Information content of systems as a physical principle

L. Czekaj,[1] M. Horodecki,[1] P. Horodecki,[2,1] and R. Horodecki[1]
[1]*Institute of Theoretical Physics and Astrophysics, National Quantum Information Centre, Faculty of Mathematics, Physics and Informatics University of Gdańsk, 80-308 Gdańsk, Poland*
[2]*Faculty of Applied Physics and Mathematics, Gdańsk University of Technology, 80-233 Gdańsk, Poland*
(Received 1 August 2016; revised manuscript received 25 October 2016; published 17 February 2017)

To explain the conceptual gap between classical and quantum and other, hypothetical descriptions of the world, several principles have been proposed. So far, all these principles have not explicitly included the uncertainty relation. Here we introduce an information content principle (*ICP*) which represents a constrained uncertainty principle. The principle, by taking into account the encoding and decoding properties of a *single* physical system, is capable of separation, both classicality and quanta from a number of potential physical theories, including hidden variable theories. The *ICP*, which is satisfied by both classical and quantum theory, states that the amount of nonredundant information which may be extracted from a given system is bounded by a perfectly decodable information content of the system. We show that *ICP* allows one to discriminate theories which do not allow for correlations stronger than Tsirelson's bound. We show also how to apply the principle to composite systems, ruling out some theories despite that their elementary constituents behave quantumly.

DOI: 10.1103/PhysRevA.95.022119

I. INTRODUCTION

It is astonishing that our best theory of the fundamental laws of physics, quantum mechanics being robust against innumerable experimental tests, is as well robust against our understanding of its physical origins. This is notoriously manifested by the variety of interpretations of quantum mechanics (e.g., [1]). One of the reasons is the way the postulates of quantum mechanics are expressed: they refer to highly abstract mathematical terms without clear physical meaning. This drives physicists to look for an alternative way of telling quantum mechanics. The problem was attacked on different levels. On one hand it has been shown that quantum theory can be derived from more intuitive axioms [2–11]. In particular, it is related to the vastly developed field aiming at reconstructing quantum theory from information properties of the system [5–11] (cf. [12]). On the other hand, an effort was made to derive some principles [13–16] which can separate quantum theory (or in a narrow sense some aspects of the theory, such as correlations) from so-called *superquantum theories* i.e., the theories that inherit from quantum theory the no-signaling principle but otherwise can offer different predictions than quantum mechanics [17].

However, in difference to those approaches, our goal is to find a criterion for physical theories which involves quantitative rather than qualitative (i.e., logical) constraints. Furthermore, in contrast to the previous information principles based on a composite system, here we define and study a principle that refers to a *single* system and represents a sort of uncertainty principle.

To this end we propose a constraint that ties together (i) the amount of *nonredundant information* which can be extracted from the system by the set of observables and (ii) the systems' informational content understood in terms of maximal number of bits that may be encoded in the system in a perfectly decodable way. We call this constraint *information content principle ICP* which represents the *constrained* uncertainty principle which holds in the classical and quantum theories for two different reasons. For classical systems it is due to lack of knowledge, while for quantum systems it reflects quantum uncertainty [18]. To demonstrate the efficiency of the principle we show how it is violated by some theories with relaxed uncertainty constraints [19] and polygon theories [20], as well as some *incomplete* classical theories akin to epistemically restricted theories [21,22].

II. INFORMATION CONTENT PRINCIPLE

Our aim is to provide a principle that would bound the information extractable from a system in a physical theory. There are two problems here to address: first, what should be the ultimate bound for such extractable information, and second, how the extractable information itself is to be defined.

Regarding the bound, it is natural to employ the following fundamental quantity, which we call *information content*. The information content is *the maximal number of bits that can be encoded in a lossless way into a given system*. We express it as $\log_2 d$, where d the maximal number of messages that can be encoded in a lossless way into a system (see Appendix A for detailed discussion). This is a quantity intrinsic to any given theory, for example, in quantum mechanics it is given by the logarithm of the dimension of the Hilbert space of the system.

The second question is more demanding. We shall first present some rough picture and then propose a concrete implementation of the idea. To begin with, information can be extracted from the system by making measurements. Rather than trying to determine the full amount of extractable information, we will consider information obtained from measuring some set of observables. We might want to add information extracted by measuring each observable; however, they may be redundant (e.g., if one observable is a function of another observable). Therefore, one has to subtract the redundancy. This can be symbolized by the following expression:

$$\sum_i I_{M_i} - I_R \leqslant I_C, \tag{1}$$

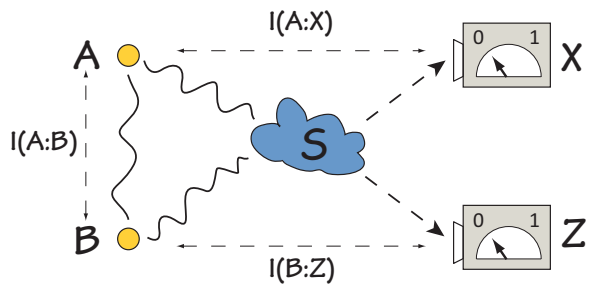

FIG. 1. Scenario of the information content principle.

where I_{M_i} denotes information obtained by measuring observable M_i, I_R represents redundant information, and I_C is the total information content as defined above. Now we would like to make the above formula more concrete so that all the quantities can be computed in a given theory.

To see the difficulties which arise when one tries to define redundant information I_R, consider two observables M_1, M_2 in classical theory. Then a natural candidate is just the mutual information of the joint probability of the outcomes of M_1 and M_2:

$$I(M_1 : M_2) = H(M_1) + H(M_2) - H(M_1, M_2). \tag{2}$$

[Here $H(\{p_i\}) = -\sum_i p_i \log_2 p_i$ denotes Shannon entropy of probability distribution $\{p_i\}$; thus $H(M_1)$ and $H(M_2)$ are entropies of distributions of M_1 and M_2, respectively, while $H(M_1, M_2)$ is the entropy of joint probability distribution of M_1 and M_2.] Indeed, the mutual information can be interpreted as a common information shared by both random variables. However, in quantum theory, such joint probability does not exist. Therefore, since our quantities are to be sensible in any theory, including the quantum one, we have to define redundancy in some indirect way.

We shall now present a setup which allows us to properly grasp the idea of nonredundant extracted information. Consider the scenario (for simplicity, just for two measurements) depicted in Fig. 1.

Consider two persons: the sender and the receiver. Let the sender hold classical information stored in two registers A and B. She wants to provide access to that information for the sender but she does not know which register is interesting for him. She prepares the system $\mathcal{S}$ in a state which depends on the content of A and B. Then she sends $\mathcal{S}$ to the receiver. After transmission of $\mathcal{S}$, the sender and the receiver share the state

$$\omega^{SAB} = \sum_{i,j} p_{i,j} \omega^S_{i,j} \otimes \sigma^A_i \otimes \sigma^B_j, \tag{3}$$

where $p_{i,j}$ is the distribution of the content of the classical registers, whose states are here labeled as σ^A_i and σ^B_j; $\omega^S_{i,j}$ denotes the state of the system, given the classical registers are in the state $\sigma^A_i \otimes \sigma^B_j$; and the classical registers A and B are in hands of the sender, while the system S is held by the receiver. The receiver extracts information from $\mathcal{S}$ by performing one of two measurements, X and Z. In this way he learns about the content of A or B, respectively. The information extracted by observables X and Z are defined, respectively, as $I(A : X)$ and $I(B : Z)$—the Shannon mutual information between classical system and outcomes of measurement. The redundant information will be the mutual information between the classical systems $I(A : B)$. The formula (1) then takes the following concrete form [23] (see also [24]):

$$I(X : A) + I(Z : B) - I(A : B) \leqslant \log_2 d, \tag{4}$$

where $\log_2 d$ is the information content of a system. Note that here all the quantities I are mutual information of classical variables. For more than two measurements $\{X_i\}$ the formula takes the form (see Appendix B for details)

$$\sum_i I(X_i : A_i) - I(A_1 : \ldots : A_n) \leqslant \log_2 d. \tag{5}$$

Now, the central postulate of the present paper is that the above formula (5) represents the information content principle which should be valid for *any* physical system (either an elementary or a composite one) in *any* physical theory. In particular, the principle holds for quantum theory, which can be proved in the spirit of [13]. Namely, any theory in which one can define a notion of entropy satisfying some natural axioms obeys the principle. In Appendix B we give the list of axioms and derive *ICP* from those axioms. The axioms are satisfied by von Neumann entropy in quantum theory and by Shannon entropy in classical theory and hence both theories obey the principle.

Note that the *ICP* incorporates idea of impossibility of encoding more information using complementary observables [25], which is a basic ingredient of information-type principles [8,9,13,26,27]. (This idea differs from the bounding capacity of quantum systems, as well as bounding the classical memory required for their simulation; see, e.g., [28].)

Below we shall show violation of (4) in two elementary examples: (i) nonlocal theories represented here by so-called sbit (square bit) [19]; (ii) epistemically restricted theories where, as an example, we consider hbit (hidden bit) [21,22] and postpone discussion of more advanced cases to the further part of the paper. To show violation for sbit and hbit, we evaluate (4) on the state of (3), with $\omega^S_{i,j}$ being such a state of sbit or hbit that the outcome i, j after measuring X, Z, respectively, is certain (i.e., $p(a = i|x = X) = 1$ and $p(a = j|x = Z) = 1$). Information encoded in the observables is completely uncorrelated, i.e., $I(A : B) = 0$ for ω^{SAB}. Since there is no uncertainty in the system, information encoded in each observable might be recovered completely; hence $I(X : A) = I(Z : B) = 1$. Taking that together we obtain violation of (4), since $I(X : A) + I(Z : B) - I(A : B) = 2 > 1$. At this point it is worth noticing that in the case of hbit, violation comes from the fact that the observed dimension d is different from what we could call an "intrinsic" system dimension: the observables available in theory have two outputs while the internal state of the system is described by two classical bits. The theory is incomplete because of lack of a fine-grained observable with four outputs that could access full information available in the system. Note that lack of such observable excludes the possibility of measuring the two observables one after another (in which case, one would eventually access both bits): indeed, then one could define the fine-grained observable as the subsequent measurement of the dichotomic observables.

For a classical bit, if the observables are nontrivial, they must be a function of one another, so that we have actually only one observable up to relabeling the outputs, and the

information is highly redundant. Indeed, if $I(A:X)=1$ and $I(A:Z)=1$, then we must have $I(A:B)=1$. Both mutual informations are maximal, but they are redundant. To discuss the quantum case, let us assume that marginal entropies $H(A)=H(B)=1$. Then for X and Z complementary, we have that $I(A:B)$ can vanish. Hence we have $I(A:X)+I(B:Z)-0\leqslant 1$. Thus, although the information is nonredundant, they are restricted. Thus, unlike in the classical case, here we have two independent "species" of information and there is room only for one of them. If we rotate the observable Z towards X, we observe that $I(X:A)+I(Z:B)$ grows up together with $I(A:B)$. The observables disclose more information; however, the information is more redundant. Extractable information cannot exceed the bound given by *ICP*.

To see that *ICP* can be interpreted as an uncertainty of a new kind, suppose that we fix $I(A:B)$ to be some number strictly less than 1, i.e.,

$$I(A:B)<1. \tag{6}$$

Then we obtain restrictions on the values $I(A:X)$ and $I(B:Z)$, namely, they cannot be both equal to 1. This would look like Hall's exclusion principle, which also bounds the sum of two mutual informations [29]. However, unlike in the exclusion principle, in the present case the restriction is the same regardless of whether the observables commute or not. Indeed, if the observables commute (i.e., in classical theory) the restriction comes from the fact that up to the relabeling of outcomes, there is only one fine-grained observable on a classical system, so that any other observable carries the same information. Therefore it is impossible to fit more than one bit into a binary system, as *ICP* states.

If the observables do not commute, the reason is less obvious because the different observables, especially if they are complementary, surely do not carry the same information. However, again the restriction posed by *ICP* holds, this time because of the quantum uncertainty.

To see more clearly the connection with quantum uncertainty, let us assume that $H(X)=1$ and $H(Z)=1$, i.e.. that the outcomes are random. Then *ICP* is written as

$$H(X|A)+H(Z|B)\geqslant 1-I(A:B), \tag{7}$$

while the standard quantum uncertainty principle is of the form [18]

$$H(X|A)+H(Z|B)\geqslant c(X,Z), \tag{8}$$

where $c(X,Z)$ quantifies the lack of a common eigenvector for X and Z. For commuting observables $c=0$, and the uncertainty relation is trivial, i.e., there is no uncertainty. In our case, when $I(A:B)<1$, the right-hand side is constant independent of observables; hence the relation is *always—both classically and quantumly*—nontrivial. Thus any theory which obeys *ICP* exhibits uncertainty of outcomes under the constraint $I(A:B)<1$. Yet, as we show below, there are theories that do not exhibit this uncertainty, and in this sense they are too "certain" to be physical. The above considerations are illustrated in Fig. 2.

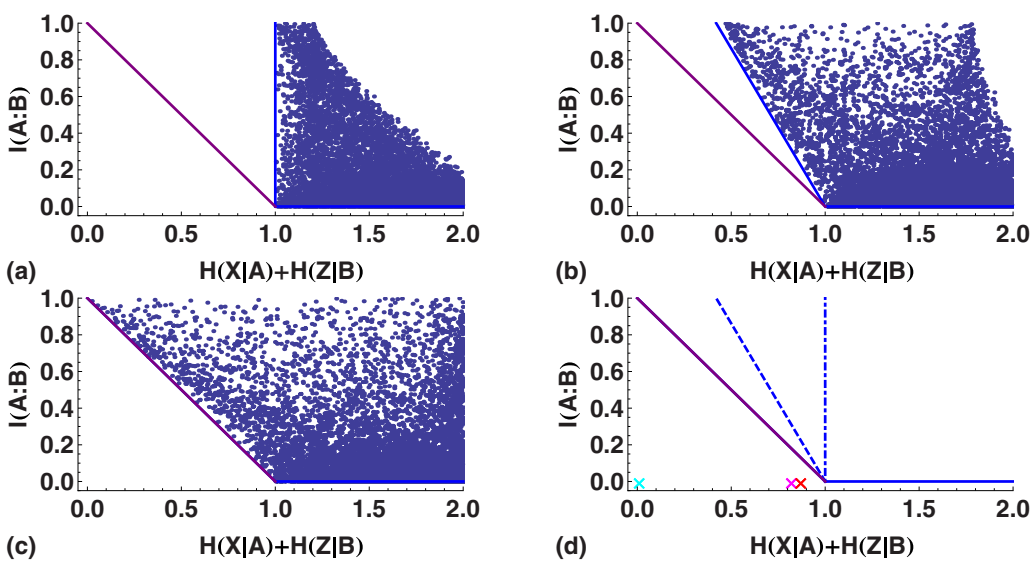

FIG. 2. *ICP* as a constrained uncertainty principle. In the picture we present on the plain $H(X|A)+H(Z|B)\approx I(A:B)$ the areas attainable by a quantum 1-qbit system for some fixed observables X and Z: (a) complementary observables; (b) observables where the angle between axes representing them on the Bloch sphere is $\pi/4$ (they are still noncommuting but are not complementary anymore); and (c) a commuting observable ($X=Z$), which we interpret as the classical case. The blue dots correspond to states ρ_{SAB} chosen randomly from a set of states satisfying $H(X)=H(Z)=1$. The blue solid line represents schematically the boundary of the area. The purple line depicts the *ICP* bound. The area attainable by quantum and classical theory is placed above this line. In item (d) we put together the areas from (a), (b), and (c). We can observe that in the setup when registers A and B keep completely independent information [i.e., $I(A:B)=0$], in both the classical and quantum case there is unavoidable uncertainty. However, when registers A and B hold the same information, uncertainty in the classical case vanishes. On the other hand, in some nonlocal theories like polygon theories, for $I(A:B)=0$ there are states that are "more certain" than classical and quantum states. These states are depicted by an "x" on (d) and correspond to polygon theories with parameters $n=4$ (cyan), $n=6$ (magenta), and $n=8$ (red).

III. VIOLATIONS OF *ICP* IN GENERAL PROBABILISTIC THEORIES

In this section we briefly discuss violation of *ICP* in two families of theories that originate from general probabilistic theories (*GPT*) (see Appendix A). We start with the *p general nonsignaling theories* (p-GNSTs) introduced in [19]. (For more details see Appendix A.) These theories violate the quantum uncertainty relation for anticommuting observables and they were originally developed to study how Tsirelson's bound for the Clauser-Horne-Shimony-Holt inequality emerges from the uncertainty relation.

The elementary system of p-GNST theory is a box with two observables X and Z. Its state space is bounded by the uncertainty relation

$$(s_x)^p + (s_z)^p \leqslant 1, \tag{9}$$

where $p \in [2,\infty]$ is a parameter of the theory and $s_x = p(a = +|x = X)_\psi - p(a = -|x = X)_\psi$ is the mean value of the observable X measured on the system in state ψ (analogically for s_z and Z). By varying the parameter p one can move from the state space of sbit to qbit.

In Appendix E 1 we show that violation of the quantum uncertainty relation by states from p-GNST (i.e., for theories with $p > 2$) not only leads to violation of Tsirelson's bound (as proved in [19]) but also to violation of *ICP* by the elementary system. Violation of *ICP* in these theories follows from the existence of sufficiently many states for which entropic uncertainty for the observables X and Z is smaller than in the quantum case. Relaxation of the uncertainty relation also was shown to increase the maximum recovery probability for so-called $2 \mapsto 1$ *random access codes* (RACs). Specifically, one can encode two bits into a system from the theory in such a way that the probability of decoding (recovery) for each bit separately reads as $p_{\rm rec} = (1/2)^{1/p}$. Therefore, excluding p-GNSTs with $p > 2$, *ICP* puts a bound on the performance of random access codes.

ICP not only applies to elementary systems but may also be used in a natural way to study composite systems. It is able to exclude theories which are nonphysical, but nevertheless their state space of the elementary system is quantum. Here an example is p-GNST with $p = 2$, where violation of *ICP* occurs for a system with at least five parties. This result is based on the existence of super-strong RAC in p-GNST.

It is interesting to ask what other geometrical constraints (e.g., other uncertainty relations, consistency constraints (cf. [19]), and local orthogonality [16]) have to be added to theory to conform with *ICP*. In the opposite direction, one may ask how *ICP* limits the strength of nonlocal correlations achievable in *GPT* for systems whose elementary subsystems obey quantum mechanics.

We move to polygon theories [20]. They are described in more detail in Appendix E 2. Here we just mention that state space in those theories is given by a polygon with n vertices. For those theories *ICP* is *more sensitive* than Tsirelson's bound, since it allows one to discriminate theories which do not allow for correlations stronger than Tsirelson's bound. Namely, *ICP* is violated in all nonphysical polygon theories. For theories with even n, it is again connected with the existence of states with lower entropic uncertainty than in the quantum case. For odd n violation of *ICP* links rather to the fact that in this case polygon theories allow for communication of more than 1 bit per elementary system in the Holevo sense (i.e., in the asymptotic limit) [30]. Interestingly, we found examples of polygon theories that are not ruled out by a principle proposed in an independent development [27] based on so-called *dimension mismatch*.

IV. SUMMARY

We have identified a constrained uncertainty informational principle (*ICP*) based on a single physical system which puts new constraints on physical theories. The principle has a form of uncertainty-type inequality with an extra information constraint. This is the feature that allows the principle to filter out both "superquantum"and "superclassical"(epistemically restricted) theories, leaving the two "modest" ones, i.e., classical and quantum, within the scope of its validity. If applied to classical theory, *ICP* reflects the fact that there is basically one type of information and all fine-grained observables in classical discrete systems are equivalent up to relabeling. On the contrary, in quantum mechanics, there are much different "species" of information, which is reflected by the presence of incompatible observables which are only partially redundant. At the same time, only one type of information may be completely present in the system [31,32], as stated in Bohr's principle of complementarity [33], which is connected to entropic uncertainty relations. Only one observable from the complementary set may be measured perfectly. However, two observables which are "less incompatible" than complementary reveal information which is redundant. *ICP* gives the tradeoff between how much information may be extracted and how redundant the information is. In particular, the power of the principle is illustrated by the fact that it rules out some theories (so-called polygonic theories) that do not violate Tsirelson bound, which therefore are not detected neither by information causality nor by local orthogonality. An intriguing feature of the *ICP* is that two thus-far-distinct fundamental concepts, uncertainty and random access coding, are inextricably and quantitatively linked within a *single* constrained uncertainty. It sheds light on the question of why quantum mechanics is so restrictive, or in other words, why it has such and only such strength. We believe that the information content principle may be a useful tool for analysis of forthcoming theories, yet to be discovered. It seems also that it may help in deeper understanding of the laws governing physical reality in general.

ACKNOWLEDGMENTS

The authors thank A. Grudka, K. Horodecki, M. Pawlowski, and R. Ramanathan for many discussions about information principles. We also thank R. Augusiak, C. Brukner, N. Brunner, A. Cabello, R. B. Griffiths, and L. Masanes for valuable comments and feedback on the first version of the manuscript. This paper was supported by the Polish Ministry of Science and Higher Education through Grant No. IdP2011 000361, ERC QOLAPS, the Polish National Science Centre through Grant No. DEC-2011/02/A/ST2/00305 and partially by the John Templeton Foundation through Grant No. 56033.

APPENDIX A: GENERALIZED PROBABILISTIC THEORIES

A generalized probabilistic theory consists of a convex state space $\Omega \subseteq \mathbb{R}^n$, i.e., the set of admissible states the system may be prepared in, and the set of measurements $\mathcal{M}$. The measurement outcome is represented by the effect e, which is a linear map $e : \Omega \to [0,1]$. $e(\omega)$ is the probability of outcome e when the measurement is performed on the system in state ω. The special effect is the unit effect u such that for every $\omega \in \Omega$ there is $u(\omega) = 1$ (here we consider only normalized states). The measurement is the set of effects $\{e_i\}$ summing up to a unit effect u.

The state of the system is entirely determined by the probabilities $p(a|x)$ it assigns to the outcomes a of every measurement x. However, there exists a subset of measurements called fiducial measurements $\mathcal{F} \subseteq \mathcal{M}$, which is enough to describe the state [6].

Particular examples of systems which may be expressed in terms of *GPT* (see Fig. 3) are classical bit, qbit, and sbit (square bit). The last one, sometimes called gbit for generalized bit, is the building block of the PR box [34].

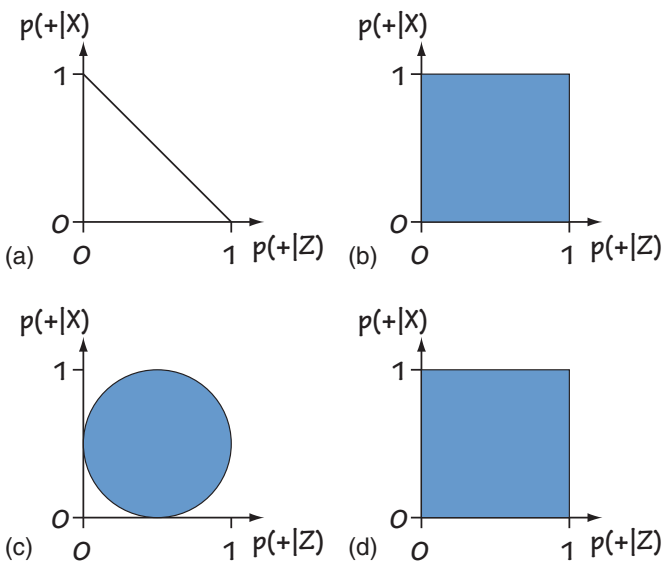

FIG. 3. Elementary systems in exemplary *GPT* from the perspective of two distinguished dichotomic observables X and Z. (a) classical bit and (b) hidden bit (hbit): This system is an example from epistemically restricted theories consisting of two classical bits where if one of them, chosen by the observer, is read out then the second necessarily disappears or becomes unreadable by any physical interaction. (c) quantum bit (qbit) and (d) square bit (sbit): This is an example from nonlocal theories and may be viewed as a building block of PR boxes. A qbit differs from the sbit and hbit by the amount of uncertainty. A classical bit admits no uncertainty; however, X and Z reveal *the same* information. This is reflected in perfect correlations of X and Z. The state space of sbit and hbit observed from the perspective of two observables may seem to be equivalent. What is missing on that picture is the issue of decomposition into pure states. Measurements X and Z have two outcomes, $+$ and $-$. The axis represents probability of outcome $+$ when measurement X or Z is performed on the system in a given state.

If the set of measurements does not reach enough, one may obtain a classical system with hidden variables. For example, the elementary system in epistemically restricted theories, consists of two classical bits. One of two observables can be measured on the system giving access to a chosen bit. After measurement, information from the complementary bit is unavoidably lost. This property reflects a lack of fine-grained observable in hidden variable theory. A more sophisticated example of the hidden variable theory may be found in [21].

Given two systems A and B, we may define in *GPT* a composite system AB. The global state of the system AB is completely determined by joint probabilities of outcomes for fiducial measurements performed at the same time on each subsystem. This is called a local tomography assumption. All effects for the composite system AB are of the form $e_A e_B$, which means that the effect e_A was measured on the subsystem system A and e_B on the subsystem B. The state space Ω_{AB} is not defined in the unique way. It contains all states of the form $\omega_A \omega_B$, i.e., states which result from preparation of states ω_A and ω_B independently of the subsystems A and B. [For $\omega_A \omega_B$ it holds that $e_A e_B(\omega_A \omega_B) = e_A(\omega_A) e_B(\omega_B)$.] Other states ω_{AB} may also belong to Ω_{AB}, provided $e_A e_B(\omega_{AB}) \geqslant 0$ is true for every pair of effects. Therefore starting from elementary systems, we may obtain different composite systems depending on the restrictions imposed on Ω_{AB} (cf. generalized nonsignaling theory and generalized local theory in [34]). In every case, Ω_{AB} contains only nonsignaling states. The dimension of state space Ω_{AB} is bounded by

$$\dim(\Omega_{AB}) + 1 \leqslant [\dim(\Omega_A) + 1][\dim(\Omega_B) + 1]. \tag{A1}$$

For a given *GPT*, we may ask for a maximal number of states that can be perfectly distinguished in a single-shot measurement [4,6]. We will call this value the *observed system dimension* and denote it by d. In terms of *GPT*, we look for the biggest set $\{\omega_i\} \in \Omega$ such that there exists a set of effects $\{e_j\}$ which obeys $e_j(\omega_i) = \delta_{i,j}$. The set of states $\{\omega_i\}$ together with set of effects $\{e_j\}$ may be interpreted as a maximal classical subsystem of *GPT*, and $\{e_j\}$ represents a generalization of the quantum projective measurement, cf. complete measurement [35]. The observed system dimension is bounded by

$$d \leqslant \dim(\Omega) + 1, \tag{A2}$$

where equality holds only for classical systems [36]. Combining Eqs. (A1) and (A2) one may obtain a bound for composite systems.

APPENDIX B: INFORMATION CONTENT PRINCIPLE FOR TWO OBSERVABLES

Here we provide detailed proof of *ICP*. To make our argumentation easier to follow, first we consider only the case with two observables $\mathcal{M} = \{X, Z\}$. Then we generalize results to a multiple observables scenario.

We start with the definition of a tripartite state of the form

$$\rho^{SAB} = \sum_{i,j} p_{i,j} \rho^S_{i,j} \otimes \sigma^A_i \otimes \sigma^B_j, \tag{B1}$$

where the state ρ^S defined on system $\mathcal{S}$ belongs to the considered theory (e.g., bit, qbit, sbit), while σ^A and σ^B are

classical registers. Their role is to keep classical information measured by observables X and Z, respectively. $\{p_{i,j}\}$ is the classical probability distribution. The state ρ^{SAB} is an analog of the quantum-classical system utilized in analysis of communication tasks.

We are in a position to prove that (B2) holds for classical and quantum systems ρ^S:

$$I(X:A)+I(Z:B)-I(A:B)=I_C\leqslant \log_2 d, \tag{B2}$$

where $I(X:A), I(Z:B), I(A:B)$ are classical mutual information and d is an observed system dimension.

Here we define mutual information and conditional entropy in the standard way as $I(A:B)=H(A)-H(A|B)$ and $H(A|B)=H(AB)-H(B)$. In the proof we make use of the following properties of classical and quantum entropies: (i) entropy of the system is bounded by $H(S)\leqslant \log_2 d$; (ii) conditional entropy of any system S correlated with the classical one C is non-negative $H(S|C)\geqslant 0$, where C is a classical system; (iii) strong subadditivity $H(SAB)+H(S)\leqslant H(SA)+H(SB)$; and (iv) information processing inequality for measurement $I(S:A)\geqslant I(X:A)$, where X denotes the measurement outcome. In Appendix D we discuss these properties on the ground of *GPT*.

First, we use (i) and (ii) to obtain the upper bound for mutual information between system $\mathcal{S}$ and $\mathcal{AB}$ for a state ρ^{SAB}:

$$\begin{aligned} I(S:AB) &= H(S)-H(S|AB)\\ &\leqslant H(S)\leqslant \log_2 d. \end{aligned} \tag{B3}$$

Using the chain rule for mutual information, we get

$$\begin{aligned} I(S:AB) &= I(S:A)+I(S:B|A),\\ I(AS:B) &= I(S:B|A)+I(A:B). \end{aligned} \tag{B4}$$

Putting this together with strong subadditivity and an information processing inequality for measurements, we obtain

$$\begin{aligned} I(S:AB) &= I(S:A)+I(AS:B)-I(A:B)\\ &\geqslant I(S:A)+I(S:B)-I(A:B)\\ &\geqslant I(X:A)+I(Z:B)-I(A:B). \end{aligned} \tag{B5}$$

In this way we proved (B2).

APPENDIX C: INFORMATION CONTENT PRINCIPLE FOR MULTIPLE OBSERVABLES

Now we prove *ICP* for the setup with n observables $\{X_i\}$. For the classically correlated state [cf. (B1)]

$$\rho^{SA_1\ldots A_n}=\sum_{i_1,\ldots,i_n} p_{i_1,\ldots,i_n}\rho^S_{i_1,\ldots,i_n}\otimes\sigma^{A_1}_{i_1}\otimes\cdots\otimes\sigma^{A_n}_{i_n}, \tag{C1}$$

where ρ_S represents system $\mathcal{S}$ belonging to considered theory and $\{\sigma^{A_j}_{i_j}\}$ denotes classical registers, we show that

$$\sum_i I(X_i:A_i)-I(A_1:\ldots:A_n)\leqslant \log_2 d. \tag{C2}$$

d is an observed system dimension [cf. Eq. (3)], and $I(A_1:\ldots:A_n)=\sum_i H(A_i)-H(A_1,\ldots,H_n)$ is multivariable mutual information. The upper bound $I(S:A_1,\ldots,A_n)\leqslant \log_2 d$ comes in exactly the same way as in (B3); hence we omit this part of the proof and focus on the left-hand side of Eq. (C2).

We start using the chain rule and write

$$\begin{aligned} I(S:A_1,\ldots,A_n) = I(S:A_1)&+I(S:A_2|A_1)\\ &+I(S:A_3|A_1,A_2)+\cdots\\ &+I(S:A_n|A_1,\ldots,A_{n-1}). \end{aligned} \tag{C3}$$

We use the chain rule once again to express express conditional mutual information in the form

$$\begin{aligned} I(S:A_2|A_1) &= I(A_1,S:A_2)\\ &\quad -I(A_1:A_2)\\ I(S:A_3|A_1,A_2) &= I(A_1,A_2,S:A_3)\\ &\quad -I(A_1,A_2:A_3)\ldots\\ I(S:A_n|A_1,\ldots,A_{n-1}) &= I(A_1,\ldots,A_{n-1},S:A_n)\\ &\quad -I(A_1,\ldots,A_{n-1}:A_n). \end{aligned}$$

Combining these together with strong subadditivity we get

$$\begin{aligned} I(S:A_1,\ldots,A_n)\geqslant I(S:A_1)&+\ldots+I(S:A_n)\\ &-I(A_1:A_2)-\ldots\\ &-I(A_1,\ldots,A_{n-1}:A_n). \end{aligned} \tag{C4}$$

From the classical mutual information properties $[I(A:B)=H(A)+H(B)-H(A,B)]$, it is easy to see that $I(A_1:A_2)+\cdots+I(A_1,\ldots,A_{n-1}:A_n)=I(A_1:\ldots:A_n)$ holds. Putting that to (C4) and applying information processing inequality for measurements, we finally get

$$I(S:A_1,\ldots,A_n)\geqslant\sum_i I(X_i:A_i)-I(A_1:\ldots:A_n).$$

That finishes the proof.

APPENDIX D: ENTROPY IN *GPT*

In this section we would like to focus on the properties (i)–(iv) of entropy, which was used in the proofs presented in Appendices B and C. We may define some general notion of entropy $\mathcal{H}$ which measures our uncertainty about the system S which belongs to *GPT*. The natural assumption is that $\mathcal{H}$ should reduce to classical or quantum entropy if we are restricted to these theories.

Moreover, as it was pointed out in [37], properties (iii) and (iv) follow from the reasonable assumption that local transformation can destroy but not create correlations. This assumption is expressed in the formal way as

$$\Delta H(AB)\geqslant \Delta H(A), \tag{D1}$$

where the transformation is performed on system A. We expect that the theory provides at least transformations like system preparation, measurement, and discarding.

Property (ii) refers to the procedure of system preparation where we randomly choose one of the several possible states of the system. Knowledge of the way the system was prepared should reduce our uncertainty.

To motivate (i), first we would like bring to attention that the general entropy $\mathcal{H}$ is often linked with the minimal output

uncertainty on the distinguished subset of measurements $\mathcal{M}_F$:

$$\mathcal{H}(S) = \inf_{M\in\mathcal{M}_F} H(M(S)), \tag{D2}$$

where M is measurement on the system S. This distinguished subset $\mathcal{M}_F$ consists of maximally informative, i.e., fine-grained measurements [38,39]. In an analogy to quantum mechanics, we may think of them as a set of rank-1 positive operator-valued measures.

In the quantum case a special role is played by the projective measurements. The von Neumann entropy is the output entropy for the measurement which consists of projectors on the eigenvectors of the state. In GPT we call a measurement a projective measurement if for every outcome e there is a state ω that the probability of the outcome e on the state ω is $e(\omega) = 1$. We observe that nonprojective measurements contain some intrinsic noise, i.e., some outcomes cannot be obtained with probability one. On the other hand, information encoded in states $\{\omega_j\}$ can be perfectly retrieved, since $e_i(\omega_j) = \delta_{i,j}$.

For those reasons, we assume that entropy should refer to the uncertainty of the outcome of fine-grained projective measurements and in this way it should be bounded by the number of bits d that may be encoded in the system in a perfectly decodable way.

Interestingly, with some additional assumptions on the postmeasurement state, if $\mathcal{H}(S)$ attains its value for projective measurement, then it has operational interpretation in terms of information compression [39].

APPENDIX E: VIOLATIONS OF *ICP* IN GENERAL PROBABILISTIC THEORIES – DETAILS

In this section we provide technical details which support the discussion of violations of *ICP* in general probabilistic theories which was presented in the main part of the paper. We will base this discussion on the fact that normalized states of $p-$GNSTs and polygon theories are real vectors $\omega \in \mathbb{R}^2$. The maximal number of perfectly distinguishable states thus satisfies $d \leqslant 3$. Equality holds if and only if the states space is a simplex [36]. Therefore any nonclassical theory has $d \leqslant 2$ so that the information content of the system satisfies $\mathcal{I}_C \leqslant 1$.

1. Violation uncertainty relation for anticommuting observables

We consider the $p-$GNST with two dichotomic observables X and Z. Admissible states fulfill the uncertainty relation

$$(s_x)^p + (s_z)^p \leqslant 1, \tag{E1}$$

where $p \in [2,\infty]$ is a parameter of the theory and $s_x = p(a = +|x = X)_\psi - p(a = -|x = X)_\psi$ is the mean value of observable X measured on the system in state ψ (analogically for s_z and Z). It is straightforward to see that (E1) is an uncertainty relation, since it bounds the probability that the state has a well-defined outcome of each observable. p-GNST is a simplified version of the model discussed in [19], since we only deal with the case of the two observables available in the elementary system. However, our results may be easily generalized to the case of three observables X,Y,Z.

The set of admissible states for $p = 2$ correspond to the set of states from the great circle of the Bloch ball. [In the case of three observables, the set of admissible states becomes a full Bloch ball and a relation of type (E1) defines the state space of a single qbit.] On the other hand, for $p \to \infty$ we approach the state space of an sbit. Therefore the increase of p leads to relaxation of the uncertainty relation.

Now we show that each theory with $p > 2$ violates *ICP*. For that purpose, (i) we show that there exists a state ψ_{++} with entropic uncertainty small enough (i.e., $H(X)_{\psi_{++}} + H(Z)_{\psi_{++}} < 1$), and (ii) then by symmetry of the state space we construct the state ρ^{SAB}, which we use to prove violation.

Let us parametrize by s_x states ψ that saturate (E1). For simplicity we assume that $s_z > 0$. Due to (E1), we have $s_z = \sqrt[p]{1 - s_x^p}$. For $s_x = 1$, the outcome of the observable X is certain but we have no knowledge on the outcome of observable Z. As s_x decreases, the knowledge of the outcome of Z increases by the cost of certainty of the outcome of X. The rate of this exchange depends on the uncertainty relation and interestingly, for $p > 2$, some states near to $s_x = 1$ have entropic uncertainty smaller than in the quantum case. Precisely, we show that there exist δ_x that any state with $1 - \delta_x < s_x < 1$ fulfills:

$$H(X)_\psi + H(Z)_\psi < 1. \tag{E2}$$

For the parametrization of state ψ by s_x, entropies of measurements take a form $H(X)_\psi = H(\frac{s_x+1}{2}), H(Z)_\psi = H(\frac{s_z+1}{2})$ and are bounded in the following way: $H(\frac{s_x+1}{2}) \leqslant (\frac{1-s_x}{2})^{1+\epsilon}$ for $\epsilon > 0$ and $\frac{1-s_x}{2} < \delta_\epsilon$ and $H(\frac{s_z+1}{2}) \leqslant 1 - (\frac{s_z}{2})^2$. This allows us to rewrite condition (E2) as

$$\left(\frac{1-s_x}{2}\right)^{1+\epsilon} < \frac{1}{4}\left(1 - s_x^p\right)^{2/p}. \tag{E3}$$

Now let us observe that

$$\left(\frac{1-s_x}{2}\right)^{1+\epsilon}\Big|_{s_x=1} = \frac{1}{4}\left(1 - s_x^p\right)^{2/p}\Big|_{s_x=1} = 0, \tag{E4}$$

and for $(1+\epsilon)p > 2$,

$$\lim_{s_x\to 1} \frac{\left(1 - s_x^p\right)^{2/p}}{(1-s_x)^{1+\epsilon}} = \infty. \tag{E5}$$

It means that that the left-hand side of (E3) converges to 0 faster than the right-hand side as $s_x \to 1$. Since both sides of (E3) are positive, it implies that (E3) holds for $1 - \delta_x < s_x < 1$ with δ_x small enough.

Since we have shown that states with the desired property exist, we can take any state ψ_{++} that $H(X) + H(Z) = \tilde{H} < 1$. The state is described by $(\tilde{s}_x,\tilde{s}_z)$. By the symmetry of (E1) and (E2), we know that states $\psi_{+-},\psi_{-+},\psi_{--}$ obtained from ψ_{++} by negation of the proper parameter are also admissible and have the same entropic uncertainty $\tilde{H}$. This allows us to construct the state

$$\rho^{SAB} = \frac{1}{4}\sum_{i,j\in\{-,+\}} \psi_{ij}^S \otimes i^A \otimes j^B. \tag{E6}$$

It is easy to observe that the outcome of X and Z for the reduced state $\frac{1}{4}\sum_{i,j\in\{-,+\}} \psi_{ij}^S$ is completely random. Hence

we may write

$$I(X:A)+I(Z:B)=H(X)+H(Z) - \sum_{i,j\in\{-,+\}}\frac{1}{4}[H(X)_{\psi_{i,j}}+H(Z)_{\psi_{i,j}}] = 2-\tilde{H}>1. \tag{E7}$$

Since, in addition, from (E6) we have $I(A:B)=0$, the above shows the expected violation and finishes the proof.

2. Polygon theories

Polygon theories (parameterized by n) were developed in [20] to study the connection between the strength of nonlocal correlations and the structure of the state spaces of individual systems. They may be viewed as a progressive relaxation of the superposition principle (cf. relaxation of the uncertainty relation in p-GNSTs) moving from the quantum case $n\to\infty$ to sbit ($n=4$) and classical trit ($n=3$). Relaxation of the superposition principle means that more restrictions are imposed on the way the states can be superposed.

The proof of violation of *ICP* by unphysical (i.e., with $n>3$ and $n<\infty$) polygon theories is quite technical and based mostly on construction of the state ρ^{SAB} with proper measurement entropies. We start with a short description of polygon theories mainly following [20]. For more details see the original paper [20].

The state space Ω of a single system in polygon theory is a regular polygon with n vertices. For fixed n, Ω is represented as a convex hull of n pure states $\{\omega_i\}_{i=1}^n$:

$$\omega_i=\begin{pmatrix} r_n\cos\left(\frac{2i\pi}{n}\right)\\ r_n\sin\left(\frac{2i\pi}{n}\right)\\ 1\end{pmatrix}\in\mathbb{R}^3, \tag{E8}$$

where $r_n=1/\sqrt{\cos(\pi/n)}$.

The set of effects is the convex hull of the unit effect, zero effect, and the extreme effects. The unit effect has the form

$$u=\begin{pmatrix}0\\0\\1\end{pmatrix}. \tag{E9}$$

Extreme effects for even n are given by

$$e_i=\frac{1}{2}\begin{pmatrix} r_n\cos\left(\frac{(2i-1)\pi}{n}\right)\\ r_n\sin\left(\frac{(2i-1)\pi}{n}\right)\\ 1\end{pmatrix}\in\mathbb{R}^3, \tag{E10}$$

and for odd n in slightly different form,

$$e_i=\frac{1}{1+r_n^2}\begin{pmatrix} r_n\cos\left(\frac{2i\pi}{n}\right)\\ r_n\sin\left(\frac{2i\pi}{n}\right)\\ 1\end{pmatrix}, e_i'=u-e_i\in\mathbb{R}^3. \tag{E11}$$

$e(\omega)=e\cdot\omega$ is the Euclidean inner product of the vectors representing the effect and the state.

Now we are in a position to construct states which violate the information content principle in the polygon theories. We will consider separately the case of even and odd n.

For even n we use the state

$$\rho^{SAB}=\tfrac{1}{4}\big(\omega_2^S\otimes\sigma_0^A\otimes\sigma_0^B+\omega_1^S\otimes\sigma_0^A\otimes\sigma_1^B +\omega_{n/2+1}^S\otimes\sigma_1^A\otimes\sigma_0^B+\omega_{n/2+2}^S\otimes\sigma_1^A\otimes\sigma_1^B\big), \tag{E12}$$

along with measurements X and Z given by the effects $\{e_2,u-e_2\}$ and $\{e_{\lfloor n/4\rfloor+2},u-e_{\lfloor n/4\rfloor+2}\}$, respectively. It is easy to see that $I(A:B)=0$, since each combination $\sigma_i^A\otimes\sigma_j^B$ occurs with the same probability 1/4. To calculate $I(X:A)$ and $I(Z:B)$ we need a conditional entropy of measurement outcome which may be obtained from the probability of given effects for a particular state [i.e., $e_j(\omega_i)$]. For $I(X:A)$ the probabilities are $e_2(\omega_1)=e_2(\omega_2)=1$ and $e_2(\omega_{n/2+1})=e_2(\omega_{n/2+2})=0$; hence $I(X:A)=1$. For $I(Z:B)$, straightforward calculations lead to

$$p(Z=0|B=0)=p(Z=1|B=1) =\frac{1}{2}\left[1+\sin\left(\frac{2\pi\lfloor\frac{n}{4}\rfloor}{n}\right)\tan\left(\frac{\pi}{n}\right)\right].$$

It shows that $I(Z:B)>0$, and hence the violation of the information content principle was proved.

For odd n we use the state

$$\rho^{SAB}=\tfrac{1}{4}\big(\omega_1^S\otimes\sigma_0^A\otimes\sigma_0^B+\omega_1^S\otimes\sigma_0^A\otimes\sigma_1^B+\omega_{\lfloor n/2\rfloor+1}^S \otimes\sigma_1^A\otimes\sigma_0^B+\omega_{\lfloor n/2\rfloor+2}^S\otimes\sigma_1^A\otimes\sigma_1^B\big). \tag{E13}$$

In this case measurements X and Z are given by the effects $\{e_1,u-e_1\}$ and $\{e_{\lfloor n/4\rfloor+1},u-e_{\lfloor n/4\rfloor+1}\}$, respectively. Once again we have that $I(A:B)=0$ and $I(X:A)=1$. Formulas for $p(Z=0|B=0)$ and $p(Z=1|B=1)$ are more complicated:

$$p(Z=0|B=0)=\frac{1}{4}\left[2\cos\left(\frac{\pi}{n}\right)+\cos\left(2\pi\frac{\lfloor\frac{n}{4}\rfloor}{n}\right) +\cos\left(2\pi\frac{\lfloor\frac{n}{4}\rfloor-\lfloor\frac{n}{2}\rfloor}{n}\right)\right]\sec\left(\frac{\pi}{2n}\right)^2, \tag{E14}$$

$$p(Z=1|B=1)=\frac{1}{4}\left[2-\cos\left(2\pi\frac{\lfloor\frac{n}{4}\rfloor}{n}\right) -\cos\left(2\pi\frac{\lfloor\frac{n}{4}\rfloor-\lfloor\frac{n}{2}\rfloor-1}{n}\right)\right]\sec\left(\frac{\pi}{2n}\right)^2. \tag{E15}$$

However, we get that $p(Z=0|B=0)>1/2$ and $p(Z=1|B=1)>1/2$; hence $I(Z:B)>0$, which proves violation also in this case.

ICP violation in polygon theories is connected, as in the case of $p-$GNST, with uncertainty relations. It is easy to see especially for even n. We notice that for $n=4m+2$, where m is the integer, noncomplementary observables are measured. In the case of odd n, the role of uncertainty is less obvious because of asymmetry of the state ρ^{SAB}.

Correlations obtained in models with odd n do not violate Tsirelson's bound [20]. It means that this class of theories cannot be separated from the quantum theory using standard

argumentation [13]. Since nonlocality is tightly connected with uncertainty relations, it might be interesting to apply *ICP* to explain the impossibility of steering to maximally certain states [40].

Very recent results on the classical information transmission in polygon theories [30] provide some more insight into this issue. It turns out that polygon theories with odd n allow for communication of more than 1 bit per elementary system in Holevo sense (i.e., in asymptotic limit). This means that (B2) is violated even in one observable setup when a nonpure measurement is performed. Therefore our result for odd n may be viewed as a simple consequence of the fact that a Holevo-like capacity exceeds the number of bits which may be encoded in the system in a perfect decodable way. This is contrary to what we observe for classical and quantum systems. For even n, the Holevo-like capacity of the elementary system is 1 bit. This emphasizes the advantage of a multiobservable rather than Holevo-like approach in discrimination of nonlocal theories. It is interesting that for odd n, the information content for multiple observables may exceed the Holevo limit.

At the end of this section we use polygon theories to compare *ICP* with criterion based on the mismatch between the measurement dimension and information dimension [27]. Here the measurement dimension denotes the number of perfectly decodable states and information dimension the number of pairwise perfectly distinguishable states. For polygon theories with $n \in \{4,\ldots,13\}$, mismatch between the measurement dimension and information dimension take place only for $n \in \{4,6\}$. Using this approach, only two cases may be ruled out while *ICP* rules out all of them. However, it cannot be excluded that the mismatch criterion will rule out these theories, if we consider composite systems, with an appropriate choice of composition rules.

3. Composite systems

At this point we go back to p-GNSTs. We will consider p-GNSTs in their original formulation from [19], i.e., where three dichotomic and anticommuting observables may be measured on an elementary system.

We show that *ICP* is able to exclude nonlocal *GPT*, even if the elementary system state space is a Bloch sphere. Namely, we show that for p-GNST with $p = 2$, a big enough multipartite system violates *ICP*. For this purpose we take advantage of a superstrong RAC present in those theories. This will also demonstrate that *ICP* can be applied to composite systems.

As it was shown in [19], p-GNST with $p = 2$ allows for encoding 3^n bits in an n-gbit state with single-bit recovery probability equal $p_{\rm rec} = \frac{1}{2} + \frac{1}{2\sqrt{2n+1}}$. Since each bit is decoded by a different observable and the bits are distributed uniformly and independently, we obtain

$$\sum_i I_{M_i} - I_R = \sum_{i=1}^{3^n} I(O_i : A_i) = 3^n[1 - H(p_{\rm rec})], \quad \text{(E16)}$$

where i denotes bit A_i, which is decoded by observable O_i and H is classical entropy. On the other hand, for an n-gbit system, the maximal number of perfectly decodable states is bounded by $d \leqslant (3+1)^n$ [cf. (A1) and (A2)]. Putting this together we get that a 5-gbit system violates *ICP* (i.e., $\sum_i I_{M_i} - I_R = 16.18 > 10 = \log_2 4^5$). This result relates to [41–45], where it was shown that a locally quantum state space with no-signaling conditions implies a fully quantum state space for bipartite systems; however, the situation changes dramatically in multipartite scenarios.

The difference $\sum_i I_{M_i} - I_R$ depends strongly on the state space of the composite system. As we have seen, the uncertainty relations for anticommuting observables do not restrict state space strongly enough to ensure that *ICP* is satisfied. Therefore it is interesting to ask what other geometrical constraints [e.g., other uncertainty relations, consistency constraints, (cf. [19]), local orthogonality] have to be added to the theory to conform with *ICP*. In the opposite direction, one may ask how *ICP* limits the strength of nonlocal correlations achievable in *GPT* for systems whose elementary subsystems obey quantum mechanics.

APPENDIX F: *ICP* VERSUS EXISTING AXIOMS AND PRINCIPLES

It should be noted here that postulating *ICP* we do not search for "physical" justification for inequality (4), but rather in the spirit of information theoretical principles such as information causality [13] (IC) or local orthogonality [16], we aim for understanding the physical reality by means of an information approach. In this context it is natural to ask how *ICP* is related to the existing principles and axioms. As we mentioned in the Introduction, there are, in substance, two paradigms within which we try to understand the peculiar role of quantum mechanics in the set of possible theories of the physical world. The first one is to derive quantum mechanics from more intuitive axioms. The other is to pose a single principle which is more complicated than simple axioms but has a chance to at least rule out many theories with different predictions than quantum ones. For the sake of clarity we will refer further to these two different paradigms as "axioms" and "principles." Note that these two paradigms are not comparable. Clearly any principle can be derived from axioms as they reproduce quantum mechanics. However, there is never a simple connection between axioms and principles. For instance, the IC, we now know, is not capable of reproducing quantum mechanics (i.e., to be "worse" than axioms), but this clearly does not mean that it is indeed less important.

Since the crux of our constrained uncertainty principle is its information theoretic flavor and usage of strong subadditivity that holds for both the quantum and classical world, it cannot be a simple consequence of axioms such as, e.g., [9,10]. But how is it related to existing principles? There is a basic difference between them. Our principle applies to a single system, while all the existing principles involve correlations between subsystems and hence cannot be applied to a single system.

[1] A. Cabello, arXiv:1509.04711.
[2] C. F. von Weizsacker, Z. Naturforsch. **7a**, 141 (1952).
[3] G.-W. Mackey, *The Mathematical Foundations of Quantum Mechanics* (W.-A. Benjamin, Inc., New York, 1963).
[4] W. K. Wootters, Found. Phys. **16**, 391 (1986).
[5] A. Zeilinger, Found. Phys. **29**, 631 (1999).
[6] L. Hardy, arXiv:quant-ph/0101012.
[7] C. Brukner and A. Zeilinger, in *Time, Quantum and Information*, edited by L. Castell and O. Ischebeck (Springer, New York, 2003).
[8] B. Dakic and C. Brukner, in *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, edited by H. Halvorson (Cambridge University Press, Cambridge, UK, 2011), pp. 365–392.
[9] L. Masanes and M.-P. Muller, New J. Phys. **13**, 063001 (2011).
[10] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **84**, 012311 (2011).
[11] L. Masanes, M. P. Mueller, R. Augusiak, and D. Perez-Garcia, Proc. Natl. Acad. Sci. USA **110**, 16373 (2013).
[12] R. B. Griffiths, Phys. Rev. A **66**, 012311 (2002).
[13] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski, Nature (London) **461**, 1101 (2009).
[14] M. Navascues and H. Wunderlich, Proc. R. Soc. London A **466**, 881 (2009).
[15] A. Cabello, Phys. Rev. Lett. **110**, 060402 (2013).
[16] T. Fritz, A. B. Sainz, R. Augusiak, J. B. Brask, R. Chaves, A. Leverrier, and A. Acin, Nat. Commun. **4**, 2263 (2013).
[17] S. Popescu, Nat. Phys. **10**, 264 (2014).
[18] H. Maassen and J. B. M. Uffink, Phys. Rev. Lett. **60**, 1103 (1988).
[19] G. V. Steeg and S. Wehner, Quantum Inf. Comput. **9**, 801 (2009).
[20] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, New J. Phys. **13**, 063024 (2011).
[21] R. W. Spekkens, Phys. Rev. A **75**, 032110 (2007).
[22] S. D. Bartlett, T. Rudolph, and R. W. Spekkens, Phys. Rev. A **86**, 012103 (2012).
[23] That form of the principle was independently derived by K. Horodecki, unpublished notes (2012).
[24] A. Grudka, K. Horodecki, M. Horodecki, W. Klobus, and M. Pawlowski, Phys. Rev. Lett. **113**, 100401 (2014).
[25] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, J. ACM **49**, 496 (2002).
[26] M. Pawłowski and V. Scarani, in *Quantum Theory: Informational Foundations and Foils*, edited by Giulio Chiribella and Robert W. Spekkens, Fundamental Theories of Physics Vol. 181 (Springer, Netherlands, 2015), pp. 423–438.
[27] N. Brunner, M. Kaplan, A. Leverrier, and P. Skrzypczyk, New J. Phys. **16**, 123050 (2014).
[28] A. Cabello, Found. Phys. **42**, 68 (2012).
[29] M. J. W. Hall, Phys. Rev. Lett. **74**, 3307 (1995).
[30] S. Massar and M. K. Patra, Phys. Rev. A **89**, 052124 (2014).
[31] P. J. Coles, L. Yu, V. Gheorghiu, and R. B. Griffiths, Phys. Rev. A **83**, 062338 (2011).
[32] R. B. Griffiths, Phys. Rev. A **76**, 062320 (2007).
[33] N. Bohr, Nature (London) **121**, 580 (1928).
[34] J. Barrett, Phys. Rev. A **75**, 032304 (2007).
[35] G. Kimura, K. Nuida, and H. Imai, Rep. Math. Phys. **66**, 175 (2010).
[36] M. P. Muller, O. C. O. Dahlsten, and V. Vedral, Commun. Math. Phys. **316**, 441 (2012).
[37] S. W. Al-Safi and A. J. Short, Phys. Rev. A **84**, 042323 (2011).
[38] H. Barnum, L. Barrett, L.-O. Clark, M. Leifer, R. Spekkens, N. Stepanik, A. Wilce, and R. Wilke, New J. Phys. **12**, 033024 (2010).
[39] A. J. Short and S. Wehner, New J. Phys. **12**, 033023 (2010).
[40] J. Oppenheim and S. Wehner, Science **330**, 1072 (2010).
[41] H. Barnum, S. Beigi, S. Boixo, M. B. Elliott, and S. Wehner, Phys. Rev. Lett. **104**, 140401 (2010).
[42] A. Acín, R. Augusiak, D. Cavalcanti, C. Hadley, J. K. Korbicz, M. Lewenstein, L. Masanes, and M. Piani, Phys. Rev. Lett. **104**, 140404 (2010).
[43] S. Muhammad, A. Tavakoli, M. Kurant, M. Pawlowski, M. Zukowski, and M. Bourennane, Phys. Rev. X **4**, 021047 (2014).
[44] H. Barnum, Electron. Notes Theor. Comput. Sci. (ENTCS) **270**, 3 (2011).
[45] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, arXiv:0805.3553.