Interference as an information-theoretic game

Sebastian Horvat[1] and Borivoje Dakić[1,2]

[1]University of Vienna, Faculty of Physics, Vienna Center for Quantum Science and Technology, Boltzmanngasse 5, 1090 Vienna, Austria
[2]Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Vienna, Austria
17 February 2021

The double slit experiment provides a clear demarcation between classical and quantum theory, while multi-slit experiments demarcate quantum and higher-order interference theories. In this work we show that these experiments pertain to a broader class of processes, which can be formulated as information-processing tasks, providing a clear cut between classical, quantum and higher-order theories. The tasks involve two parties and communication between them with the goal of winning certain parity games. We show that the order of interference is in one-to-one correspondence with the parity order of these games. Furthermore, we prove the order of interference to be additive under composition of systems both in classical and quantum theory. The latter result can be used as a (semi)device-independent witness of the number of particles in the quantum setting. Finally, we extend our game formulation within the generalized probabilistic framework and prove that tomographic locality implies the additivity of the order of interference under composition. These results shed light on the operational meaning of the order of interference and can be important for the identification of the information-theoretic principles behind second-order interference in quantum theory.

I Introduction

As Richard Feynman famously put it, "the double slit experiment is absolutely impossible to explain in any classical way and has in it the heart of quantum mechanics. In reality, it contains the only mystery" [1]. Indeed, the most common way of introducing quantum theory and its basic tenets is via the double-slit experiment, in which a single particle is shown to produce an interference pattern when sent through two slits, in contrast to classical theory in which this effect is missing. Much later, Sorkin [2] analyzed multi-slit experiments (i.e. generalizations of the double slit experiment to three or more slits) and noticed that quantum mechanics exhibits second-order interference only, meaning that any measurement pattern produced by a quantum system is reducible to the combination of double-slit interferences. The latter served

Sebastian Horvat: sebastian.horvat@univie.ac.at
Borivoje Dakić: borivoje.dakic@univie.ac.at

as a motivation for introducing higher-order interference theories which are defined with respect to the order of interference produced in generalized multi-slit experiments [3]. Such theories are usually formulated within the framework of *generalized probabilistic theories* (GPT-s) [4, 5], with quantum theory being only a particular example. This further motivated the search for a set of intuitive physical principles which could explain why experiments should exhibit at most second-order interference [6–9].

In this work we study higher-order interference from the information-theoretic perspective. Firstly, we reconsider the definitions of multi-slit experiments and Sorkin's classification and notice that: (a) the framework is defined only for single particles (or single systems), (b) the interference order is defined with respect to simple operations of blocking/opening the slits, and (c) the (final) measurement refers exclusively to intensity measurements (average number of particles at a particular location on the screen). We proceed by generalizing this setting to an arbitrary number of particles (systems), arbitrary set of (local) operations and arbitrary final measurements. The framework for higher-order interference is formulated as an information-theoretic game where we analyze generic scenarios involving classical and quantum systems. Our generalization involves many particles (systems) and their ability to win certain parity games. The order of these games and their winning probabilities are directly related to the order of interference (in Sorkin's classification) and to the number of systems involved. This generalization offers a shift in the perspective: higher order interference theories are not defined by which phenomena are allowed (e.g. by the structure of interference patterns), but by which tasks can or cannot be accomplished within the theory (i.e. their information-processing capacity). Furthermore, our study provides a direct relation between the order of interference and number of systems used in the protocol, and can thus provide a (semi) device-independent way of witnessing the number of particles (systems) present in the process.

In the final section we provide the information-theoretic formulation within the GPT formalism and we study how the order of interference behaves under the composition of systems. We construct a lower bound on the interference order which is additive under composition. In classical and quantum theory, this lower bound coincides with the upper bound, which in turn shows an interesting connection between the order of interference and the

composition law. Finally, motivated by the latter, we prove that in generic local-tomographic theories [28] the order of interference is indeed additive (under certain additional assumptions that we deem reasonable).

Together with an information-theoretic perspective, our findings can be seen in light of paving an alternative way towards understanding the physical principles behind the order of interference of quantum theory: we suspect that an important clue might be provided by the composition of systems and by the tensor-product structure. The latter would thus contribute to the operational and physical understanding of quantum theory [10–13] and provide indications for potential developments of post-quantum theories. Moreover, our work deepens the connection between interference and information processing which has already been alluded to in various contexts involving two-way communication [14–16], information speed [17], quantum acausal processes [18], superposition of orders [19] and directions [20], quantum combs [21], quantum switch [22] and quantum causal models [23].

II Information-theoretic formulation: Parity games

In the standard double-slit experiment a particle is sent on a plate pierced by two parallel slits. After passing through the plate, the particle can be detected on a screen. Each slit can be either open (which we denote with 0), or blocked (which we denote with 1). The figure of merit is the interference term

$$I_2 = P_{00} - P_{10} - P_{01} + P_{11}, \tag{1}$$

where $P_{x_1 x_2}$ denotes the probability of detecting the particle at a point on the screen given that the slits are in states x_1 and x_2. Notice that we explicitly included the term which corresponds to the situation in which both slits are closed, despite it being necessarily zero (i.e. $P_{11} = 0$). Classical mechanics predicts $I_2 = 0$, while quantum theory allows the particle to be in spatial superposition, thereby enabling the possibility of generating a non-vanishing I_2.

In order to reformulate this experiment as an information-theoretic task, let us consider the scenario involving two parties, Alice and Bob, located at opposite sides of the plate, as shown in Figure 1. Let us suppose that the slits are in a certain state (i.e. each slit is either open or blocked) which is completely unknown both to Alice and to Bob; we can thus imagine that an external party (e.g. a referee) has control of the slits and decides whether to block them or not. Moreover, Alice possesses a single particle that she can send towards Bob. On the other side, Bob receives (or not) the particle, performs an arbitrary measurement and outputs a bit $b \in \{0, 1\}$. We introduce the redefined sec-
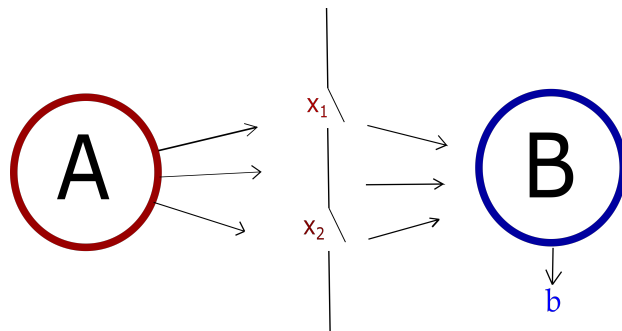

Figure 1: Alice and Bob are located at opposite sides of a pierced plate with two slits, each of which can be either open or blocked depending on their input bits $\{x_1, x_2\}$. Alice sends her particle through the slits towards Bob, who, upon receiving the particle, generates an output bit b.

ond order interference term as follows

$$\tilde{I}_2 = \frac{1}{4} \left[P(0|11) - P(0|10) - P(0|01) + P(0|00) \right] =$$

$$= \frac{1}{4} \sum_{x_1, x_2 = 0}^{1} P(b = x_1 \oplus x_2 | x_1 x_2) - \frac{1}{2}, \tag{2}$$

where $P(b|x_1 x_2)$ is the probability that Bob outputs b given that the two slits are in states x_1 and x_2. Notice that the term corresponding to both slits being closed can now be non-zero, since Bob's measurement is arbitrary, i.e. does not necessarily refer to measuring the probability of the particle inflicting on a point of the screen (intensity measurement). The probability $P(b|x_1 x_2)$ is thus a generalization of $P_{x_1 x_2}$ from (1) and reduces to the latter only in the case in which Bob performs the intensity measurement. The redefined interference term $\tilde{I}_2$ measures the probability of successfully accomplishing the following task (parity game):

(a) in each run, the slits are set in some state $\{x_1, x_2\}$ unknown both to Alice and to Bob;

(b) Alice prepares her particle and sends it through the slits towards Bob;

(c) Bob receives (or not) the particle, and, in order to complete the task, must output the parity $s_{\vec{x}} \equiv x_1 \oplus x_2$.

Bob does not have any prior information about the inputs, which is why a uniform average is taken. We stress that the state of the particle sent by Alice does not depend on the inputs x_1 and x_2, since the latter are unknown to her. Notice that expression (2) is formulated in a device-independent way [24], relating only inputs and outputs, without any mention of their underlying physical realization. It is therefore natural to generalize the scenario by replacing slit operations with generic black boxes, which implement arbitrary local operations depending on their inputs. The boxes can thus perform any operation (e.g. in quantum theory, general completely-positive maps): the only restriction stems from the operations being local. Moreover, instead of constraining Alice to sending only single particles (as it is also the case for standard interference experiments), we can generalize her resources to an

arbitrary number of systems and analyze how the winning probability depends on the number of systems used in the process (as we will see in the next section, the game can be won even in classical theory, by using two particles). Additionally, the particles/systems sent by Alice can have any internal structure (e.g. spin) that can be potentially accessed by the black boxes.

Analogously to the generalization of the standard double-slit experiment to multi-slit experiments, we can further generalize the scenario to an arbitrary number of boxes m. In this case, Alice sends her resources, which consist of k systems, towards Bob, whose task is to output the overall parity of the boxes' inputs, as depicted in Figure 2. The figure of merit is thus

$$\tilde{I}_m(k) = \frac{1}{2^m} \sum_{x_1,...,x_m=0}^{1} P(b = s_{\vec{x}}|x_1...x_m) - \frac{1}{2}, \quad (3)$$

where $\{x_1, ..., x_m\}$ are input bits encoded in the m boxes and $s_{\vec{x}} \equiv \oplus_{i=1}^{m} x_i$ is the overall parity. In the case in which Alice is constrained to sending only single particles, the boxes are implemented as slits and Bob's final operation consists of intensity measurements, the generalized higher-order interference term $\tilde{I}_m(k)$ reduces to the standard higher-order term defined in [2] (up to normalization):

$$I_m = \sum_{x_1,...,x_m=0}^{1} (-1)^{\sum_j x_j} P_{\vec{x}}, \quad (4)$$

where $x_j = 0(1)$ corresponds to j-th slit being open(closed) and $P_{\vec{x}}$ is the probability of detecting the particle on the screen given that the slits are in state $\vec{x}$.

Juxtaposed to the standard definition of m-th order interference theories involving the structure of interference patterns produced by single particles in m-slit experiments, our information-theoretic extension provides a definition in terms of the probability of winning the parity game involving m boxes by using a finite amount of resources. In addition, the latter formulation enables an analysis of the relation of the order of interference and the number of systems involved in the process (e.g., as we will see in the next section, the m-th order parity game can be won using m classical particles). We therefore introduce the characterization of higher-order theories in terms of functions $n(k)$, where n refers to the maximum order of interference that can be exhibited using k systems, i.e. $\tilde{I}_{m=n(k)}(k) \neq 0$ and $\tilde{I}_{m>n(k)}(k) = 0$.

To recapitulate, the modified multi-slit experiment consists of the following generalizations: (a) instead of one particle, Alice can possess arbitrarily many particles/systems of arbitrary internal structure, (b) the slits are replaced by black boxes which implement generic local operations depending on their pertaining input bits, (c) the screen is replaced by Bob who is supposed to generate an output corresponding to the overall parity of the inputs encoded by the boxes.

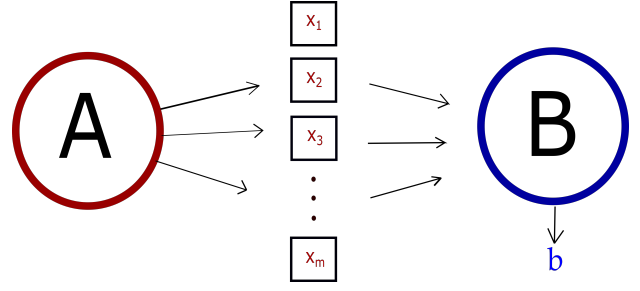Finally, we can generalize the input bits $\{x_1...,x_m\}$ to

Figure 2: Alice possesses a finite amount of resources that she sends towards Bob through m black boxes. The latter implement arbitrary local operations depending on their pertaining inputs $\{x_1, ..., x_m\}$. Upon receiving the resources, Bob performs an arbitrary operation and generates an output bit b.

dits, i.e. elements of a set with cardinality d. Throughout this manuscript we will assume that d is a prime number, the reason of which will be clearer later. In this case we introduce a class of games in which Bob is supposed to decode one of the *weighted sums modulo d* of the inputs, i.e. $s_{\vec{x},\vec{\nu}}^{(d)} \equiv (\sum_{i=1}^{m} \nu_i x_i) \mod d$, where $\vec{\nu}$ is a dit-string with components $\nu_i \in \{1...,d-1\}$. Given that Alice is using k systems, the generalized interference terms are then

$$\tilde{I}_{m,\vec{\nu},f}^{(d)}(k) = \frac{1}{d^m} \sum_{x_1,...,x_m=0}^{d-1} P\left(b = f(s_{\vec{\nu},\vec{x}}^{(d)})|x_1...x_m\right) - \frac{1}{d},$$
$$(5)$$

which are defined for all *reversible* functions f that map dits into dits. Operationally, these functions take into account the potential relabelling of Bob's outputs. Notice that for $d = 2$ there was no need for specifying this, since relabelling Bob's output in Eq. (3) introduces only a minus sign. The games defined in (5) offer a definition of higher order interference theories for arbitrary prime d.

Definition. *We define $n(k)$-theories as theories that satisfy the following two properties:*
(i) all processes produce $\tilde{I}_{m>n(k),\vec{\nu},f}^{(d)}(k) = 0$ for all reversible functions f, all dit-strings $\vec{\nu}$ with components $\nu_i \in \{1...,d-1\}$, and for all prime $d \geq 2$;
(ii) for some prime $d \geq 2$, there exists at least one process that produces $\tilde{I}_{m=n(k),\vec{\nu},f}^{(d)}(k) \neq 0$, for at least one dit-string $\vec{\nu}$ with components $\nu_i \in \{1...,d-1\}$.

Notice that the (non)existence of the required process for a specific outputs' labelling fixed by f implies the (non)existence of the analogous process for any other labelling f' (since Bob can always relabel his outcomes independently of the process). Therefore, throughout this manuscript, we will only construct proofs of existence of processes which exhibit $\tilde{I}_{m=n(k),\vec{\nu},f}^{(d)}(k) \neq 0$ for f fixed to be the identity map (i.e. no final relabelling) and we will thereby omit the index f. Moreover, whenever we omit the index $\vec{\nu}$, we refer to the unit dit-string, i.e. $\nu_i = 1, \forall i$. Therefore, the term $\tilde{I}_m^{(d)}(k)$ will correspond to the specific game in which Bob's goal is to output $s_{\vec{x}} = (\sum_i x_i) \mod$

d. We acknowledge that the additional games specified by the labels $\vec{\nu}$ and f may at the moment seem redundant; however, we will need to take them into account in the proofs regarding higher-order interference and we will thus keep them in the definition of $n(k)$-theories. As we will show later, all conclusions regarding the order of interference of classical, quantum and higher-order theories will hold for arbitrary prime $d \geq 2$.

Winning the parity games can be seen as a truly "global" task, in the sense that Bob needs to produce an output that depends on all the local inputs in *each run*. For suppose that Bob receives all but one input, say the j-th one: in this case, the overall modulo $s_{\vec{x}} = (\sum_i x_i) \bmod d$ is completely unknown to Bob. The same conclusion holds if Bob does not receive more inputs or even if he does not receive any information at all. Therefore, the interference term $\tilde{I}^{(d)}_{m,\vec{\nu},f}(k)$ is zero irrespectively of whether Bob lacks knowledge about one or about all inputs. Here we can see the importance of assuming that d is a prime number: if it was not, then, in some cases, Bob would be able to guess the overall modulo with a probability greater than $\frac{1}{d}$ even with having no information about one or more inputs. A simple example is the following one: assume that $d = 4$, $m = 2$, and that Bob is supposed to output the value $(x_1 + 2x_2) \bmod d$ (i.e. $\nu_1 = 1, \nu_2 = 2$); by knowing the value of x_1, Bob's winning probability will be $\frac{1}{2}$, which is larger than $\frac{1}{d}$, essentially due to d being divisible by ν_2. This would cause several complications in our definitions and proofs, which is why we will simply restrict d to be prime (while still letting it to be arbitrarily large).

In the subsequent sections, we are going to focus on classical and quantum theory and investigate their power to win the parity games. We will show that the order of interference $n(k)$ satisfies the relation

$$n(k) = kn(1), \tag{6}$$

where $n(1)$ is the order of single-system interference (for classical theory, $n(1) = 1$, while for quantum theory, $n(1) = 2$) and $n(k)$ is the interference order achievable using k systems. Moreover, we will show that this relation holds in generic GPT-s that satisfy the local tomography principle and certain additional assumptions.

Before proceeding further, we will first show an important mathematical property that relates the order of interference to the algebraic order of the probability distributions $P(b|\vec{x})$.

III Algebraic order of the probability distributions

In this section we are going to show that any distribution which does not exhibit higher than n-th order interference can be written as a linear combination of functions of at most n different inputs, or, in other words, that the algebraic order of $P(b|\vec{x})$ is at most n. Here we will consider the $d = 2$ case, while the general case is addressed in Appendix A.

Let us consider the parity game involving m boxes and assume that Alice uses one system of order $n < m$. By definition, this means that the interference terms (3) vanish for all parity games involving more than n boxes, which can be written in a concise form as follows:

$$\tilde{I}^{(2)}_{m,\vec{\sigma}} = \frac{1}{2^m} \sum_{\vec{x}} (-1)^{\sum_j \sigma_j x_j} P(0|\vec{x}) = 0,$$
$$\forall \sigma_j \in \{0, 1\}, \quad \text{s.t.} \quad \sum_j \sigma_j > n. \tag{7}$$

For compactness, we introduced the m-component bit string $\vec{\sigma}$, that specifies which interference term the latter equation refers to: if $\sigma_j = 1$ for $j = i_1, ..., i_l$, the equation states that the order of interference involving boxes $i_1, ..., i_l$ is equal to zero (l can be any integer between $n+1$ and m).

Let us regard $P(0|\vec{x})$ as one component of a vector $\vec{P}$ in a 2^m-dimensional vector space formed by the tensor product of m two-dimensional spaces, i.e. $P(0|\vec{x}) = \vec{e}_{x_1} \otimes ... \otimes \vec{e}_{x_m} \cdot \vec{P}$, where $\vec{e}_{x_j=0,1}$ span the j-th two-dimensional space. Equations (7) then imply that for all $\sum_j \sigma_j > n$:

$$\tilde{I}^{(2)}_{m,\vec{\sigma}}(k=1) = \frac{1}{2^m} \sum_{\vec{x}} (-1)^{\sum_j \sigma_j x_j} \vec{e}_{x_1} \otimes ... \otimes \vec{e}_{x_m} \cdot \vec{P} =$$
$$= \frac{1}{2^{m/2}} \vec{f}_{\sigma_1} \otimes ... \otimes \vec{f}_{\sigma_m} \cdot \vec{P} = \frac{1}{2^{m/2}} \lambda_{\vec{\sigma}} = 0, \tag{8}$$

where we introduced the rotated vectors

$$\vec{f}_{\sigma_j} \equiv \frac{1}{\sqrt{2}} \left(\vec{e}_0 + (-1)^{\sigma_j} \vec{e}_1 \right). \tag{9}$$

The terms $\lambda_{\vec{\sigma}}$ are components of vector $\vec{P}$ in the new basis spanned by the rotated vectors $\vec{f}_{\sigma_j}$. Equation (8) then states that the components $\lambda_{\vec{\sigma}}$ are zero if $\sum_j \sigma_j > n$. The probabilities $P(0|\vec{x})$ can thus be expressed as

$$P(0|\vec{x}) = \left[\bigotimes_{i=1}^m \vec{e}_{x_i} \right] \cdot \left[\sum_{\vec{\sigma}} \lambda_{\vec{\sigma}} \bigotimes_{j=1}^m \vec{f}_{\sigma_j} \right] =$$
$$= \frac{1}{2^{m/2}} \sum_{\substack{\sigma_1, ..., \sigma_m \\ \sum_j \sigma_j \leq n}} (-1)^{\sum_j \sigma_j x_j} \lambda_{\vec{\sigma}}, \tag{10}$$

where we used $\vec{e}_{x_i} \cdot \vec{f}_{\sigma_i} = 2^{-1/2}(-1)^{\sigma_i x_i}$. Therefore, $P(0|\vec{x})$ is a linear combination of functions of at most n different inputs:

$$P(0|\vec{x}) = \sum_{l=1}^n \sum_{i_1, ..., i_l} c^{(l)}_{i_1, ..., i_l} g^{(l)}(x_{i_1}, ..., x_{i_l}), \tag{11}$$

where we introduced the functions

$$g^{(l)}(x_{i_1}, ..., x_{i_l}) = (-1)^{\sum_{a=1}^l x_{i_a}}, \tag{12}$$

and the coefficients

$$c^{(l)}_{i_1, ..., i_l} = \frac{1}{2^{m/2}} \vec{\lambda}_{\vec{\sigma}^{(l)}_{(i_1, ..., i_l)}}, \quad \left(\vec{\sigma}^{(l)}_{(i_1, ..., i_l)} \right)_s = \sum_{r=1}^l \delta_{s, i_r}. \tag{13}$$

Next, in order to tackle the problem for arbitrary d, for each set of generalized interference terms $\left\{\tilde{I}^{(d)}_{m,\vec{\nu},f}, \forall \vec{\nu}, f\right\}$ we introduce its *dual* terms $\left\{\tilde{J}^{(d)}_{m,\vec{\nu},\alpha,b}, \forall \vec{\nu}; \forall b \in \{0,...,d-1\}\right\}$ as

$$\tilde{J}^{(d)}_{m,\vec{\nu},b} = \frac{1}{d^m} \sum_{\vec{x}} (\omega_d)^{\sum_a \nu_a x_a} P(b|\vec{x}), \qquad (14)$$

where $\omega_d = e^{i2\pi/d}$ is the d-th root of unity. For the $d = 2$ case, the dual terms coincide with the ones defined in terms of the game formulation (3); however, this is not the case for $d > 2$. Nevertheless, in Appendix A we show that the generalized interference term $\tilde{I}^{(d)}_{m,\vec{\nu},f}$ from (5) vanishes for all relabellings f and all dit-strings $\vec{\nu}$ if and only if $\tilde{J}^{(d)}_{m,\nu,b} = 0$, for all $b = 0,...,d-1$ and for all $\vec{\nu}$. This allows us to characterize the order of interference by analysing the behaviour of its dual, which can often be more tractable. Indeed, by this method we show in Appendix B that the conclusion derived in (12) holds for arbitrary d, i.e. that any distribution that exhibits at most m-th order interference can be written as a linear sum of functions which depend on at most m dits.

IV Classical resources

Now we are going to analyze the generalized interference terms achievable within classical theory. Here we will keep the analysis at an informal and intuitive level; a more formal treatment will be given at the end of Section V, where classical theory will be regarded as a special case of quantum theory.

Single systems. For a start, let us focus on the simplest scenario, i.e. the generalization of the double slit experiment, in which Alice sends a single particle through two boxes towards Bob (for now we stick to binary inputs, i.e. $d = 2$). Recall that the two boxes can implement arbitrary local operations (not only blocking/opening the slits) labelled by x_1 and x_2. If the particle is classical, i.e. has a definite trajectory, Bob's output can in each run depend on the state of at most one box. The conditional probabilities describing the process can thus be decomposed as follows:

$$P(b|x_1 x_2) = q_1 P_1(b|x_1) + q_2 P_2(b|x_2), \qquad (15)$$

where q_i is the probability of Alice sending the particle through i-th box and $P_i(b|x_i)$ is the conditional probability of Bob outputting b given that the particle was sent through i-th box. The modified interference term therefore vanishes:

$$\tilde{I}^{(2)}_2(k=1) = \frac{1}{4} \sum_{x_1,x_2=0}^{1} P(b = x_1 \oplus x_2 | x_1 x_2) - \frac{1}{2}$$
$$= \frac{1}{4} \sum_{x_1,x_2=0}^{1} (-1)^{x_1+x_2} P(0|x_1 x_2) = 0.$$
$$(16)$$

Intuitively, the knowledge about the state of one box does not increase the probability of correctly guessing the inputs' parity with respect to a random guess. The same reasoning holds for arbitrary prime d, implying that $\tilde{I}^{(d)}_2(k=1) = 0$ for one classical particle.

Multiple systems. Now, what if Alice possesses two classical particles? Let us for the moment assume that the boxes are implemented by slits (as in the original double-slit experiment) and that Alice in each run sends deterministically one particle towards the first slit and the other towards the second slit. If the parity of the slits' states is $x_1 \oplus x_2 = 0$, then Bob either receives both particles on his side or he receives no particles at all; alternatively, if the parity is 1, Bob receives exactly one particle. Thus, by simply counting the number of received particles, Bob can in each run determine the inputs' parity and can perfectly accomplish the required task, thereby generating $\tilde{I}^{(2)}_2(k=2) = 1/2$. On the other hand, the standard interference definition (1) remains null also for two particles, since the average number of particles received by Bob (or inflicted on the screen) is equal to 1 regardless of the inputs' parity. This shows the difference between the standard formulation and our game formulation, as the latter allows Bob to measure coincidences, and not only the average particle number (i.e. intensity).

We proceed by analysing the fully general scenario with m boxes that implement arbitrary local transformations and assume that Alice's resources consist of k classical objects, be it particles, conglomerates of particles or localized wave packets. For this reason, it is clear that Alice's resources cannot interact with more than k boxes, which means that Bob's output can depend on at most k inputs: if $k < m$, Bob's output is equivalent to a random guess and $\tilde{I}^{(d)}_m(k) = 0$, while if $k \geq m$, he can in principle deterministically accomplish the required task and thus generate $\tilde{I}^{(d)}_m(k) = 1 - 1/d$ (if Alice sends one system per box, Bob can in principle retrieve all inputs). The same reasoning holds for all the other interference terms $\tilde{I}^{(d)}_{m,\vec{\nu},f}$ (as already noted, this reasoning would not be valid if d was not a prime number).

Therefore, classical theory satisfies $n(k) = k$, meaning that k classical systems can exhibit at most k-th order interference.

V Quantum resources

Contrasted to classical mechanics, quantum theory allows spatial superpositions of physical objects, which can generate a non-zero interference term I_2, even with a single particle. On the other hand, higher order interference terms $I_{j>2}$ defined in multi-slit experiments remain null even in quantum theory (see for instance [2]). In this section we show that analogous statements hold for the generalized interference terms $\tilde{I}^{(d)}_{m,\vec{\nu},f}(k=1)$ and provide an extension to more particles (systems). Let us first look into the simplest case, i.e. one particle and two boxes.

1 Two boxes

We start by considering the case $d = 2$. Let us suppose that Alice possesses a single quantum particle and sends it in spatial superposition towards the two boxes, which implement binary inputs x_1, x_2. The quantum state is given by

$$|\psi\rangle_0 = \frac{1}{\sqrt{2}} \left(|1\rangle + |2\rangle \right), \qquad (17)$$

where $|1\rangle$ and $|2\rangle$ are states corresponding to the two paths. Next, suppose that each box interacts with the particle by applying a local phase-shift $\phi_i = x_i\pi$. After passing through the boxes, the state is

$$|\psi\rangle_{x_1, x_2} = \frac{1}{\sqrt{2}} \left(e^{ix_1\pi} |1\rangle + e^{ix_2\pi} |2\rangle \right). \qquad (18)$$

Therefore, Bob receives the particle in the following states (up to global phases)

$$|\psi\rangle = \begin{cases} \frac{1}{\sqrt{2}} \left(|1\rangle + |2\rangle \right), & \text{if } x_1 \oplus x_2 = 0, \\ \frac{1}{\sqrt{2}} \left(|1\rangle - |2\rangle \right), & \text{if } x_1 \oplus x_2 = 1, \end{cases} \qquad (19)$$

which are orthogonal and thus perfectly distinguishable, thereby enabling Bob to deterministically decode the parity $x_1 \oplus x_2$ and to produce $\tilde{I}_2^{(2)}(k=1) = 1/2$.

The previous result holds for binary inputs; now suppose that the parties are playing the modulo game (5) specified by dit string $\vec{\nu} = (\nu_1, \nu_2)$, i.e. Bob's goal is to output $s_{\vec{x}, \vec{\nu}}^{(d)} = (\nu_1 x_1 + \nu_2 x_2) \bmod d$ (we assume there is no final relabelling, i.e. that f is the identity; as we already argued, all other cases follow automatically). The players can employ the following strategy.

As in the binary case, let Alice prepare her particle in the state

$$|\psi\rangle_0 = \frac{1}{\sqrt{2}} \left(|1\rangle + |2\rangle \right), \qquad (20)$$

where $|1\rangle$ and $|2\rangle$ are states corresponding to the two paths, and let the boxes apply local phase shifts as follows

$$\begin{aligned} |\psi\rangle_{x_1, x_2} &= \frac{1}{\sqrt{2}} \left(\omega_d^{-\nu_1 x_1} |1\rangle + \omega_d^{\nu_2 x_2} |2\rangle \right) \\ &= \frac{1}{\sqrt{2}} \left(|1\rangle + (\omega_d)^{s_{\vec{x}, \vec{\nu}}^{(d)}} |2\rangle \right), \end{aligned} \qquad (21)$$

where the last equality is valid up to a global phase. Upon receiving the latter state, Bob performs a d-outcome POVM $\{\Pi_0, ..., \Pi_{d-1}\}$ corresponding to the d moduli; the interference term is then

$$\begin{aligned} \tilde{I}_{2,\vec{\nu}}^{(d)} &= \frac{1}{d^2} \sum_{x_1 x_2} P\left(b = s_{\vec{x}, \vec{\nu}} | x_1 x_2 \right) - \frac{1}{d} \\ &= \frac{1}{d} \sum_s \text{Tr}\left(\Pi_s \rho^{(s)} \right) - \frac{1}{d}, \end{aligned} \qquad (22)$$

where

$$\rho^{(s)} \equiv \frac{1}{d} \sum_{\substack{x_1, x_2 \\ s_{\vec{x}, \vec{\nu}}^{(d)} = s}} \rho_{\vec{x}}, \qquad (23)$$

and $\rho_{\vec{x}} = |\psi\rangle_{x_1, x_2} \langle\psi|$.

Let us choose Bob's POVM to be a "pretty good measurement" [25], which in our case reduces to a set of projectors up to normalization, i.e.

$$\Pi_s = \alpha \rho^{(s)}, \qquad (24)$$

where α is a normalization constant. The latter can be easily calculated by requiring $\sum_s \Pi_s = \mathbb{1}$, thereby obtaining $\alpha = \frac{2}{d}$.

The interference term is then

$$\tilde{I}_{2,\vec{\nu}}^{(d)} = \frac{1}{d} \sum_s \text{Tr}\left(\frac{2}{d} \rho^{(s)} \rho^{(s)} \right) - \frac{1}{d} = \frac{1}{d} > 0. \qquad (25)$$

We have thus shown that a single quantum particle can be used to exhibit second order interference for any d. Notice that in this protocol we have not used the internal degree of freedom of the particle and we thus expect that larger interference values may arise in more sophisticated protocols.

2 General case

Here we analyze the fully general scenario involving arbitrary local operations and arbitrary measurements. We start by constraining Alice's resources to a single quantum system and afterwards we extend our considerations to the case of multiple systems.

2.1 Single systems

For a start, let us fix the resources to one quantum system, without restricting its internal degrees of freedom. The resource can thus be an electron, an atom or any localized quantum system that can be prepared in coherent spatial superposition using a beam splitter or some more sophisticated mechanism. The local operations implemented by the m boxes will be represented as CP maps acting on the internal degrees of freedom of the system.

The most general state Alice can prepare is given by

$$\rho_0 = \sum_{i, j, k, l} c_{ijkl} |i\rangle \langle j| \otimes |\phi_k\rangle \langle\phi_l|, \qquad (26)$$

where $\{|i\rangle, i = 1, ..., m\}$ denote states representing defined paths directed towards the m boxes, while $\{|\phi_k\rangle\}$ span the Hilbert space of the internal degrees of freedom of the system (the dimension of which is arbitrary). The matrix elements c_{ijkl} need to satisfy certain conditions in order for ρ_0 to be a legitimate quantum state, but we will not specify them since this constraint is not relevant for what follows.

Each box implements a local CP-map that depends on its corresponding input. We denote the Kraus operators representing the action of the i-th box with

$$\left\{ B_i^{(k)}(x_i), \quad \sum_k B_i^{(k)\dagger}(x_i) B_i^{(k)}(x_i) \le \mathbb{1} \right\}, \qquad (27)$$

where the operators $B_i^{(k)}(x_i)$ act on the internal Hilbert space of the system. Notice that we allow the transformations to be non-trace-preserving, since they can destroy the system, as for instance in the case of slits. Since the boxes act locally on the system, the total transformation implemented by the boxes is a CP-map with the following Kraus operators:

$$\left\{ M_{\vec{x}}^{(k)} = \sum_{i=1}^{m} |i\rangle \langle i| \otimes B_i^{(k)}(x_i) \right\}. \qquad (28)$$

The Kraus operators of the total map are thus obtained by conditioning the Kraus operators of the local maps on the path degree of freedom. For example, in the case that the local operators are unitary transformations, for each configuration of inputs there is only one Kraus operator, i.e.

$$M_{\vec{x}} = \sum_{i=1}^{m} |i\rangle \langle i| \otimes U_i(x_i), \qquad (29)$$

where $U_i(x_i)$ are unitary operators. Let us now inspect how the general CP-maps in (28) act on the input state (26). After passing through the boxes, the state of the system is

$$\rho_{\vec{x}} = \sum_k M_{\vec{x}}^{(k)} \rho_0 M_{\vec{x}}^{(k)\dagger} = \sum_{i,j} |i\rangle \langle j| \otimes C_{ij}(x_i, x_j), \qquad (30)$$

where, in order to simplify the expression, we introduced the following operators

$$C_{ij}(x_i, x_j) \equiv \sum_{knm} c_{ijnm} B_i^{(k)}(x_i) |\phi_n\rangle \langle\phi_m| B_j^{(k)\dagger}(x_j). \qquad (31)$$

Next, let us suppose that Alice and Bob are playing the modulo game (5) specified by a dit-string $\vec{\nu}$ and with no final relabelling (all other cases follow automatically). Notice that for $m > 2$ and for any $\alpha \in \{1, ..., d-1\}$, the following property holds

$$\sum_{\vec{x}} (\omega_d)^{\alpha \sum_k \nu_k x_k} \rho_{\vec{x}} =$$
$$= \sum_{i,j} |i\rangle \langle j| \otimes \sum_{\vec{x}} (\omega_d)^{\alpha \sum_k \nu_k x_k} C_{ij}(x_i, x_j) = 0, \qquad (32)$$

where $\omega_d \equiv e^{i2\pi/d}$. This is so because the tensors C_{ij} depend on at most two inputs and because $\sum_{x_k=0}^{d-1} (\omega_d)^{\alpha \nu_k x_k} = 0$.

Let us define the average taken over states with equal modulo as follows

$$\rho^{(S)} \equiv \frac{1}{d^{m-1}} \sum_{\substack{x_1, ..., x_m \\ (\sum_k \nu_k x_k) \bmod d = S}} \rho_{\vec{x}}. \qquad (33)$$

As we show in Appendix C, equations (32) then imply

$$\rho^{(S)} = \rho^{(S')} \equiv \bar{\rho}, \quad \forall S, S' = 0, ..., d-1. \qquad (34)$$

Since all average states $\rho^{(S)}$ are equal, Bob cannot distinguish them by any means. Formally, Bob performs a measurement consisting of d outcomes and represented by a generic POVM $\{\Pi_0, ..., \Pi_{d-1}\}$. If Alice sends $k = 1$ quantum systems, the generalized interference term specified by the dit-string $\vec{\nu}$ necessarily vanishes for $m > 2$:

$$\tilde{I}_{m>2,\vec{\nu}}^{(d)}(k=1) = \frac{1}{d} \text{Tr}\left[\sum_S \Pi_S \rho^{(S)} \right] - 1/d$$
$$= \frac{1}{d} \text{Tr}[\mathbb{1}\bar{\rho}] - 1/d = 0. \qquad (35)$$

A resource consisting of one quantum system is enough to produce non-vanishing second order interference (for binary inputs, it can even raise the interference term to its maximum possible value); on the other hand, for $m > 2$, there is no difference between sending one quantum system and sending no resource at all. This drastic difference can be traced down to (32), where the tensors C_{ij} couple at most two inputs x_i and x_j. This is because quantum states are described by $(1,1)$ tensors, i.e. density matrices. The latter constraint arises through the Born rule, which essentially sets quantum interference to the second order [3].

2.2 Multiple systems and additivity of interference

Next, let us suppose that Alice's resources consist of more than one quantum system, say, in general, k of them. We assume that the systems are distinguishable and thus described by the tensor product of the single-system Hilbert spaces. The most general state prepared by Alice is then

$$\rho_0 = \sum_{\mathbf{i,j,r,l}} c_{\mathbf{ijrl}} \bigotimes_{p=1}^{k} \left[|n_{i_p}^{(p)}\rangle \langle n_{j_p}^{(p)}| \otimes |\phi_{r_p}^{(p)}\rangle \langle \phi_{l_p}^{(p)}| \right], \quad (36)$$

where $\mathbf{i}$ is short for $(i_1, ..., i_k)$ and analogously for the other indices. The vectors $\left\{ |n_{i_p}^{(p)}\rangle, \forall i_p = 1, ..., m \right\}$ span the spatial Hilbert space of the p-th system, while $\left\{ |\phi_{r_p}^{(p)}\rangle, \forall r_p \right\}$ span its internal degrees of freedom (the dimensionality of which is arbitrary).

After passing through the boxes, the state is transformed to

$$\rho_{\vec{x}} = \sum_{\mathbf{i,j}} \bigotimes_{p=1}^{k} \left[|n_{i_p}^{(p)}\rangle \langle n_{j_p}^{(p)}| \right] \otimes C_{\mathbf{ij}}(x_{i_1}, x_{j_1}, ..., x_{i_k}, x_{j_k}), \qquad (37)$$

where $\{C_{\mathbf{ij}}\}$ are operators depending on at most $2k$ inputs (defined in the same fashion as in the single-system case, see Eq. (31)). Here we again assumed that the action of the boxes is represented by local CP-maps conditioned on the path degrees of freedom of the systems (see Appendix D for details). The crucial difference with respect to the single system case is that the expressions (32) are now valid only if $m > 2k$. Analogously to the single-system case, one comes to the following conclusion

$$\tilde{I}_{m>2k,\vec{\nu}}^{(d)}(k) = 0. \qquad (38)$$

We therefore showed that k quantum systems *cannot* produce more than $2k$-th order interference. Now we are going to show that k systems *can* produce $2k$-th order interference. For binary inputs (i.e. $d = 2$), Alice and Bob can partition the $2k$ boxes into pairs $\{(x_1, x_2), ..., (x_{2k-1}, x_{2k})\}$ and for each pair use the protocol described in Section 1; Bob can thus perfectly decode the parity of each pair, which enables him to win the game with unit probability. For $d > 2$, the proof is not so straightforward, since the protocol involving one particle and two boxes does not raise the interference term to its maximum value as it does for binary inputs. Nevertheless, in Appendix E we derive a general result which shows that k generic systems of single-system orders $\{n_1, ..., n_k\}$ *can* produce $(\sum_{i=1}^{k} n_i)$-th order interference, for any $d \geq 2$ (this statement is independent of the underlying physical theory). Therefore, k quantum systems (i.e. systems of order two) can produce $2k$-th order interference.

According to our classification, the latter results imply that quantum theory satisfies $n(k) = 2k$, meaning that k quantum systems can produce at most $2k$-th order interference[1].

Moreover, since the order of interference depends sharply on the number of particles used in the process, it can be used as a semi-device-independent witness of the number of particles. More precisely, one can extract a lower bound on the number of particles present in the process from purely operational quantities (i.e. conditional probability distributions $P(b|\vec{x})$). This witness is not fully device-independent, since one must still assume that the underlying mechanism obeys quantum theory and that the particle number is well defined (the latter assumption can be violated if the resources consist of e.g. coherent states of light). Additionally, even though the boxes' operations can be mutually spacelike separated (and therefore the boxes cannot mutually signal their local inputs to each other), each box's operation must still be timelike separated from Bob (since the systems must in the end be sent from the boxes towards Bob). Thus, one cannot exclude the possibility that a box signals its local input towards Bob by sending some additional system, which had not been previously sent by Alice. Our particle witness can be seen as similar to device independent dimension witnesses, where instead of the number of particles, one aims to certify the Hilbert space dimension of the underlying system [26, 27]. *Classical theory revisited.* Even though we already showed that classical theory satisfies $n(k) = k$ in Section IV, here we want to briefly explicate how the latter can be seen by treating classical theory as a special

case of quantum theory. We simply need to restrict the one-particle-states sent by Alice to be block diagonal in the spatial basis, and the multi-particle-states to be block-diagonal in the configuration space basis. The internal degrees of freedom can remain quantum (since it does not affect our result): here we thus regard "classical theory" to mean "quantum theory with no spatial coherence". A general classical k-particle-state is then:

$$\rho_0 = \sum_{\mathbf{i}, \mathbf{r}, \mathbf{l}} c_{\mathbf{irl}} \bigotimes_{p=1}^{k} \left[|n_{i_p}^{(p)}\rangle \langle n_{i_p}^{(p)}| \otimes |\phi_{r_p}^{(p)}\rangle \langle \phi_{l_p}^{(p)}| \right], \quad (39)$$

where the notation is the same as before. Analogously to the quantum case, the transformed state is

$$\rho_{\vec{x}} = \sum_{\mathbf{i}} \bigotimes_{p=1}^{k} \left[|n_{i_p}^{(p)}\rangle \langle n_{i_p}^{(p)}| \right] \otimes C_{\mathbf{i}}(x_{i_1}, ..., x_{i_k}), \quad (40)$$

where $C_{\mathbf{i}}(x_{i_1}, ..., x_{i_k})$ are tensors as the ones defined in the quantum case. Since the transformed state is a linear combination of terms, each of which depends on at most k inputs, we recover the result that classical theory (i.e. quantum theory with no spatial coherence) satisfies $n(k) = k$.

VI Higher-order interference theories

1 Independent operations

In Appendix E we constructed a proof that k generic systems of orders $\{n_1, ..., n_k\}$ can exhibit $\sum_i n_i$-th order interference, for any $d \geq 2$. The latter provides a lower bound which is additive in the interference order. On the other hand, we saw that in classical and quantum theory, this lower bound coincides with the upper bound, i.e. these theories satisfy $n(k) = kn(1)$. In what follows, we will discuss the extendability of this relation to a broader class of theories. In this subsection we will assume that the boxes act independently on each subsystem and show that additivity follows straightforwardly from the principle of local tomography [28]; in the next subsection we will drop this restriction, but in order to prove our statement we will need to introduce further assumptions.

We adopt the standard formalism of *generalized probabilistic theories* (GPTs) [28], and suppose that Alice possesses k systems. The state space of the j-th system is a convex set that will be denoted with $\text{St}(A_j)$, and we regard it as a subset of a real vector space A_j. The set of transformations consists of linear maps acting on A_j, and will be denoted with $\text{Transf}(A_j)$. Finally, the set of effects $\text{Eff}(A_j)$ consists of linear functionals that map states into probabilities and can thus be regarded as subset of the dual vector space A_j^*.

Now we proceed with specifying how systems compose. We assume the principle of *local tomography* [28], which states that the state of a composite system can be fully characterized by measurements performed separately on

---

[1]Notice that the same result holds also for indistinguishable quantum systems. In that case, the composite Hilbert space is the (anti)symmetric subspace of the tensor product of the single-system Hilbert spaces. Therefore, the order of interference cannot be higher than the one achievable by indistinguishable systems, since further constraints are imposed on the allowed states and transformations. Moreover, the additivity of the lower bound proved in Appendix E holds for any systems, including also indistinguishable quantum systems. This implies that all quantum systems satisfy $n(k) = 2k$.

each subsystem. Under this assumption, the composite state space of Alice's systems, which we denote with St(A), is a subset of the tensor product of the subsystems' vector spaces, i.e. St(A) $\subset A_1 \otimes ... \otimes A_k$. Moreover, the composite space of effects Eff(A) is a subset of the tensor product of the single-systems' dual vector spaces, i.e. Eff(A) $\subset A_1^* \otimes ... \otimes A_k^*$. Furthermore, as mentioned above, in this subsection we will assume that the boxes act independently on the subsystems; we will discuss the general scenario in the next subsection. We want to show that under the above assumptions, k systems of orders $\{n_1, ..., n_k\}$ cannot exhibit more than $\sum_i n_i$-th order interference. Together with the lower bound derived in Appendix E, this will imply that k systems of orders $\{n_1, ..., n_k\}$ constitute a $\sum_i n_i$-th order system. Let us first focus on binary inputs, i.e. $d = 2$.

An arbitrary state $\vec{s} \in$ St(A) prepared by Alice can be written as a linear combination of the tensor product of single system states: $\vec{s} = \sum_{\mathbf{i}} a_{\mathbf{i}} \vec{s}_{i_1}^{(1)} \otimes ... \otimes \vec{s}_{i_k}^{(k)}$, where $\mathbf{i}$ is short for $i_1, ..., i_k$, and $a_{\mathbf{i}}$ are real coefficients. The same holds for an arbitrary effect $\vec{f} \in$ Eff(A) representing Bob's final measurement, i.e. $\vec{f} = \sum_{\mathbf{i}} b_{\mathbf{i}} \vec{f}_{i_1}^{(1)} \otimes ... \otimes \vec{f}_{i_k}^{(k)}$. Since we consider only single-system operations, the overall transformation of the boxes on the composite system is given by $T_{\vec{x}} = T_{\vec{x}}^{(1)} \otimes ... \otimes T_{\vec{x}}^{(k)}$, where $T_{\vec{x}}^{(j)} \in$ Transf(A_j) are input-dependent transformations acting on the j-th system. For $d = 2$, the m-th order interference term for an arbitrary process is then:

$$\tilde{I}_m^{(2)}(k) = \frac{1}{2^m} \sum_{\vec{x}} (-1)^{\sum_a x_a} P(0|\vec{x})$$

$$= \frac{1}{2^m} \sum_{\mathbf{i},\mathbf{j}} a_{\mathbf{i}} b_{\mathbf{j}} \sum_{\vec{x}} (-1)^{\sum_a x_a} \left[\bigotimes_{p=1}^{k} \vec{f}_{j_p}^{(p)} \right] \cdot \left[\bigotimes_{p'=1}^{k} T_{\vec{x}}^{(p')} \vec{s}_{i_{p'}}^{(p')} \right]$$

$$= \frac{1}{2^m} \sum_{\mathbf{i},\mathbf{j}} a_{\mathbf{i}} b_{\mathbf{j}} \sum_{\vec{x}} (-1)^{\sum_a x_a} \prod_{p=1}^{k} P_{\vec{x}}^{(p)},$$

$$(41)$$

where $P_{\vec{x}}^{(p)} \equiv \vec{f}_{j_p}^{(p)} \cdot T_{\vec{x}}^{(p)} \vec{s}_{i_p}^{(p)}$ are probabilities arising from the single system processes involving p-th system (for simplicity we omitted indices $\mathbf{i}, \mathbf{j}$). Therefore, we see that due to local tomography and the restriction to single-system operations, the interference term decouples into a linear combination of products of single-system processes. As it was shown in Section III, the probability distribution pertaining to any process involving a system of order n_p can be written as a linear combination of functions of at most n_p inputs:

$$P_{\vec{x}}^{(p)} = \sum_{i_1, ..., i_{n_p}} c_{i_1, ..., i_{n_p}}^{(p)} f^{(p)}(x_{i_1}, ..., x_{i_{n_p}}). \qquad (42)$$

Therefore, if $N \equiv \sum_p n_p < m$, the interference term necessarily vanishes:

$$\tilde{I}_m^{(2)}(k) = \sum_{i_1, ..., i_N} a_{i_1, ..., i_N} \sum_{\vec{x}} (-1)^{\sum_a x_a} g(x_{i_1}, ..., x_{i_N}) = 0, \qquad (43)$$

where we introduced for simplicity the coefficients $a_{i_1, ..., i_N}$ and functions $g(x_{i_1}, ..., x_{i_N})$, the exact forms of which are of no relevance.

In order to show that the latter results holds for arbitrary prime d, we just need to take a look at the dual interference terms defined in Section III. Given our assumptions, one arrives to the generalization of (41):

$$\tilde{J}_{m,\vec{\nu},b}^{(d)}(k) = \frac{1}{d^m} \sum_{\mathbf{i},\mathbf{j}} a_{\mathbf{i}} b_{\mathbf{j}} \sum_{\vec{x}} (\omega_d)^{\sum_a \nu_a x_a} \prod_{p=1}^{k} P_{b,\vec{x}}^{(p)}.$$

$$(44)$$

By assumption, the probabilities $P_{b,\vec{x}}^{(p)} = \vec{f}_{b,j_p}^{(p)} \cdot T_{\vec{x}}^{(p)} \vec{s}_{i_p}^{(p)}$ pertaining to p-th system do not exhibit more than n_p-th order interference and can thus be expressed as a linear combination of functions depending on at most n_p inputs, as we showed in Appendix B. Therefore, if $N \equiv \sum_p n_p < m$, then $\tilde{J}_{m,\vec{\nu},b}^{(d)}(k)$ vanishes for all $\vec{\nu}, b$. Since the dual formulation is equivalent to the game formulation (see Appendix A), this implies that the k systems cannot be used to achieve more than $\sum_j n_j$-th order interference for any prime d.

On a side note, one may wonder whether the above proof could be applied to general non-independent transformations, since local tomography implies that all transformations can be written as a linear combination of independent transformations, i.e. $T_{\vec{x}} = \sum_{\mathbf{i}} c_{\mathbf{i}}(\vec{x}) \bigotimes_{p=1}^{k} T_{i_p,\vec{x}}^{(p)}$. However, due to the coefficients $c_{\mathbf{i}}(\vec{x})$ being explicitly dependent on $\vec{x}$, our proof would not be valid in general. Therefore, in order to drop the assumption of independent transformations, we will characterize the locality of the boxes' operations in the next subsection.

2   Non-independent operations and locality

In the previous section we proved that under the assumptions of (i) local tomography, and (ii) independent operations, a system composed of k systems of orders $(n_1, ..., n_k)$ is a $(\sum_j n_j)$-th order system. Notice that within the proof we did not need to specify anything about the transformations $T_{\vec{x}}$ and we did not impose the notion of locality of the boxes' actions. On the other hand, in quantum (and classical) theory we proved a stronger statement: additivity follows without assumption (ii); however, in that case we explicitly implemented the notion of locality by conditioning the boxes' actions on the path degrees of freedom of the systems. In what follows, we will provide an analogous analysis in the context of GPTs, thereby dropping assumption (ii). We will start by defining the relevant concepts for single systems.

In order to generalize the quantum proof from Section V to GPTs, we need to introduce controlled operations, which means that we will treat the degrees of freedom of the systems as composed of the "path" degree of freedom (which will serve as the control) and the "internal" degrees of freedom (on which the information will be encoded). Let us start by considering a single-system

A. We denote the state space of the control degrees of freedom with $\text{St}(A^{(c)}) \subset A^{(c)}$, and the state space of the internal degrees of freedom with $\text{St}(A^{(int)}) \subset A^{(int)}$, where $A^{(c)}$ and $A^{(int)}$ are real vector spaces. Assuming local tomography, the total state space of system *A* will then be $\text{St}(A) \subset A^{(c)} \otimes A^{(int)}$. Analogously, we can define the spaces of effects as subspaces of the dual vector spaces, i.e. $\text{Eff}(A^{(c)}) \subset A^{(c)*}$, $\text{Eff}(A^{(int)}) \subset A^{(int)*}$, and $\text{Eff}(A) \subset A^{(c)*} \otimes A^{(int)*}$. Now we will characterize the state space of the control degree of freedom.

Assumption 0. Let us assume that there exists an m-outcome measurement for the control degree of freedom corresponding to the following set of effects $\left\{ \vec{f}^{(1)}, ..., \vec{f}^{(m)} \in \text{Eff}(A^{(c)}); \quad \sum_i \vec{f}^{(i)} = \vec{u}^{(c)} \right\}$, where $\vec{u}^{(c)}$ is the unit effect, i.e. the effect that satisfies $\vec{u}^{(c)} \cdot \vec{v} = 1, \forall \vec{v} \in \text{St}(A^{(c)})$. The physical meaning of the latter is that each effect $\vec{f}^{(i)}$ gives the probability of the system being found at box i. We will thus refer to this measurement as the "which-box measurement".

Let us also define subsets of boxes $J = \left\{ j_1, , ..., j_{|J|} \right\} \subseteq I$, where $I \equiv \{1, 2, ..., m\}$, and introduce the coarse-grained effects $\vec{f}^{(J)} \equiv \sum_{i=1}^{|J|} \vec{f}^{(j_i)}$, with $|J|$ being the cardinality of J. Effect $\vec{f}^{(J)}$ thus outputs the probability of the system being found at the subset of boxes $J \subseteq I$. We will say that an arbitrary state $\vec{v} \in \text{St}(A^{(c)})$ is *localized at subset J* if:

- $\vec{f}^{(J)} \cdot \vec{v} = 1$[2];

- $\forall J', \quad J' \subset J \rightarrow \vec{f}^{(J')} \cdot \vec{v} > 0$.

The first point of this definition simply say that "a system being localized at some particular subset" means that "the probability of finding the system outside of this subset is null". The second point is needed in order to select the minimum possible subset which satisfies the first point and thus select a unique subset.

We will now consider a basis $S = \{\vec{s}_1, ..., \vec{s}_{r^{(c)}}\}$ of $A^{(c)}$, with $\vec{s}_i \in \text{St}(A^{(c)})$, and $r^{(c)}$ being the dimension of $A^{(c)}$. If the theory describing our system was classical, then $r^{(c)} = m$, and we could choose a basis consisting of states localized at each box, i.e. $\vec{f}^{(i)} \cdot \vec{s}_j = \delta_{ij}$. On the other hand, if we consider quantum theory, then $r^{(c)} = m^2$, and we can construct a basis $S = \{\vec{s}_{j_1 j_2}, \forall j_1, j_2 = 1, ..., m\}$ such that each of its elements $\vec{s}_{j_1 j_2}$ is localized at subset $\{j_1, j_2\}$. One example of such a basis is the following: states $\vec{s}_{jj}$ correspond to the classical states localized at each box, i.e. $\vec{s}_{jj} = |j\rangle \langle j|$, whereas, for $j_1 < j_2$, states $\vec{s}_{j_1 j_2}$ and $\vec{s}_{j_2 j_1}$ correspond to the superposition states $\vec{s}_{j_1 j_2} = |j_1 j_2\rangle_x \langle j_1 j_2|$ and $\vec{s}_{j_2 j_1} = |j_1 j_2\rangle_y \langle j_1 j_2|$, where $|j_1 j_2\rangle_x \sim (|j_1\rangle + |j_2\rangle)$ and $|j_1 j_2\rangle_y \sim (|j_1\rangle + i |j_2\rangle)$ (see for instance [10]). Here we see that the state-space of quantum theory can be spanned by states, each of which is localized at most at two boxes. On the other hand, all states in classical theory can be spanned by states,

[2]Or equivalently: $\forall J', J' \cap J = \emptyset \rightarrow \vec{f}^{(J')} \cdot \vec{v} = 0$.

each of which is localized at one box. Tentatively, one may guess from the latter that there is a strong relation between the "localizability" of the state-space of a theory with respect to the "which-box measurement" and the order of interference of the theory[3]. In what follows, we will analyze this relation and show that it does indeed hold under certain assumptions. However, we will first formalize the notion of the "localizability" with respect to the "which-box measurement" by introducing the notion of "locality order".

Locality order. Let us focus on a generic state-space of the control degree of freedom $\text{St}(A^{(c)})$ (while still assuming that there exists the which-box measurement, i.e. Assumption 0), and consider an arbitrary basis $S = \{\vec{s}_1, ..., \vec{s}_{r^{(c)}}\}$ of $A^{(c)}$, with $\vec{s}_i \in \text{St}(A^{(c)})$. The elements of S can be localized at various boxes' subsets, and these subsets consist in general of different numbers of boxes (i.e. the subsets have different cardinalities). Let us define $l_S \in \mathbb{N}$ to be the maximum among these numbers, meaning that each element of S is localized within at most l_S different boxes. Formally, for each basis $S = \{\vec{s}_1, ..., \vec{s}_{r^{(c)}}\}$ we define the parameter l_S in the following way:

$$\forall i, \quad \exists (J_i \subseteq I), \quad \text{s.t.} \quad (\vec{s}_i \text{ is localized at } J_i) \,\&\, (|J_i| \leq l_S). \tag{45}$$

Naturally, different basis choices S correspond to different values of l_S. We will define the *locality order L* as the *minimum l_S* among all possible basis, i.e. $L \equiv \min_S l_S$, and denote a basis that achieves this value of l_S with S^* (there can be many different basis that achieve the minimum value; we will simply take S^* to be one of them). Therefore, the parameter l_S is associated to each basis S, whereas the locality order L is associated to the theory, together with the "which-box measurement". Quantum theory satisfies $L = 2$ (with respect to the specific which-box measurement as defined in Assumption 0), since there exists a complete basis whose elements consist of superpositions of at most two boxes (see the example from the previous paragraph), and there is no basis with all of its elements localized at single boxes (classical theory of course satisfies $L = 1$). Note that a hypothetical theory could also satisfy $L = m$, meaning that any basis must contain at least one "completely de-localized" state. Now we are ready to state our main definitions and assumptions, which will formalize the notion of local operations.

Assumption 1. Here we will introduce the notion of locality of the boxes' operations. As before, we

[3]A similar reasoning motivated the construction of the density cube theory which is a third-order-interference theory obtained as a modification of quantum theory, by replacing density matrices (tensors with two indices) with tensors with three indices (the so called density cubes). Translated to our framework, such a state space is spanned by states, each of which is localized at most at three boxes [3].

will label with $S^* = \{\vec{s}_1, ..., \vec{s}_{r^{(c)}}\}$ one of the basis of the control degree of freedom, which satisfy $l_{S^*} = \min_S l_S = L$. Each of the states $\vec{s}_n$ is localized at subset $J_n = \left\{ j_1^{(n)}, ..., j_{|J_n|}^{(n)} \right\}$, with $|J_n| \leq L$. Let us now suppose that a system A is initially prepared in the product state $\vec{s}_i \otimes \vec{g}$, where $\vec{s}_i \in S^*$, and $\vec{g} \in \text{St}(A^{(int)})$. The initial state is thus localized within the subset of boxes $J_i \subseteq I$. We deem it natural to assume that the boxes act in such a way that (i) the transformed state is still localized within subset J_i (i.e. the boxes cannot "delocalize" the system), and (ii) the transformed state cannot depend on inputs pertaining to boxes outside of J_i (i.e. if the probability of finding the system at some box is null, then the system should not be affected by that box). We will formalize point (i) by requiring that the boxes implement transformations $T_{\vec{x}}$ on the input state $(\vec{s}_i \otimes \vec{g})$ in such a way that the output state satisfies

$$(\vec{f}^{(J)} \otimes \vec{e}) \cdot T_{\vec{x}}(\vec{s}_i \otimes \vec{g}) = 0, \tag{46}$$

for all coarse-grained effects $\vec{f}^{(J)}$ with $J \cap J_i = \emptyset$, and arbitrary effects $\vec{e} \in \text{Eff}(A^{(int)})$. Equation (46) thus states that one cannot steer the control degree of freedom of the output state into a state which is localized outside of the initial subset J_i (i.e. this is what we mean that the output state is localized within J_i). The most general form of the action of the transformations $T_{\vec{x}}$ that satisfies the above requirements is the following:

$$T_{\vec{x}}(\vec{s}_i \otimes \vec{g}) = \sum_{\substack{n \\ J_n \subseteq J_i}} \vec{s}_n \otimes T_{\vec{x}^{(J_i)}}^{(i,n)} \vec{g}, \tag{47}$$

where the sum runs over those basis states pertaining to S^* which are localized within the initial subset, i.e. each basis state $\vec{s}_n \in S^*$ is localized within subset J_n, with $J_n \subseteq J_i$. The transformations $T_{\vec{x}^{(J_i)}}^{(i,n)}$ are linear operators on $A^{(int)}$ and depend only on inputs contained within subset J_i, i.e. $\vec{x}^{(J_i)} \equiv \left(x_{j_1^{(i)}}, ..., x_{j_{|J_i|}^{(i)}} \right)$, where we used the notation $J_i = \left\{ j_1^{(i)}, ..., j_{|J_i|}^{(i)} \right\}$. Moreover, all additional coefficients that may appear in front of each term of the sum are for simplicity absorbed in the operators $T_{\vec{x}^{(J_i)}}^{(i,n)}$ (also, note that the vectors $T_{\vec{x}^{(J_i)}}^{(i,n)} \vec{g} \in A^{(int)}$ may generally lie outside of the internal state space $\text{St}(A^{(int)})$).

Relation between the locality order and the order of interference. Now we will show that Assumption 1 and the principle of local tomography imply a strong relation between the order of interference and the locality order. Consider a general instance of our game which consists of Alice preparing her system in an arbitrary state, sending it through the boxes and Bob performing a generic measurement, thereby producing the conditional probabilities $P(b|\vec{x})$. Let us introduce a basis for the internal degrees of freedom $\{\vec{g}_1, ..., \vec{g}_{r^{(int)}}\}$, where $r^{(int)}$ is the dimension of the internal vector space $A^{(int)}$.

According to the principle of local tomography, any initial state $\vec{v} \in \text{St}(A)$ can be written as

$$\vec{v} = \sum_{i,j} c_{ij} \vec{s}_i \otimes \vec{g}_j, \tag{48}$$

where c_{ij} are real coefficients, and $\vec{s}_i \in S^*$. The boxes implement a transformation $T_{\vec{x}}$, and Bob's d-outcome measurement is modelled via a generic set of effects $\left\{ \vec{E}_0, ..., \vec{E}_{d-1} \in \text{Eff}(A), \sum_i \vec{E}_i = \vec{u} \right\}$, where $\vec{u}$ is the unit effect on the total state-space. The conditional probability is thus

$$P(b|\vec{x}) = \sum_{i,j} c_{ij} \vec{E}_b \cdot (T_{\vec{x}} \vec{s}_i \otimes \vec{g}_j) =$$
$$= \sum_{\substack{i,j,n \\ J_n \subseteq J_i}} c_{ij} \vec{E}_b \cdot \left(\vec{s}_n \otimes T_{\vec{x}^{(J_i)}}^{(i,n)} \vec{g}_j \right), \tag{49}$$

where in the second equality we used Assumption 1. Each transformation $T_{\vec{x}^{(J_i)}}^{(i,n)}$ depends only on inputs within subset J_i. Therefore, for a given theory with locality order L, the probabilities $P(b|\vec{x})$ are at most of algebraic order L. Now we can see a direct connection to the order of interference: indeed, as proven in Section III and Appendix B, if a distribution exhibits k-th order interference, then its algebraic order is k. Thus, if a system *can* exhibit k-th order interference, then Assumption 1 implies that the theory satisfies $L \geq k$.

Next, we will introduce a further assumption which will relate the possible transformations implementable by the boxes to the locality order of the theory, and show that together with the previous assumptions, it implies $L = k$.

Assumption 2. Let us suppose that the locality order of our system is L, meaning that all states can be written as linear combinations of states, each of which is localized in subsets of at most L boxes. Moreover, there exist states which cannot be decomposed into linear combinations of states, each of which is localized at *less* than L boxes. We assume the following: there exists at least one state $\vec{v} \in \text{St}(A)$ and one set of transformations $\{T_{\vec{x}}, \forall \vec{x}\}$, such that the transformed state $T_{\vec{x}} \vec{v}$ is of algebraic order L, i.e. it cannot be decomposed into a combination of terms, each of which depends on less than L inputs. Let us heuristically illustrate the meaning and motivation of this assumption with a simple example involving two boxes and a system described by a theory with locality order $L = 2$. Let us take for the moment that our assumption *does not* hold, i.e. that all states and transformations in the theory give rise only to states with algebraic order 1. This means that for any state $\vec{v} \in \text{St}(A)$ and any set of transformations $\{T_{x_1 x_2}, \forall x_1, x_2\}$, the following holds:

$$T_{x_1, x_2} \vec{v} = \sum_i \vec{s}_i \otimes T_{x(i)}^{(v,i)} \vec{g}_i, \tag{50}$$

where $\vec{s}_i \in S^*$, $\vec{g}_i \in \text{St}(A^{(int)})$. The inputs $x(i)$ can be either x_1 or x_2 depending on i, which makes each of the

operators $T_{x(i)}^{(v,i)}$ depend only on one input. On the other hand, if we considered a system of locality order $L = 1$ (e.g. classical theory), then a general output state $T_{x_1,x_2}\vec{v}$ within that theory would have a similar form to the one in Equation (50): it would still be a linear combination of terms, each of which depends only on one input. Thus, if our assumption did not hold, a system of locality order 2 would encode the inputs in the qualitatively same way as a system of order 1. In order to introduce a difference between systems of different orders, we will assume that a system of order L can produce a state with algebraic order L. This assumption is rather technical and admittedly requires further physical justification; however, in this work we will take it at face value and show that, together with the other assumptions, it implies additivity of interference under composition. Moreover, note that the assumption is satisfied by quantum theory: namely, if the system is prepared in a superposition of two boxes, there exist local transformations that encode the two inputs on the state (e.g. local phase gates) in such a way that the transformed state cannot be decomposed into a linear combination of terms, each of which depends only on one input. On the other hand, superpositions of more than two boxes still give rise only to distributions of algebraic order 2, which is related to the fact that the quantum density matrix is a tensor with two indices (which essentially sets the locality order to $L = 2$).

$L = k$. Let us now briefly show how the conjunction of our assumptions implies $L = k$. Namely, Assumption 2 directly implies that if the system *cannot* exhibit more than k-th order interference, then $L \leq k$, since for $L > k$ there would exist states and transformations that produce probabilities of algebraic order larger than k which would exhibit more than k-th order interference. On the other hand, as we have seen before, Assumption 1 and the principle of local tomography imply $L \geq k$. Hence, the conjunction of the latter statements implies that the locality order of a k-th order system is $L = k$.

Assumption 3. Up until now, the discussion has been focused only on single systems; now we will specify how the boxes act on composite systems. In sub-section 1 we had assumed that the boxes act independently on the subsystems; here we want to drop this assumption. Let us consider a system composed of two subsystems. We will label the basis that achieve the minimal locality order for the first and second subsystem respectively with S_1^* and S_2^*. Suppose that the system is prepared in the following product state $\vec{s}_{i_1} \otimes \vec{s}_{i_2} \otimes \vec{g}_1 \otimes \vec{g}_2 \in A_1^{(c)} \otimes A_2^{(c)} \otimes A_1^{(int)} \otimes A_2^{(int)}$, where $\vec{g}_1$ and $\vec{g}_2$ are arbitrary internal states of the two subsystems, and $\vec{s}_{i_1} \in S_1^*$, $\vec{s}_{i_2} \in S_2^*$, are localized respectively at subsets $J_{i_1} = \left\{ j_1^{(i_1)}, ..., j_{|J_{i_1}|}^{(i_1)} \right\} \subseteq I$ and $J_{i_2} = \left\{ j_1^{(i_2)}, ..., j_{|J_{i_2}|}^{(i_2)} \right\} \subseteq I$. Since the two subsystems are localized respectively at J_{i_1} and J_{i_2}, we will say that the composite system is localized at $J_{i_1} \cup J_{i_2}$. Now we

will introduce the action of the boxes on the composite system as a direct generalization of Assumption 1. Namely, we will demand that (i) the composite control degree of freedom of the transformed state can be steered only into states localized within $J_{i_1} \cup J_{i_2}$, and (ii) the transformed state depends only on inputs contained within subset $J_{i_1} \cup J_{i_2}$. Analogously to Assumption 1., the first part states that the boxes cannot further "delocalize" the system, while the second part states that if the probability of finding either of the subsystems at some box is 0, then the composite system should not be affected by the action of that box. Thus we see that this assumption states that if two subsystems of a composite system are localized at subsets J_{i_1} and J_{i_2}, then the transformations acts on the composite system, as if the latter was a single system localized at subset $J_{i_1} \cup J_{i_2}$. The most general form of the transformations $T_{\vec{x}}$ that satisfies the above requirements is

$$
\begin{aligned}
&T_{\vec{x}} \left(\vec{s}_{i_1} \otimes \vec{s}_{i_2} \otimes \vec{g}_1 \otimes \vec{g}_2 \right) \\
&= \sum_{\substack{n_1 n_2 \\ (J_{n_1} \cup J_{n_2}) \\ \subseteq (J_{i_1} \cup J_{i_2})}} \vec{s}_{n_1} \otimes \vec{s}_{n_2} \otimes T_{\vec{x}^{(J_{i_1}, J_{i_2})}}^{(\mathbf{in})} (\vec{g}_1 \otimes \vec{g}_2), \quad (51)
\end{aligned}
$$

where we introduced the input strings

$$
\vec{x}^{(J_{i_1}, J_{i_2})} \equiv \left(x_{j_1^{(i_1)}}, ..., x_{j_{|J_{i_1}|}^{(i_1)}}, x_{j_1^{(i_2)}}, ..., x_{j_{|J_{i_2}|}^{(i_2)}} \right), \quad (52)
$$

and the bold index $\mathbf{n}$ stands for $n_1 n_2$. Each of the transformations $T_{\vec{x}^{(J_{i_1}, J_{i_2})}}^{(\mathbf{in})}$ is a linear operator on $A_1^{(int)} \otimes A_2^{(int)}$ that depends on inputs $\vec{x}^{(J_{i_1}, J_{i_2})}$. The sum in Equation (51) runs over states which are localized within the union of the initial subsets, i.e. states $\vec{s}_{n_1}$ and $\vec{s}_{n_2}$ are localized respectively at subsets J_{n_1} and J_{n_2}, where the latter satisfy $(J_{n_1} \cup J_{n_2}) \subseteq (J_{i_1} \cup J_{i_2})$. Notice that transformations that entangle the subsystems are not excluded and are to be expected if the initial states of the control systems of the two subsystems are localized within overlapping subsets of the boxes, i.e. if $J_{i_1} \cap J_{i_2} \neq \emptyset$.

Additivity of interference. Now we will show that the conjunction of Assumption 3 with the previous assumptions implies that interference is additive under composition of systems. Let us consider a system composed of two k-th order subsystems. As we have shown before, the principle of local tomography and Assumptions 1&2 imply that the locality orders of both subsystems are $L = k$. Consequently, this enables us to choose the following basis for the composite control state space: $\left\{ \vec{s}_{i_1} \otimes \vec{s}_{i_2}, \forall i_1, i_2 = 1, ..., r^{(c)} \right\}$, where the single-system states $\vec{s}_{i_1}$ and $\vec{s}_{i_2}$ are localized respectively at subsets $J_{i_1} = \left\{ j_1^{(i_1)}, ..., j_{|J_{i_1}|}^{(i_1)} \right\}$ and $J_{i_2} = \left\{ j_1^{(i_2)}, ..., j_{|J_{i_2}|}^{(i_2)} \right\}$, with $|J_{i_1}| \leq k$ and $|J_{i_2}| \leq k$. From here it is simple to show that two k-th order systems cannot exhibit more than $2k$-th order interference. Namely, due to the assumption of local tomography, the most general initial state prepared by Alice is

$$
\vec{v} = \sum_{\mathbf{ik}} c_{\mathbf{ik}} \vec{s}_{i_1} \otimes \vec{s}_{i_2} \otimes \vec{g}_{k_1} \otimes \vec{g}_{k_2}, \quad (53)
$$

where $\{\vec{g}_{k_1} \otimes \vec{g}_{k_2}, \forall k_1, k_2\}$ forms a basis for the internal state space of the composite system. The bold indices $\mathbf{i}, \mathbf{k}$ stand for i_1, i_2, k_1, k_2, and $c_{\mathbf{ik}}$ are real coefficients. The composite system is then sent through the boxes, which implement the input-dependent transformation $T_{\vec{x}}$. Finally, Bob performs a measurement $\left\{\vec{E}_0, ..., \vec{E}_{d-1} \in \mathrm{Eff}(A), \sum_i \vec{E}_i = \vec{u}\right\}$, where $\vec{u}$ is the unit effect on the total state-space. The conditional probabilities thus follow:

$$
\begin{aligned}
P(b|\vec{x}) &= \sum_{\mathbf{ik}} c_{\mathbf{ik}} \vec{E}_b \cdot [T_{\vec{x}} (\vec{s}_{i_1} \otimes \vec{s}_{i_2} \otimes \vec{g}_{k_1} \otimes \vec{g}_{k_2})] \\
&= \sum_{\mathbf{ikn}} c_{\mathbf{ik}} \vec{E}_b \cdot \left[\vec{s}_{n_1} \otimes \vec{s}_{n_2} \otimes T^{(\mathbf{in})}_{\vec{x}^{(J_{i_1} J_{i_2})}} (\vec{g}_{k_1} \otimes \vec{g}_{k_2}) \right].
\end{aligned}
\tag{54}
$$

Since $\vec{x}^{(J_{i_1} J_{i_2})}$, as defined in Equation (52), is a string of at most $2k$ different inputs, the algebraic order of the conditional probabilities arising from such processes is bounded by $2k$, which implies that two k-th order systems produce at most $(2k)$-th order interference. The procedure can be trivially generalized to an arbitrary number of systems, by composing them two-by-two, thus arriving to the conclusion that the upper bound on interference is additive under the composition of systems. Together with the result in Appendix E which proves the additivity of the lower bound of interference under composition, this completes the proof that a system composed of subsystems of orders $\{n_1, ..., n_p\}$ is a $(\sum_{i=1}^{p} n_i)$-th order system. Note that this result is valid for the inputs x_i being elements of a set of cardinality d, with d being an arbitrary prime number.

We will now briefly summarize and emphasize the main assumptions and results of this section. In Subsection 1 we had proved the additivity of interference under the assumptions of local tomography and independent operations, without referring at all to the locality of the boxes' operations. In Subsection 2, in order to drop the second assumption, we implemented the notion of locality via controlled operations, as a natural generalization of the quantum and classical cases. Besides the principle of local tomography and the assumption of the existence of the which-box measurement (Assumption 0), we introduced three further assumptions on single-(Assumptions 1 & 2) and composite-system transformations (Assumption 3). Finally, notice that in this section and overall in the whole manuscript, we treated locality via controlled operations; one may however adopt a different view and from the very start associate to each space(time) region a system, the transformations on which would automatically be regarded as local (in quantum theory, such a formulation would correspond to second-quantization and to local quantum physics [29], which makes it more compatible with relativity).

VII   Conclusion and Outlook

In this work we introduced a class of information-theoretic games, which generalize standard multi-slit interference experiments. The order of interference of a theory is then seen as the (im)possibility of accomplishing certain information-processing tasks using finite resources. These games essentially characterize how much information can be decoded from a physical system given that the information was encoded in a global property (parity or modulo of the inputs) of local pieces of information (local inputs). We showed that within quantum theory, the order of interference of k systems is $2k$; it would be interesting to inspect potential connections of this result to superdense coding, which states that k qubits can be used to send at most $2k$ bits. Moreover, the game formulation can provide a (semi)-device-independent witness of the particle number. Finally, even though our work is mainly focused on conceptual issues, it might inspire novel experimental tests of higher-order interference (albeit now redefined in terms of our game); for instance, one could replace slits with boxes implementing generic unitary operations and observe how the order of interference scales with the number of systems. This could then complement the vast experimental search for higher-order interference [30–33]. Another interesting direction is to take the boxes' inputs to be concrete spatio-temporal variables, as in Ref. [34], and see how this constrains the order of interference achievable with a fixed amount of resources (e.g. one may perhaps find connections between the order of interference and the dimensionality of space).

So far there have been several attempts at explaining the order of interference of quantum theory; however, we deem that a physically intuitive explanation has not yet been obtained. The reason for this might be that the sole question is leading us in the wrong direction. Instead of asking ourselves why does quantum theory behave in a particular way in multi-slit experiments (or GPT generalizations thereof), we could ask why does quantum theory restrict the amount of globally-encoded information (parity/modulo of locally encoded bits/dits) that can be retrieved from a system. Moreover, what is the relation between the (im)possibility of such a decoding and the number of systems used for encoding the information? Is there any physical argument for why interference should be additive under composition (e.g. violation of the no-signalling principle)? How does one even define the notion of *number of systems* in a device-independent scenario (or ought one to quantify the resources in an alternative manner)? We plan to investigate these and further questions in future works.

Acknowledgements

Bibliography

[1] Feynman, R.P., Leighton, R.B. and Sands, M., 2005. The Feynman Lectures on Physics: Definitive Edition. Pearson Addison-Wesley. ISBN 0-8053-9046-4

[2] Sorkin, R.D., 1994. Quantum mechanics as quantum measure theory. Modern Physics Letters A, 9(33), pp.3119-3127. DOI: https://doi.org/10.1142/S021773239400294X

[3] Dakic, B., Paterek, T. and Brukner, C., 2014. Density cubes and higher-order interference theories. New Journal of Physics, 16(2), p.023028. DOI: http://dx.doi.org/10.1088/1367-2630/16/2/023028

[4] Barnum, H., Mueller, M.P. and Ududec, C., 2014. Higher-order interference and single-system postulates characterizing quantum theory. New Journal of Physics, 16(12), p.123029. DOI: http://dx.doi.org/10.1088/1367-2630/16/12/123029

[5] Lee, C.M. and Selby, J.H., 2017. Higher-order interference in extensions of quantum theory. Foundations of Physics, 47(1), pp.89-112. DOI: https://doi.org/10.1007/s10701-016-0045-4

[6] Barnum, H., Lee, C.M., Scandolo, C.M. and Selby, J.H., 2017. Ruling out higher-order interference from purity principles. Entropy, 19(6), p.253. DOI: https://doi.org/10.3390/e19060253

[7] Ududec, C., Barnum, H. and Emerson, J., 2011. Three slit experiments and the structure of quantum theory. Foundations of Physics, 41(3), pp.396-405. DOI: https://doi.org/10.1007/s10701-010-9429-z

[8] Niestegge, G., 2013. Three-slit experiments and quantum nonlocality. Foundations of Physics, 43(6), pp.805-812. DOI: https://doi.org/10.1007/s10701-013-9719-3

[9] Lee, C.M. and Selby, J.H., 2016. Generalised phase kick-back: the structure of computational algorithms from physical principles. New Journal of Physics, 18(3), p.033023. DOI: http://dx.doi.org/10.1088/1367-2630/18/3/033023

[10] Hardy, L., 2001. Quantum theory from five reasonable axioms. arXiv preprint quant-ph/0101012.

[11] Dakic, B. and Brukner, C. Quantum theory and beyond: is entanglement special? Deep Beauty: Understanding the Quantum World Through Mathematical Innovation (ed. Halvorson, H.) (Cambridge Univ. Press, 2011). DOI: https://doi.org/10.1017/CBO9780511976971.011

[12] Masanes, L. and Mueller, M.P., 2011. A derivation of quantum theory from physical requirements. New Journal of Physics, 13(6), p.063001. DOI: http://dx.doi.org/10.1088/1367-2630/13/6/063001

[13] Chiribella, G., D Ariano, G.M. and Perinotti, P., 2011. Informational derivation of quantum theory. Physical Review A, 84(1), p.012311. DOI: http://dx.doi.org/10.1103/PhysRevA.84.012311

[14] Del Santo, F. and Dakić, B., 2018. Two-way communication with a single quantum particle. Physical review letters, 120(6), p.060503. DOI: https://doi.org/10.1103/PhysRevLett.120.060503

[15] Massa, F., Moqanaki, A., Baumeler, Ä., Del Santo, F., Kettlewell, J.A., Dakić, B. and Walther, P., 2019. Experimental two-way communication with one photon. Advanced Quantum Technologies, 2(11), p.1900050. DOI: https://doi.org/10.1002/qute.201900050

[16] Hsu, L.Y., Lai, C.Y., Chang, Y.C., Wu, C.M. and Lee, R.K., 2020. Infinite information can be carried using a single quantum particle. arXiv preprint arXiv:2002.10374.

[17] Horvat, S. and Dakić, B., 2019. Quantum enhancement to information speed. arXiv preprint arXiv:1911.11803.

[18] Oreshkov, O., Costa, F. and Brukner, C., 2012. Quantum correlations with no causal order. Nature communications, 3, p.1092. DOI: https://doi.org/10.1038/ncomms2076

[19] Feix, A., Araujo, M. and Brukner, C., 2016. Causally nonseparable processes admitting a causal model. New Journal of Physics, 18(8), p.083040. DOI: http://dx.doi.org/10.1088/1367-2630/18/8/083040

[20] Guerin, P.A., Feix, A., Araujo, M. and Brukner, C., 2016. Exponential communication complexity advantage from quantum superposition of the direction of communication. Physical review letters, 117(10), p.100502. DOI: https://doi.org/10.1103/PhysRevLett.117.100502

[21] Chiribella, G., D Ariano, G.M. and Perinotti, P., 2008. Quantum circuit architecture. Physical review letters, 101(6), p.060401. DOI: https://doi.org/10.1103/PhysRevLett.101.060401

[22] Chiribella, G., D Ariano, G.M., Perinotti, P. and Valiron, B., 2013. Quantum computations without definite causal structure. Physical Review A, 88(2), p.022318. DOI: https://doi.org/10.1103/PhysRevA.88.022318

[23] Allen, J.M.A., Barrett, J., Horsman, D.C., Lee, C.M. and Spekkens, R.W., 2017. Quantum common causes and quantum causal models. Physical Review X, 7(3), p.031021. DOI: https://doi.org/10.1103/PhysRevX.7.031021

[24] Brunner, N., Cavalcanti, D., Pironio, S., Scarani, V. and Wehner, S., 2014. Bell nonlocality. Reviews of Modern Physics, 86(2), p.419. DOI: https://doi.org/10.1103/RevModPhys.86.419

[25] Hausladen, P. and Wootters, W.K., 1994. A 'pretty good' measurement for distinguishing quantum states. Journal of Modern Optics, 41(12), pp.2385-2390. DOI: https://doi.org/10.1080/09500349414552221

[26] Cong, W., Cai, Y., Bancal, J.D. and Scarani, V., 2017. Witnessing irreducible dimension. Physical review letters, 119(8), p.080401. DOI: https://doi.org/10.1103/PhysRevLett.119.080401

[27] Aguilar, E.A., Farkas, M., Martinez, D., Alvarado, M., Carine, J., Xavier, G.B., Barra, J.F., Canas, G., Pawlowski, M. and Lima, G., 2018. Certifying an irreducible 1024-dimensional photonic state using refined dimension witnesses. Physical Review Letters, 120(23), p.230503. DOI: https://doi.org/10.1103/PhysRevLett.120.230503

[28] Barrett, J., 2007. Information processing in generalized probabilistic theories. Physical Review A, 75(3), p.032304. DOI: https://doi.org/10.1103/PhysRevA.75.032304

[29] Haag, R., 2012. Local quantum physics: Fields, particles, algebras. Springer Science & Business Media. DOI: https://doi.org/10.1007/978-3-642-61458-3

[30] Sinha, U., Couteau, C., Jennewein, T., Laflamme, R. and Weihs, G., 2010. Ruling out multi-order interference in quantum mechanics. Science, 329(5990), pp.418-421. DOI: https://doi.org/10.1126/science.1190545

[31] Park, D.K., Moussa, O. and Laflamme, R., 2012. Three path interference using nuclear magnetic resonance: a test of the consistency of Born's rule. New Journal of Physics, 14(11), p.113025. DOI: http://dx.doi.org/10.1088/1367-2630/14/11/113025

[32] Kauten, T., Keil, R., Kaufmann, T., Pressl, B., Brukner, C. and Weihs, G., 2017. Obtaining tight bounds on higher-order interferences with a 5-path interferometer. New Journal of Physics, 19(3), p.033017. DOI: https://doi.org/10.1088/1367-2630/aa5d98

[33] Lee, K.S., Zhuo, Z., Couteau, C., Wilkowski, D. and Paterek, T., 2020. Atomic test of higher-order interference. Physical Review A, 101(5), p.052111. DOI: https://doi.org/10.1103/PhysRevA.101.052111

[34] Garner, A.J., Krumm, M. and Mueller, M.P., 2020. Semi-device-independent information processing with spatiotemporal degrees of freedom. Physical Review Research, 2(1), p.013112. DOI: https://doi.org/10.1103/PhysRevResearch.2.013112

[35] Lomonosov, V. and Rosenthal, P., 2004. The simplest proof of Burnside's theorem on matrix algebras. Linear algebra and its applications, 383, pp.45-47. DOI: https://doi.org/10.1016/j.laa.2003.08.012

A Equivalence between the game formulation and its dual

In this section we will show that the formulation of higher order interference theories in terms of winning probabilities of the "modulo" games

$$\tilde{I}_{m,\vec{\nu},f}^{(d)}(k) = \frac{1}{d^m} \sum_{x_1,...,x_m=0}^{d-1} P\left(b = f(s_{\vec{\nu},\vec{x}}^{(d)})|x_1...x_m\right) - \frac{1}{d} = 0, \quad \forall \vec{\nu}, \forall f \tag{55}$$

is equivalent to its dual formulation in terms of the dual interference terms

$$\tilde{J}_{m,\vec{\nu},b}^{(d)} = \frac{1}{d^m} \sum_{\vec{x}} (\omega_d)^{\sum_a \nu_a x_a} P(b|\vec{x}) = 0, \quad \forall \vec{\nu}, \quad \forall b \in \{0,...,d-1\}, \tag{56}$$

where $\omega_d = e^{i2\pi/d}$ is the d-th root of unity.
For convenience, let us first introduce the dual terms with an added index $\alpha \in \{1,...,d-1\}$:

$$\tilde{J}_{m,\vec{\nu},b,\alpha}^{(d)} \equiv \frac{1}{d^m} \sum_{\vec{x}} (\omega_d)^{\alpha \sum_a \nu_a x_a} P(b|\vec{x}). \tag{57}$$

We will show that for any dit-string $\vec{\nu}$, the following equivalence holds:

$$\left\{ \tilde{I}_{m,\vec{\nu},f}^{(d)} = 0, \quad \forall f \right\} \longleftrightarrow \left\{ \tilde{J}_{m,\vec{\nu},b,\alpha}^{(d)} = 0, \forall b \in \{0,...,d-1\}, \forall \alpha \in \{1,...,d-1\} \right\} \tag{58}$$

Since the order of interference is defined as the vanishing of the interference terms (55) for all $\vec{\nu}$, (58) would imply that the dual formulation of interference (56) is equivalent to the game formulation (55) (this is so because if the RHS of equivalence (58) holds *for all* $\vec{\nu}$, then the dual conditions (56) also hold, since the index α becomes redundant). In what follows, we will prove that (58) does hold indeed.
In order to simplify the notation, let us introduce the $d \times d$ matrix P, whose elements are defined as[4]

$$P_{bs} = \frac{1}{d^{m-1}} \sum_{\substack{\vec{x} \\ s_{\vec{\nu},\vec{x}}^{(d)}=s}} P(b|\vec{x}). \tag{59}$$

---

[4]Notice that the matrix P depends on the dit-string $\vec{\nu}$, but we do not write it explicitly in the notation in order to simplify the expressions.

The normalization of probabilities implies that P is a stochastic matrix, i.e.

$$\sum_b P_{bs} = 1, \quad \forall s. \tag{60}$$

The game formulation equations on the left side of the equivalence in (58) can be written as

$$\frac{1}{d^m} \sum_{x_1,\dots,x_m=0}^{d-1} P\left(b = f(s_{\vec{\nu},\vec{x}}^{(d)})|x_1\dots x_m\right) = \frac{1}{d}, \quad \forall f \quad \to \quad \sum_s P_{f(s)s} = 1, \quad \forall f, \tag{61}$$

which we rewrite succinctly as

$$\mathrm{Tr}[\Pi_f P] = 1, \quad \forall f, \tag{62}$$

where $\Pi_f \equiv \sum_s |f(s)\rangle \langle s|$ ranges over all d-dimensional permutation matrices (the most general finite dimensional reversible transformations are indeed permutations).

On the other hand, the dual conditions on the right side of equivalence (58) assume the following form

$$(PF)_{lk} = \delta_{k,0} \frac{1}{\sqrt{d}} \sum_j P_{lj}, \tag{63}$$

where F is the d-dimensional Fourier matrix with elements

$$(F)_{lk} = \frac{1}{\sqrt{d}} (\omega_d)^{lk}. \tag{64}$$

We will first show that the dual conditions (63) imply the information-theoretic conditions (62). Assuming that the the dual conditions (63) hold, we evaluate the trace

$$\mathrm{Tr}[\Pi_f P] = \mathrm{Tr}[F^\dagger \Pi_f PF] = \sum_{kl} (F^\dagger \Pi_f)_{kl} (PF)_{lk} = \sum_l (F^\dagger \Pi_f)_{0l} (PF)_{l0} = 1, \forall f. \tag{65}$$

The first step follows from the unitarity of F and the ciclicity of the trace, the third step is due to the dual conditions (63), and the last step follows from $(F^\dagger \Pi_f)_{0l} = 1/\sqrt{d}$. Therefore, conditions (63) imply (62).

In order to prove the converse statement, let us introduce the following new basis $\beta = \{|E\rangle, |e_1\rangle, \dots, |e_{d-1}\rangle\}$, where $|E\rangle$ has the following form

$$|E\rangle = 1/\sqrt{d}(1, 1, \dots, 1)^T, \tag{66}$$

and the remaining vectors span the orthogonal subspace (they can for instance be chosen as the rows/columns of the Fourier matrix). The normalization conditions (60) can then be written as

$$\langle E| P = \langle E|, \tag{67}$$

which implies that the matrix P in the new basis has the following form

$$P = |E\rangle \langle E| + \sum_j q_j |e_j\rangle \langle E| + \bar{E}, \tag{68}$$

where q_j are arbitrary complex numbers and $\bar{E}$ is a matrix that has support only on the subspace orthogonal to $|E\rangle$.

In the new basis β, all permutation matrices have the following form

$$\Pi_f = |E\rangle \langle E| + \Delta_f, \tag{69}$$

where Δ_f is again a matrix with no support on $|E\rangle$. The latter follows from the fact that the representation of the permutation group that we are using is reducible to the direct sum of a one-dimensional representation (spanned by the vector $|E\rangle$ which is invariant under all permutations) and a $(d-1)$-dimensional irreducible representation (here given by Δ_f). Assuming that the game-conditions (62) hold, we obtain the following

$$1 = \mathrm{Tr}[\Pi_f P] = 1 + \mathrm{Tr}[\Delta_f \bar{E}] \to \mathrm{Tr}[\Delta_f \bar{E}] = 0, \quad \forall f. \tag{70}$$

Since the matrices Δ_f provide an irreducible representation of the permutation group, the Burnside Theorem [35] implies that they span the set of all $(d-1)$-dimensional matrices; therefore, Equation (70) implies that $\bar{E}$ is necessarily zero. The matrix P is thus

$$P = |E\rangle \langle E| + \sum_j q_j |e_j\rangle \langle E| = |q\rangle \langle E|, \tag{71}$$

where we introduced the vector $|q\rangle \equiv |E\rangle + \sum_j q_j |e_j\rangle$. Therefore, the sought matrix PF is

$$(PF)_{kl} = \langle l|q\rangle \langle E| F |k\rangle = \delta_{k,0} \frac{1}{\sqrt{d}} \sum_j P_{lj}, \tag{72}$$

where we used $\sum_{k=0}^{d-1}(\omega_d)^k = 0$ and $\{|k\rangle; k = 0, ..., d-1\}$ are the original basis vectors. Hence, equivalence (58) holds, and thus the game formulation (55) is equivalent to the dual formulation (56). Besides being an interesting mathematical result, the exact equivalence between these two formulations will be useful for showing that any distribution that exhibits at most m-th order interference can be written as a linear combination of functions with at most m inputs. This will consequently be the necessary ingredient for proving that Local Tomography implies the additivity of the order of interference under the composition of systems.

B   Algebraic order for arbitrary d

In this Appendix we are going to extend the mathematical property of Section III to arbitrary d.
Let us consider the modulo games involving m boxes and assume that Alice uses a system of order $n < m$. In Appendix A we showed that this is equivalent to the following constraints

$$\tilde{J}_{m,\vec{\nu},b}^{(d)} = \frac{1}{d^m} \sum_{\vec{x}} (\omega_d)^{\sum_j \nu_j x_j} P(b|\vec{x}) = 0,$$
$$\forall b \in \{0, 1, ..., d-1\}, \nu_j \in \{0, 1, ..., d-1\}, \quad \text{s.t.} \quad \sum_j \delta_{\nu_j, 0} < (m-n). \tag{73}$$

For compactness, we allow the dit string $\vec{\nu}$ to range over all dit-strings with less than $(m-n)$ null components. This notation specifies which interference terms the equation refers to: if $\nu_j \neq 0$ for $j = i_1, ..., i_l$, the equation states that the order of interference involving boxes $i_1, ..., i_l$ is equal to zero (l can be any integer between $n+1$ and m).
Let us regard $P(b|\vec{x})$ as one component of a vector $\vec{P}_b$ in a d^m-dimensional vector space formed by the tensor product of m d-dimensional spaces, i.e. $P(b|\vec{x}) = \bigotimes_{i=1}^m \vec{e}_{x_i} \cdot \vec{P}_b$, where $\{\vec{e}_{x_j}, x_j = 0, ..., d-1\}$ span the j-th d-dimensional space. Equations (73) then imply:

$$\tilde{J}_{m,\vec{\nu},b}^{(d)} = \frac{1}{d^m} \sum_{\vec{x}} (\omega_d)^{\sum_j \nu_j x_j} \bigotimes_{i=1}^m \vec{e}_{x_i} \cdot \vec{P}_b = \frac{1}{d^{m/2}} \bigotimes_{i=1}^m \vec{f}_{\nu_i} \cdot \vec{P}_b = \frac{1}{d^{m/2}} \lambda_{\vec{\nu},b} = 0, \tag{74}$$

where the rotated vectors (rows/columns of the d-dimensional Fourier matrix) are defined as

$$\vec{f}_{\nu_j} \equiv \frac{1}{\sqrt{d}} \sum_{x_j=0}^{d-1} (\omega_d)^{\nu_j x_j} \vec{e}_{x_j}, \tag{75}$$

and the coefficients $\lambda_{\vec{\nu},b}$ are the components of the vector $\vec{P}_b$ in the corresponding basis. Equation (74) states that the components $\lambda_{\vec{\nu},b}$ vanish for all $\vec{\nu}$ with more than n non-zero components. The probabilities $P(b|\vec{x})$ can thus be expressed as

$$P(b|\vec{x}) = \left[\bigotimes_{i=1}^m \vec{e}_{x_i} \right] \cdot \left[\sum_{\vec{\nu}} \lambda_{\vec{\nu},b} \bigotimes_{j=1}^m \vec{f}_{\nu_j} \right] = \frac{1}{d^{m/2}} \sum_{\substack{\nu_1, ..., \nu_m \\ \sum_j \delta_{\nu_j, 0} \geq (m-n)}} (\omega_d)^{\sum_j \nu_j x_j} \lambda_{\vec{\nu},b}, \tag{76}$$

where we used $\vec{e}_{x_i} \cdot \vec{f}_{\nu_i} = \frac{1}{\sqrt{d}} (\omega_d)^{\nu_i x_i}$. Since the sum runs only over strings ν with at most n non-zero components, $P(b|\vec{x})$ can be written as a function of at most n inputs, for every b.
Therefore we have proved that the n-th order interference dual conditions (73), or equivalently, the information-theoretic formulation (5), provide a set of necessary and sufficient conditions for a distribution to be decomposable into a linear combination of at most n inputs. Hence, the algebraic order of the distribution $P(b|\vec{x})$ is at most n.

C   Proof of Equation (34)

In this Appendix we will show that equations

$$\sum_{\vec{x}} (\omega_d)^{\alpha \sum_k \nu_k x_k} \rho_{\vec{x}} = 0, \quad \forall \alpha \in \{1, ..., d-1\} \tag{77}$$

imply the following relations

$$\rho^{(S)} = \rho^{(S')}, \quad \forall S, S' = 0, ..., d-1, \tag{78}$$

where $\rho^{(S)}$ is defined as

$$\rho^{(S)} \equiv \frac{1}{d^{m-1}} \sum_{\substack{x_1, ..., x_m \\ (\sum_k \nu_k x_k) \bmod d = S}} \rho_{\vec{x}}. \tag{79}$$

For a start, Equations (77) can be written as

$$\sum_{S=0}^{d-1} (\omega_d)^{\alpha S} \rho^{(S)} = 0, \quad \forall \alpha \in \{1, ..., d-1\}. \tag{80}$$

Next, if we introduce the vector $\vec{\rho} = (\rho^{(0)}, \rho^{(1)}, ..., \rho^{(d-1)})^T$, the latter equations can be cast in the following form

$$F\vec{\rho} = \sqrt{d}\bar{\rho}(1, 0, ..., 0)^T, \tag{81}$$

where $F_{kl} = 1/\sqrt{d}(\omega_d)^{kl}$ is the d-dimensional Fourier matrix and $\bar{\rho} = 1/d \sum_S \rho^{(S)}$. Taking the inverse of Equation 81 (i.e. applying $F^\dagger$ from the left) we obtain

$$\vec{\rho} = \bar{\rho}(1, 1, ..., 1)^T, \tag{82}$$

which is equivalent to

$$\rho^{(S)} = \rho^{(S')} = \bar{\rho}, \quad \forall S, S' = 0, ..., d-1. \tag{83}$$

D Details of the proof of additivity in quantum theory

Here we will write explicitly the expressions for the transformations invoked in the proof of the additivity of interference under composition in quantum theory (Section 2.2). We will denote the "control" degree of freedom of j-th system with $H^{(c_j)}$ (the control is simply the path degree of freedom), and its "internal" degree of freedom with $H^{(int_j)}$. The total Hilbert space H of the k systems is a tensor product of the single-system Hilbert spaces, i.e.

$$H = H^{(c)} \otimes H^{(int)}, \tag{84}$$

where $H^{(c)} \equiv \bigotimes_{j=1}^k H^{(c_j)}$ is the composite of the control/path degrees of freedom, and $H^{(int)} \equiv \bigotimes_{j=1}^k H^{(int_j)}$ is the composite of the internal degrees of freedom.

Analogously to the single-system case, we introduce the following set of Kraus operators that act on the composite internal Hilbert space $H^{(int)}$:

$$\left\{ B_{\mathbf{i}}^{(s)}(x_{i_1}, ..., x_{i_k}), \quad \sum_s B_{\mathbf{i}}^{(s)\dagger}(x_{i_1}, ..., x_{i_k}) B_{\mathbf{i}}^{(s)}(x_{i_1}, ..., x_{i_k}) \leq \mathbb{1} \right\}, \tag{85}$$

where $\mathbf{i}$ stands for $i_1, ..., i_k$, and $\mathbb{1}$ is the identity operator on $H^{(int)}$. We do not impose any further restrictions on the latter operators, thereby including also operations that entangle the internal degrees of freedom of the subsystems. The total transformation implemented by the boxes for a given set of inputs $\vec{x}$ will be represented with the following Kraus operators that act on the total Hilbert space H:

$$\left\{ M_{\vec{x}}^{(s)} = \sum_{\mathbf{i}} \bigotimes_{p=1}^k |n_{i_p}^{(p)}\rangle \langle n_{i_p}^{(p)}| \otimes B_{\mathbf{i}}^{(s)}(x_{i_1}, ..., x_{i_k}) \right\}, \tag{86}$$

where $\left\{ |n_{i_p}^{(p)}\rangle, \forall i_p = 1, ..., m \right\}$ forms a basis of the control-space of p-th system $H^{(c_p)}$.

Let us now suppose that the total system is prepared in the following state

$$\rho_0 = \sum_{\mathbf{i}, \mathbf{j}, \mathbf{r}, \mathbf{l}} c_{\mathbf{i}\mathbf{j}\mathbf{r}\mathbf{l}} \bigotimes_{p=1}^k \left[|n_{i_p}^{(p)}\rangle \langle n_{j_p}^{(p)}| \otimes |\phi_{r_p}^{(p)}\rangle \langle \phi_{l_p}^{(p)}| \right], \tag{87}$$

where $\mathbf{i}$ is short for $i_1, ..., i_k$ (and analogously for the other indices), and vectors $\left\{ |\phi_{r_p}^{(p)}\rangle, \forall r_p \right\}$ span the internal space of p-th system $H^{(int_p)}$.

The boxes then act on the system as follows:

$$\rho_{\vec{x}} = \sum_s M_{\vec{x}}^{(s)} \rho_0 M_{\vec{x}}^{(s)\dagger} = \sum_{\mathbf{i}, \mathbf{j}} \bigotimes_{p=1}^k \left[|n_{i_p}^{(p)}\rangle \langle n_{j_p}^{(p)}| \right] \otimes C_{\mathbf{i}\mathbf{j}}(x_{i_1}, x_{j_1}, ..., x_{i_k}, x_{j_k}), \tag{88}$$

Accepted in 〈 Quantum 2021-02-15, click title to verify. Published under CC-BY 4.0.

18

where

$$C_{\mathbf{ij}}(x_{i_1}, x_{j_1}, ..., x_{i_k}, x_{j_k}) = \sum_{s,\mathbf{r},\mathbf{l}} c_{\mathbf{ijrl}} B_{\mathbf{i}}^{(s)}(x_{i_1}, ..., x_{i_k}) \left(\bigotimes_{p=1}^{k} |\phi_{r_p}^{(p)}\rangle \langle \phi_{l_p}^{(p)}| \right) B_{\mathbf{j}}^{(s)\dagger}(x_{j_1}, ..., x_{j_k}). \tag{89}$$

As stated in Section 2.2, this implies that k quantum systems constitute a $2k$-th order system, since the probability distributions $P(b|\vec{x})$ arising from processes involving k systems will necessarily be at most of algebraic order $2k$.

E   A lower bound on interference of generic composite systems

In this Appendix we are going to show that two systems, labelled by (A, B) and respectively of single-system orders (n_A, n_B), can be used to achieve $(n_A + n_B)$-th order interference.

By definition, there exist two processes involving single-systems A and B that can win the modulo games with n_A and n_B boxes with some probabilities $q_A > 1/d$ and $q_B > 1/d$. For concreteness, suppose that Alice and Bob play the modulo game for the unit dit-string $\nu_i = 1, \forall i$, with $(n_A + n_B)$ boxes, i.e. Bob is supposed to output $s_{\vec{x}}^{(d)} = (\sum_j x_j) \bmod d$. The proof for any other $\vec{\nu}$ is completely analogous. The protocol proceeds as follows. Alice sends system A to the first n_A boxes (i.e. containing inputs $\{x_1, ..., x_{n_A}\}$) and system B through the other n_B boxes (which encode inputs $\{x_{n_A+1}, ..., x_{n_A+n_B}\}$). Upon receiving the systems, Bob performs two separate measurements with outcomes $b^{(A)}$ and $b^{(B)}$ and produces a final output $b = (b^{(A)} + b^{(B)}) \bmod d$. Here the two systems are treated completely independently: the only "mixing" between them happens in the final step, since Bob's output depends on both systems' processes. The $(n_A + n_B)$-th order interference term is then

$$\tilde{I}_{n_A+n_B}^{(d)} = \frac{1}{d^{n_A+n_B}} \sum_{x_1, ... x_{n_A+n_B}} P\left(b = s_{\vec{x}}^{(d)}|\vec{x}\right) - \frac{1}{d} = \frac{1}{d^{n_A+n_B}} \sum_{x_1, ... x_{n_A+n_B}} \sum_l P\left(b_l^{(A)}, b_l^{(B)}|\vec{x}\right) - \frac{1}{d}, \tag{90}$$

where the sum runs over all two-outcome measurements which produce the correct modulo, i.e. $\left(b_l^{(A)} + b_l^{(B)}\right) \bmod d = \left(\sum_{j=1} x_j\right) \bmod d$. Intuitively, Bob can produce the correct overall modulo, even if he measures wrong single-system moduli: e.g. for $d = 2$, if Bob's single-system outcomes are $s_A \oplus 1$ and $s_B \oplus 1$, where s_A and s_B are the correct parities, he will still produce the correct overall parity, since $(s_A \oplus 1) \oplus (s_B \oplus 1) = s_A \oplus s_B$. Indeed, for arbitrary d, there are in total $(d-1)$ wrong single-system outcomes which provide the correct overall output. The interference term is thus

$$\tilde{I}_{n_A+n_B}^{(d)} = q_A q_B + \sum_{i=1}^{d-1} p_i^{(A)} p_{d-i}^{(B)} - \frac{1}{d}, \tag{91}$$

where q_A and q_B are average winning probabilities for the two systems, while the distribution averages over wrong outcomes are defined as

$$p_i^{(A)} \equiv \frac{1}{d^{n_A}} \sum_{x_1, ..., x_{n_A}} P^{(A)}\left(b^{(A)} = (s_{\vec{x}}^{(d)} + i) \bmod d|\vec{x}\right), \tag{92}$$

and analogously for system B. Next, we define weights α_j and β_j in the following way

$$\begin{aligned} p_i^{(A)} &= (1 - q_A)\alpha_i, \\ p_{d-i}^{(B)} &= (1 - q_B)\beta_i, \end{aligned} \tag{93}$$

which implies $\sum_j \alpha_j = \sum_j \beta_j = 1$[5]. Plugging expressions (93) into the interference term we obtain

$$\tilde{I}_{n_A+n_B}^{(d)} = q_A q_B + (1 - q_A)(1 - q_B)\lambda - \frac{1}{d}, \quad \lambda \equiv \sum_j \alpha_j \beta_j. \tag{94}$$

By a straightforward application of the method of Lagrange multipliers with constraints $\sum_j \alpha_j = 1$ and $\sum_j \beta_j = 1$, one obtains that the minimum value of λ is $\frac{1}{d-1}$ (which is achieved for uniform α_j, β_j). Therefore the following inequality holds:

$$\tilde{I}_{n_A+n_B}^{(d)} \geq q_A q_B + \frac{1}{d-1}(1 - q_A)(1 - q_B) - \frac{1}{d}. \tag{95}$$

The RHS of the latter inequality is equal to zero for $q_A = q_B = 1/d$ and is monotonically increasing for $q_A > 1/d, q_B > 1/d$. Thus we constructed a process such that, if the two systems exhibit n_A-th and n_B-th order interference, the composite

---

[5]In order for the coefficients α_i and β_i to be well defined, we assume that $q_A \neq 1$ and $q_B \neq 1$. In the case that either q_A or q_B (or both) are equal to 1, the first term in (91) is $q_A q_B > 1/d$, while the second term is non-negative, which makes the overall interference term positive.

system exhibits $(n_A + n_B)$-th order interference.

Using the same reasoning, the discussion can be generalized to an arbitrary number of systems, since systems can be composed two-by-two. Therefore, by the same construction, k systems of orders $\{n_1, ..., n_k\}$ can produce $(\sum_{i=1}^{k} n_i)$-th order interference.