HAL
open science

Measurement-dependence cost for Bell nonlocality: causal vs retrocausal models

Michael J. W. Hall, Cyril Branciard

To cite this version:

Michael J. W. Hall, Cyril Branciard. Measurement-dependence cost for Bell nonlocality: causal vs retrocausal models. Physical Review A, 2020, 102 (5), pp.052228. 10.1103/PhysRevA.102.052228 . hal-02997720

HAL Id: hal-02997720

https://hal.science/hal-02997720v1

Submitted on 31 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Measurement-dependence cost for Bell nonlocality: Causal versus retrocausal models

Michael J. W. Hall [1] and Cyril Branciard [2]
[1] *Department of Theoretical Physics, Research School of Physics, Australian National University, Canberra ACT 0200, Australia*
[2] *Université Grenoble Alpes, CNRS, Grenoble INP, Institut Néel, 38000 Grenoble, France*

(Received 23 July 2020; accepted 19 October 2020; published 30 November 2020)

Device-independent protocols based on Bell nonlocality, such as quantum key distribution and randomness generation, must ensure no adversary can have prior knowledge of the measurement outcomes. This requires a measurement independence assumption: that the choice of measurement is uncorrelated with any other underlying variables that influence the measurement outcomes. Conversely, relaxing measurement independence allows for a fully "causal" simulation of Bell nonlocality. We construct the most efficient such simulation, as measured by the mutual information between the underlying variables and the measurement settings, for the Clauser-Horne-Shimony-Holt (CHSH) scenario, and find that the maximal quantum violation requires a mutual information of just ∼0.080 bits. Any physical device built to implement this simulation allows an adversary to have full knowledge of a cryptographic key or "random" numbers generated by a device-independent protocol based on violation of the CHSH inequality. We also show that a previous model for the CHSH scenario, requiring only ∼0.046 bits to simulate the maximal quantum violation, corresponds to the most efficient "retrocausal" simulation, in which future measurement settings necessarily influence earlier source variables. This may be viewed either as an unphysical limitation of the prior model or as an argument for retrocausality on the grounds of its greater efficiency. Causal and retrocausal models are also discussed for maximally entangled two-qubit states, as well as superdeterministic, one-sided, and zigzag causal models.

DOI: 10.1103/PhysRevA.102.052228

I. INTRODUCTION

Quantum information protocols that promise the secure distribution of cryptographic keys, or the generation of guaranteed randomness, must rely on the validity of certain physical assumptions. The strongest protocols, in the sense of requiring the weakest assumptions, are both device and theory independent—they have no reliance on internal details of preparation and measurement devices, nor even on whether quantum mechanics is valid. Instead, their promise is based on witnessing the phenomenon of Bell nonlocality, independently of the physical means by which it is generated [1,2].

In particular, if some set of statistical correlations between spacelike separated observers violates a Bell inequality [3], then any model of the correlations must either (1) permit superluminal influences between the observation regions, (2) include underlying variables that influence both the measurement outcomes and the choice of measurements, or (3) have intrinsically random and unpredictable measurement outcomes [1]. Thus, if one rules out the first two options by assumption, then intrinsic randomness is guaranteed. There can, therefore, be no eavesdropper or adversary holding a list of predetermined measurement outcomes, leading to the promised security of protocols based on such correlations [1,2,4–9].

It follows that the assumptions ruling out options (1) and (2) above require very careful scrutiny. Can observers who wish to make secure transactions trust the assertion of a device manufacturer that there are no (possibly hidden) superluminal influences nor any subtle influences on the choice of measurements ("measurement independence")—particularly if these choices are determined by random number generators supplied by the manufacturer? Or is there a loophole that an adversary can exploit to generate lists of predetermined measurement outcomes?

In regard to no superluminal influences the observers can be easily reassured: it is simply not known how to build classical devices that can influence each other superluminally (even if this might one day be possible, e.g., via some future wormhole technology). However, the case of measurement independence is not so reassuring: it is certainly possible to build devices that exhibit Bell nonlocality by violating this assumption, with no superluminal effects and with all measurement outcomes being predetermined [10–13].

Surprisingly, however, not all measurement-dependent models of Bell nonlocality can be implemented "causally"—where by "causal" we mean here that it should be the underlying variables (which we consider to originate from a source in the past) that influence the (future) choice of measurement settings, rather than the other way around. Indeed, we will show here that the most efficient models can be implemented only "retrocausally," with future measurement selections influencing past source variables. Thus, similarly to superluminal models, they cannot be implemented with known technology. We are therefore led to seek models suitable for implementing Bell nonlocality in a fully causal manner, and to find the minimum information-theoretic resources that they require. An adversary with such

 ©2020 American Physical Society

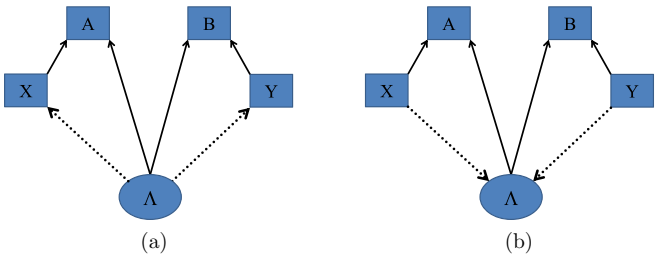

FIG. 1. Examples of causal structures, for separable models of Bell nonlocality as per Eq. (2). Each box is labeled by a corresponding random variable, with arrows indicating allowed causal influences between the boxes. Thus, the values of a source variable Λ and measurement selection variable X (Y) can causally influence measurement outcome A (B) (solid arrows). Two types of measurement dependence are indicated by the dotted arrows. (a) Causal measurement dependence: the source variable can influence the measurement selections. (b) Retrocausal measurement dependence: the source variable can be influenced by the measurement selections. Note that while causal structures do not inherently have a direction of time [21], it is convenient here (particularly when considering practical implementations of causal structures), to regard them as being embedded in a relativistic spacetime, so that the upwards and horizontal directions in the figure are timelike and spacelike, respectively. This allows "causal," "retrocausal," and "superluminal" causal influences to be unambiguously distinguished. Other examples are given in Sec. III.

resources can then build physical devices that subvert device-independent protocols.

In Sec. II we review models for Bell nonlocal correlations between two observers, the quantification of measurement dependence via mutual information, and the fundamental distinction between causal and retrocausal models of measurement dependence (see Fig. 1). In Sec. III we specialize to the Clauser-Horne-Shimony-Holt (CHSH) scenario, in which each observer has two unbiased measurement choices with binary outcomes [1,14]. We determine measurement-dependent models that require the least possible mutual information to model a given arbitrary violation of the CHSH Bell inequality, for several causal structures of interest. These models have deterministic measurement outcomes and no superluminal influences. The corresponding minimal informations are found to increase progressively for retrocausal, causal, one-sided, zigzag, and superdeterministic measurement dependence as we consider here. In particular, retrocausal measurement dependence requires strictly less mutual information than any other type. The results also verify a recent optimality conjecture made in Ref. [15].

In Sec. IV we show, by explicit construction, that biased measurement choices in the CHSH scenario reduce the minimum amount of mutual information required, in comparison to the unbiased case, for each of the causal structures considered. Hence the mutual informations calculated in Sec. III are sufficient to simulate CHSH Bell nonlocality irrespective of the probabilities with which each observer makes a given measurement choice. We also obtain upper bounds on the required mutual information for arbitrarily biased measurement choices, that approach zero in the limit of extreme bias.

In Sec. V we discuss the ordering of minimal mutual information with causal structure for maximally entangled two-qubit states. Conclusions are presented in Sec. VI.

II. SEPARABLE MODELS OF CORRELATIONS

Consider a given set of statistical correlations, represented by a set of joint probabilities $\{p(a,b|x,y)\}$, where (a,b) labels the outcomes of a joint experiment (x,y) by two parties. We will use upper case letters to denote the corresponding random variables A, B, X, Y. Any underlying model of the correlations introduces a further random variable, Λ (with specific values denoted by λ), on which the correlations depend. From Bayes' theorem we have the identity

$$p(a,b|x,y) = \sum_{\lambda} p(\lambda|x,y)\, p(a,b|x,y,\lambda), \tag{1}$$

where summation is replaced by integration if one considers a continuous range of λ. Various classes of models are defined by imposing conditions on $p(\lambda|x,y)$ and $p(a,b|x,y,\lambda)$, and have been of deep interest in quantum foundations and information since Bell's seminal paper on local hidden variable models [3].

This paper is concerned with models that satisfy the separability condition

$$p(a,b|x,y,\lambda) = p(a|x,\lambda)\, p(b|y,\lambda). \tag{2}$$

Thus, conditioned on λ, the local measurement outcomes for each party are statistically independent of each other ("outcome independence") and of the choice of measurement made by the other party ("parameter independence") [16]. We use the term "separable" rather than "local" for models satisfying Eq. (2), for reasons discussed at the end of this section.

It is well known that if, further, there is no correlation between the underlying variable λ and the joint measurement (x,y) ("measurement independence"),

$$p(\lambda|x,y) = p(\lambda), \tag{3}$$

then the model is Bell local, and the correlations $\{p(a,b|x,y)\}$ satisfy a corresponding set of Bell inequalities [1,3]. The predicted (and observed [17,18]) violation of Bell inequalities by some quantum systems implies that quantum mechanics is Bell nonlocal.

The assumption of measurement independence, as per Eq. (3), is crucial to the security of device-independent protocols in quantum information theory that rely on Bell nonlocality [1,2,4–9], as discussed in the Introduction. In particular, if measurement independence is relaxed, then there are underlying models which satisfy the separability condition (2) but which violate Bell inequalities, in principle allowing an adversary to determine the cryptographic key or random numbers generated by such protocols [12,13].

The degree of measurement dependence in a given model is conveniently quantified by the mutual information shared between the measurement settings and the underlying variable [19],

$$I(X,Y{:}\Lambda) := H(X,Y) - \sum_{\lambda} p(\lambda) H_{\lambda}(X,Y), \tag{4}$$

where $H(X,Y) := -\sum_{x,y} p(x,y)\log_2 p(x,y)$ denotes the entropy (in bits) of the distribution $p(x,y)$ of measurement settings (with summations replaced by integration over continuous ranges of X, Y and/or Λ), and $H_\lambda(X,Y)$ denotes the entropy of the conditional distribution $p(x,y|\lambda)$. The mutual information measures the information cost of the correlation between the settings (X,Y) and the underlying variable Λ, and vanishes only for models satisfying measurement independence as per Eq. (3).

It has previously been shown that no more than ~0.066 bits of measurement dependence is required for a separable model of a maximally entangled two-qubit state [20]. Surprisingly, however, we will show that the corresponding model can in general be implemented only retrocausally, i.e., with future measurement settings affecting past source variables. Thus, particularly for practical implementations, it becomes necessary to distinguish different possible causal structures for measurement-dependent models, and the information costs associated with them.

Two examples of causal structures for measurement-dependent models are depicted in Fig. 1, with the random variables denoted by corresponding boxes, and causal influences by arrows. Figures 1(a) and 1(b) correspond to what we call "causal" and "retrocausal" measurement dependence, respectively. Such structures form directed acyclic graphs [21–23], and the examples in Fig. 1 also appear in Figs. 26(b) and 27(b) of Ref. [22]. There is a natural prescription for the form of the corresponding joint probability distribution of random variables $Z_1, Z_2, \ldots,$ connected by a causal structure [21]:

$$p(Z_1, Z_2, \ldots) = \prod_j p(Z_j|\mathrm{Pa}(Z_j)), \tag{5}$$

where $\mathrm{Pa}(Z_j)$ denotes the "parents" of Z_j, i.e., those random variables with arrows directly pointing to Z_j. This prescription is motivated by Reichenbach's principle that correlations arise only from direct or common causes [24,25]. For the examples in Fig. 1 it implies, in particular, the separability condition (2). Note that it is also natural, in the implementation of quantum information protocols, for the measurement settings to appear to be selected randomly and independently. Hence, we will further typically impose the factorizability constraint

$$p(x,y) = p(x)\,p(y) \tag{6}$$

in what follows, unless explicitly indicated otherwise.

A main focus of this paper is to determine the minimum amount of measurement dependence needed for separable models of Bell nonlocality, under various causality constraints. Fortunately, the task of finding an optimal separable model is substantially simplified by noting we can, without any loss of generality, assume that the underlying outcome probabilities are deterministic, i.e., that

$$p(a,b|x,y,\lambda) \in \{0,1\}. \tag{7}$$

In particular, for any given nondeterministic model of a given set of joint correlations $\{p(a,b|x,y)\}$, one can construct a corresponding deterministic model which has the same degree of measurement dependence [20]. This construction, a generalization of a one-party model by Bell [26], has two additional variables, Λ_A and Λ_B, say, held locally by each party so as to make their response functions deterministic. These additional variables are independent of the measurement (x,y) and λ, implying that both models have the same value of the mutual information in Eq. (4), i.e., $I(X,Y{:}\Lambda,\Lambda_A,\Lambda_B) = I(X,Y{:}\Lambda)$ [27]. Thus, we can restrict attention to the class of deterministic separable models.

Finally, let us come back to our choice of terminology. We have preferred to use the neutral term "separable" for models satisfying condition (2), rather than terminology involving "local" (e.g., "measurement-dependent locality" [28]), for two main reasons. First, some models of this type can be implemented only retrocausally, as remarked above, which does not mesh with standard notions of locality. Second, there are models of this type for *signaling* correlations [28–30], i.e., for correlations which do *not* satisfy

$$p(a|x,y) = p(a|x),\ p(b|x,y) = p(b|y) \tag{8}$$

for all $a, b, x,$ and y, which leads to a conceptual tension for "locality" at the underlying and observable levels of such models (this is related to the subtlety of "signaling" for measurement-dependent models; see discussions in Refs. [29–31]).

III. OPTIMAL SEPARABLE MODELS FOR THE CHSH SCENARIO

In the CHSH scenario two observers, Alice and Bob, say, can each make one of two binary measurements. For convenience we will label their choice of measurement by $x, y = 0, 1$, and their outcomes by $a, b = \pm 1$. For any separable model, it follows from Eqs. (1) and (2) that the average joint correlation for the pair of measurements (x,y) is given by

$$\begin{aligned}\langle AB\rangle_{xy} &:= \sum_{a,b,\lambda} ab\, p(a,b,\lambda|x,y)\\ &= \sum_\lambda p(\lambda|x,y)\,A_x(\lambda)B_y(\lambda),\end{aligned} \tag{9}$$

where

$$A_x(\lambda) := \sum_a a\, p(a|x,\lambda),\ B_y(\lambda) := \sum_b b\, p(b|y,\lambda) \tag{10}$$

denote the average expectation values of a and b for a given value of λ and measurement settings x, y, and summation over λ is replaced by integration over any continuous range of λ.

If measurement independence is also satisfied, as per Eq. (3), then the correlations are Bell local and satisfy the well-known CHSH Bell inequalities, given by [1,14]

$$S := \langle AB\rangle_{00} + \langle AB\rangle_{01} + \langle AB\rangle_{10} - \langle AB\rangle_{11} \leqslant 2 \tag{11}$$

and the seven distinct permutations obtained therefrom by swapping the signs of the outcomes corresponding to one or more measurement settings. However, if measurement independence is relaxed, then there are separable models for which the CHSH parameter S in Eq. (11) can be as large as 4 [19,32], corresponding to the maximum possible algebraic value of S [obtained for $\langle AB\rangle_{xy} = (-1)^{xy}$ [33]]. We will thus be interested in $S \in [2,4]$, with any value $S > 2$ corresponding to a violation of the CHSH inequality (11). Of particular interest is the value $S_Q := 2\sqrt{2}$, which is the maximal possible value

one can reach quantum mechanically (in a standard Bell test, with an independent choice of measurement settings)—the so-called Tsirelson bound [34].

Our goal in this section is to find the minimum degree of measurement dependence required to violate the CHSH inequality (11) by a given amount, for various causal structures. As noted in the previous section, we may restrict attention to the class of deterministic models as per Eq. (12), without any loss of generality, corresponding to deterministic outcomes,

$$A_x(\lambda), B_y(\lambda) = \pm 1, \tag{12}$$

in Eq. (10). Here $A_x(\lambda)$ and $B_y(\lambda)$ can be understood as the (deterministic) "response functions" of Alice and Bob.

It is natural to further restrict attention to models that generate nonsignaling correlations, as per Eq. (8). However, this is in fact a trivial constraint in the scenario we consider here: for any given deterministic separable model, one can always construct a corresponding deterministic separable model that is nonsignaling, and which has the same values of the CHSH parameter S and mutual information I. This is achieved by taking an equal mixture of the given model with essentially the same model but with the outcomes flipped. More formally, introduce an additional (unbiased) source variable, $\Lambda' \in \{0, 1\}$, and define $p(x, y, \lambda, \lambda') := \frac{1}{2}p(x, y, \lambda)$ and

$$A_x(\lambda, \lambda') := (-1)^{\lambda'} A_x(\lambda), B_y(\lambda, \lambda') := (-1)^{\lambda'} B_y(\lambda). \tag{13}$$

This leaves the correlators $\langle AB\rangle_{xy}$ unchanged, which together with the independence of Λ' from X and Y implies that S and $I(X, Y{:}\Lambda, \Lambda') = I(X, Y{:}\Lambda)$ are invariant. Moreover, nonsignaling as per Eq. (8) is satisfied, with $p(a|x) = p(b|y) = \frac{1}{2}$. Hence, in our search for optimal models of Bell nonlocality having minimal measurement dependence we both can and will ignore the no-signaling constraint, since it can be trivially imposed via the above construction.

In the following subsections it is convenient to assume that the measurement settings appear (when not conditioned on λ) to be selected randomly and independently,

$$p(x, y) = \tfrac{1}{4}, \tag{14}$$

for all x, y (so that $H(X, Y) = -\log_2 \frac{1}{4} = 2$), consistent with Eq. (6). We will relax this condition in Sec. IV.

A. Optimal separable model with arbitrary measurement dependence

We first consider the case where no constraints are placed on measurement dependence, and determine the separable model that requires the least mutual information for a given CHSH violation. This model is related to the separable deterministic model of the singlet state given in Ref. [32] (see Sec. V below), and its optimality confirms a conjecture made in Ref. [15]. It turns out, somewhat surprisingly, that this model can be implemented only via retrocausal measurement dependence, as per Fig. 1(b).

From Bayes' theorem we have that

$$p(\lambda|x, y) = \frac{p(\lambda)p(x, y|\lambda)}{p(x, y)}. \tag{15}$$

Noting that S in Eq. (11) can be written as

$$S = \sum_{x,y} (-1)^{xy} \langle AB\rangle_{xy}, \tag{16}$$

it then follows via Eqs. (9) and (14) that

$$S = 4 \sum_{x,y,\lambda} (-1)^{xy} p(\lambda)\, p(x, y|\lambda) A_x(\lambda) B_y(\lambda). \tag{17}$$

It is convenient here to classify the possible values of λ, depending on whether $A_x(\lambda)$ takes the same value for $x = 0, 1$, and similarly for $B_y(\lambda)$. Let us thus define the sets

$$\mathfrak{L}_{\mu\nu} := \{\lambda{:}\, A_1(\lambda) = (-1)^{\mu} A_0(\lambda), B_1(\lambda) = (-1)^{\nu} B_0(\lambda)\}, \tag{18}$$

for $\mu, \nu = 0, 1$. Hence, for $\lambda \in \mathfrak{L}_{\mu\nu}$ and $x, y = 0, 1$,

$$A_x(\lambda) = (-1)^{\mu x} A_0(\lambda) \text{ and } B_y(\lambda) = (-1)^{\nu y} B_0(\lambda). \tag{19}$$

The expression for S above can then be written as

$$S = 4 \sum_{\mu,\nu} \sum_{\lambda \in \mathfrak{L}_{\mu\nu}} p(\lambda) A_0(\lambda) B_0(\lambda) \sum_{x,y} (-1)^{xy+\mu x+\nu y} p(x, y|\lambda). \tag{20}$$

We may now substitute $\sum_{x,y} (-1)^{xy+\mu x+\nu y} p(x, y|\lambda) = (-1)^{\mu\nu} \sum_{x,y} (-1)^{(x+\nu)(y+\mu)} p(x, y|\lambda) = (-1)^{\mu\nu}[1 - 2p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda)]$ (using the identity $\sum_{x,y} p(x, y|\lambda) = 1$), with $\bar{\mu} := 1-\mu$ and $\bar{\nu} := 1-\nu$, to give

$$S = 4 \sum_{\mu,\nu} \sum_{\lambda \in \mathfrak{L}_{\mu\nu}} p(\lambda) A_0(\lambda) B_0(\lambda) (-1)^{\mu\nu} \times [1 - 2p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda)]. \tag{21}$$

It immediately follows that one has the tight bound

$$S \leqslant 4 \sum_{\mu,\nu} \sum_{\lambda \in \mathfrak{L}_{\mu\nu}} p(\lambda) |1 - 2p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda)| \tag{22}$$

for S, with saturation achieved by choosing $A_0(\lambda)$ and $B_0(\lambda)$ such that

$$A_0(\lambda) B_0(\lambda) = (-1)^{\mu\nu} \operatorname{sgn}[1 - 2p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda)] \tag{23}$$

for $\lambda \in \mathfrak{L}_{\mu\nu}$. Inequality (22) is central to obtaining the main results of this and the following subsections.

Further, if $p_{\min}$ ($\leqslant \frac{1}{4}$) denotes the infimum of $p(x, y|\lambda)$ over all values of x, y, and λ, then we have $|1 - 2p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda)| \leqslant 1 - 2p_{\min}$, which yields, via Eq. (22), the tight bound

$$S \leqslant 4 \sum_{\lambda} p(\lambda)(1 - 2p_{\min}) = 4 - 8p_{\min}, \tag{24}$$

with saturation achieved by choosing

$$p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda) = p_{\min} \text{ or } 1 - p_{\min} \tag{25}$$

for $\lambda \in \mathfrak{L}_{\mu\nu}$ and the measurement outcomes as per Eq. (23). Note that Eq. (24) implies the relaxed CHSH inequalities in Eq. (4) of Ref. [13] [with $G = 1$ and $P = \max_{x,y,\lambda} p(x, y|\lambda) \geqslant (1 - p_{\min})/3$], and is equivalent to Eq. (11) of Ref. [28] (with $\ell' = p_{\min}$).

To find the minimal mutual information cost for a given value S of the CHSH parameter, note first that Eq. (24) above implies that

$$p_{\min} \leqslant (4 - S)/8. \tag{26}$$

Further, the mutual information simplifies via Eqs. (4) and (14) to

$$I(X, Y{:}\Lambda) = 2 - \sum_{\lambda} p(\lambda) H_{\lambda}(X, Y), \tag{27}$$

and hence is minimized by maximizing $H_{\lambda}(X, Y)$ for each λ, i.e., by making $p(x, y|\lambda)$ as uniform as possible. In turn, this is achieved by taking $p_{\min}(\in [0, \frac{1}{4}])$ as large as possible, i.e., saturating Eq. (26), and $p(x, y|\lambda)$ to be a distribution of the form

$$\{p_{\min}, (1 - p_{\min})/3, (1 - p_{\min})/3, (1 - p_{\min})/3\} \tag{28}$$

up to permutations [35]. Note this corresponds to taking the first choice in Eq. (25), since the second choice forces a distribution of the form $\{p_{\min}, 1 - p_{\min}, 0, 0\}$. By considering all four permutations of the distribution in Eq. (28), corresponding to the first choice in Eq. (25) for the four respective cases $\lambda \in \mathfrak{L}_{\mu\nu}$ [with each giving the same value of $H_{\lambda}(X, Y)$ in Eq. (27)], one can further ensure that the average distribution of settings $p(x, y) = \sum_{\lambda} p(\lambda) p(x, y|\lambda) = \frac{1}{4}$ as per Eq. (14): see the explicit model below.

The corresponding minimum mutual information possible, for a given value S, follows from Eq. (27) as

$$I_R(S) := 2 - h\left(\frac{4 - S}{8}\right) - \frac{4 + S}{8} \log_2 3, \tag{29}$$

where

$$h(p) := -p \log_2 p - (1 - p) \log_2 (1 - p) \tag{30}$$

is the binary entropy function. Here the subscript R stands for "retrocausal" because, as will be seen, this minimum value can be obtained only by retrocausal models. This is plotted in Fig. 2 and ranges from 0 bits for $S = 2$ (no violation of the CHSH Bell inequality) to $\log_2 \frac{4}{3} \sim 0.415$ bits for $S = 4$ (maximum algebraic violation). For the maximum quantum violation, $S_Q = 2\sqrt{2}$, one has

$$I_R(S_Q) \sim 0.046 \text{ bits}, \tag{31}$$

i.e., just a little more than 1/22 of a bit of measurement dependence is required to reach the Tsirelson bound.

The simplest optimal separable model reaching the minimal mutual information of Eq. (29), satisfying the saturating conditions (23) and (25), with equality in Eq. (26) and with uniform settings x, y as in Eq. (14), is obtained by taking the four sets $\mathfrak{L}_{\mu\nu}$ to contain just one element $\lambda_{\mu\nu}$ each, with equal probabilities $p(\lambda_{\mu\nu}) = 1/4$. This model is given explicitly in Table I. It is equivalent to the separable model in Tables I and II of Ref. [15] (with $p_1 = p_2 = (1 - 4p)/3$ and $p_3 = 0$), which was conjectured to be optimal; our analysis above thus allows us to prove that conjecture. Note that this model gives signaling correlations; but as explained earlier with reference to Eq. (13), it can easily be extended to a (still optimal) nonsignaling model that gives the same CHSH value S for the same amount of measurement dependence.

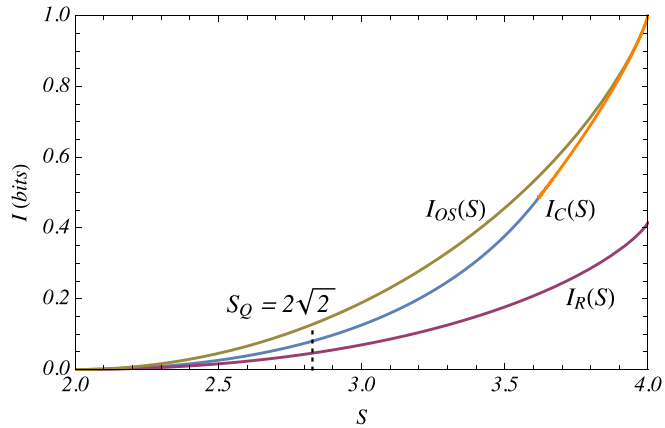

FIG. 2. Plot of the minimal mutual information between source and measurement setting variables required to reach a value $S \in [2, 4]$ of the CHSH parameter, for various causal structures. Retrocausal models are bounded below by the magenta curve [$I_R(S)$ in Eq. (29)], causal models by the blue and orange curve [$I_C(S)$ in Eq. (43); see also Fig. 4 further below], and one-sided models by the olive curve [$I_{\rm OS}(S)$ in Eq. (51)]. Note that $I_R(S) \leqslant I_C(S) \leqslant I_{\rm OS}(S)$, i.e., retrocausal models are more efficient than causal models, which in turn are more efficient than one-sided models. The value $S_Q = 2\sqrt{2}$ indicates the Tsirelson bound, i.e., the maximal quantum violation of the CHSH inequality (with independent measurement settings).

There is clearly a retrocausal implementation of the model in Table I (and indeed of any measurement-dependent model), as per Fig. 1(b), where the source receives the values x and y of X and Y with some prior factorized probability $p(x, y) = p(x)p(y)$ as per Eq. (6), and generates $\Lambda = \lambda$ with probability $p(\lambda|x, y) = p(x, y|\lambda)p(\lambda)/p(x, y)$ [36]. However, it is of interest to ask whether the model in Table I also

TABLE I. Optimal separable model, with the minimal required measurement dependence required to obtain a given value of the CHSH parameter S. The probabilities $p_{xy|\lambda} := p(x, y|\lambda)$ are defined for $\lambda = \lambda_{\mu\nu}$ such that $p_{xy|\lambda} = p \in [0, 1/4]$ if $x = \bar{\nu}$ and $y = \bar{\mu}$, and $p_{xy|\lambda} = \frac{1-p}{3} \in [\frac{1}{4}, \frac{1}{3}]$ otherwise. Thus, $p \equiv p_{\min}$ in Eq. (28). With $p(\lambda) \equiv \frac{1}{4}$, the average probability distribution of the joint measurement setting (x, y) is $p(x, y) = \sum_{\lambda} p(\lambda) p_{xy|\lambda} = \frac{1}{4}$, as required by Eq. (14), and the "retrocausal" conditional probabilities $p(\lambda|x, y)$ are $p(\lambda|x, y) = p(x, y|\lambda)p(\lambda)/p(x, y) = p_{xy|\lambda}$. The parameters s, t, u, v defining the measurement outcomes of the model can take any value ± 1. These response functions are defined, in accordance with Eqs. (19) and (23) (noticing that all $p_{xy|\lambda} \leqslant \frac{1}{2}$), such that $A_x(\lambda_{\mu\nu}) B_y(\lambda_{\mu\nu}) = (-1)^{\mu x + \nu y + \mu\nu}$. One thus obtains $\langle AB \rangle_{xy} = \sum_{\lambda} p(\lambda|x, y) A_x(\lambda) B_y(\lambda) = (-1)^{xy}(1 - 2p)$. The corresponding value of the CHSH parameter is $S = 4 - 8p$, with corresponding mutual information given by Eq. (29). As shown in this paper, this model has inherent retrocausal measurement dependence.

λ	$p(\lambda)$	$p_{00\|\lambda}$	$p_{01\|\lambda}$	$p_{10\|\lambda}$	$p_{11\|\lambda}$	$A_0(\lambda)$	$A_1(\lambda)$	$B_0(\lambda)$	$B_1(\lambda)$
λ_{00}	1/4	$\frac{1-p}{3}$	$\frac{1-p}{3}$	$\frac{1-p}{3}$	p	s	s	s	s
λ_{10}	1/4	$\frac{1-p}{3}$	$\frac{1-p}{3}$	p	$\frac{1-p}{3}$	t	$-t$	t	t
λ_{01}	1/4	$\frac{1-p}{3}$	p	$\frac{1-p}{3}$	$\frac{1-p}{3}$	u	u	u	$-u$
λ_{11}	1/4	p	$\frac{1-p}{3}$	$\frac{1-p}{3}$	$\frac{1-p}{3}$	v	$-v$	$-v$	v

has a causal implementation, with the future measurement settings influenced by past source variables? As shown in the next section, the answer is negative: the model in Table I is *inherently retrocausal*.

B. Optimal separable model with causal measurement dependence

The structure of causal measurement-dependent models is shown in Fig. 1(a), where the value of the source variable λ can causally influence the subsequent selection of measurements by Alice and Bob. Thus, all arrows in Fig. 1(a) are implemented in timelike directions. Such models are characterized by the condition

$$p(x, y|\lambda) = p(x|\lambda)p(y|\lambda), \tag{32}$$

following from the prescription in Eq. (5). This is analogous to the separability condition in Eq. (2) (see also Ref. [13]). Note that no further generality is gained by writing

$$p(x, y|\lambda) = \sum_{\lambda'} p(\lambda'|\lambda)p(x|\lambda, \lambda')p(y|\lambda, \lambda'), \tag{33}$$

via an additional source variable Λ' (which is always formally possible), as this merely generates a correlation model of the same form as Eqs. (9) and (32) with respect to the extended source variable $\Lambda'' := (\Lambda, \Lambda')$.

We call the condition in Eq. (32) "causal measurement dependence" (it has also been called "independent sources" [31]). It is a nontrivial constraint, leading to the requirement of a higher degree of mutual information to achieve a given Bell inequality violation than is the case for general separable models. In particular, it will be seen that the optimal model in Table I cannot be implemented via causal measurement dependence, but requires *retrocausal* measurement dependence as per Fig. 1(b) (or, among the other causal structures we shall consider, the supplemented "zigzag" measurement dependence of Fig. 7 below, which also involves a retrocausal influence).

To determine the minimum mutual information $I(X, Y{:}\Lambda)$ required to simulate a violation of the CHSH inequality, under causal measurement dependence, necessitates a little more work than for the general case in Sec. III A. First, following the previous analysis, one obtains a tight bound as per Eq. (22) for the CHSH parameter as before, which is still saturated by deterministic measurement outcomes satisfying Eq. (23). Second, denoting by $p^X_{\min}(\lambda)$ the minimal value of $p(x|\lambda)$ for $x = 0, 1$ and by $p^Y_{\min}(\lambda)$ the minimal value of $p(y|\lambda)$ for $y = 0, 1$ (both for a given λ), we have, under the assumption that $p(x, y|\lambda)$ decomposes as in Eq. (32), that $p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda), 1 - p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda) \geqslant p^X_{\min}(\lambda)\, p^Y_{\min}(\lambda)$, so that $|1 - 2p(x{=}\bar{\nu}, y{=}\bar{\mu}|\lambda)| \leqslant 1 - 2p^X_{\min}(\lambda)\, p^Y_{\min}(\lambda)$. Substituting this into Eq. (22) then leads to the tight bound

$$S \leqslant \bar{S}_{\max} := \sum_{\lambda} p(\lambda)S_{\max}(\lambda)$$
$$\text{with} \quad S_{\max}(\lambda) := 4 - 8\, p^X_{\min}(\lambda)\, p^Y_{\min}(\lambda), \tag{34}$$

for causal models [rather than Eq. (24) for general models], with saturation achieved when

$$p(x{=}\bar{\nu}|\lambda) = p^X_{\min}(\lambda), \quad p(y{=}\bar{\mu}|\lambda) = p^Y_{\min}(\lambda) \tag{35}$$

for $\lambda \in \mathfrak{L}_{\mu\nu}$.

The mutual information then follows via Eqs. (4), (14), and (32) as

$$I(X, Y{:}\Lambda) = 2 - \sum_{\lambda} p(\lambda)[H_\lambda(X) + H_\lambda(Y)] = \sum_{\lambda} p(\lambda)I(\lambda), \tag{36}$$

where

$$I(\lambda) := 2 - h\big(p^X_{\min}(\lambda)\big) - h\big(p^Y_{\min}(\lambda)\big). \tag{37}$$

Note that by letting one party introduce some local noise, one can obtain any value of S between 0 and $\bar{S}_{\max}$ above, without changing the mutual information. This implies, in particular, that the minimal mutual information required to get a given value S increases with S, and that it is indeed obtained by saturating the upper bound in Eq. (34), i.e., by taking $S = \bar{S}_{\max} = \sum_\lambda p(\lambda)S_{\max}(\lambda)$. This, together with Eq. (36), then implies that the optimal pairs of values (S, I) are obtained as convex combinations of optimal pairs $(S_{\max}(\lambda), I(\lambda))$, for some fixed λ [and with the weights $p(\lambda)$ in the combination taken so as to satisfy $p(x, y) = \sum_\lambda p(\lambda)p(x, y|\lambda) = \frac{1}{4}$ as in Eq. (14)].

Note that both $S_{\max}(\lambda)$ and $I(\lambda)$ are expressed above as functions of $p^X_{\min}(\lambda)$ and $p^Y_{\min}(\lambda)$, which we shall simply denote here by p_X and p_Y, respectively, so as to lighten the notations: $S_{\max}(\lambda) = S_{\max}(p_X, p_Y) = 4 - 8p_X p_Y$ and $I(\lambda) = I(p_X, p_Y) = 2 - h(p_X) - h(p_Y)$. To find the optimal pairs $(S_{\max}(\lambda), I(\lambda))$, our goal is thus to calculate

$$I_C(S) := \min_{p_X, p_Y \in [0, \frac{1}{2}]} I(p_X, p_Y)$$
$$\text{s.t. } S_{\max}(p_X, p_Y) = S. \tag{38}$$

For this we introduce a Lagrange multiplier κ and define the Lagrangian

$$\mathcal{L}(p_X, p_Y, \kappa) := I(p_X, p_Y) - \kappa\,[S_{\max}(p_X, p_Y) - S]. \tag{39}$$

Setting $\partial\mathcal{L}/\partial p_X = \partial\mathcal{L}/\partial p_Y = 0$ and eliminating κ gives $f(p_X) = f(p_Y)$, where the function

$$f(p) := p \log_2 \tfrac{1-p}{p} \tag{40}$$

is plotted in Fig. 3, for $p \in [0, \frac{1}{2}]$.

As one can see, for any given value of p_X there are generally two solutions to $f(p_X) = f(p_Y)$: (1) $p_Y = p_X$ and (2) $p_Y = p^*_X$, where p^*_X denotes the abscissa of the second point of intersection of the curve $f(p)$ with the horizontal line passing through $(p_X, f(p_X))$. The first solution gives, via Eq. (34), $S_{\max}(\lambda) = 4 - 8\,(p_X)^2 \in [2, 4]$, and yields via Eq. (37) the associated mutual information $I(\lambda) = I_1(S_{\max}(\lambda))$, with

$$I_1(S) := 2 - 2h\left(\sqrt{\frac{4-S}{8}}\right). \tag{41}$$

For the second solution, one has $S_{\max}(\lambda) = 4 - 8\, p_X\, p^*_X$ from Eq. (34). We find that this can give only values $S_{\max}(\lambda) \geqslant S_0 \sim 3.620$, where $S_0 := 4 - 8p_0^2$ is obtained for the value $p_0 \sim 0.218$ that gives the maximum of the function $f(p)$ [i.e., the solution of $\frac{df}{dp} = \log_2\left(\frac{1-p}{p}\right) - \frac{1}{(1-p)\log_e 2} = 0$, such that $p^*_0 = p_0$; see Fig. 3]. For a value of $S \geqslant S_0$, the solutions $p_X(S)$ and $p^*_X(S)$ to $S = 4 - 8p_X\, p^*_X$ must in general (except for the extremal values $S = S_0$ and $S = 4$) be found

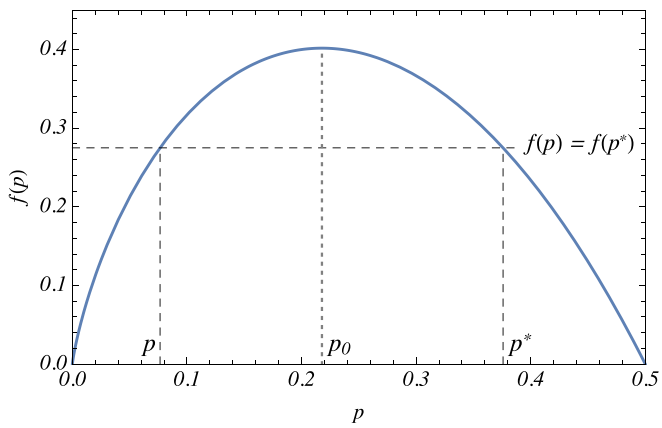

FIG. 3. Function $f(p) := p \log_2 \frac{1-p}{p}$ plotted for $p \in [0, \frac{1}{2}]$. There are in general two different values p, p^* giving the same $f(p) = f(p^*)$.

numerically. The corresponding mutual information is then, according to Eq. (37), $I(\lambda) = I_2(S_{\max}(\lambda))$, where

$$I_2(S) := 2 - h(p_X(S)) - h(p_X^*(S)). \tag{42}$$

The curves $I_1(S)$ (defined for any $S \in [2, 4]$) and $I_2(S)$ (defined only for $S \geqslant S_0$) are plotted in Fig. 4, for the relevant values of S. One can see that $I_2(S) \leqslant I_1(S)$ for $S \geqslant S_0$, so that the second solution to our optimization problem provides a lower mutual information in that range. Furthermore, both $I_1(S)$ and $I_2(S)$ are seen to be convex, and one can prove that they share the same tangent at $S = S_0$ (see the Appendix), so that $\min[I_1(S), I_2(S)]$ is also convex. Taking convex combinations of $S_{\max}(\lambda)$ and $I(\lambda)$ as in Eqs. (34) and (36) hence does not provide a lower mutual information. We will see with the explicit model below, however, that one can still combine different λs having the same values of $S_{\max}(\lambda)$ and $I(\lambda)$, so as to satisfy $p(x, y) = \sum_\lambda p(\lambda)p(x, y|\lambda) = \frac{1}{4}$ as required.

We conclude from this analysis that the minimum mutual information possible for a given value S of the CHSH parameter, via a causally measurement-dependent separable model,

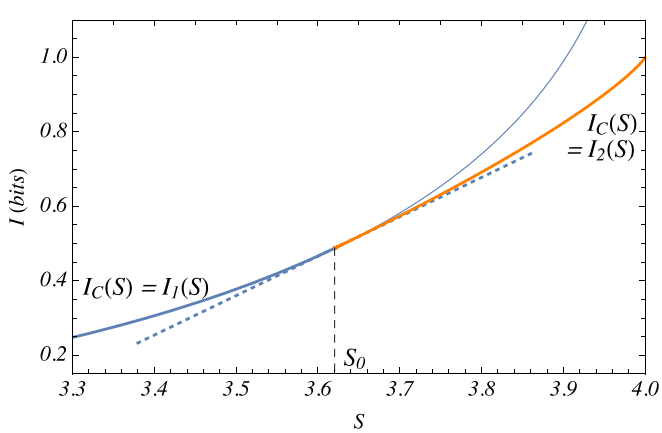

FIG. 4. Plot of the functions $I_1(S)$ [blue curve, Eq. (41)], $I_2(S)$ [orange curve, Eq. (42), defined for $S \geqslant S_0 \sim 3.620$], and $I_C(S)$ [Eq. (43)], for S in the region around S_0. The dotted line represents the common tangent to $I_1(S)$ and $I_2(S)$ at $S = S_0$. A full plot of $I_C(S)$ is given in Fig. 2.

TABLE II. Optimal separable model with minimal causal measurement dependence, for any value $S \in [2, 4]$ of the CHSH parameter. The conditional probabilities $p(x|\lambda)$ and $p(y|\lambda)$ are defined, in terms of the parameter $p \in [0, \frac{1}{2}]$, such that for $\lambda = \lambda_{\mu\nu}$, $p(x{=}0|\lambda) = 1-p$ if $\nu = 0$, $p(x{=}0|\lambda) = p$ otherwise, and $p(y{=}0|\lambda) = 1-\tilde{p}$ if $\mu = 0$, $p(y{=}0|\lambda) = \tilde{p}$ otherwise [as in Eq. (35)]. Here $\tilde{p} := p$ if one aims at obtaining $S \leqslant S_0$ (which includes any quantum violation), while $\tilde{p} := p^*$ for $S \geqslant S_0$ (as defined in this paper). The values of $p(x, y|\lambda)$ may be determined from $p(x{=}0|\lambda)$ and $p(y{=}0|\lambda)$ via Eq. (32) [and with $p(x(y){=}1|\lambda) = 1 - p(x(y){=}0|\lambda)$]. The average probability distribution of x, y is $p(x, y) = \sum_\lambda p(\lambda)p(x, y|\lambda) = \frac{1}{4}$, as required. The outcome parameters $s, t, u, v = \pm 1$ are taken as in Table I, such that $A_x(\lambda_{\mu\nu})B_y(\lambda_{\mu\nu}) = (-1)^{\mu x+\nu y+\mu\nu}$. Noting that $p(\lambda|x, y) = p(x, y|\lambda)p(\lambda)/p(x, y) = p(x|\lambda)p(y|\lambda)$ here, one obtains $\langle AB\rangle_{xy} = (-1)^{xy}(1 - 2p\tilde{p})$, and in turn $S = 4 - 8p\tilde{p}$, with corresponding mutual information given by Eq. (43).

λ	$p(\lambda)$	$p(x{=}0\|\lambda)$	$p(y{=}0\|\lambda)$	$A_0(\lambda)$	$A_1(\lambda)$	$B_0(\lambda)$	$B_1(\lambda)$
λ_{00}	1/4	$1-p$	$1-\tilde{p}$	s	s	s	s
λ_{10}	1/4	$1-p$	$\tilde{p}$	t	$-t$	t	t
λ_{01}	1/4	p	$1-\tilde{p}$	u	u	u	$-u$
λ_{11}	1/4	p	$\tilde{p}$	v	$-v$	$-v$	v

is given by

$$I_C(S) = \begin{cases} I_1(S) & \text{for } 2 \leqslant S \leqslant S_0, \\ I_2(S) & \text{for } S_0 \leqslant S \leqslant 4. \end{cases} \tag{43}$$

This is plotted in Fig. 2, together with $I_R(S)$ from Eq. (29). One finds that

$$I_C(S) > I_R(S) \tag{44}$$

for all $S > 2$. Thus, as claimed earlier, the optimal separable model in Table I [whose input distribution $p(x, y|\lambda)$ indeed does not satisfy the causality condition of Eq. (32)] cannot be implemented via causal measurement dependence.

An explicit optimal model that reaches the lower bound $I_C(S)$ above is given in Table II. Again, this model is signaling, but can easily be turned into a nonsignaling causal model which gives the same value of the CHSH parameter S, as per the construction in Eq. (13), with the same amount of measurement dependence.

Note that the maximum quantum violation, $S_Q = 2\sqrt{2}$, is strictly less than S_0. Hence, for any quantum violation the minimum mutual information can be calculated analytically from Eq. (41). One finds in particular that a minimum value

$$I_C(S_Q) = I_1(S_Q) \sim 0.080 \text{ bits} \tag{45}$$

is required for any fully causal model reproducing the maximal quantum violation. Note this is nearly twice as much as $I_R(S_Q)$ in Eq. (31). Further, a fully causal model for the maximal algebraic value $S = 4$ requires a mutual information of $I_C(4) = I_2(4) = 1$ bit [obtained for $p_X = 0$, $p_X^* = \frac{1}{2}$ in Eq. (42)], i.e., more than twice as much as the value $I_R(4) = \log_2 \frac{4}{3} \sim 0.415$ bits required for a retrocausal model.

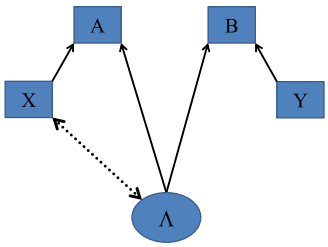

FIG. 5. Two-party Bell scenario with one-sided measurement dependence (dotted line), which may be either causal or retrocausal.

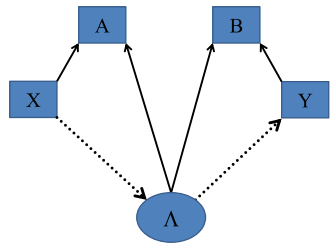

FIG. 6. Two-party Bell scenario with zigzag measurement dependence (dotted arrows). As discussed in this paper, this is formally equivalent to causal measurement dependence as per Fig. 1(a) (and it reduces to one-sided measurement dependence, as per Fig. 5, if only one dotted arrow is permitted).

C. Optimal separable model with one-sided measurement dependence

We now consider the case of one-sided measurement dependence, where the source variable is correlated with just one of the measurement selections, as depicted in Fig. 5. Taking this correlation to be on Alice's side, it follows that Y is independent of X and Λ, yielding

$$p(x, y|\lambda) = p(x|\lambda)\, p(y) \tag{46}$$

for one-sided models. Thus, one-sided models are formally a special case of causal models as per Eq. (32), implying they will, in general, require a greater degree of measurement dependence to achieve a given CHSH violation. Note that one-sided models can equivalently be implemented either causally, via the earlier source variable λ influencing the later measurement selection x, or retrocausally, with x influencing λ, as indicated by the double arrowhead in Fig. 5.

For the CHSH scenario, assuming unbiased measurement selection probabilities as per Eq. (14), summation over x in Eq. (46) gives

$$p(y|\lambda) = p(y) = \tfrac{1}{2}. \tag{47}$$

Hence, we have $p^Y_{\min}(\lambda) = \min_y p(y|\lambda) = \frac{1}{2}$, and can directly use Eq. (34) for causal models to write the maximum possible value of the CHSH parameter as

$$S \leqslant \sum_\lambda p(\lambda)\big[4 - 4\, p^X_{\min}(\lambda)\big] \tag{48}$$

[with saturation achieved when $p(x{=}\bar{\nu}|\lambda) = p^X_{\min}(\lambda)$ for $\lambda \in \mathfrak{L}_{\mu\nu}$], so that

$$\sum_\lambda p(\lambda) p^X_{\min}(\lambda) \leqslant 1 - S/4. \tag{49}$$

The mutual information on the other hand is given via Eq. (46) by

$$\begin{aligned} I(X, Y{:}\Lambda) = I(X{:}\Lambda) &= 1 - \sum_\lambda p(\lambda)\, h\big(p^X_{\min}(\lambda)\big) \\ &\geqslant 1 - h\left(\sum_\lambda p(\lambda)\, p^X_{\min}(\lambda)\right) \geqslant 1 - h(1 - S/4), \end{aligned} \tag{50}$$

where the first inequality follows from the concavity of the binary entropy function h and is saturated when $p^X_{\min}(\lambda)$ is independent from λ, while the second one follows from Eq. (49) above, and the fact that $h(p)$ increases monotonically for $p \in [0, \frac{1}{2}]$. We thus find that the minimal required mutual information for one-sided measurement dependence is

$$I_{\rm OS}(S) := 1 - h\left(\frac{S}{4}\right). \tag{51}$$

A corresponding optimal model [which further satisfies $p(x) = \sum_\lambda p(\lambda) p(x|\lambda) = \frac{1}{2}$, as required to recover Eq. (14)], for all values of $S \in [2, 4]$, is obtained by replacing $\tilde{p}$ with $\frac{1}{2}$ in Table II.

The mutual information in Eq. (51) is plotted in Fig. 2 and is seen to be larger than both $I_R(S)$ and $I_C(S)$, as expected. For the maximum quantum violation $S_Q = 2\sqrt{2}$ one finds in particular that

$$I_{\rm OS}(S_Q) \sim 0.128 \text{ bits}. \tag{52}$$

Note that this is approximately half of the mutual information of $\sim$0.247 bits required in the Banik *et al.* one-sided model [15,37] (although the latter model is optimal for the measure of measurement independence M introduced in Ref. [32]). For the algebraic maximum $S = 4$, one obtains $I_{\rm OS}(4) = 1$ bit, which is the same value as $I_C(4) = 1$ bit for optimal causal models.

D. Zigzag measurement dependence

A further type of causal structure that has been considered for the explanation of quantum correlations is zigzag causality, as depicted in Fig. 6. This type of causal structure was introduced by Costa de Beauregard [38,39] and has recently been examined by Price and Wharton [40]. As shown in Fig. 6, one of the measurement selections, Alice's, say, influences the source variable, which can in turn influence the other measurement selection.

Applying the prescription in Eq. (5) to zigzag measurement dependence yields $p(x, y, \lambda) = p(x)p(y|\lambda)p(\lambda|x) = p(x, \lambda)p(y|\lambda)$, reflecting the lack of a direct causal influence from X to Y in Fig. 6. Thus, dividing by $p(\lambda)$,

$$p(x, y|\lambda) = p(x|\lambda)\, p(y|\lambda). \tag{53}$$

Comparing with Eq. (32) (and since these are, in both cases, the only further constraints imposed on separable models), it follows that zigzag measurement dependence is formally equivalent to causal measurement dependence, and so may be analyzed precisely as in Sec. III B. In particular, for the CHSH scenario with $p(x, y) = \frac{1}{4}$, the corresponding minimal

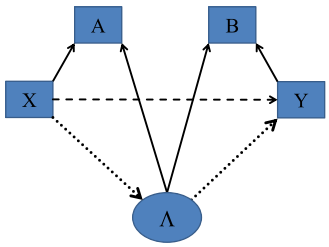

FIG. 7. Zigzag measurement dependence supplemented with a superluminal influence from X to Y (dashed line). As discussed in this paper, this scenario is formally equivalent to retrocausal measurement dependence as per Fig. 1(b).

information is given by

$$I_Z(S) := I_C(S), \tag{54}$$

for zigzag models, where $I_C(S)$ is defined in Eq. (43). We note this equivalence can be seen as strengthening de Beauregard's analogy between zigzag causality and particle-antiparticle pair creation [38,39], for which causal propagation of the pair is formally equivalent to a single particle first propagating retrocausally and then causally.

It is also of interest to note that if one supplements zigzag measurement dependence with a superluminal influence from X to Y, as depicted in Fig. 7, then the prescription in Eq. (5) places no restrictions on the form of $p(x, y, \lambda)$. Hence, this modification is formally equivalent to the general case discussed in Sec. III A, which we have in turn seen is formally equivalent to retrocausal measurement dependence as per Fig. 1(b). For the CHSH scenario this implies that adding a superluminal arrow to zigzag measurement dependence as per Fig. 7 decreases the minimal mutual information required, from $I_Z(S)$ to $I_R(S)$, where the latter is defined in Eq. (29).

E. Superdeterministic models

Superdeterminism is a rather strong form of measurement dependence, in which both the measurement settings and measurement outcomes are fully determined by underlying variables [41–45]. The latter variables themselves must still retain a statistical quality (corresponding to ignorance of initial conditions), for superdeterministic models to be able to reproduce quantum predictions. The causal structure of a superdeterministic model corresponds, therefore, to causal measurement dependence as in Fig. 1(a), but with the values of each of A, B, X, and Y being fully determined by the value of Λ.

For the CHSH scenario it follows that $p(x, y|\lambda) \in \{0, 1\}$ for superdeterministic models. Hence, the mutual information of any such model follows via Eq. (4) as

$$I(X, Y{:}\Lambda) = H(X, Y) - \sum_\lambda p(\lambda) H_\lambda(X, Y) = H(X, Y). \tag{55}$$

Thus, for $p(x, y) = 1/4$ as per Eq. (14), the minimum mutual information required to achieve any value S of the CHSH parameter is given by

$$I_{\mathrm{SD}}(S) := 2 \text{ bits} \tag{56}$$

for superdeterministic models. It follows that such models have the highest possible degree of measurement dependence [since $I(X, Y{:}\Lambda) \leqslant H(X, Y) \leqslant 2$ for any model of the CHSH scenario].

More generally, it is straightforward to write down a superdeterministic and separable model for any prior distribution $p(x, y)$ and any set of statistical correlations $\{p(a, b|x, y)\}$, which may or may not satisfy the no-signaling condition (8), by generalizing the outcome deterministic model in Eq. (11.24) of Ref. [30] to the superdeterministic case. In particular, define $\lambda := (\alpha, \beta, \xi, \zeta)$, where $\alpha, \beta, \xi, \zeta$ range over the possible values of a, b, x, y, respectively, and define

$$\begin{aligned} p(\lambda) &:= p(a{=}\alpha, b{=}\beta|x{=}\xi, y{=}\zeta)\, p(x{=}\xi, y{=}\zeta), \\ p(a, b, x, y|\lambda) &:= \delta_{a,\alpha}\, \delta_{b,\beta}\, \delta_{x,\xi}\, \delta_{y,\zeta} \end{aligned} \tag{57}$$

(where δ denotes the Kronecker or Dirac delta as appropriate). The desired correlations are then easily recovered via $p(x, y) = \sum_{\lambda,a,b} p(\lambda) p(a, b, x, y|\lambda)$ and $p(a, b|x, y) = \sum_\lambda p(\lambda) p(a, b, x, y|\lambda)/p(x, y)$.

Finally, note that the type of superdeterminism we considered here might more precisely be referred to as "causal superdeterminism," since all causal influences are from the past to the future, as in Fig. 1(a). One could also consider, for example, "retrocausal superdeterminism," in which the past source variable Λ is fully determined by the future measurement selections X and Y. However, we will not pursue this possibility here.

IV. BIASED MEASUREMENT CHOICES IN THE CHSH SCENARIO

The analysis of the minimum information costs of measurement dependence in the CHSH scenario, for different causal structures, was restricted in Sec. III to the case of unbiased measurement choices, i.e., to $p(x, y) = \frac{1}{4}$ as per Eq. (14). However, one may also consider different distributions for X and Y. A question then immediately arises: if Alice and Bob are allowed to make any choice of $p(x, y)$, how does the information cost behave?

A simple example suggests that this cost will go down, i.e., that unbiased measurement choices correspond to the worst-case scenario. In particular, starting with the form of the causal model given in Table II, define a new model by replacing each of p and $\tilde{p}$ by 0 and the column for $p(\lambda)$ by the distribution $\{q^2, q(1{-}q), q(1{-}q), (1-q)^2\}$, for some $0 < q < 1$. This yields a causal model having the maximum possible value of the CHSH parameter, $S = 4$ (note that we cannot take q strictly equal to 0 or 1, as otherwise not all settings are possible and one cannot calculate S). Further, by construction, this model has $p(x = 0) = p(y = 0) = q$ for the distributions of Alice and Bob's measurement choices, and a mutual information $I(X, Y{:}\Lambda) = H(X, Y) = H(\Lambda) = 2h(q)$. Thus, as $p(x)$ and $p(y)$ become more biased, i.e., as q approaches 0 or 1, the mutual information required for this model becomes arbitrarily small, while still maximally violating the CHSH inequality.

We will show here, more generally, that unbiased measurement choices do indeed correspond to the worst-case scenario: for any factorizable distribution $p(x, y) = p(x)p(y)$ as per

Eq. (6), the minimal informations calculated in Sec. III are sufficient to achieve any given value of the CHSH parameter, for each of the causal structures considered [46]. We also obtain upper bounds on the amount of mutual information required for arbitrary $p(x)$ and $p(y)$, which allow us to significantly strengthen the above example by showing that the minimal mutual information required can always be made arbitrarily close to zero in the limit of highly biased choices, for any value of S.

A. Less mutual information is needed for biased choices

To demonstrate the results above, we construct explicit models, for any $p(x,y) = p(x)p(y)$, that require no more mutual information to implement than the unbiased models for retrocausal, causal, zigzag, one-sided, and superdeterministic measurement dependence in Sec. III. These constructions are closely related to the latter models, and make use of an entropic property peculiar to those models. In particular, we will rely on the following lemma.

Lemma. If two measurement-dependent models, M and M', with the same range of X, Y, Λ, satisfy $p_{M'}(\lambda|x,y) = p_M(\lambda|x,y)$ for all λ, x, y, and the entropy of $p_M(\lambda|x,y)$ is independent of x and y, then their mutual informations satisfy

$$I_{M'}(X,Y{:}\Lambda) = I_M(X,Y{:}\Lambda) + H_{M'}(\Lambda) - H_M(\Lambda), \tag{58}$$

where $H(\Lambda)$ denotes the entropy (in bits) of Λ.

Proof. Equation (4) for mutual information can be rewritten as $I(X,Y:\Lambda) = H(\Lambda) - \sum_{x,y} p(x,y)H_{xy}(\Lambda)$, where $H_{xy}(\Lambda)$ denotes the entropy of $p(\lambda|x,y)$. Hence, under the assumptions of the lemma, one has

$$\begin{aligned}
&I_{M'}(X,Y:\Lambda) - I_M(X,Y:\Lambda)\\
&\quad= H_{M'}(\Lambda) - H_M(\Lambda)\\
&\qquad- \sum_{x,y}[p_{M'}(x,y)H_{M',xy}(\Lambda) - p_M(x,y)H_{M,xy}(\Lambda)]\\
&\quad= H_{M'}(\Lambda) - H_M(\Lambda)\\
&\qquad- \sum_{x,y}[p_{M'}(x,y) - p_M(x,y)]H_{M,xy}(\Lambda)\\
&\quad= H_{M'}(\Lambda) - H_M(\Lambda)\\
&\qquad- H_{M,x_0y_0}(\Lambda)\sum_{x,y}[p_M(x,y) - p_{M'}(x,y)]\\
&\quad= H_{M'}(\Lambda) - H_M(\Lambda),
\end{aligned} \tag{59}$$

as desired, where x_0 and y_0 are arbitrary values in the ranges of X and Y. ■

We first consider the retrocausal model in Table I, which has mutual information $I_R(S)$ as per Eq. (29) for a given value of the Bell parameter S. It is convenient to denote the various probabilities and other quantities appearing in this model by the subscript R. Thus, for example, $p_R(x,y) = \frac{1}{4}$. We then construct, for an arbitrary prior distribution $p_{R'}(x,y) = p_{R'}(x)p_{R'}(y)$, a corresponding model R' defined by $p_{R'}(\lambda|x,y) := p_R(\lambda|x,y)$, $A_{R',x}(\lambda) := A_{R,x}(\lambda)$, and $B_{R',y}(\lambda) := B_{R,y}(\lambda)$. Note that all other properties of R' can be calculated via Bayes' theorem, e.g., $p_{R'}(\lambda) = \sum_{x,y} p_{R'}(x,y)p_{R'}(\lambda|x,y) = \sum_{x,y} p_{R'}(x)p_{R'}(y)p_R(\lambda|x,y)$. It follows immediately from Eqs. (9) and (11) that these models have the same value of the Bell parameter S, i.e., $S = S_{R'} = S_R$ ($= 4 - 8p$ from Table I). Further, the conditions of the lemma are satisfied (with $H_{R,xy}(\Lambda) = h(p) + (1-p)\log_2 3$ for all x, y and $p_R(\lambda) = \frac{1}{4}$), yielding

$$I_{R'}(X,Y{:}\Lambda) = I_R(S) + H_{R'}(\Lambda) - 2 \leqslant I_R(S), \tag{60}$$

noting that the entropy of $p_{R'}(\lambda)$ is bounded above by two bits (since λ takes only four possible values). Thus, as claimed, the minimum mutual information required to implement any given violation of the CHSH inequality via a retrocausal model is never greater than $I_R(S)$ in Eq. (29), irrespective of the choice of $p(x)$ and $p(y)$.

We proceed similarly for causal and zigzag models. In particular, consider the optimal causal model in Table II, and denote all quantities appearing in this model via the subscript C. A corresponding model C' with arbitrary prior distribution $p_{C'}(x,y) = p_{C'}(x)p_{C'}(y)$ is then defined via $p_{C'}(\lambda|x,y) := p_C(\lambda|x,y)$, $A_{C',x}(\lambda) := A_{C,x}(\lambda)$, and $B_{C',y}(\lambda) := B_{C,y}(\lambda)$. Note that this model is also causal, as it follows via repeated applications of Bayes' theorem, and using $p_{C'}(\lambda|x,y) = p_C(\lambda|x,y)$ and $p_C(x,y|\lambda) = p_C(x|\lambda)p_C(y|\lambda)$, that

$$\begin{aligned}
p_{C'}(x,y|\lambda) &= \frac{p_{C'}(\lambda|x,y)p_{C'}(x)p_{C'}(y)}{p_{C'}(\lambda)}\\
&= \frac{p_C(\lambda)}{p_{C'}(\lambda)}\,\frac{p_C(x|\lambda)p_{C'}(x)}{p_C(x)}\,\frac{p_C(y|\lambda)p_{C'}(y)}{p_C(y)}\\
&= p_{C'}(x|\lambda)\,p_{C'}(y|\lambda),
\end{aligned} \tag{61}$$

in agreement with Eq. (32) [47]. It follows that $S = S_{C'} = S_C$ ($= 4 - 8p\tilde{p}$ from Table II) and, noting the conditions of the lemma are satisfied [with $H_{C,xy}(\Lambda) = h(p) + h(\tilde{p})$ and $p_C(\lambda) = \frac{1}{4}$], that

$$I_{C'}(X,Y:\Lambda) = I_C(S) + H_{C'}(\Lambda) - 2 \leqslant I_C(S) \tag{62}$$

for any value of the CHSH parameter S. Thus, in analogy to the retrocausal case above, no more than $I_C(S)$ in Eq. (43) is required to implement any given violation of the CHSH inequality via causal measurement dependence, irrespective of the choice of $p(x)$ and $p(y)$. A similar result immediately follows for zigzag measurement dependence via the equivalence discussed in Sec. III D.

For the case of one-sided causal measurement dependence we again consider the model in Table II, but for the choice $\tilde{p} = \frac{1}{2}$ (see Sec. III C). Denoting the quantities in this model by the subscript OS, a corresponding model OS' with arbitrary prior distribution $p_{OS'}(x,y) = p_{OS'}(x)p_{OS'}(y)$ is then defined via $p_{OS'}(\lambda|x,y) := p_{\rm OS}(\lambda|x)$ [implying OS' is also one-sided as per Eq. (46)], and $A_{OS',x}(\lambda) := A_{OS,x}(\lambda)$, $B_{OS',y}(\lambda) := B_{OS,y}(\lambda)$. It follows again from Eqs. (9) and (11) that $S = S_{OS'} = S_{\rm OS}$ ($= 4 - 4p$ from Table II with $\tilde{p} = \frac{1}{2}$). Further, the conditions of the lemma are satisfied [with $H_{OS,xy}(\Lambda) = 1 + h(p)$], yielding

$$I_{OS'}(X,Y{:}\Lambda) = I_{\rm OS}(S) + H_{OS'}(\Lambda) - 2 \leqslant I_{\rm OS}(S) \tag{63}$$

for any value of the CHSH parameter S. Thus, no more than $I_{\rm OS}(S)$ in Eq. (51) is always sufficient for a one-sided model of Bell nonlocality in the CSHS scenario.

Finally, for superdeterministic models it follows trivially from Eq. (55) that any such model, SD', say, with prior distri-

bution $p_{SD'}(x,y)$, requires a mutual information

$$I_{SD'}(X,Y{:}\Lambda) = H_{SD'}(X,Y) \leqslant 2 \text{ bits} = I_{\rm SD}(S), \tag{64}$$

independently of the value of S, where the inequality is an immediate consequence of $H(X,Y) \leqslant 2$ for any joint distribution $p(x,y)$, and $I_{\rm SD}(S)$ is defined in Eq. (56). Thus, again, the unbiased prior $p(x,y) = \frac{1}{4}$ is the worst-case scenario.

B. Explicit bounds on mutual information for biased choices

The left-hand sides of Eqs. (60) and (62)–(64) can be explicitly evaluated for any given distributions $p(x)$ and $p(y)$, via calculation of the corresponding distribution $p(\lambda)$ and entropy $H(\Lambda)$. This provides corresponding upper bounds on the mutual informations required to implement a given violation of the CHSH Bell inequality for retrocausal, causal (or zigzag), one-sided, and superdeterministic measurement dependence, respectively. We explicitly calculate these bounds here and show that they approach zero for sufficiently biased $p(x)$ and $p(y)$.

It is convenient to define the biases of the distributions $p(x)$ and $p(y)$ via

$$\epsilon_X := p(x=0) - p(x=1),\ \epsilon_Y := p(y=0) - p(y=1), \tag{65}$$

respectively. Thus, $\epsilon_X, \epsilon_Y \in [-1,1]$ (although for the calculation of S we cannot take ϵ_X, ϵ_Y strictly equal to ± 1, as otherwise not all settings are possible). Further, for the specific models in Tables I and II we have $p(\lambda|x,y) = p(x,y|\lambda)$, yielding

$$\begin{aligned} p_R(\lambda_{\mu\nu}|x,y) &= \frac{1-p}{3} + \delta_{x,\bar\nu}\,\delta_{y,\bar\mu}\,\frac{4p-1}{3}, \\ p_C(\lambda_{\mu\nu}|x,y) &= \frac{1+(-1)^{x+\nu}(1-2p)}{2}\,\frac{1+(-1)^{y+\mu}(1-2\tilde p)}{2}, \\ p_{\rm OS}(\lambda_{\mu\nu}|x,y) &= \frac{1+(-1)^{x+\nu}(1-2p)}{2} \end{aligned} \tag{66}$$

(recall that the latter is obtained from the preceding case by taking $\tilde p = \frac{1}{2}$). The distributions of Λ for the models R', C', and OS' can then be calculated, using $p_{M'}(\lambda) = \sum_{x,y} p(x)p(y)p_M(\lambda|x,y)$, as

$$\begin{aligned} p_{R'}(\lambda_{\mu\nu}) &= \frac{1-p}{3} + \frac{1-(-1)^\nu\epsilon_X}{2}\,\frac{1-(-1)^\mu\epsilon_Y}{2}\,\frac{4p-1}{3}, \\ p_{C'}(\lambda_{\mu\nu}) &= \frac{1+(-1)^\nu\epsilon_X(1-2p)}{2}\,\frac{1+(-1)^\mu\epsilon_Y(1-2\tilde p)}{2}, \\ p_{OS'}(\lambda_{\mu\nu}) &= \frac{1+(-1)^\nu\epsilon_X(1-2p)}{4}, \end{aligned} \tag{67}$$

with corresponding entropies

$$\begin{aligned} H_{R'}(\Lambda) &= H\left(\frac{1-p}{3} + \frac{1\pm\epsilon_X}{2}\,\frac{1\pm\epsilon_Y}{2}\,\frac{4p-1}{3}\right), \\ H_{C'}(\Lambda) &= h\left(\frac{1+\epsilon_X[1-2p]}{2}\right) + h\left(\frac{1+\epsilon_Y[1-2\tilde p]}{2}\right), \\ H_{OS'}(\Lambda) &= 1 + h\left(\frac{1+\epsilon_X[1-2p]}{2}\right). \end{aligned} \tag{68}$$

Substitution of these entropies into Eqs. (60), (62), and (63) leads to explicit expressions for the mutual informations $I_{R'}(X,Y{:}\Lambda)$, $I_{C'}(X,Y{:}\Lambda)$ and $I_{OS'}(X,Y{:}\Lambda)$, as desired. For example, using Eq. (29) and recalling that $S = 4 - 8p$ for the retrocausal model in Table I, we have

$$\begin{aligned} I_{R'}(X,Y{:}\Lambda) = H&\left(\left\{\frac{4+S}{24} + \frac{1\pm\epsilon_X}{2}\,\frac{1\pm\epsilon_Y}{2}\,\frac{2-S}{6}\right\}\right) \\ &- h\left(\frac{4-S}{8}\right) - \frac{4+S}{8}\log_2 3. \end{aligned} \tag{69}$$

Similarly, one finds

$$\begin{aligned} I_{C'}(X,Y{:}\Lambda) = h&\left(\frac{1+\epsilon_X[1-2p]}{2}\right) - h(p) \\ &+ h\left(\frac{1+\epsilon_Y[1-2\tilde p]}{2}\right) - h(\tilde p) \end{aligned} \tag{70}$$

(where $S = 4 - 8p\tilde p$), and

$$I_{OS'}(X,Y{:}\Lambda) = h\left(\frac{1+\epsilon_X[S/2-1]}{2}\right) - h(S/4). \tag{71}$$

One also has directly from Eqs. (64) and (65) that

$$I_{SD'}(X,Y{:}\Lambda) = h\left(\frac{1+\epsilon_X}{2}\right) + h\left(\frac{1+\epsilon_Y}{2}\right). \tag{72}$$

These mutual informations reduce to $I_R(S)$, $I_C(S)$, $I_{\rm OS}(S)$, and $I_{\rm SD}(S)$ for the unbiased case, $\epsilon_X = \epsilon_Y = 0$, as expected. Equations (69)–(72) also provide upper bounds for the minimum mutual informations required to implement retrocausal, causal one-sided, and superdeterministic models, respectively, for any given biases ϵ_X, ϵ_Y and CHSH parameter S [48]. It may be verified that they decrease monotonically to zero as $\epsilon_X, \epsilon_Y \to \pm 1$. Thus, an arbitrarily small amount of mutual information is required in the limit of extreme bias.

V. GENERAL AND SINGLET-STATE MODELS

In Sec. III we determined optimal models for Bell nonlocality in the CHSH scenario, under various causal constraints on measurement dependence, and found the ordering

$$I_R(S) < I_C(S) = I_Z(S) < I_{\rm OS}(S) < I_{\rm SD}(S) \tag{73}$$

for the minimum informations required to model any given value $S \in (2,4)$ of the CHSH parameter. Recall that the subscripts denote retrocausal, causal, zigzag, one-sided, and superdeterministic measurement dependence, respectively.

It would be of interest to determine whether an analogous ordering holds more generally, beyond the CHSH scenario. This however appears to be a difficult problem, as optimization of models becomes harder for increasing numbers of inputs and outputs (including for continuous ranges). In contrast, nevertheless, it is straightforward to demonstrate the result

$$I_R \leqslant I_C = I_Z \leqslant I_{\rm OS} \leqslant I_{\rm SD}, \tag{74}$$

for separable models of any given set of joint correlations $\{p(a,b|x,y)\}$, where I_C, for example, denotes the minimum mutual information required to generate the set $\{p(a,b|x,y)\}$ under the constraint of causal measurement dependence. This ordering is an immediate logical consequence of the definitions of the corresponding types of measurement dependence in Sec. III. In particular, all separable models

have retrocausal implementations [36]; causal and zigzag measurement dependence are formally equivalent; one-sided measurement dependence is formally equivalent to a special case of causal measurement dependence; and superdeterministic models have the maximum possible value of mutual information.

Hence, the source of the difficulty in generalizing Eq. (73) lies in determining whether a strict ordering obtains. In the remainder of this section we consider the case of separable models of spin measurements on maximally entangled two-qubit states, which allow at least two possible measurement selections for each observer, and argue that

$$I_R^{\rm me} < I_C^{\rm me} \leqslant I_{\rm OS}^{\rm me} < I_{\rm SD}^{\rm me} \tag{75}$$

for this case. In particular, we give evidence that the first inequality is strict for this case, and formally prove strictness for the last inequality.

First, note that we can restrict attention to the singlet state, since all other such states differ by local rotations, corresponding to a simple relabelling of the measurement settings [where the mutual information in Eq. (4) is invariant under any such relabeling, including for continuous ranges of measurement settings X, Y and of Λ]. For the singlet state, the measurement selections x and y correspond to spin directions on the unit sphere; the measurement results $a, b = \pm 1$ correspond to spin "up" and spin "down"; and any separable model must reproduce the correlations

$$p(a, b|x, y) = \tfrac{1}{4}(1 - ab\, x \cdot y) \tag{76}$$

for some prior distribution $p(x, y)$.

Second, there is a known separable model of spin correlations for the singlet state [32], with a mutual information no greater than $\sim$0.066 bits for any choice of $p(x, y)$ [20]. Since this model trivially has a retrocausal implementation [36], it follows that

$$I_R^{\rm me} \lesssim 0.066 \text{ bits} \tag{77}$$

for any $p(x, y)$. For the special case of the CHSH scenario, in which x and y are each restricted to two orthogonal directions, $x = x_0, x_1$ and $y = y_0, y_1$, lying in a common plane with y_0 bisecting x_0 and x_1, and with prior distribution $p(x, y) = 1/4$, it can be checked that this model reduces to the model in Table I with $p = \frac{1-1/\sqrt{2}}{2}$ and $S = S_Q$ [15], and has the same mutual information, $I_R(S_Q) \sim 0.046$ bits in Eq. (31) [15,20]. It follows via Eq. (44) that, for the CHSH scenario at least, this singlet-state model can be implemented only retrocausally, suggesting more generally that $I_R^{\rm me} < I_C^{\rm me}$ as per the first inequality in Eq. (75).

Third, there is a known one-sided separable model of spin correlations for the singlet state, with a mutual information no greater than $\log_2(2/\sqrt{e})$ bits for any choice of $p(x, y)$ [19,30,49]. Thus, using Eq. (74),

$$I_C^{\rm me} \leqslant I_{\rm OS}^{\rm me} \leqslant \log_2 \frac{2}{\sqrt{e}} \sim 0.279 \text{ bits}. \tag{78}$$

Fourth, for any superdeterministic model of the singlet state the settings are fully determined by λ, implying that $p(x, y|\lambda) = \delta_{x, f_X(\lambda)}\, \delta_{y, f_Y(\lambda)}$ for two functions f_X and f_Y. Hence, if $p(x, y)$ is supported on a set of nonzero measure, then $H_\lambda(X, Y) = -\infty$ and $H(X, Y) \leqslant 2\log_2(4\pi)$, yielding $I_{\rm SD}^{\rm me} = \infty$ via Eq. (4). Alternatively, if $p(x, y)$ is supported only on a discrete set of directions (e.g., as in the CHSH scenario), then $H_\lambda(X, Y) = 0$, yielding $I_{\rm SD}^{\rm me} = H(X, Y)$. But for the latter case $I_{\rm OS} \leqslant H(X) < H(X) + H(Y) = H(X, Y)$ for any one-sided model, where the strict inequality follows since equality can hold only for the trivial case $H(Y) = 0$, i.e., only one possible measurement selection for Bob (and we have assumed that the factorizability constraint of Eq. (6) is satisfied). Hence, for either alternative we have

$$I_{\rm OS}^{\rm me} < I_{\rm SD}^{\rm me} \tag{79}$$

as per the final inequality in Eq. (75).

It would be interesting to find a separable model of the singlet state with causal measurement dependence and a mutual information lying strictly between the values in Eq. (77) and (78). Such a model would provide evidence that the central inequality in Eq. (75) is in fact also strict. Finally, note that the above models can be easily generalized to models of spin measurements on noisy singlet states ("Werner states" [50]), by mixing them with a model having random outcomes independently of the measurement directions.

VI. CONCLUSIONS

We have determined the minimum mutual information required to implement separable models of Bell nonlocality, under various causal constraints on measurement dependence, for the case of unbiased measurement choices in the CHSH scenario, (Sec. III). This leads to a monotonic ordering of the information-theoretic resources required to implement each of retrocausal, causal, zigzag, one-sided, and superdeterministic measurement dependence, as per Eq. (73) (see also Fig. 2). In particular, retrocausal models require strictly less mutual information to implement than causal models, for any given violation of the CHSH Bell inequality. A similar result holds for models of maximally entangled two-qubit states (Sec. V). It follows that some measurement-dependent models in the literature have no causal implementation [15,32].

The underlying reason for why causal measurement dependence, as per Fig. 1(a), inherently requires more mutual information than retrocausal measurement dependence, as per Fig. 1(b), is the independence condition (32) for the former case, corresponding physically to causal propagation of influences from the source to the measurement devices. In general, the more independence conditions imposed by a causal structure [21], the higher the required mutual information is expected to be.

It is worth noting that while there is no known means of implementing retrocausal models as in Fig. 1(b) (with the depicted timelike and spacelike separations), the result that such models require strictly less resources than causal models is of some theoretical interest in itself. In particular, the relative efficiency of such models for simulating Bell nonlocality provides a further argument for retrocausality, in addition to existing arguments in the literature [51–57].

We have also constructed corresponding optimal models for each of the considered causal constraints, which have fully deterministic measurement outcomes and no superluminal signaling (Sec. III). Thus, an implementation of any of these models subverts the security of device-independent

information protocols based on Bell nonlocality in the CHSH scenario, with an adversary able to in principle have full knowledge of a cryptographic key or sequence of random numbers generated by such a protocol (see also Sec. I). This is of particular interest for the case of causal measurement dependence, for which the corresponding models can be easily implemented in practice by an adversarial device manufacturer (see Sec. III B) and which require at most $I_C(S_Q) \sim 0.080$ bits of mutual information to achieve the maximum quantum violation of the CHSH inequality, as per Eqs. (45) and (62).

We have further constructed explicit separable models that achieve violation of the CHSH Bell inequality by any specified amount, for any biased distribution of measurement settings $p(x, y) = p(x)p(y)$, and used these models to show that the case of unbiased settings, $p(x, y) = \frac{1}{4}$, has the highest information cost irrespective of the causal structure, and that the information cost approaches zero in the case of extreme bias (Sec. IV). It would be of interest to improve on the upper bounds given by $I_{R'}(X, Y{:}\Lambda)$, $I_{C'}(X, Y{:}\Lambda)$, and $I_{OS'}(X, Y{:}\Lambda)$ in Eqs. (69)–(71), by calculating the minimum possible mutual information required for a given choice of $p(x)$ and $p(y)$.

We note that the above models can easily be extended to include some further measurement choices on either side (or both), whose corresponding outputs are perfectly correlated (or anticorrelated) with some other outputs, e.g., to add $x = 2$ with $A_2(\lambda) := B_0(\lambda)$. Such correlations are useful in establishing cryptographic keys [1,12,58], and it would be of interest to calculate the corresponding mutual information requirements for such extended models, under various causal constraints and for relevant distributions of settings $p(x, y)$, and to determine the maximum key rates possible for a given degree of measurement dependence.

It would also be of significant interest, as noted in Sec. V, to determine whether there is a causal separable model for the spin correlations of maximally entangled two-qubit states that requires strictly less mutual information than any one-sided separable model.

Finally, note that the various causal structures considered in this paper (Figs. 1, 5, 6, and 7) differ only in the causal relations between X, Y, and Λ: in all cases A and B are directly influenced by X and Λ, and by Y and Λ, respectively—subject to the separability condition of Eq. (2). A possible direction for future work would be to also consider structures with different causal relations involving A and B, e.g., to also consider retrocausal influences from A and B to X, Y and/or Λ.

ACKNOWLEDGMENTS

We thank N. Gisin for early discussions. M.H. is grateful to V Vijendran for discussions and for raising the question of zigzag causality.

APPENDIX: PROPERTIES OF $I_1(S)$ AND $I_2(S)$

Here we give a partial formal proof of some properties of the curves $I_1(S)$ and $I_2(S)$ depicted in Fig. 4. In particular, we show that these curves have a common tangent at the intersection point $S = S_0$, that $I_1(S)$ is convex, and that convexity of $I_2(S)$ corresponds to the monotonicity of $f(p)/(4 - S)$ with respect to S, which can be verified numerically.

Note first from Eqs. (34) and (37) that

$$I_1'(S) = \frac{(d/dp)[2 - 2h(p)]}{(d/dp)[4 - 8p^2]} = \frac{h'(p)}{8p}, \tag{A1}$$

and that, using the identities $f(p) = p\,h'(p)$ and $f(p) = f(p^*)$ (and considering p^* to be a function of p),

$$\begin{aligned} I_2'(S) &= \frac{(d/dp)[2 - h(p) - h(p^*)]}{(d/dp)[4 - 8pp^*]} = \frac{h'(p) + p^{*'}h'(p^*)}{8(p^* + pp^{*'})} \\ &= \frac{f(p)/p + p^{*'}f(p^*)/p^*}{8(p^* + pp^{*'})} = \frac{f(p)}{8pp^*} = \frac{h'(p)}{8p^*}. \end{aligned} \tag{A2}$$

Hence, it directly follows for $p = p^* = p_0$ that $I_1'(S_0) = I_2'(S_0) = \frac{h'(p_0)}{8p_0} \sim 1.059$, i.e., that the curves $I_1(S)$ and $I_2(S)$ in Fig. 4 have a common tangent at $S = S_0$. Note the right-hand sides of Eqs. (A1) and (A2) are positive, verifying that $I_1(S)$ and $I_2(S)$, and accordingly $I_C(S)$ in Eq. (43), are monotonically increasing with S.

Via a similar calculation as in Eq. (A1), one further finds that

$$I_1''(S) = \frac{1}{128p^3}\left[\log_2 \frac{1-p}{p} + \frac{1}{(1-p)\log_e 2}\right] > 0 \tag{A3}$$

(for $0 < p \leqslant \frac{1}{2}$), which shows that $I_1(S)$ is convex.

Finally, using $S = 4 - 8pp^*$ for $S \geqslant S_0$, it follows from Eq. (A2) that the curve $I_2(S)$ is convex for $S \geqslant S_0$ if and only if $f(p)/(4 - S)$ is a monotonic increasing function of S (where p is obtained from S as the solution to $4 - 8pp^* = S$, which in general has to be found numerically). This can be verified via a numerical plot. Alternatively, calculating $I_2''(S)$ in a similar way to Eq. (A2) one finds

$$\begin{aligned} I_2''(S) &= -\frac{(d/dp)\left[\frac{f(p)}{pp^*}\right]}{64(d/dp)[pp^*]} \\ &= \frac{1}{64(pp^*)^2}\frac{(p^* + pp^{*'})f(p) - pp^*f'(p)}{p^* + pp^{*'}}, \end{aligned} \tag{A4}$$

which can be numerically verified to be positive (one finds that both numerator and denominator in the last fraction above are positive for $p < p_0$, and negative for $p > p_0$).

[1] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Rev. Mod. Phys. **86**, 419 (2014).
[2] A. Ekert and R. Renner, Nature (London) **507**, 443 (2014).
[3] J. S. Bell, Physics **1**, 195 (1964).

[4] D. Mayers and A. Yao, in *Proc. of the 39th IEEE Symposium on Foundations of Computer Science* (IEEE Computer Society, Los Alamitos, CA, 1998), p. 503.

[5] J. Barrett, L. Hardy, and A. Kent, Phys. Rev. Lett. **95**, 010503 (2005).

[6] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).

[7] S. Pironio, A. Acín, S. Massar, A. B. de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, Nature (London) **464**, 1021 (2010).

[8] R. Colbeck and A. Kent, J. Phys. A: Math. Theor. **44**, 095305 (2011).

[9] V. Scarani, *Bell Nonlocality* (Oxford University Press, Oxford, 2019).

[10] J. S. Bell, J. F. Clauser, M. A. Horne, and A. Shimony, Dialectica **39**, 85 (1985).

[11] C. H. Brans, Int. J. Theor. Phys. **27**, 219 (1988).

[12] J. Kofler, T. Paterek, and Č. Brukner, Phys. Rev. A **73**, 022104 (2006).

[13] D. E. Koh, M. J. W. Hall, Setiawan, J. E. Pope, C. Marletto, A. Kay, V. Scarani, and A. Ekert, Phys. Rev. Lett. **109**, 160404 (2012).

[14] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).

[15] A. S. Friedman, A. H. Guth, M. J. W. Hall, D. I. Kaiser, and J. Gallicchio, Phys. Rev. A **99**, 012121 (2019).

[16] A. Shimony, in *Quantum Concepts in Space and Time*, edited by R. Penrose and C. J. Isham (Clarendon Press, Oxford, 1986), pp. 182–203.

[17] A. Aspect, J. Dalibard, and G. Roger, Phys. Rev. Lett. **49**, 1804 (1982).

[18] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán *et al.*, Nature (London) **526**, 682 (2015).

[19] J. Barrett and N. Gisin, Phys. Rev. Lett. **106**, 100406 (2011).

[20] M. J. W. Hall, Phys. Rev. A **84**, 022102 (2011).

[21] J. Pearl, *Causality: Models, Reasoning, and Inference*, 2nd ed. (Cambridge University Press, Cambridge, 2009).

[22] C. J. Wood and R. W. Spekkens, New. J. Phys. **17**, 033002 (2015).

[23] R. Chaves, R. Kueng, J. B. Brask, and D. Gross, Phys. Rev. Lett. **114**, 140403 (2015).

[24] H. Reichenbach, *The Direction of Time* (University of California Press, Berkeley, 1956).

[25] B. C. Van Fraassen, Synthese **52**, 25 (1982).

[26] J. S. Bell, Rev. Mod. Phys. **38**, 447 (1966).

[27] A similar result holds for any good quantifier of measurement-dependent correlations, including the variational distance quantifier M introduced in Ref. [32].

[28] G. Pütz, D. Rosset, T. J. Barnea, Y.-C. Liang, and N. Gisin, Phys. Rev. Lett. **113**, 190402 (2014).

[29] L. P. Thinh, L. Sheridan, and V. Scarani, Phys. Rev. A **87**, 062121 (2013).

[30] M. J. W. Hall, The significance of measurement independence for Bell inequalities and locality, in *At the Frontier of Spacetime*, edited by T. Asselmeyer-Maluga (Springer, Cham, 2016), Chap. 11.

[31] G. Pütz and N. Gisin, New J. Phys. **18**, 055006 (2016).

[32] M. J. W. Hall, Phys. Rev. Lett. **105**, 250404 (2010).

[33] P. Rastall, Found. Phys. **15**, 963 (1985).

[34] B. S. Csirel'son, Lett. Math. Phys. **4**, 93 (1980).

[35] More explicitly, if one writes $p(x, y|\lambda) = \{p, (1-p)q_1, (1-p)q_2, (1-p)q_3\}$, with $p := p_{\min}$, $q_i \geqslant 0$, and $q_1 + q_2 + q_3 = 1$, then one has $H_\lambda(X, Y) = h(p) + (1-p)H(\{q_1, q_2, q_3\}) \leqslant h(p) + (1-p)\log_2 3$ where $H(\{\cdot\})$ denotes the Shannon entropy of the distribution with components in $\{\cdot\}$. This bound is saturated for $q_1 = q_2 = q_3 = 1/3$, and increases monotonically with $p \in [0, 1/4]$.

[36] More generally, a retrocausal model can be constructed for any given set of probabilities $\{p(\lambda|x, y), p(a|x, \lambda), p(b|y, \lambda)\}$ that generates separable correlations $\{p(a, b|x, y)\}$ as per Eqs. (1) and (2), even for the case of an arbitrary given prior distribution $p(x, y)$. In particular, one adds a new random variable, $\Lambda_0 = \lambda_0$, say, that causally influences X and Y in Fig. 1(b) but is not directly connected to any other variable, such that $p(x, y) = \sum_{\lambda_0} p(\lambda_0)p(x|\lambda_0)p(y|\lambda_0)$ [this corresponds to generating the desired $p(x, y)$ via shared randomness]. The measurement choices x and y obtained via Λ_0 then retrocausally influence Λ, to generate the desired distribution $p(\lambda|x, y)$, which in turn causally influences A and B, to generate the desired outcome probabilities $p(a|x, \lambda)$, $p(b|y, \lambda)$. Note that such a retrocausal implementation of measurement dependence generates a corresponding distribution $p(\lambda) = \sum_{x,y} p(\lambda|x, y)p(x, y)$ for Λ, and that the measurement dependence, as quantified by $I(X, Y{:}\Lambda)$, is unchanged by the additional variable Λ_0 (since X and Y "screen off" Λ_0 from Λ).

[37] M. Banik, M. Rajjak Gazi, S. Das, A. Rai, and S. Kunkri, J. Phys. A **45**, 205301 (2012).

[38] O. Costa de Beauregard, C. R. Math. Acad. Sci. **236**, 1632 (1953).

[39] O. Costa de Beauregard, Nuovo Cimento B **42**, 41 (1977).

[40] H. Price and K. Wharton, Entropy **17**, 7752 (2015).

[41] "Suppose the world is superdeterministic, with not just inanimate nature running on behind-the-scenes clockwork, but with our behavior, including our belief that we are free to choose to do one experiment rather than another, absolutely predetermined, including the "decision" by the experimenter to carry out one set of measurements rather than another".—J. S. Bell, in *The Ghost in the Atom*, edited by P. C. W. Davies and J. R. Brown (Cambridge University Press, Cambridge, 1983), Chap. 3.

[42] P. G. Kwiat, P. H. Eberhard, A. M. Steinberg, and R. Y. Chiao, Phys. Rev. A **49**, 3209 (1994).

[43] G. 't Hooft, *The Cellular Automaton Interpretation of Quantum Mechanics* (Springer, Cham, 2016).

[44] The BIG Bell Test Collaboration, Nature (London) **557**, 212 (2018).

[45] W. Myrvold, M. Genovese and A. Shimony, Bell's Theorem, in *The Stanford Encyclopedia of Philosophy*, Spring 2019 edition, edited by E. N. Zalta (Metaphysics Research Lab, Stanford University, Stanford, CA, 2019).

[46] The factorizability assumption for $p(x, y)$ can be dropped here for the cases of retrocausal and superdeterministic measurement dependence.

[47] The last line in Eq. (61) follows via rewriting the second line as $p_{C'}(x, y|\lambda) = u(\lambda)v(x, \lambda)w(y, \lambda)$, so that summation over y and x gives $p_{C'}(x|\lambda) = u(\lambda)v(x, \lambda)W(\lambda)$, $p_{C'}(y|\lambda) = u(\lambda)V(\lambda)w(y, \lambda)$, and $1 = u(\lambda)V(\lambda)W(\lambda)$, where V and W denote the summation of v and w over x and y, respectively.

[48] The upper bound given by Eq. (72) is in fact tight, noting Eq. (55), whereas the upper bounds given by Eqs. (69)–(71) are not. In particular, one finds that $I_{R'}(X, Y{:}\Lambda)$, $I_{C'}(X, Y{:}\Lambda)$, and $I_{OS'}(X, Y{:}\Lambda)$ do not satisfy the ordering of Eq. (74) (which should hold for optimal models) for all values of S, ϵ_X, ϵ_Y [e.g., for $S = 4$, $\epsilon_X \to \pm 1$ and $\epsilon_Y = 0$, one has $I_{R'}(X, Y{:}\Lambda) \to 1/3$ while $I_{OS'}(X, Y{:}\Lambda) \to 0$].
[49] J. Degorre, S. Laplante, and J. Roland, Phys. Rev. A **72**, 062314 (2005).
[50] R. F. Werner, Phys. Rev. A **40**, 4277 (1989).
[51] H. Price, Mind **103**, 303 (1994).
[52] N. Argaman, Am. J. Phys. **78**, 1007 (2010).
[53] H. Price, Stud. Hist. Philos. Mod. Phys. **43**, 75 (2012).
[54] Y. Aharonov, E. Cohen, and A. C. Elitzur, Ann. Phys. **355**, 258 (2015).
[55] M. S. Leifer and M. F. Pusey, Proc. R. Soc. A **473**, 20160607 (2017).
[56] N. Argaman, Entropy **20**, 294 (2018).
[57] P. D. Drummond and M. D. Reid, Phys. Rev. Res. **2**, 033266 (2020).
[58] A. Acín, S. Massar, and S. Pironio, New J. Phys **8**, 126 (2006).