 = \sqrt{\tfrac{1}{2}-\epsilon}|0\rangle\langle 0| + \sqrt{\tfrac{1}{2}+\epsilon}|1\rangle\langle 1|. \tag{A9}$$

In order to show that with such definitions of sharp and gentle measurements, the two assumptions mentioned above are satisfied, we consider a procedure where the gentle measurement is followed by the sharp one.

The marginal probability distributions for outcomes $b_1^g \in \{0,1\}$ are given by

$$\begin{aligned}\tilde{p}\big(b_1^g = 0\big|B_1^g\big) &= \mathrm{Tr}(\hat{E}_0|\Psi\rangle\langle\Psi|\hat{E}_0^\dagger) \\ &= \big(\tfrac{1}{2}+\epsilon\big)\beta^2 + \big(\tfrac{1}{2}-\epsilon\big)(1-\beta^2),\end{aligned} \tag{A10}$$

$$\begin{aligned}\tilde{p}\big(b_1^g = 1\big|B_1^g\big) &= \mathrm{Tr}(\hat{E}_1|\Psi\rangle\langle\Psi|\hat{E}_1^\dagger) \\ &= \big(\tfrac{1}{2}-\epsilon\big)\beta^2 + \big(\tfrac{1}{2}+\epsilon\big)(1-\beta^2),\end{aligned} \tag{A11}$$

where $|\Psi\rangle$ is described in Eq. (A5), and $\hat{E}_0$, $\hat{E}_1$ in Eqs. (A8) and (A9).

After obtaining the outcomes $b_1^g = 0$ and $b_1^g = 1$, the postmeasurement states are given by

$$\begin{aligned}\big|\Psi_0^g\big\rangle &= \frac{\hat{E}_0|\Psi\rangle}{\sqrt{\langle\Psi|\hat{E}_0^\dagger\hat{E}_0|\Psi\rangle}} \\ &= \frac{\sqrt{\frac{1}{2}+\epsilon}\beta}{\sqrt{\big(\frac{1}{2}+\epsilon\big)\beta^2 + \big(\frac{1}{2}-\epsilon\big)(1-\beta^2)}}|0\rangle \\ &\quad + \frac{\sqrt{\frac{1}{2}-\epsilon}\sqrt{1-\beta^2}}{\sqrt{\big(\frac{1}{2}+\epsilon\big)\beta^2 + \big(\frac{1}{2}-\epsilon\big)(1-\beta^2)}}|1\rangle,\end{aligned} \tag{A12}$$

$$\begin{aligned}\big|\Psi_1^g\big\rangle &= \frac{\hat{E}_1|\Psi\rangle}{\sqrt{\langle\Psi|\hat{E}_1^\dagger\hat{E}_1|\Psi\rangle}} \\ &= \frac{\sqrt{\frac{1}{2}-\epsilon}\beta}{\sqrt{\big(\frac{1}{2}-\epsilon\big)\beta^2 + \big(\frac{1}{2}+\epsilon\big)(1-\beta^2)}}|0\rangle \\ &\quad + \frac{\sqrt{\frac{1}{2}+\epsilon}\sqrt{1-\beta^2}}{\sqrt{\big(\frac{1}{2}-\epsilon\big)\beta^2 + \big(\frac{1}{2}+\epsilon\big)(1-\beta^2)}}|1\rangle.\end{aligned} \tag{A13}$$

The second measurement is thus performed on the above postmeasurement states, and leads to the following conditional probabilities for the outcome $b_1' = 0$:

$$\begin{aligned}\tilde{p}\big(b_1' = 0\big|b_1^g = 0,B_1,B_1^g\big) &= \mathrm{Tr}\big(\hat{P}_0\big|\Psi_0^g\big\rangle\big\langle\Psi_0^g\big|\hat{P}_0^\dagger\big) \\ &= \frac{\big(\frac{1}{2}+\epsilon\big)\beta^2}{\big(\frac{1}{2}+\epsilon\big)\beta^2 + \big(\frac{1}{2}-\epsilon\big)(1-\beta^2)},\end{aligned} \tag{A14}$$

$$\begin{aligned}\tilde{p}\big(b_1' = 0\big|b_1^g = 1,B_1,B_1^g\big) &= \mathrm{Tr}\big(\hat{P}_0\big|\Psi_1^g\big\rangle\big\langle\Psi_1^g\big|\hat{P}_0^\dagger\big) \\ &= \frac{\big(\frac{1}{2}-\epsilon\big)\beta^2}{\big(\frac{1}{2}-\epsilon\big)\beta^2 + \big(\frac{1}{2}+\epsilon\big)(1-\beta^2)},\end{aligned} \tag{A15}$$

where $\hat{P}_0$ is given in Eq. (A3), and $|\Psi_0^g\rangle$ and $|\Psi_1^g\rangle$ in Eqs. (A12) and (A13).

Using $\tilde{p}(b_1' = j|B_1,B_1^g) = \sum_{i=0,1} \tilde{p}(b_1' = j|b_1^g = i,B_1,B_1^g)\tilde{p}(b_1^g = i|B_1^g)$ together with Eqs. (A10),(A11) and Eqs. (A14),(A15), we obtain that $\tilde{p}(b_1' = 0|B_1,B_1^g) = \beta^2$; hence it is equal to $p(b_1 = 0|B_1)$ of Eq. (A6). The same reasoning applies to the case of $b_1' = 1$. Therefore, we have showed that our first assumption works for quantum mechanics.

To show that our second assumption holds [i.e., that Eq. (A2) holds] we write

$$\tilde{p}\big(b_1^g = i\big|b_1' = j,B_1,B_1^g\big) = \tilde{p}\big(b_1' = j\big|b_1^g = i,B_1,B_1^g\big)\frac{\tilde{p}\big(b_1^g = i\big|B_1^g\big)}{\tilde{p}\big(b_1' = j\big|B_1,B_1^g\big)}. \quad (A16)$$

Now, replacing $\tilde{p}(b_1' = j|B_1,B_1^g)$ with $p(b_1 = j|B_1)$ (since they are equal), and inserting Eqs. (A10),(A11) and Eqs. (A14),(A15), we obtain the required identity.

APPENDIX B: DISTURBANCE

In this section we examine the relation between change of nonlocality and the disturbance caused in the system. In particular, we prove Eq. (6) from the main text given in the following form:

$$n\mathcal{D} \geqslant |\beta(p) - \beta(\tilde{p})|, \quad (B1)$$

where n is the number of Alice's observables and $\beta(p),\beta(\tilde{p})$ are the values of the Bell quantity evaluated on initial $p(a,b|A_x,B_y)$ and final statistics $\tilde{p}(a,b|A_x,B_y,B_1^g) = \sum_{b_1^g} \tilde{p}(a,b,b_1^g|A_x,B_y,B_1^g)$, respectively. We prove that the change in nonlocality, quantified by the change of an arbitrary Bell quantity (with coefficients bounded by 1), inevitably leads to nontrivial disturbance.

Proof. First, note that any Bell inequality can be written (up to a constant factor) as

$$\sum_{a,b,x,y} c(a,b,A_x,B_y)p(a,b|A_x,B_y) \leqslant \beta_{cl}, \quad (B2)$$

where

$$|c(a,b,A_x,B_y)| \leqslant 1. \quad (B3)$$

We then have

$$\begin{aligned}
|\beta(p) - \beta(\tilde{p})| &= \left|\sum_{a,b,x,y} c(a,b,A_x,B_y)\big(p(a,b|A_x,B_y) - \tilde{p}\big(a,b\big|A_x,B_y,B_1^g\big)\big)\right. \\
&\leqslant \sum_{a,b,x,y} \big|(p(a,b|A_x,B_y) - \tilde{p}\big(a,b\big|A_x,B_y,B_1^g\big)\big| \\
&= \sum_{a,b,x,y\neq 1} \big|(p(a,b|A_x,B_y) - \tilde{p}\big(a,b\big|A_x,B_y,B_1^g\big)\big| \\
&= n\sum_{y\neq 1}\left(\sum_{a,x}\frac{1}{n}p(a|A_x)\sum_b \big|p(b|B_y,a,A_x) - \tilde{p}\big(b\big|B_y,B_1^g,a,A_x\big)\big|\right) \\
&= n\sum_{y\neq 1}\left(\sum_{a,x}\frac{1}{n}p(a|A_x)D_{a,x}(B_y)\right) = n\mathcal{D},
\end{aligned} \quad (B4)$$

where in the first equality we used Eq. (B2), in the first inequality, Eq. (B3), and, in the second equality, Eq. (2) from the main text, i.e., that $p(b_1|B_1,a,A_x) = \tilde{p}(b_1|B_1,B_1^g,a,A_x),\forall a,x$. In the last equality we assume that all the choices of Alice's observable are equiprobable, i.e., $p(A_x) = \frac{1}{n}\forall x$. ■

APPENDIX C: RELEVANCE OF BELL INEQUALITIES FOR OBSERVABLE

In the main text, for a chosen observable B_1 we defined the relevance $w(B_1) \equiv w$ given by

$$w = \beta^{\max} - \beta_1^{\max}, \quad (C1)$$

with $\beta^{\max}$ standing for maximal algebraic value of the Bell quantity and $\beta_1^{\max}$ for maximal value of Bell quantity with deterministic observable B_1.

1. Monogamy relation with relevance w

In this section, we consider the situation where Alice and Bob measure $|\mathcal{X}| = n$ and $|\mathcal{Y}| = m$ number of binary observables A_x and B_y, respectively. Let us first prove the following monogamy relation (related to some Bell quantity β) whose strength is determined by the relevance w (C1)

$$\beta + w\big\langle B_1^g B_1\big\rangle \leqslant \beta^{\max}, \quad (C2)$$

where $\langle B_1^g B_1\rangle$ describes the correlations between observables B_1^g and B_1.

Proof. Let us consider the tripartite box $\tilde{p}(a,b,b_1^g|A_x,B_y,B_1^g)$ and convex decompose it as

$$\tilde{p}\big(a,b,b_1^g\big|A_x,B_y,B_1^g\big) = \sum_i r_i p_i(a,b|A_x,B_y)\otimes q_i\big(b_1^g\big|B_1^g\big), \quad (C3)$$

with $r_i \geqslant 0$, $\sum_i r_i = 1$. This can be done owing to the fact that Grace measures a single observable B_1^g. By convexity, it is sufficient to restrict the analysis to boxes of the form $p(a,b|A_x,B_y) \otimes q(b_1^g|B_1^g)$. Let us further decompose the bipartite box $p(a,b|A_x,B_y)$ shared by Alice and Bob into two types of extremal boxes. The extremal boxes in the two-party scenario for arbitrary number of inputs and binary outputs were classified in [28]. From this classification, we see that with probability p_N we have a box with fully random observable B_1, and with probability p_D, a box where the observable B_1 is deterministic. In the first case, the statistics of B_1 is fully correlated with other observables of Alice-Bob's box producing a fully random output which gives $\langle B_1^g B_1\rangle_N = 0$, whereas in the second case, the statistics of B_1 being uncorrelated with other observables is deterministic which for an appropriate choice of $q(b_1^g|B_1^g)$ gives $\langle B_1^g B_1\rangle_D = 1$. Then $\langle B_1^g B_1\rangle = p_N\langle B_1^g B_1\rangle_N + p_D\langle B_1^g B_1\rangle_D$, where $\langle B_1^g B_1\rangle_N = 0$ and $\langle B_1^g B_1\rangle_D = 1$. Therefore, $p_D = \langle B_1^g B_1\rangle$. Now, for any Bell quantity β

$$\begin{aligned}\beta \leqslant p_N\beta^{\max} + p_D\beta_1^{\max} &= \left(1-\langle B_1^g B_1\rangle\right)\beta^{\max} + \langle B_1^g B_1\rangle\beta_1^{\max}\\ &= \beta^{\max} - \left(\beta^{\max} - \beta_1^{\max}\right)\langle B_1^g B_1\rangle\end{aligned} \tag{C4}$$

and we recover (C2) with substitution (C1). ■

2. Examples of relevance w

(1) For total function XOR games with uniform probabilities of inputs, i.e., correlation Bell inequalities of binary outputs with ± 1 coefficients.

The relevance w is defined in Eq. (C1). Let us restrict the analysis to extremal boxes [28]. In order to obtain $\beta_1^{\max}$ we must consider all extremal boxes with B_1 being deterministic. In general, such boxes can have more than one deterministic observable. Suppose then that the box is defined by having k_A deterministic observables on Alice's side and k_B deterministic observables on Bob's side. In such a case, the matrix of correlators $C = \langle A_x B_y\rangle$, where $x = 1,\ldots,n$ and $y = 1,\ldots,m$, takes the form

$$C = \left[\begin{array}{c|c} \overbrace{[\beta_{\rm ns}]}^{m\text{-}k_B} & \overbrace{[0]}^{k_B} \\ \hline [0] & [\beta_{cl}] \end{array}\right] \begin{array}{l} \}\, n-k_A, \\ \}\, k_A, \end{array} \tag{C5}$$

where [0] denotes the zero matrix with respective dimensions and $[\beta_{\rm ns}]$ ($[\beta_{cl}]$) the matrix of correlators for the no-signaling (classical) part of the box. Analyzing the nonzero part of the matrix C (C5), we conclude that the Bell quantity for such a box depends on the number of deterministic observables, such that

$$\beta_1 \leqslant \max\left\{(n-k_A)(m-k_B) + k_A k_B, \beta_{cl}^{\max}\right\}. \tag{C6}$$

Notice that the value $(n-k_A)(m-k_B) + k_A k_B$ is maximized only if $k_A = n$ and $k_B = m$, in which case $k_A k_B = \beta_{cl}^{\max}$, or if $k_A = 0$ and $k_B = 1$ where the correlation matrix becomes

$$C' = \left[\begin{array}{c|c} \overbrace{[\beta_{\rm ns}]}^{m\text{-}1} & \overbrace{[0]}^{1} \end{array}\right] \Big\}\, n. \tag{C7}$$

Hence we obtain

$$\beta_1 \leqslant \max\left\{n(m-1), \beta_{cl}^{\max}\right\}. \tag{C8}$$

Eventually, substituting the RHS of Eq. (C8) to the definition of relevance w (C1), we have

$$w_{\rm tot} \geqslant \min\left(n, \beta^{\max} - \beta_{cl}^{\max}\right), \tag{C9}$$

where we derived the first term in the bracket by taking $\beta^{\max} = nm$. Note that $w_{\rm tot} = n$ for a generic total function XOR game, when the coefficient matrix C is a random Bernoulli matrix, i.e., each entry C_{ij} takes value ± 1 with probability $\frac{1}{2}$ independent of other entries. This can be seen for example from the bound on $\|\cdot\|_{\infty\to 1}$ shown in [29] which translates to the statement that, for such random XOR games, the expected classical value is bounded as

$$\beta_{cl}^{\max} \leqslant 2(n\sqrt{m} + m\sqrt{n}). \tag{C10}$$

(2) For Bell-CHSH inequality.

Directly from the value of relevance w obtained for total function XOR games in Eq. (C9) with the substitution $n = 2$, $\beta^{\max} = 4$, and $\beta_{cl}^{\max} = 2$, we obtain

$$w_{\rm CHSH} = 2. \tag{C11}$$

(3) For chain Bell inequality.

Since the box with one deterministic observable cannot violate the chain inequality, we obtain $\beta_1^{\max} = 2n - 2$. Therefore, from Eq. (C1) we get

$$w_{\rm chain} = 2 \tag{C12}$$

with the substitution $\beta^{\max} = 2n$.

APPENDIX D: INFORMATION GAIN VERSUS DISTURBANCE TRADE-OFF

In this section we prove our main result [Eq. (9) in the main text]

$$n\mathcal{D} \geqslant w\mathcal{I} - \mathcal{L}, \tag{D1}$$

where w is given by Eq. (C1), $\mathcal{I} = \langle B_1^g B_1\rangle$ denotes the information gain, and $\mathcal{L} = \beta^{\max} - \beta$ is the degree of locality.

First, let us show that

$$\mathcal{I} \equiv \left\langle B_1^g B_1\right\rangle = 2\epsilon. \tag{D2}$$

Proof.

$$\begin{aligned}\langle B_1^g B_1\rangle &= p\big(b_1^g = b_1\big) - p\big(b_1^g \neq b_1\big)\\ &= \frac{1}{2} + \epsilon - \left(\frac{1}{2} - \epsilon\right) = 2\epsilon,\end{aligned} \tag{D3}$$

where in the second equality we used the formula (3) from the main text. ■

Now, we can prove our main result (D1).

Proof.

$$n\mathcal{D} \geqslant \beta(p) - \beta(\tilde{p}) \geqslant \beta - \beta^{\max} + w2\epsilon, \tag{D4}$$

where in the first inequality we used Eq. (B1) and in the second inequality we used Eq. (C2) for $\beta = \beta(\tilde{p})$.

Therefore

$$n\mathcal{D} \geqslant w2\epsilon - (\beta^{\max} - \beta) \tag{D5}$$

and we obtain Eq. (D1) with $\mathcal{I} = 2\epsilon$ (D2) and $\mathcal{L} = \beta^{\max} - \beta$. ■

APPENDIX E: GENERALIZED CHAIN INEQUALITY

Suppose that Alice and Bob receive inputs $x,y \in [n]$ and output $a,b \in \{0,1\}$. We consider the correlation Bell inequality (partial function XOR game) $\mathcal{I}_{n,k}$ described by the coefficient matrix $C = (t_{y-x})_{x,y=1}^n$ with

$$t_l = \begin{cases} 1, & \text{if } |l| \leqslant k-1 \;\;\vee\;\; l = k,\\ -1, & \text{if } |l| \geqslant n-k+2 \;\;\vee\;\; l = -(n-k+1),\\ 0 & \text{else}\end{cases} \tag{E1}$$

for a fixed parameter $k \leqslant n/2$. The coefficient matrix thus has the following banded Toeplitz form

$$C = \begin{array}{c} k\left\{\begin{array}{c} \\ \\ \end{array}\right. \\ \\ \\ \\ \\ \\ \\ k\left\{\begin{array}{c} \\ \\ \end{array}\right. \end{array} \begin{bmatrix} \overbrace{1 \quad 1 \quad 1 \quad 1}^{k+1} & 0 & \dots & 0 & \overbrace{-1 \quad -1}^{k-1} \\ 1 \quad 1 \quad 1 \quad 1 & 1 & 0 & \dots & 0 \quad -1 \\ 1 \quad 1 \quad \dots \quad 1 & 1 & 1 & 0 & \dots \quad 0 \\ 0 \quad 1 \quad 1 \quad \dots & 1 & 1 & 1 & 0 \quad 0 \\ \vdots \quad 0 \quad 1 \quad 1 & \vdots & 1 & 1 & 1 \quad 0 \\ \cdots\cdots\cdots\cdots & \cdots & \cdots & \cdots & \cdots\cdots \\ 0 \quad \vdots \quad 0 \quad 1 & 1 & \vdots & 1 & 1 \quad 1 \\ 0 \quad \vdots \quad \vdots \quad 0 & 1 & 1 & \vdots & 1 \quad 1 \\ -1 \quad 0 \quad \dots \quad \dots & 0 & 1 & 1 & 1 \quad 1 \\ -1 \quad -1 \quad 0 \quad \dots & \dots & 0 & 1 & 1 \quad 1 \\ -1 \quad -1 \quad -1 \quad 0 & \dots & \dots & 0 & 1 \quad 1 \end{bmatrix} \tag{E2}$$

Proposition 1: The relevance $w(B_i)$ of observable B_i for the inequality $\mathcal{I}_{n,k}$ given by the coefficient matrix in (E1) with parameter $k \leqslant n/2$ is $w(B_i) = 2k$ for any $i \in [n]$. The no-signaling value of the inequality is given by $\beta_{\rm ns} = 2kn$. The quantum value of the inequality is given by

$$\beta_q = n \csc\left(\frac{\pi}{2n}\right) \sin\left(\frac{k\pi}{n}\right). \tag{E3}$$

For n divisible by k, the classical value of the inequality is given by

$$\beta_{cl} = 2k(n-k). \tag{E4}$$

Proof. Recall that the relevance $w(B_i)$ is defined by $w(B_i) = \beta^{\max} - \beta_i^{\max}$ with $\beta_i^{\max}$ being the maximum no-signaling value of the Bell quantity when observable B_i is forced to be deterministic. Now, the maximal no-signaling value of the Bell quantity is evidently equal to the maximal algebraic value (the inequality being an XOR game for which there always exists a no-signaling strategy that wins), and is given by

$$\beta_{\rm ns} = \beta^{\max} = 2kn, \tag{E5}$$

since for every input x of Alice, there are $2k$ inputs y of Bob such that the coefficients $C_{x,y}$ obey $|C_{x,y}| = 1$.

Now, we follow an analogous argument to the total function XOR games by setting observable B_i to be deterministic, and considering all the extremal no-signaling boxes from [28]. Let k_A denote the number of Alice's observables for which she returns a deterministic output in the extremal no-signaling box and let k_B denote the number of Bob's observables set to be deterministic. For $k_A,k_B \leqslant 2k$, the value achieved by this no-signaling strategy is given by

$$\beta_i^{\max} \leqslant 2k(n - k_A - k_B) + 2k_A k_B. \tag{E6}$$

The other strategy to check is the fully deterministic (classical) strategy. We claim that for n divisible by k

$$\beta_{cl} = 2kn - 2k^2. \tag{E7}$$

This value is achieved when Alice and Bob deterministically output $a,b = 0$ for all x,y.

We will prove Eq. (E7) by writing the coefficient matrix C as a sum of k^2 chain Bell expressions, each with n/k inputs so that the classical value of the individual chain expressions is $2(n/k - 1)$. Accordingly, the corresponding chain expressions are given by

$$\begin{aligned}&\sum_{i=0}^{(n/k)-2} A_{j+ik+l-1}(B_{j+ik} + B_{j+(i+1)k})\\ &\quad + A_{j+n-k+l-1}(B_{j+n-k} - B_j)\\ &\quad \leqslant 2(n/k - 1) \;\; \forall j \in [k], l \in [k],\end{aligned} \tag{E8}$$

with $A_{n+m} := -A_m$ for all $m \in [k]$. The classical value (E7) then follows from the sum of the classical value of the chain inequalities, i.e., $(k^2)(2(n/k - 1)) = 2nk - 2k^2$. Evidently, the optimal value for w is then given from (E6) by $k_A = 1, k_B = 0$ which achieves the value $2kn - 2k$ leading to $w(B_i) = 2k$.

We now show the optimal quantum strategy for the game. Consider the strategy given by measuring the state

$$|\phi_+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle) \tag{E9}$$

with observables

$$\begin{aligned}A_x &= \sin(\theta_x)\sigma_x + \cos(\theta_x)\sigma_z,\\ B_y &= \sin(\theta_y)\sigma_x + \cos(\theta_y)\sigma_z,\end{aligned} \tag{E10}$$

where σ_x,σ_z are the standard Pauli matrices and the measurement angles are given by

$$\theta_x = (x-1)\frac{\pi}{n}, \quad \theta_y = (2y-1)\frac{\pi}{2n}. \tag{E11}$$

This strategy gives the following correlations:

$$\begin{aligned} \langle A_{x+j} B_x \rangle &= \cos\left(\frac{(2j+1)\pi}{2n}\right), \\ \langle A_x B_{x+j} \rangle &= \cos\left(\frac{(2j-1)\pi}{2n}\right) \ \forall \ 0 \leqslant j \leqslant n-1. \end{aligned} \tag{E12}$$

It therefore achieves the value $\beta_q \geqslant \sum_{j=1}^{k} 2n \cos\left(\frac{(2j-1)\pi}{2n}\right)$ for the Bell quantity. Let us now show that this strategy is in fact optimal.

To do this, we show that the strategy achieves the upper bound on β_q given as $\beta_q \leqslant n\|C\|$ [24,30,31], where $\|C\|$ denotes the spectral norm, i.e., the maximal singular value of the coefficient matrix C. While C given in (E1) is a Toeplitz matrix, it is not circulant, but a "sign-flipped circulant matrix" with each row obtained from the previous row by a shift to the right and a sign change on the corresponding entry. Still, we consider as an ansatz the system of eigenvectors $|\lambda_j\rangle$ with $j \in \{0,\dots,n-1\}$ with entries

$$|\lambda_j\rangle_i = \omega_j^{n-i}, \tag{E13}$$

with $\omega_j = \exp\left(\frac{-i\pi(2j+1)}{n}\right)$. The corresponding eigenvalues of C are then given by

$$\lambda_j = \frac{\sum_{i=1}^{k+1} \omega_j^{n-i} - \sum_{i=n-k+2}^{n} \omega_j^{n-i}}{\omega_j^{n-1}}. \tag{E14}$$

It is readily seen that the eigenvalue equations are satisfied; the mth eigenvalue equation being, for $m \leqslant k-1$,

$$\left(\sum_{i=1}^{k+m} \omega_j^{n-i} - \sum_{i=n-k+m+1}^{n} \omega_j^{n-i}\right)|\lambda_j\rangle_m = \lambda_j \omega_j^{n-m}, \tag{E15}$$

which is satisfied by (E13) and (E14) by applying multiple times the identity $\exp(-i\pi(2j+1)) = -1$. Similarly, for $k \leqslant m \leqslant n-k$,

$$\sum_{i=m-k+1}^{k+m} \omega_j^{n-i} |\lambda_j\rangle_m = \lambda_j \omega_j^{n-m}, \tag{E16}$$

and for $n-k+1 \leqslant m \leqslant n$,

$$\left(\sum_{i=m-k+1}^{n} \omega_j^{n-i} - \sum_{i=1}^{m-n+k} \omega_j^{n-i}\right)|\lambda_j\rangle_m = \lambda_j \omega_j^{n-m}. \tag{E17}$$

The singular values of C are then given from (E14) by $|\lambda_j|$, so that the upper bound $n\|C\|$ is given after simplification by

$$\begin{aligned} \beta_q &\leqslant \sum_{j=1}^{k} 2n \cos\left(\frac{(2j-1)\pi}{2n}\right) \\ &= n \csc\left(\frac{\pi}{2n}\right) \sin\left(\frac{k\pi}{n}\right). \end{aligned} \tag{E18}$$

The qubit strategy achieving this bound shows that the strategy is optimal. ■

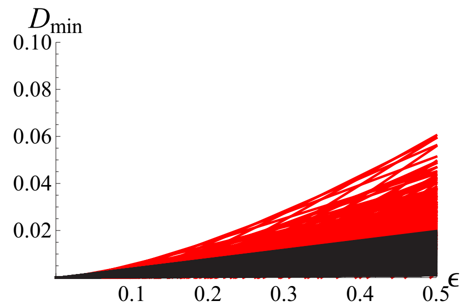

FIG. 3. The comparison of lower bound $\mathcal{D}_{\min}$ on average total disturbance implied by chain Bell inequality (red lines), $\mathcal{D}_{\min} = \frac{4}{n}\epsilon - \frac{1}{n}(2n - \beta_{\text{chain}})$, where $\beta_{\text{chain}} = 2n\cos(\frac{\pi}{2n})$, with that implied by generalized chain inequality (black lines), $\mathcal{D}_{\min} = \frac{4k\epsilon}{n} - 2[k - \sum_{j=1}^{k}\cos(\frac{(2j-1)\pi}{2n})]$ given in Eq. (E19), for $n = 100,\dots,1000$.

For the inequality given by (E1), the information gain versus disturbance trade-off is given as

$$\mathcal{D} \geqslant \frac{4k\epsilon}{n} - 2\left(k - \sum_{j=1}^{k}\cos\left(\frac{(2j-1)\pi}{2n}\right)\right). \tag{E19}$$

The second term tends to zero for appropriate choice of k. With $\cos\left(\frac{(2j-1)\pi}{2n}\right) = 1 - \left(\frac{(2j-1)\pi}{2n}\right)^2 + O\left(\frac{(2j-1)^4}{n^4}\right)$, and $\sum_{j=1}^{k}(2j-1)^2 = (4k^2-1)k/3$, we see that one may choose up to $k = O(n^{1/2-\delta})$ for any $\delta > 0$ such that $n^{2\delta} > \pi^2/(6\epsilon)$ to get a nontrivial information gain versus disturbance relation, with $\mathcal{D} = O(n^{-1/2-\delta})$.

In Fig. 3 we compare the obtained trade-off in Eq. (E19) with the trade-off for chain Bell inequality depicted in the main text in Eq. (13), and show the case where the former outperforms the latter. To this end, we choose the number of Alice's measurement choices in a range $n = 100,\dots,1000$.

APPENDIX F: QUANTUM MONOGAMY RELATION FOR CHSH INEQUALITY

Here, we prove a quantum monogamy relation for the case of CHSH in the following form [Eq. (14) in the main text]:

$$(\beta_{\text{CHSH}})^2 + 4\left|\langle B_1^g B_1\rangle\right|^2 \leqslant 8. \tag{F1}$$

Proof. To this end, we use the result of [32] that

$$\left(\beta_{\text{CHSH}}^{AB}\right)^2 + \left(\beta_{\text{CHSH}}^{BC}\right)^2 \leqslant 8, \tag{F2}$$

where

$$\beta_{\text{CHSH}}^{AB} = \langle A_1 B_1\rangle + \langle A_1 B_2\rangle + \langle A_2 B_1\rangle - \langle A_2 B_2\rangle, \tag{F3}$$

$$\beta_{\text{CHSH}}^{BC} = \langle B_1 C_1\rangle + \langle B_1 C_2\rangle + \langle B_2 C_1\rangle - \langle B_2 C_2\rangle. \tag{F4}$$

Now, let us choose $C_1 = C_2 = B_1^g$. Therefore, from (F4) we get $\beta_{\text{CHSH}}^{BC} = 2|\langle B_1 B_1^g\rangle|$. Substituting this into Eq. (F2), we obtain (F1). ■

[1] S. Popescu and D. Rohrlich, Found. Phys. **24**, 379 (1994).
[2] J. S. Bell, Physics **1**, 195 (1964).
[3] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[4] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).
[5] W. Heisenberg, Z. Phys. **43**, 172 (1927).
[6] H. Robertson, Phys. Rev. **34**, 163 (1929).
[7] S. Wehner and A. Winter, New J. Phys. **12**, 025009 (2010).
[8] I. Bialynicki-Birula and Ł. Rudnicki, Entropic uncertainty relations in quantum physics, in *Statistical Complexity: Applications in Electronic Structure*, edited by K. Sen (Springer Netherlands, Dordrecht, 2011), pp. 1–34.
[9] P. J. Coles, M. Berta, M. Tomamichel, and S. Wehner, Rev. Mod. Phys. **89**, 015002 (2017).
[10] M. Ozawa, Phys. Rev. A **67**, 042105 (2003).
[11] M. Ozawa, Phys. Lett. A **320**, 367 (2004).
[12] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, Nat. Phys. **8**, 185 (2012).
[13] P. Busch, P. Lahti, and R. F. Werner, Phys. Rev. Lett. **111**, 160405 (2013).
[14] M. Tomamichel and E. Hänggi, J. Phys. A **46**, 055301 (2013).
[15] J. Oppenheim and S. Wehner, Science **330**, 1072 (2010).
[16] C. A. Fuchs and A. Peres, Phys. Rev. A **53**, 2038 (1996).
[17] G. D'Ariano, Fortschr. Phys. **51**, 318 (2003).
[18] L. Maccone, Phys. Rev. A **73**, 042307 (2006).
[19] K. Banaszek, Open Syst. Inf. Dyn. **13**, 1 (2006).
[20] F. Buscemi, M. Hayashi, and M. Horodecki, Phys. Rev. Lett. **100**, 210504 (2008).
[21] F. Buscemi, M. J. W. Hall, M. Ozawa, and M. M. Wilde, Phys. Rev. Lett. **112**, 050401 (2014).
[22] C. H. Bennett and G. Brassard, *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing, Bangalore, India* (IEEE, New York, 1984), p. 175.
[23] C. H. Bennett, Phys. Rev. Lett. **68**, 3121 (1992).
[24] S. Wehner, Phys. Rev. A **73**, 022110 (2006).
[25] S. L. Braunstein and C. M. Caves, Ann. Phys. (NY) **202**, 22 (1990).
[26] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, Phys. Rev. A **71**, 022101 (2005).
[27] R. Ramanathan, D. Goyeneche, P. Mironowicz, and P. Horodecki, arXiv:1506.05100.
[28] N. S. Jones and L. Masanes, Phys. Rev. A **72**, 052312 (2005).
[29] A. Gittens and J. A. Tropp, arXiv:0911.4108.
[30] R. Ramanathan, R. Augusiak, and G. Murta, Phys. Rev. A **93**, 022333 (2016).
[31] N. Linden, S. Popescu, A. J. Short, and A. Winter, Phys. Rev. Lett. **99**, 180502 (2007).
[32] B. Toner and F. Verstraete, arXiv:quant-ph/0611001.