Simulability of observables in general probabilistic theories

Sergey N. Filippov[*]

Institute of Physics and Technology, Russian Academy of Sciences, Moscow 117218, Russia
and Moscow Institute of Physics and Technology, Dolgoprudny, Moscow Region 141700, Russia

Teiko Heinosaari[†] and Leevi Leppäjärvi[‡]

QTF Centre of Excellence, Department of Physics and Astronomy, University of Turku, Turku 20014, Finland

The existence of incompatibility is one of the most fundamental features of quantum theory and can be found at the core of many of the theory's distinguishing features, such as Bell inequality violations and the no-broadcasting theorem. A scheme for obtaining new observables from existing ones via classical operations, the so-called simulation of observables, has led to an extension of the notion of compatibility for measurements. We consider the simulation of observables within the operational framework of general probabilistic theories and introduce the concept of simulation irreducibility. While a simulation irreducible observable can only be simulated by itself, we show that any observable can be simulated by simulation irreducible observables, which in the quantum case correspond to extreme rank-1 positive-operator-valued measures. We also consider cases where the set of simulators is restricted in one of two ways: in terms of either the number of simulating observables or their number of outcomes. The former is seen to be closely connected to compatibility and k compatibility, whereas the latter leads to a partial characterization for dichotomic observables. In addition to the quantum case, we further demonstrate these concepts in state spaces described by regular polygons.

I. INTRODUCTION

Recently, the concept of measurement simulability of quantum observables (modeled as positive-operator-valued measures) has been introduced and studied [1,2]. It can be seen as a natural generalization of the concept of compatibility, and it allows one to study how one can implement a set of target observables from some chosen set of observables. This kind of concept naturally arises in the studies of local hidden variable models [3] as well as proposals to test fundamentally binary or n-ary theories [4,5].

The framework of general probabilistic theories (GPTs) is natural platform to investigate foundational aspects of quantum theory. Features of quantum theory, such as incompatibility and nonlocality, can be explored in a wider class of theories, allowing one to compare theories to one another and quantify how restricted these features are in different theories. GPTs are based on operational notions of states and measurements so that, for example, an observable is any affine function that maps states into probability distributions. This is the exact analog of positive-operator-valued measures (POVMs) in the case of quantum theory. The incompatibility of observables in GPTs has been recently studied in several works [6–11]. The purpose of the present paper is to formulate measurement simulability in the framework of GPTs and to further investigate the properties of this concept.

The difficulty or complexity of simulating a given collection of observables can be quantified by studying two types of limitations on the set of simulator observables. First, we can look for the minimal set of simulator observables that can produce the target observables. From this point of view, a target set is compatible if and only if it can be obtained with a single simulator observable. Another quantification is obtained by allowing an arbitrary number of simulator observables but restricting them to have fewer outcomes than some threshold value.

We will demonstrate these two quantifications of simulability by comparing quantum theory to polygon theories [12]. It is interesting to recall that the so-called box world (i.e., square bit state space) [13,14] possesses more incompatibility than any finite-dimensional quantum state space [6,15] if incompatibility is quantified as the global robustness under noise. However, in both quantifications of simulability, the box world is closest to classical theory among all nonclassical theories.

The key concept in our investigation is *simulation irreducibility*. An observable has this property if it cannot be obtained from some essentially different simulator observables. We present a general characterization of simulation irreducible observables and explicitly give them in several theories. In particular, we show that the set of all observables on state spaces described by regular polygons can be simulated by a finite number of trichotomic simulation irreducible observables with the only exception being the square bit state space, where simulation irreducible observables are dichotomic.

II. OBSERVABLES, POSTPROCESSING, AND MIXING

A. States, effects, and observables

We first recall some of the basic concepts of general probabilistic theories. The state space $\mathcal{S}$ is a compact convex subset

---
[*]sergey.filippov@phystech.edu
[†]teiko.heinosaari@utu.fi
[‡]leille@utu.fi

of a finite-dimensional real vector space $\mathcal{V}$. The convexity arises from the probabilistic mixing of states so that for $p \in [0,1]$ and states $s_1, s_2 \in \mathcal{S}$ the convex sum $ps_1 + (1-p)s_2$ represents a state where we prepare the state s_1 with probability p and state s_2 with probability $1 - p$.

An effect e is given as a function $e : \mathcal{S} \to [0,1]$ on states such that

$$e(ps_1 + (1-p)s_2) = pe(s_1) + (1-p)e(s_2). \qquad (1)$$

Then $e(s) \in [0,1]$ is interpreted as the probability that the measurement event that the effect e represents happens when the system is in the state $s \in \mathcal{S}$. A functional $f : \mathcal{S} \to \mathbb{R}$ with property (1) is called affine on $\mathcal{S}$ and we denote by $F(\mathcal{S})$ the set of affine functionals on $\mathcal{S}$. We can define a partial order in $F(\mathcal{S})$ by denoting $e \leqslant f$ for $e, f \in F(\mathcal{S})$ if $e(s) \leqslant f(s)$ for all $s \in \mathcal{S}$. The effect space can then be expressed as

$$\mathcal{E}(\mathcal{S}) = \{e \in F(\mathcal{S}) \, | \, o \leqslant e \leqslant u\}, \qquad (2)$$

where o and u are the zero and unit effects respectively, i.e., $o(s) = 0$ and $u(s) = 1$ for all $s \in \mathcal{S}$.

Sometimes it is useful consider the state space $\mathcal{S}$ as being embedded in an ordered vector space $\mathcal{A}$ such that $\mathcal{S}$ is a compact base for a generating positive cone $\mathcal{A}_+ = \{x \in \mathcal{A} \, | \, x \geqslant 0\}$ [16]. Hence, the state space can be expressed as

$$\mathcal{S} = \{x \in \mathcal{A} \, | \, x \geqslant 0, \, u(x) = 1\}, \qquad (3)$$

i.e., as an intersection of the positive cone $\mathcal{A}_+$ and an affine hyperplane determined by (the extension of) the unit effect u on $\mathcal{A}$. Furthermore, if $\dim(\mathrm{aff}(\mathcal{S})) = d$, where $\mathrm{aff}(\mathcal{S})$ denotes the affine span of $\mathcal{S}$, then we can take $\dim(\mathcal{A}) = d + 1$. It follows that, by adopting this approach, the effects can be expressed as linear functionals on $\mathcal{A}$ so that

$$\mathcal{E}(\mathcal{S}) = \{e \in \mathcal{A}^* \, | \, o \leqslant e \leqslant u\}, \qquad (4)$$

where the partial order in the dual space $\mathcal{A}^*$ is the dual order defined by the positive dual cone $\mathcal{A}_+^* = \{f \in \mathcal{A}^* \, | \, f(x) \geqslant 0 \text{ for all } x \in \mathcal{A}_+\}$ of $\mathcal{A}_+$, and also $\dim(\mathcal{A}^*) = d + 1$. In fact, $\mathcal{E}(\mathcal{S})$ is then just the intersection of the positive dual cone $\mathcal{A}_+^*$ and the set $u - \mathcal{A}_+^*$.

A nonzero effect e is *indecomposable* if a decomposition $e = e_1 + e_2$ is possible only when e_1 and e_2 are scalar multiples of e; otherwise they are decomposable. It has been shown in Ref. [17] that in any GPT there exist indecomposable effects and, further, any effect can be written as a finite sum of indecomposable effects. It is easy to see that the indecomposable effects are exactly the ones laying on the extreme rays of the cone $\mathcal{A}_+^*$.

Let $\mathcal{S}$ be a state space. An observable A with a finite number of outcomes is a map $\mathsf{A} : x \mapsto \mathsf{A}_x$ from a finite (outcome) set X to $\mathcal{E}(\mathcal{S})$ with the normalization $\sum_{x \in X} \mathsf{A}_x(s) = 1$ for all $s \in \mathcal{S}$. The normalization condition, which is equivalent to the requirement that $\sum_{x \in X} \mathsf{A}_x = u$, guarantees that we detect with certainty one of the events corresponding to one of the effects A_x of the observable. We denote the set of all observables with outcome set X by $\mathcal{O}_X$ and the set of all observables with a finite number of outcomes on $\mathcal{S}$ by $\mathcal{O}$.

An observable is called *indecomposable* if all of its nonzero effects are indecomposable; otherwise it is *decomposable*. From the decomposition of the unit effect into indecompos-

able effects, it follows that indecomposable observables do exist [17].

Example 1 (quantum theory). In finite-dimensional quantum theory, the state space $\mathcal{S}_q$ is given by the set of positive trace-1 self-adjoint operators on a finite-dimensional Hilbert space $\mathcal{H}$:

$$\mathcal{S}_q = \mathcal{S}(\mathcal{H}) = \{\varrho \in \mathcal{L}_s(\mathcal{H}) \, | \, \varrho \geqslant O, \, \mathrm{tr}[\varrho] = 1\}, \qquad (5)$$

where $\mathcal{L}_s(\mathcal{H})$ is the set of self-adjoint operators on $\mathcal{H}$ and O is the zero operator. The set of positive operators forms a generating positive cone in the vector space of self-adjoint operators $\mathcal{L}_s(\mathcal{H})$ with $\mathcal{S}(\mathcal{H})$ as its compact base. The effect space is given by the set of operators

$$\mathcal{E}(\mathcal{H}) = \{E \in \mathcal{L}_s(\mathcal{H}) \, | \, O \leqslant E \leqslant \mathbb{1}\}, \qquad (6)$$

where $\mathbb{1}$ is the identity operator, so that the one-to-one correspondence with the effect functionals in $\mathcal{E}(\mathcal{S}_q)$ can be given by the equation $e(\varrho) = \mathrm{tr}[\varrho E]$. An observable A with a finite outcome set X then corresponds to a POVM $\mathsf{A} : x \mapsto A(x)$ such that $\sum_{x \in X} A(x) = \mathbb{1}$. An effect E is indecomposable if and only if E has rank equal to 1, or equivalently, E is a scalar multiple of a one-dimensional projection [17].

B. Postprocessing of observables

A classical channel between outcome spaces X and Y is given by a (right) stochastic linear map $\nu : X \to Y$, i.e., map with matrix elements $\nu_{xy}, x \in X, y \in Y$ with $0 \leqslant \nu_{xy} \leqslant 1$ and $\sum_{x \in X} \nu_{xy} = 1$. The matrix element ν_{xy} gives the transition probability that outcome x is mapped into outcome y. In addition to being used as a transformation between outcome spaces, classical channels are most commonly used to describe noise.

For an observable A with an outcome set X and a classical channel $\nu : X \to Y$ between X and some other outcome space Y, we denote by $\nu \circ \mathsf{A}$ a new observable defined as

$$(\nu \circ \mathsf{A})_y = \sum_{x \in X} \nu_{xy} \mathsf{A}_x \qquad (7)$$

for all outcomes $y \in Y$. Physically, the observable $\nu \circ \mathsf{A}$ can be implemented by first measuring A and then using the classical channel ν on each measurement outcome.

For two observables A and B, we say that B *is a postprocessing of* A, denoted by $\mathsf{A} \to \mathsf{B}$, if there exists a classical channel ν such that $\mathsf{B} = \nu \circ \mathsf{A}$. In the context of quantum observables, this relation was introduced in Ref. [18]. We follow the terminology of Ref. [19] and say that an observable A is *postprocessing clean* if, for any observable B, the relation $\mathsf{B} \to \mathsf{A}$ implies that $\mathsf{A} \to \mathsf{B}$. We have the following characterization.

Proposition 1. An observable is postprocessing clean if and only if it is indecomposable.

Proof. Let A be a postprocessing clean observable with an outcome set Ω. In Ref. [17], it was shown that each nonzero effect A_x has a decomposition $\mathsf{A}_x = \sum_{i=1}^{r_x} a_i^{(x)}$ into $r_x < \infty$ indecomposable effects $a_i^{(x)}$. We denote $r = \max_{x \in \Omega} r_x$ and define an observable B with an outcome set $\{1, \ldots, r\} \times \Omega$ by $\mathsf{B}_{(i,x)} = a_i^{(x)}$ if $i \leqslant r_x$ and $\mathsf{B}_{(i,x)} = o$ otherwise. Now we see

that

$$\mathsf{A}_x = \sum_{i=1}^{r_x} a_i^{(x)} = \sum_{i=1}^{r} \mathsf{B}_{(i,x)} = \sum_{i,x'} \nu_{(i,x')x} \mathsf{B}_{(i,x')},$$

where we have defined the postprocessing $\nu : \{1,\ldots,r\} \times \Omega \to \Omega$ by $\nu_{(i,x')x} = \delta_{x',x}$ for all $i = 1,\ldots,r$ and $x,x' \in \Omega$. Thus, $\mathsf{B} \to \mathsf{A}$. Since A is postprocessing clean, it follows that also $\mathsf{A} \to \mathsf{B}$; hence there exists a postprocessing $\mu : \Omega \to \{1,\ldots,r\} \times \Omega$ such that

$$\mathsf{B}_{(i,x)} = \sum_{y \in \Omega} \mu_{y(i,x)} \mathsf{A}_y$$

for all $i = 1,\ldots,r$ and $x \in \Omega$. Each nonzero effect $\mathsf{B}_{(i,x)} = a_i^{(x)}$ is indecomposable, and so for all $x,y \in \Omega$ and $i = 1,\ldots,r_x$ there exists a real number $p_{xy}^i > 0$ such that $\mu_{y(i,x)} \mathsf{A}_y = p_{xy}^i a_i^{(x)}$. From the normalization $\sum_{i,x} \mu_{y(i,x)} = 1$ for all $y \in \Omega$ it follows that for each $y \in \Omega$ (such that $\mathsf{A}_y \neq o$) there exists an element $\mu_{y(i,x)} \neq 0$ for some $i_y = 1,\ldots,r_{x_y}$ and $x_y \in \Omega$. Hence, for each $y \in \Omega$ (such that $\mathsf{A}_y \neq o$), there exists an outcome (i_y,x_y) for the observable B with an indecomposable effect $\mathsf{B}_{(i_y,x_y)} = a_{i_y}^{(x_y)}$ such that

$$\mathsf{A}_y = \frac{p_{x_y y}^{i_y}}{\mu_{y(i_y,x_y)}} a_{i_y}^{(x_y)}.$$

Thus, each nonzero effect of A is indecomposable.

Let then A be an indecomposable observable with an outcome set Ω. We consider an observable C with an outcome set Ω' such that $\mathsf{C} \to \mathsf{A}$, i.e., there exists a postprocessing $\eta : \Omega' \to \Omega$ such that $\mathsf{A}_x = \sum_{z \in \Omega'} \eta_{zx} \mathsf{C}_z$ for all $x \in \Omega$. Without loss of generality, observable C has only nonzero outcomes. Thus, each effect C_z has a decomposition $\mathsf{C}_z = \sum_{j=1}^{r_z} c_j^{(z)}$ into r_z indecomposable effects $c_j^{(z)}$, and

$$\mathsf{A}_x = \sum_{z \in \Omega'} \sum_{j=1}^{r_z} \eta_{zx} c_j^{(z)}$$

for all $x \in \Omega$. Hence, for each $z \in \Omega'$, $j \in \{1,\ldots,r_z\}$ and $x \in \Omega$ such that $\mathsf{A}_x \neq o$ there exists a real number $q_{xz}^j \geq 0$ such that $\eta_{zx} c_j^{(z)} = q_{xz}^j \mathsf{A}_x$. By summing over all $z \in \Omega'$ and $j = 1,\ldots,r_z$ we have that

$$\left(\sum_{z \in \Omega'} \sum_{j=1}^{r_z} q_{xz}^j \right) \mathsf{A}_x = \sum_{z \in \Omega'} \sum_{j=1}^{r_z} \eta_{zx} c_j^{(z)} = \sum_{z \in \Omega'} \eta_{zx} \mathsf{C}_z = \mathsf{A}_x$$

for all $x \in \Omega$ such that $\mathsf{A}_x \neq o$. Thus, for such $x \in \Omega$ we have that $\sum_{z \in \Omega'} \sum_{j=1}^{r_z} q_{xz}^j = 1$. From this it also follows that $0 \leq \sum_j q_{xz}^j \leq 1$ for all $z \in \Omega'$ and $x \in \Omega$ such that $\mathsf{A}_x \neq o$.

Since $\eta_{zx} = 0$ for all $x \in \Omega$ such that $\mathsf{A}_x = o$, we have from the normalization of the postprocessing η that

$$c_j^{(z)} = \sum_{x \in \Omega} \eta_{zx} c_j^{(z)} = \sum_{\substack{x \in \Omega \\ \mathsf{A}_x \neq o}} \eta_{zx} c_j^{(z)} + \sum_{\substack{x \in \Omega \\ \mathsf{A}_x = o}} \eta_{zx} c_j^{(z)}$$

$$= \sum_{\substack{x \in \Omega \\ \mathsf{A}_x \neq o}} q_{xz}^j \mathsf{A}_x.$$

Thus,

$$\mathsf{C}_z = \sum_{j=1}^{r_z} c_j^{(z)} = \sum_j \sum_{\substack{x \in \Omega \\ \mathsf{A}_x \neq o}} q_{xz}^j \mathsf{A}_x = \sum_{x \in \Omega} \lambda_{xz} \mathsf{A}_x,$$

where we have defined $\lambda_{xz} = \sum_{j=1}^{r_z} q_{xz}^j$ when $\mathsf{A}_x \neq o$ and $\lambda_{xz} = 1/\#\Omega'$ otherwise. From the observations made above, we have that $0 \leq \lambda_{xz} \leq 1$ for all $x \in \Omega$ and $z \in \Omega'$, and furthermore $\sum_{z \in \Omega'} \lambda_{xz} = 1$ for all $x \in \Omega$ so that the map $\lambda : \Omega \to \Omega'$ defined by matrix element λ_{xz} is a postprocessing. Hence, $\mathsf{A} \to \mathsf{C}$ for all observables C such that $\mathsf{C} \to \mathsf{A}$ and so A is postprocessing clean. ∎

The postprocessing relation is a preorder on $\mathcal{O}$, i.e., a transitive and symmetric relation. Two observables A and B are *postprocessing equivalent* if both $\mathsf{A} \to \mathsf{B}$ and $\mathsf{B} \to \mathsf{A}$, and in this case we denote $\mathsf{A} \leftrightarrow \mathsf{B}$. This is an equivalence relation, and the set $\mathcal{O}$ therefore splits into equivalence classes. Two postprocessing equivalent observables do not differ in any physically relevant way.

Example 2 (Minimally sufficient representative). In every equivalence class, one has an observable for which all effects are pairwise linearly independent. This was proven for quantum observables with a finite number of outcomes in Ref. [18]. A generalization of this property was introduced and studied in Ref. [20], where such observables were called *minimally sufficient*.

To see that an observable with pairwise linearly independent effects exists for each equivalence class in our setting, let us consider an observable $\mathsf{A} : X \to \mathcal{E}(\mathcal{S})$. Suppose that two effects $\mathsf{A}_{x'}$ and $\mathsf{A}_{x''}$ are linearly dependent (proportional to each other). Consider the outcome set $Y = X \setminus \{x''\}$ and a postprocessing $\nu : X \to Y$ such that $\nu_{xy} = \delta_{xx'} + \delta_{xx''}$ if $y = x'$ and $\nu_{xy} = \delta_{xy}$ otherwise. In the resulting observable $\mathsf{B} = \nu \circ \mathsf{A}$, the effects $\mathsf{A}_{x'}$ and $\mathsf{A}_{x''}$ are merged into $\mathsf{B}_{x'} = \mathsf{A}_{x'} + \mathsf{A}_{x''}$. Thus, by construction $\mathsf{A} \to \mathsf{B}$.

Note that $\mathsf{A}_{x'} = p' \mathsf{B}_{x'}$, $\mathsf{A}_{x''} = p'' \mathsf{B}_{x'}$, where $p',p'' \geq 0$ and $p' + p'' = 1$. By defining the postprocessing $\mu : Y \to X$ such that $\mu_{yx} = p' \delta_{xx'} + p'' \delta_{xx''}$ if $y = x'$ and $\mu_{yx} = \delta_{xy}$ otherwise, we see that $\mathsf{A} = \mu \circ \mathsf{B}$, and hence $\mathsf{B} \to \mathsf{A}$. Therefore, $\mathsf{A} \leftrightarrow \mathsf{B}$. By continuing this kind of merging of linearly dependent pairs of effects, we will eventually obtain an observable $\hat{\mathsf{A}}$ with pairwise linearly independent effects which is postprocessing equivalent with A.

Furthermore, it can be shown that the observable $\hat{\mathsf{A}}$ is essentially unique: If $\widetilde{\mathsf{A}}$ is another pairwise linearly independent observable in the equivalence class of A, then the postprocessing equivalence between $\hat{\mathsf{A}}$ and $\widetilde{\mathsf{A}}$ is given by permutation matrices so that the observables are only bijective relabellings of each other. In Ref. [18], this was proved for quantum observables but since the proof is analogous in the GPT framework it is omitted here.

C. Mixing of observables

A mixing of observables means a procedure where, in each measurement round, we randomly pick an observable from a finite collection and measure it. Thus, if we have m observables $\mathsf{B}^{(1)}, \ldots, \mathsf{B}^{(m)}$ with respective outcome sets $X_1, \ldots, X_m$, then for any probability distribution $p : i \mapsto p_i$ on $\{1,\ldots,m\}$ we

can form an observable B with the outcome set $X \equiv \cup_{i=1}^m X_i$ by

$$\mathsf{B}_x = \sum_{i=1}^m p_i \mathsf{B}_x^{(i)}, \tag{8}$$

where each observable $\mathsf{B}^{(i)}$ is extended onto X by setting $\mathsf{B}_x^{(i)} = o$ if $x \notin X_i$. We can therefore assume that the outcome sets of the mixed observables are the same.

There is another way of forming mixtures, found by keeping track of the measured observable in each round of the measurement. Just as above, we take the outcome sets of the observables $\mathsf{B}^{(1)}, \ldots, \mathsf{B}^{(m)}$ to be equal, say X, but now the outcome in each measurement round is a pair (k,x), where $k \in \{1, \ldots, m\}$ labels the measured observable $\mathsf{B}^{(k)}$ and $x \in X$ is the obtained outcome. To formulate the mixing procedure mathematically, the mixture of these observables is an observable $\tilde{\mathsf{B}}$ with the outcome set $X_0 = \{1, \ldots, m\} \times X$ defined as

$$\tilde{\mathsf{B}}_{(k,x)} = p_k \mathsf{B}_x^{(k)} \tag{9}$$

for each $k \in \{1, \ldots, m\}$ and $x \in X$, where p_k is the probability of measuring the observable $\mathsf{B}^{(k)}$.

It is clear that the latter way of mixing leads to a finer observable than the first one; by postprocessing $\tilde{\mathsf{B}}$ we can obtain the observable that corresponds to mixing without keeping track of the measured observable. Namely, we define a function $f : X_0 \to X$ by $f(k,x) = x$ and define the relabelling $\nu^f : X_0 \to X$ by $\nu^f_{(k,x)y} = 1$ if $f(k,x) = x = y$ and $\nu^f_{(k,x)y} = 0$ if $f(k,x) = x \neq y$. Then

$$(\nu^f \circ \tilde{\mathsf{B}})_x = \sum_{i,x'} \nu^f_{(i,x')x} p_i \mathsf{B}_{x'}^{(i)} = \sum_i p_i \mathsf{B}_x^{(i)} = \mathsf{B}_x. \tag{10}$$

In the following definition, we will understand mixing in the sense that the outcome sets are the same and we do not keep track of the mixed observables.

Definition 1. An observable $\mathsf{A} \in \mathcal{O}_X$ is called *extreme* if a convex sum decomposition $\mathsf{A} = \lambda \mathsf{B} + (1 - \lambda)\mathsf{C}$, $0 < \lambda < 1$ with $\mathsf{B}, \mathsf{C} \in \mathcal{O}_X$ implies $\mathsf{B} = \mathsf{C} = \mathsf{A}$.

The following is a well-known fact for quantum observables; see, e.g., Ref. [21]. It is proven analogously in the GPT framework and a proof is given in the appendix.

Proposition 2. Nonzero effects of an extreme observable are linearly independent.

Also the following statement is well-known for quantum observables; see, e.g., Ref. [22]. We find it useful to give a proof that is valid in the GPT framework.

Proposition 3. A postprocessing clean observable A is extreme if and only if its nonzero effects are linearly independent.

Proof. The necessity of linear independence follows from Proposition 2. To prove sufficiency, let the nonzero effects $\mathsf{A}_1, \ldots, \mathsf{A}_n$ of a postprocessing clean observable A be linearly independent. Let $\mathsf{A} = \sum_k p_k \mathsf{B}^{(k)}$ for some probability distribution $\{p_k\}_k$ and some set of observables $\{\mathsf{B}^{(k)}\}_k$. We define an observable $\tilde{\mathsf{B}}$ as $\tilde{\mathsf{B}}_{(k,x)} = p_k \mathsf{B}_x^{(k)}$. Then

$$\mathsf{A}_y = \sum_{k,z} \nu_{(k,z)y} \tilde{\mathsf{B}}_{(k,z)}, \tag{11}$$

where the postprocessing matrix ν has the form

$$\nu_{(k,z)y} = \begin{cases} 1 & \text{if } z = y, \\ 0 & \text{if } z \neq y. \end{cases} \tag{12}$$

Thus, $\tilde{\mathsf{B}} \to \mathsf{A}$. Since A is postprocessing clean, the latter relation implies $\mathsf{A} \to \tilde{\mathsf{B}}$; i.e., there exists a postprocessing matrix $\mu_{x(k,z)}$ such that

$$\tilde{\mathsf{B}}_{(k,z)} = \sum_x \mu_{x(k,z)} \mathsf{A}_x. \tag{13}$$

Combining (11) with (13), we get

$$\mathsf{A}_y = \sum_x \left(\sum_{k,z} \mu_{x(k,z)} \nu_{(k,z)y} \right) \mathsf{A}_x. \tag{14}$$

Since the effects $\mathsf{A}_1, \ldots, \mathsf{A}_n$ are linearly independent, the term in parentheses must be equal to δ_{xy}. Further, since ν has the specific form of (12), we have that $\sum_k \mu_{x(k,x)} = 1$ for all x. Since $\sum_{k,z} \mu_{x(k,z)} = 1$ for all x and k, it follows that all the elements $\mu_{x(k,z)}$ with $z \neq x$ are zero. Thus, only one term in the sum (13) contributes and hence

$$p_k \mathsf{B}_z^{(k)} = \tilde{\mathsf{B}}_{(k,z)} = \mu_{z(k,z)} \mathsf{A}_z,$$

i.e., $\mathsf{B}_z^{(k)} = \frac{\mu_{z(k,z)}}{p_k} \mathsf{A}_z$. Finally, as $\sum_z \mathsf{B}_z^{(k)} = u$ and $\sum_z \mathsf{A}_z = u$, we get

$$\sum_z \left(\mathsf{B}_z^{(k)} - \mathsf{A}_z \right) = \sum_z \left(\frac{\mu_{z(k,z)}}{p_k} - 1 \right) \mathsf{A}_z = 0. \tag{15}$$

Since the effects $\mathsf{A}_1, \ldots, \mathsf{A}_n$ are linearly independent, the latter equation implies $\mu_{z(k,z)} = p_k$ for all z, so $\mathsf{B}_z^{(k)} = \mathsf{A}_z$ and $\mathsf{B}^{(k)} = \mathsf{A}$ for all k. This means that A is extreme. ∎

III. SIMULATION OF OBSERVABLES

A. Simulation scheme

Let us consider a subset $\mathcal{B} \subseteq \mathcal{O}$ of observables. Following Ref. [1], we consider the set of observables that can be obtained from $\mathcal{B}$ by means of classical manipulations, namely by mixing and postprocessing. The simulation scheme consists of two steps: (i) for any finite subset $\{\mathsf{B}^{(i)}\}_{i=1}^m \subseteq \mathcal{B}$ of observables with outcome set X we choose an observable $\mathsf{B}^{(i)}$ with some probability p_i and measure it, and (ii) after obtaining an outcome (i,x) by keeping track of the measured observable we perform some postprocessing $\nu : \cup_{k=1}^m \{k\} \times X \to Y$, outputting an outcome $y \in Y$ for some outcome space Y with a probability $\nu_{(i,x)y}$. Thus, the result is an observable A with an outcome set Y such that

$$\mathsf{A}_y = (\nu \circ \tilde{\mathsf{B}})_y = \sum_{(i,x)} \nu_{(i,x)y} \tilde{\mathsf{B}}_{(i,x)} \tag{16}$$

for all $y \in Y$, where $\tilde{\mathsf{B}}_{(i,x)} = p_i \mathsf{B}_x^{(i)}$ is the observable used to define the mixture where we keep track of the outcomes. The scheme is depicted in Fig. 1.
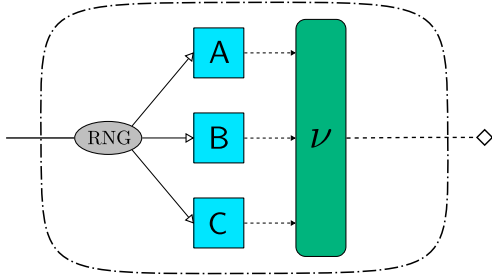
FIG. 1. Simulation of a new observable by using observables A, B, and C. In every individual measurement, the random number generator (RNG) chooses one of observables A, B, and C. Then, the classical outcome is affected by a classical channel ν.

We see that we can write A in two equivalent ways. By expanding (16), we see that

$$
\mathsf{A}_y = \sum_{(i,x)} \nu_{(i,x)y} \tilde{\mathsf{B}}_{(i,x)} = \sum_{(i,x)} \nu_{(i,x)y} p_i \mathsf{B}_x^{(i)}
$$

$$
= \sum_i p_i \left(\sum_x \nu_{(i,x)y} \mathsf{B}_x^{(i)} \right). \tag{17}
$$

Now we may split the postprocessing ν into m parts by defining postprocessings $\nu^{(i)} : X \to Y$ by

$$
\nu_{xy}^{(i)} = \nu_{(i,x)y} \tag{18}
$$

for all $x \in X$, $y \in Y$ and $i = 1, \ldots, m$. Hence, we can express A as

$$
\mathsf{A}_y = \sum_i p_i \left(\sum_x \nu_{xy}^{(i)} \mathsf{B}_x^{(i)} \right) = \sum_i p_i (\nu^{(i)} \circ \mathsf{B}^{(i)})_y. \tag{19}
$$

Thus, we can think of either first mixing the observables and then postprocessing the mixture, or first postprocessing the observables individually and then mixing the postprocessed observables. The scheme in Fig. 1 is therefore equivalent to the scheme in Fig. 2.

As an additional remark, let us consider the case where some of the observables used in the simulation are the same. Suppose we have an observable A with an outcome set Y that can be simulated by observables $\mathsf{B}^{(1)}, \ldots, \mathsf{B}^{(m)}, \mathsf{B}^{(m+1)}, \ldots, \mathsf{B}^{(n)}$ with outcome sets X such that $\mathsf{B}^{(m+1)} = \mathsf{B}^{(m+2)} = \cdots = \mathsf{B}^{(n)}$; i.e.,
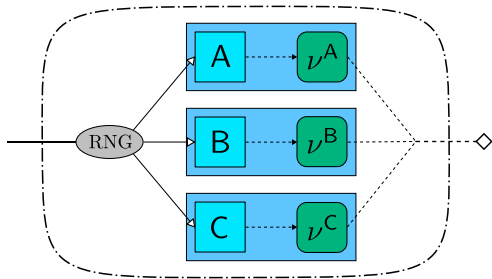

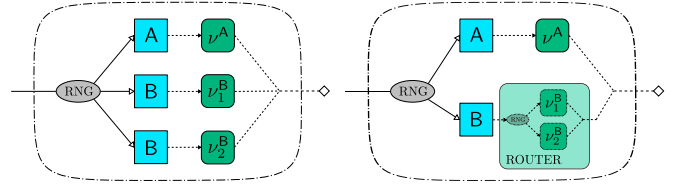FIG. 2. Simulation can be equivalently seen as a mixture of postprocessed observables.

FIG. 3. Multiple uses of a same observables are allowed but do not alter the generality of simulability. The scheme on the left-hand side can be reduced to the scheme on the right-hand side.

we can express A as

$$
\mathsf{A}_y = \sum_{(i,x)} \nu_{(i,x)y} p_i \mathsf{B}_x^{(i)}
$$

$$
= \sum_{i=1}^m \sum_x \nu_{(i,x)y} p_i \mathsf{B}_x^{(i)} + \sum_{i=m+1}^n \sum_x \nu_{(i,x)y} p_i \mathsf{B}_x^{(m+1)} \tag{20}
$$

for all $y \in Y$ with some probability distribution $(p_i)_{i=1}^n$ and postprocessing $\nu : \cup_{k=1}^n \{k\} \times X \to Y$. We can then form a new probability distribution $(p_i')_{i=1}^{m+1}$ by

$$
p_i' = p_i \quad \forall i = 1, \ldots, m,
$$

$$
p_{m+1}' = \sum_{j=m+1}^n p_j. \tag{21}
$$

and a new postprocessing $\nu' : \cup_{k=1}^{m+1} \{k\} \times X \to Y$ by setting

$$
\nu_{(i,x)y}' = \nu_{(i,x)y} \quad \forall i = 1, \ldots, m,
$$

$$
\nu_{(m+1,x)y}' = \sum_{j=m+1}^n \frac{p_j}{p_{m+1}'} \nu_{(j,x)y}, \tag{22}
$$

for all $x \in X$ and $y \in Y$, so that

$$
\mathsf{A}_y = \sum_{i=1}^m \sum_x \nu_{(i,x)y} p_i \mathsf{B}_x^{(i)} + \sum_x p_{m+1}' \sum_{i=m+1}^n \frac{p_i}{p_{m+1}'} \nu_{(i,x)y} \mathsf{B}_x^{(m+1)}
$$

$$
= \sum_{i=1}^m \sum_x \nu_{(i,x)y}' p_i' \mathsf{B}_x^{(i)} + \sum_x \nu_{(m+1,x)y}' p_{m+1}' \mathsf{B}_x^{(m+1)}
$$

$$
= \sum_{i=1}^{m+1} \sum_x \nu_{(i,x)y}' p_i' \mathsf{B}_x^{(i)}
$$

for all $y \in Y$. Hence, instead of using multiple instances of the same observable in the simulating scheme, by modifying the mixing and postprocessing we can reduce the multiplicity so that only one instance of each different observable is used. The intuitive reason for this is that when there are multiple instances of the same observable in the simulator, a *router* can be used to direct the outcomes to the individual postprocessings with some (weighted) probabilities resulting in a reduction of multiplicity; see Fig. 3. Looking from the other way round, we can think of using the same simulator observable several times, even if we would have only a single device to hand.

B. The simulation map

Consider a subset of observables $\mathcal{B} \subseteq \mathcal{O}$. Following the terminlogy from Ref. [1], we say that an observable A is $\mathcal{B}$
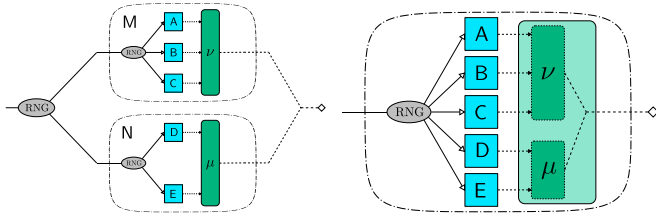
FIG. 4. The set of simulable observables is convex.

simulable if it can be implemented with a simulation scheme by using some finite number of observables from $\mathcal{B}$. Further, we denote by $\mathfrak{sim}(\mathcal{B})$ the set of all observables that are $\mathcal{B}$ simulable, and we treat $\mathfrak{sim}(\cdot)$ as a map on the power set $2^{\mathcal{O}}$. In the case of a singleton set $\{B\}$, we simply denote $\mathfrak{sim}(B) \equiv \mathfrak{sim}(\{B\})$.

For any subsets $\mathcal{B}, \mathcal{C} \subseteq \mathcal{O}$, the map $\mathfrak{sim}(\cdot)$ satisfies the following basic properties:

(sim1) $\mathcal{B} \subseteq \mathfrak{sim}(\mathcal{B})$,

(sim2) $\mathfrak{sim}(\mathfrak{sim}(\mathcal{B})) = \mathfrak{sim}(\mathcal{B})$,

(sim3) $\mathcal{B} \subseteq \mathcal{C} \Rightarrow \mathfrak{sim}(\mathcal{B}) \subseteq \mathfrak{sim}(\mathcal{C})$.

These properties are easy to verify and they mean that $\mathfrak{sim}(\cdot)$ is a *closure operator* on $\mathcal{O}$. It is commonly known that the closure operator properties (sim1)–(sim3) are equivalent to the single condition

(sim4) $\mathcal{B} \subseteq \mathfrak{sim}(\mathcal{C}) \Leftrightarrow \mathfrak{sim}(\mathcal{B}) \subseteq \mathfrak{sim}(\mathcal{C})$.

In the definition of simulability we are requiring that the simulation scheme consists of a finite number of observables. It thus follows that

(sim5) $\mathfrak{sim}(\mathcal{B}) = \bigcup \{ \mathfrak{sim}(\mathcal{B}') : \mathcal{B}' \subseteq \mathcal{B} \text{ and } \mathcal{B}' \text{ is finite} \}$.

This property means that $\mathfrak{sim}(\cdot)$ is an *algebraic* closure operator.

The map $\mathfrak{sim}(\cdot)$ also has the following two properties:

(sim6) $\mathfrak{sim}(\mathcal{B})$ is convex, i.e., closed under mixing,

(sim7) $\mathfrak{sim}(\mathcal{B})$ is closed under postprocessing.

The properties (sim6) and (sim7) are straightforward to verify by noticing the equivalent ways to write mixtures and postprocessings; see Figs. 4 and 5. Complete proofs are presented in the appendix.

C. Simulability and noise content

For an observable A and a set $\mathcal{N} \subset \mathcal{O}$ of noisy observables, we define [10]

$$w(A; \mathcal{N}) = \sup\{0 \leqslant \lambda \leqslant 1 \mid A = \lambda N + (1 - \lambda) B$$
$$\text{for some } N \in \mathcal{N} \text{ and } B \in \mathcal{O}\}$$

as the *noise content of* A *with respect to* $\mathcal{N}$. The noise content $w(A; \mathcal{N})$ thus quantifies how much of A is in $\mathcal{N}$, which is taken to describe noise in the measurements. Contrary to external
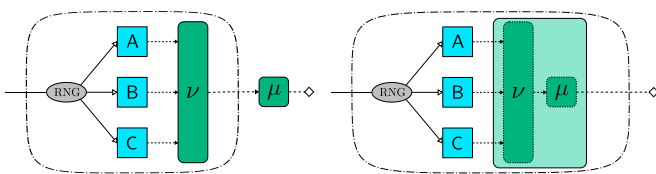

FIG. 5. The set of simulable observables is closed under postprocessings.

noise, i.e., noise that is added to the observables, the noise content gives us the amount of intrinsic noise that is already contained in the observable. The typical choice for the set of noisy observables is $\mathcal{N} = \mathcal{T}$, the set of trivial observables.

The noise content satisfies the following two properties [10]:

(a) If $\mathcal{N}$ is closed under postprocessings, then $w(\nu \circ A; \mathcal{N}) \geqslant w(A; \mathcal{N})$ for all observables A and postprocessings ν,

(b) If $\mathcal{N}$ is convex, then $w(\sum_i p_i A^{(i)}; \mathcal{N}) \geqslant \sum_i p_i w(A^{(i)}; \mathcal{N})$ for all mixtures of any set of observables $\{A^{(i)}\}_i \subset \mathcal{O}$.

We can now prove the intuitive result that we cannot simulate a less noisy observable from noisier ones:

Proposition 4. Let $\mathcal{B}$ be a set of simulators. If the set of noisy observables $\mathcal{N}$ is closed under postprocessings and mixing, then any observable in $\mathfrak{sim}(\mathcal{B})$ has a noise content greater or equal than the smallest noise content of its simulating observables in $\mathcal{B}$.

Proof. Let $A \in \mathfrak{sim}(\mathcal{B})$ so that

$$A_x = \sum_{(i,x)} p_i \nu_{(i,y)x} B_y^{(i)} = \sum_i p_i (\nu^{(i)} \circ B^{(i)})_x \qquad (23)$$

for some set of simulators $\{B^{(i)}\}_i \subset \mathcal{B}$, probability distribution $(p_i)_i$ and postprocessing ν. Now from properties (a) and (b) of the noise content it follows that

$$w(A; \mathcal{N}) = w\left(\sum_i p_i (\nu^{(i)} \circ B^{(i)}); \mathcal{N} \right)$$
$$\geqslant \sum_i p_i w(\nu^{(i)} \circ B^{(i)}; \mathcal{N})$$
$$\geqslant \sum_i p_i w(B^{(i)}; \mathcal{N})$$
$$\geqslant \sum_i p_i \min_k w(B^{(k)}; \mathcal{N})$$
$$= \min_k w(B^{(k)}; \mathcal{N}).$$

If there is an observable $B \in \mathcal{B}$ such that $w(B; \mathcal{N}) \leqslant w(B^{(i)}; \mathcal{N})$ for all i, then $w(A; \mathcal{N}) \geqslant w(B; \mathcal{N})$.

We note that the set of trivial observables $\mathcal{T}$ is indeed convex and closed under postprocessings. ∎

D. Simulation irreducible observables

Clearly, an observable A can be simulated by a subset $\mathcal{B}$ whenever $\mathcal{B}$ contains A, or more generally, if there is $B \in \mathcal{B}$ such that B is postprocessing equivalent to A. Those observables for which this is the only way that they can be simulated we call simulation irreducible:

Definition 2. An observable A is *simulation irreducible* if for any subset $\mathcal{B} \subset \mathcal{O}$, we have $A \in \mathfrak{sim}(\mathcal{B})$ only if there is $B \in \mathcal{B}$ such that $A \leftrightarrow B$.

Simulation irreducibility thus means that the only way we can simulate such observable is essentially with the observable itself. We obtain a following characterization for the simulation irreducible observables.

Proposition 5. An observable is simulation irreducible if and only if it is postprocessing clean and postprocessing equivalent to an extreme observable.

Proof. Let A be postprocessing clean and postprocessing equivalent to an extreme observable $\widetilde{A}$ so that there exist postprocessings μ and η such that $\widetilde{A} = \mu \circ A$ and $A = \eta \circ \widetilde{A}$. Suppose that $A \in \mathfrak{sim}(\mathcal{B})$ for some set of simulators $\mathcal{B}$, i.e., there exists a probability distribution $(p_i)_i$ and postprocessings $\nu^{(i)}$ such that $A_y = \sum_{i,x} p_i \nu_{xy}^{(i)} B_x^{(i)}$ for some $B^{(i)}$'s in $\mathcal{B}$. We can assume that $p_i \neq 0$ for every i as if this is not the case, we simply drop those terms away. We can now write

$$\widetilde{A}_z = \sum_y \mu_{yz} A_y = \sum_{i,x,y} p_i \nu_{xy}^{(i)} \mu_{yz} B_x^{(i)}$$

$$= \sum_i p_i \sum_x \left(\sum_y \nu_{xy}^{(i)} \mu_{yz} \right) B_x^{(i)}$$

$$= \sum_i p_i \sum_x (\mu \circ \nu^{(i)})_{xz} B_x^{(i)}$$

$$= \sum_i p_i (\mu \circ \nu^{(i)} \circ B^{(i)})_z$$

for all outcomes z. From the extremality of $\widetilde{A}$ it follows that $\mu \circ \nu^{(i)} \circ B^{(i)} = \widetilde{A}$ for all i, and therefore $A = \eta \circ \mu \circ \nu^{(i)} \circ B^{(i)}$ for all i. Since A is postprocessing clean, this means that $B^{(i)} \leftrightarrow A$ for all i. Therefore, A is simulation irreducible.

Now let A be a simulation irreducible observable. First, A has to be postprocessing clean; otherwise there exists an observable B such that B is not a postprocessing of A but $A \in \mathfrak{sim}(B)$. Second, if A is extreme, we are done, so let us consider the case when A is not extreme. Then there exists some set of extreme observables $\mathcal{B} = \{B^{(i)}\}_i$ such that A has a convex decomposition $A = \sum_i \lambda_i B^{(i)}$. In particular, $A \in \mathfrak{sim}(\mathcal{B})$ and since A is simulation irreducible, there exists some k such that $A \leftrightarrow B^{(k)}$, where now $B^{(k)}$ is extreme. ∎

We see that both postprocessing cleanness and postprocessing equivalence to an extreme observable are truly needed for simulation irreducibility.

Example 3 (*Postprocessing clean but not simulation irreducible quantum observable*). There are postprocessing clean quantum observables that are not simulation irreducible. For instance, the four-outcome qubit observable A, related to the POVM $A(\pm 1) = \frac{1}{4}(\mathbb{1} \pm \sigma_x)$ and $A(\pm 2) = \frac{1}{4}(\mathbb{1} \pm \sigma_y)$, consists of linearly dependent but pairwisely linearly independent effects. Therefore, A is not simulation irreducible even though it is postprocessing clean. In fact, A can be obtained from two dichotomic observables X and Y as a mixture, where the corresponding POVMs are $X(\pm 1) = \frac{1}{2}(\mathbb{1} \pm \sigma_x)$ and $Y(\pm 1) = \frac{1}{2}(\mathbb{1} \pm \sigma_y)$, respectively.

We recall from the end of Sec. II B that for each observable A, we can form an observable $\hat{A} \leftrightarrow A$ such that the effects of $\hat{A}$ are pairwisely linearly independent. By using the previous propositions, we find a more practical characterization of simulation irreducibility.

Corollary 1. An observable A is simulation irreducible if and only if $\hat{A}$ is indecomposable and extreme, i.e., it consists of linearly independent indecomposable effects.

Proof. First, let A be simulation irreducible. By Proposition 5, A is postprocessing clean and postprocessing equivalent to an extreme observable B. From Proposition 1, we see that A is indecomposable from which it follows that also the pairwise linearly independent observable $\hat{A}$ is indecomposable. What remains to show is that the effects of $\hat{A}$ are actually linearly independent. Since B is extreme, by Proposition 2 the nonzero effects of B are linearly independent. As $\hat{B}$ is formed by combining the pairwise linearly dependent effects of B, we have that $\hat{B} = B$ (without the possible zero effects of B). Thus, $\hat{B}$ is extreme. Since $\hat{B} = B \leftrightarrow A$ is pairwise linearly independent, we have by the uniqueness of $\hat{A}$ that $\hat{B}$ is a bijective relabelling of $\hat{A}$. Hence, $\hat{A}$ is extreme. Because $\hat{A}$ is also postprocessing clean, by Proposition 3 it consists of linearly independent effects.

Second, suppose that $\hat{A}$ consists of linearly independent indecomposable effects. By Proposition 1, $\hat{A}$ is postprocessing clean so that by taking into account that the effects of $\hat{A}$ are linearly independent we have by Proposition 3 that $\hat{A}$ is extreme. Since also A is postprocessing clean and $\hat{A} \leftrightarrow A$ is extreme, from Proposition 5 we conclude that A is simulation irreducible. ∎

We would expect that an observable that is not simulation irreducible is reducible in the sense that it can be simulated by some simulation irreducible observables. Indeed, we can show that this is the case and even holds with a finite number of simulators [21].

Proposition 6. For every observable A, there is a finite collection $\mathcal{B}^A$ of simulation irreducible observables such that $A \in \mathfrak{sim}(\mathcal{B}^A)$.

Proof. Let A be an observable with an outcome set Ω. Each effect A_x can be decomposed into indecomposable effects $a_i^{(x)}$ such that $A_x = \sum_{i=1}^{r_x} a_i^{(x)}$ for some finite r_x. As in the proof of Proposition 1, we denote $r = \max_{x \in \Omega} r_x$ and define an indecomposable observable B with an outcome set $\{1, \ldots, r\} \times \Omega$ by $B_{(i,x)} = a_i^{(x)}$ if $i \leqslant r_x$ and $B_{(i,x)} = o$ otherwise. Let us consider the pairwise linearly independent observable $\hat{B}$. Observable A is then a postprocessing of $\hat{B}$ (as it is of B).

If $\hat{B}$ is extreme, we are done. Otherwise, $\hat{B}$ is not extreme so its effects are linearly dependent; i.e., there exist numbers $\beta_i \in \mathbb{R}$ such that $\sum_i \beta_i \hat{B}_i = o$ with $\sum_i |\beta_i| > 0$. Note that we must have both positive and negative β_i's.

We denote $\kappa_+ = \max_i \beta_i > 0$ and $\kappa_- = \min_i \beta_i < 0$ and consider two observables C and D defined as

$$C_i = (1 - \beta_i/\kappa_+)\hat{B}_i, \tag{24}$$

$$D_i = (1 - \beta_i/\kappa_-)\hat{B}_i, \tag{25}$$

for all outcomes i. We note that both C and D have one nonzero outcome less than $\hat{B}$ since for some indices j and k we have that $\beta_j = \kappa_+$ and $\beta_k = \kappa_-$ so that $C_j = D_k = o$. Since the effects of $\hat{B}_i$ are indecomposable, the observables C and D are also indecomposable. By setting $\lambda = \kappa_+/(\kappa_+ - \kappa_-)$, we find that

$$\hat{B}_i = \lambda C_i + (1 - \lambda) D_i \tag{26}$$

for all i.

Thus, $\hat{B}$ can be expressed as a mixture of two indecomposable observables with one less nonzero outcome. If the nonzero effects of C and D are still linearly dependent, we continue this procedure until we eventually have reduced the outcomes

with finite steps in such a way that the resulting observables, denoted by the set $\mathcal{B}^A$, have linearly independent effects. Since the indecomposability is preserved over the procedure, the observables in $\mathcal{B}^A$ are simulation irreducible. ∎

Example 4 (Simulation irreducible quantum observables). As explained before, a quantum observable A is postprocessing clean if and only if each operator $A(x)$ is rank 1. To check if such an observable is simulation irreducible, we can construct a minimally sufficient representative $\hat{A}$ of the postprocessing equivalence class of A as explained in Sec. II B and then check the linear independence of the effects of $\hat{A}$. In d-dimensional quantum theory $\mathcal{Q}_d$, the maximal number of linearly independent operators is d^2. For any integer $d, \ldots, d^2$, one can construct an extreme postprocessing clean observable [21]. Further, two POVMs A and B consisting of rank 1 operators are seen to be postprocessing equivalent if and only if the set of ranges $\cup_x \{\text{ran}(A(x))\}$ and $\cup_y \{\text{ran}(B(y))\}$ are the same. There is therefore a continuum of postprocessing inequivalent simulation irreducible observables in $\mathcal{Q}_d$ for any $d \geqslant 2$.

IV. LIMITATION ON THE NUMBER OF OBSERVABLES

A. Minimal simulation number

A set of observables $\mathcal{A}$ is *compatible* if there exists an observable G such that every observable in $\mathcal{A}$ can be postprocessed from G. Thus, if $\mathcal{A} = \{A^{(1)}, \ldots, A^{(m)}\}$ is a collection of m observables with outcome sets $X_1, \ldots, X_m$, then $\mathcal{A}$ is compatible if there exists an observable G with an outcome set Y and postprocessings $\nu^{(i)} : Y \to X_i$, $i = 1, \ldots, m$, such that

$$A^{(i)} = \nu^{(i)} \circ G \tag{27}$$

for all $i = 1, \ldots, m$. This means that by measuring only G we can implement a measurement of any observable in $\mathcal{A}$ just by choosing a suitable postprocessing.

As explained in Ref. [1], simulability can be seen as an extension of compatibility. In fact, if we consider an observable G and the set of G-simulable observables $\mathfrak{sim}(G)$, we see that every simulation in $\mathfrak{sim}(G)$ comprises mixing a single observable G so that by reducing the multiplicity the mixing becomes trivial. Then $\mathfrak{sim}(G)$ is seen to be just the set of postprocessings of G, and so $\mathfrak{sim}(G)$ is a compatible set of observables and every subset $\mathcal{A} \subseteq \mathfrak{sim}(G)$ is compatible. On the other hand, if there is a compatible set $\mathcal{A}$ such that every observable can be postprocessed from G, then clearly $\mathcal{A} \subseteq \mathfrak{sim}(G)$.

If a subset $\mathcal{A}$ is not compatible, then there is no single observable G such that $\mathcal{A} \subseteq \mathfrak{sim}(G)$. But we can still search for the minimal collection of simulators that can produce $\mathcal{A}$. This leads to the following definition.

Definition 3. For a subset $\mathcal{A} \subseteq \mathcal{O}$, we denote by $\mathfrak{s}_{\min}(\mathcal{A})$ the minimal number of observables $B^{(1)}, \ldots, B^{(n)}$, if they exist, such that $\mathcal{A} \subseteq \mathfrak{sim}(B^{(1)}, \ldots, B^{(n)})$. Otherwise we denote $\mathfrak{s}_{\min}(\mathcal{A}) = \infty$. We call $\mathfrak{s}_{\min}(\mathcal{A})$ *the minimal simulation number for $\mathcal{A}$.*

Let us consider a finite set $\mathcal{A} = \{A^{(1)}, \ldots, A^{(m)}\} \subset \mathcal{O}$. Clearly, $\mathfrak{s}_{\min}(A^{(1)}, \ldots, A^{(m)}) \leqslant m$. Further, if k observables among $A^{(1)}, \ldots, A^{(m)}$ are compatible, then $\mathfrak{s}_{\min}(A^{(1)}, \ldots, A^{(m)}) \leqslant m - k + 1$. This indicates that the hypergraph structure of the compatibility relation of the set

$\mathcal{A}$, as defined in Ref. [23], relates to $\mathfrak{s}_{\min}(\mathcal{A})$; by identifying the largest subset of compatible observables we get an upper bound for $\mathfrak{s}_{\min}(\mathcal{A})$. This connection is, however, only in one direction, as observed in Ref. [1]. Namely, there exists a set $\{A, B, C\}$ of three quantum observables such that no pair is compatible, but still $\mathfrak{s}_{\min}(A, B, C) = 2$. The following example is slightly different from Example 1 in Ref. [1], which consisted of four observables.

Example 5 (There exist three pairwisely incompatible quantum observables A, B, C such that $\mathfrak{s}_{\min}(A, B, C) = 2$). We denote $A(\pm) = \frac{1}{2}(\mathbb{1} \pm \sigma_x)$, $B(\pm) = \frac{1}{2}(\mathbb{1} \pm \sigma_y)$, and $C_t(\pm) = \frac{1}{2}[\mathbb{1} \pm t(\sigma_x + \sigma_y)/\sqrt{2}]$, where $0 < t < 1$ is a parameter to be specified. Since A (resp. B) consists of projections, any observable compatible with it must commute with it (see, e.g., Ref. [24]). Hence, C_t is incompatible with both for any $0 < t \leqslant 1$. We clearly have $A, B \in \mathfrak{sim}(A, B)$, and we also have $C_t \in \mathfrak{sim}(A, B)$ whenever $t \leqslant 1/\sqrt{2}$. Namely, by taking the equal mixture of A and B, we get a POVM $C_{1/2}(\pm) = \frac{1}{2}[\mathbb{1} \pm (\sigma_x + \sigma_y)/2]$. By using a postprocessing matrix

$$\frac{1}{2}\begin{pmatrix} 1 + \sqrt{2}t & 1 - \sqrt{2}t \\ 1 - \sqrt{2}t & 1 + \sqrt{2}t \end{pmatrix},$$

we get C_t from $C_{1/2}$ for any $t \leqslant 1/\sqrt{2}$. (The fact $C \notin \mathfrak{sim}(A, B)$ for $t > 1/\sqrt{2}$ will be shown in Example 9.)

B. Connection to k compatibility

A joint measurement of observables $A^{(1)}, \ldots, A^{(n)}$ means that we can simultaneously implement their measurements using a single observable, even if only one input system is available. In the context of quantum observables, this notion has been recently generalized to the case where it is assumed that we have access to k copies [25]. We can then make a collective measurement on a state $s^{\otimes k}$. After obtaining a measurement outcome, we can make copies of the outcome and postprocess each copy in a preferred way. This leads to the following notion: Observables $A^{(1)}, \ldots, A^{(n)}$ on sets $\Omega_1, \ldots, \Omega_n$, respectively, are k *compatible* if there exists an observable G with an outcome set Ω_0 acting on the state space $\mathcal{S}^{\otimes k}$, and stochastic matrices $\nu_1, \ldots, \nu_n$ with $\nu_i : \Omega_i \times \Omega_0 \to [0, 1]$, such that

$$\sum_{y \in \Omega_0} \nu_i(x_i, y) G_y(s^{\otimes k}) = A^{(i)}_{x_i}(s) \tag{28}$$

for all $i = 1, \ldots, n$, $x_i \in \Omega_i$, and $s \in \mathcal{S}$. This definition obviously requires that we have specified the tensor product of two state spaces.

As in the usual case of compatibility, we can restrict to a special kind of observables and postprocessings when deciding whether a collection of observables is k compatible. Namely, suppose that Ω is the Cartesian product $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and that C is an observable with this outcome set. One particular type of postprocessing comes from ignoring all but the ith component x_i of a measurement outcome $(x_1, \ldots, x_n)$. This kind of postprocessing gives the ith marginal of C, which we denote as $C^{[i]}$, i.e.,

$$C^{[i]}_x = \sum_{\ell \neq i} \sum_{x_\ell} C_{x_1, \ldots, x_{i-1}, x, x_{i+1} \ldots, x_n}. \tag{29}$$

Suppose there exits an observable G and stochastic matrices $\nu_1, \ldots, \nu_n$ such that (28) holds. We define C as

$$\mathsf{C}_{x_1, \ldots, x_n} = \sum_{y \in \Omega_0} \nu_1(x_1, y) \ldots \nu_n(x_n, y) \mathsf{G}_y, \qquad (30)$$

in which case C is an observable with the outcome set $\Omega_1 \times \cdots \times \Omega_n$ and $\mathsf{C}^{[i]} = \mathsf{A}^{(i)}$.

Proposition 7. If observables $\mathsf{A}^{(1)}, \ldots, \mathsf{A}^{(n)}$ can be simulated by k observables (i.e., $\mathfrak{s}_{\min}(\mathsf{A}^{(1)}, \ldots, \mathsf{A}^{(n)}) \leqslant k$), then they are k compatible.

Proof. Let Λ_i denote the outcome set of an observable $\mathsf{A}^{(i)}$ for all $i = 1, \ldots, n$ and let $\mathsf{B}^{(1)}, \ldots, \mathsf{B}^{(k)}$ be observables with an outcome set Ω such that $\{\mathsf{A}^{(1)}, \ldots, \mathsf{A}^{(n)}\} \subseteq \mathfrak{sim}(\{\mathsf{B}^{(1)}, \ldots, \mathsf{B}^{(k)}\})$. Thus, there exist n probability distributions $(p_j^{(i)})_{j=1}^k$, $i = 1, \ldots, n$ and n postprocessings $\nu^{(i)}$: $\{1, \ldots, k\} \times \Omega \to \Lambda_i$, $i = 1, \ldots, n$ such that

$$\mathsf{A}_{y_i}^{(i)} = \sum_{j=1}^k \sum_{x \in \Omega} p_j^{(i)} \nu_{(j,x)y_i}^{(i)} \mathsf{B}_x^{(j)} \qquad (31)$$

for all $y_i \in \Lambda_i$ and all $i = 1, \ldots, n$.

We define an observable G with an outcome set Ω^k on $\mathcal{S}^{\otimes k}$ as

$$\mathsf{G}_{x_1, \ldots, x_k} = \mathsf{B}_{x_1}^{(1)} \otimes \cdots \otimes \mathsf{B}_{x_k}^{(k)} \qquad (32)$$

for all $(x_1, \ldots, x_k) \in \Omega^k$, and postprocessings $\mu^{(i)} : \Omega^k \to \Lambda_i$ for all $i = 1, \ldots, n$ by

$$\mu_{\vec{x} y_i}^{(i)} = \sum_{j=1}^k p_j^{(i)} \nu_{(j, x_j)y_i}^{(i)} \qquad (33)$$

for all $\vec{x} = (x_1, \ldots, x_k) \in \Omega^n$ and $y_i \in \Lambda_i$. We now see that

$$\begin{aligned}
\sum_{\vec{x} \in \Omega^k} \mu_{\vec{x} y_i}^{(i)} \mathsf{G}_{\vec{x}}(s^{\otimes k}) &= \sum_{x_1 \in \Omega} \cdots \sum_{x_k \in \Omega} \mu_{(x_1, \ldots, x_k)y_i}^{(i)} \prod_{l=1}^k \mathsf{B}_{x_l}^{(l)}(s) \\
&= \sum_{j=1}^k \sum_{x_1 \in \Omega} \cdots \sum_{x_k \in \Omega} p_j^{(i)} \nu_{(j, x_j)y_i}^{(i)} \prod_{l=1}^k \mathsf{B}_{x_l}^{(l)}(s) \\
&= \sum_{j=1}^k \sum_{x_j \in \Omega} p_j^{(i)} \nu_{(j, x_j)y_i}^{(i)} \mathsf{B}_{x_j}^{(j)}(s) \\
&= \mathsf{A}_{y_i}^{(i)}(s)
\end{aligned}$$

for all states $s \in \mathcal{S}$, outcomes $y_i \in \Lambda_i$ and $i = 1, \ldots, n$. Hence, the observables $\mathsf{A}^{(1)}, \ldots, \mathsf{A}^{(n-1)}$, and $\mathsf{A}^{(n)}$ are k compatible. ∎

Example 6 (Triplet of orthogonal qubit observables). We denote $X_t(\pm) = \frac{1}{2}(\mathbb{1} \pm t\sigma_x)$, $Y_t(\pm) = \frac{1}{2}(\mathbb{1} \pm t\sigma_y)$, and $Z_t(\pm) = \frac{1}{2}(\mathbb{1} \pm t\sigma_z)$, where $t \in [0,1]$ is a noise parameter. For $t = 1$ these observables are simulation irreducible and therefore $\mathfrak{s}_{\min}(\mathsf{X}_1, \mathsf{Y}_1, \mathsf{Z}_1) = 3$. The triplet is compatible if and only if $0 \leqslant t \leqslant 1/\sqrt{3}$, so for exactly those values $\mathfrak{s}_{\min}(\mathsf{X}_t, \mathsf{Y}_t, \mathsf{Z}_t) = 1$. It was proved in Ref. [25] that this triplet is 2 compatible if and only if $0 \leqslant t \leqslant \sqrt{3}/2$; hence we conclude that $\mathfrak{s}_{\min}(\mathsf{X}_t, \mathsf{Y}_t, \mathsf{Z}_t) = 2$ for $1/\sqrt{3} < t \leqslant \sqrt{3}/2$. From these results, we cannot conclude the minimal simulation number for values $\sqrt{3}/2 < t < 1$.

V. LIMITATION ON THE NUMBER OF OUTCOMES

A. Effective number of outcomes

We denote by $\mathcal{O}_n$ the set of observables with the outcome set $\{1, \ldots, n\}$.

Definition 4. An observable A *has effectively n outcomes* if n is the least number such that A can be simulated by $\mathcal{O}_n$. We denote by $\mathcal{O}_n^{\mathrm{eff}}$ the set of all those observables that have effectively n or less outcomes. Further, we say that a subset $\mathcal{O}' \subseteq \mathcal{O}$ is *effectively n-tomic* if $\mathcal{O}' \subseteq \mathcal{O}_n^{\mathrm{eff}}$.

Clearly, if an observable A can be simulated by $\mathcal{O}_n$, then also any postprocessing of A can be simulated by $\mathcal{O}_n$. Further, a mixture of two n-tomic observables is at most n-tomic. Therefore, the sets $\mathcal{O}_1^{\mathrm{eff}} \subseteq \mathcal{O}_2^{\mathrm{eff}} \subseteq \cdots$ are convex and closed under postprocessing. The set $\mathcal{O}_1^{\mathrm{eff}}$ consists exactly of all trivial observables, i.e., observables of the form $\mathsf{T}(x) = t(x)u$.

As explained at the end of Sec. II B, for each observable A we can form an observable $\hat{\mathsf{A}} \leftrightarrow \mathsf{A}$ such that the effects of $\hat{\mathsf{A}}$ are pairwisely linearly independent. It follows from the construction of $\hat{\mathsf{A}}$ that the number of outcomes of $\hat{\mathsf{A}}$ is at most the number of outcomes of A. If A is simulation irreducible, then the effective number of outcomes of A is equal to the number of outcomes of $\hat{\mathsf{A}}$.

By Proposition 6, every observable can be simulated with simulation irreducible observables. Therefore, the maximal effective number of outcomes in a given theory can be concluded by looking at the extreme simulation irreducible observables. For instance, for a set of quantum observables $\mathcal{Q}_d$ in a d-dimensional quantum theory, the maximal effective number of outcomes is d^2. We will calculate the maximal effective number of outcomes for some other states spaces in Sec. VI.

Example 7 (Informationally complete quantum observables). An observable A is called *informationally complete* if $\mathsf{A}(s_1) \neq \mathsf{A}(s_2)$ for any two states $s_1 \neq s_2$. A quantum observable A is informationally complete if and only if the respective set of POVM elements $\{A(x) : x \in \Omega\}$ spans the vector space $\mathcal{L}_s(\mathcal{H})$ of all self-adjoint operators [26]. It follows that an informationally complete observable on $\mathcal{Q}_d$ has at least d^2 outcomes. However, it is easy to construct an informationally complete observable which is effectively dichotomic. For this purpose, fix linearly independent operators $B_1, \ldots, B_{d^2} \in \mathcal{L}_s(\mathcal{H})$. For each j, we define a dichotomic POVM $A^{(j)}$ as

$$A^{(j)}(\pm) = \frac{1}{2}(\mathbb{1} \pm B_j / \|B_j\|).$$

The equal mixture of these POVMs is then

$$A(\pm, j) = \frac{1}{2d^2}(\mathbb{1} \pm B_j / \|B_j\|),$$

and then the span of the elements of A is clearly $\mathcal{L}_s(\mathcal{H})$. Therefore, the corresponding observable A is informationally complete but effectively dichotomic.

The mathematical criterion for an observable to be informationally complete is the same in every general probabilistic theory [27], and one can show that the previous conclusion is valid in any general probabilistic theory: There exists an informationally complete observable which is effectively dichotomic.

B. Dichotomic observables

As a particular example, we will take a closer look at dichotomic and effectively dichotomic observables. We will see that in many cases they have a simple geometrical characterization.

Proposition 8. Let $\mathcal{B} = \{\mathsf{B}^{(i)}\}_{i=1}^m$ be a collection of m observables with an outcome set Ω. For a dichotomic observable A with effects A_+ and A_- the following implication holds:

$$\mathsf{A}_+ \in \mathrm{conv}\left(\left\{\{\mathsf{B}_x^{(i)}\}_{i,x}, o, u\right\}\right) \quad \Rightarrow \quad \mathsf{A} \in \mathfrak{sim}(\mathcal{B}).$$

Proof. Let $\mathsf{A}_+ \in \mathrm{conv}(\{\{\mathsf{B}_x^{(i)}\}_{i,x}, o, u\})$ so that

$$\mathsf{A}_+ = \sum_{i=1}^m \sum_{x \in \Omega} \eta_{ix} \mathsf{B}_x^{(i)} + \lambda u + \mu o \tag{34}$$

for some positive numbers $\eta_{ix}, \lambda, \mu \in \mathbb{R}$ for all $i = 1, \dots, m$ and $x \in \Omega$ such that $\sum_{i,x} \eta_{ix} + \lambda + \mu = 1$. From the normalization of the observables in $\mathcal{B}$, it follows that for any probability distribution $(q_i)_{i=1}^m$ we have

$$u = \sum_{i,x} q_i \mathsf{B}_x^{(i)}. \tag{35}$$

By plugging the previous expression in Eq. (34) and neglecting the term with the zero effect o, we have that

$$\mathsf{A}_+ = \sum_{i,x} (\eta_{ix} + \lambda q_i) \mathsf{B}_x^{(i)} = \sum_{i,x} \tilde{\eta}_{ix} \mathsf{B}_x^{(i)}, \tag{36}$$

where we have denoted $\tilde{\eta}_{ix} = \eta_{ix} + \lambda q_i$ for all $i = 1, \dots, m$ and $x \in \Omega$. We can now introduce a probability distribution $(p_i)_{i=1}^m$ by

$$p_i = \max_{x \in \Omega} \tilde{\eta}_{ix}, \quad i = 1, \dots, m - 1,$$

$$p_m = 1 - \sum_{i=1}^{m-1} p_i.$$

It is straightforward to check that $(p_i)_i$ actually forms a probability distribution.

We define a postprocessing $\nu : \{1, \dots, m\} \times \Omega \to \{+, -\}$ by

$$\nu_{(i,x)+} = \begin{cases} \frac{\tilde{\eta}_{ix}}{p_i} & \text{if } p_i \neq 0, \\ 0, & \text{if } p_i = 0, \end{cases} \tag{37}$$

$$\nu_{(i,x)-} = 1 - \nu_{(i,x)+}, \tag{38}$$

for all $i = 1, \dots, m$ and $x \in \Omega$. We see that indeed $\nu_{(i,x)\pm} \in [0,1]$ and $\nu_{(i,x)+} + \nu_{(i,x)-} = 1$ for all $i = 1, \dots, m$ and $x \in \Omega$, so ν is a legitimate postprocessing. Hence, there exists a probability distribution $(p_i)_i$ and a postprocessing ν such that

$$\mathsf{A}_\pm = \sum_{i,x} \nu_{(i,x)\pm} p_i \mathsf{B}_x^{(i)} \tag{39}$$

so that $\mathsf{A} \in \mathfrak{sim}(\mathcal{B})$. ∎

The previous proposition only considers simulated observables which have only two outcomes. We see that the proposition can in fact be extended to cover simulated observables with more outcomes at the expense of the form of the simulator observables.

Proposition 9. Let $\mathcal{B} = \{\mathsf{B}^{(i)}\}_{i=1}^m$ be a collection of m dichotomic observables such that the set $\{u, \{\mathsf{B}_+^{(i)}\}_{i=1}^m\}$ is linearly independent. For an observable A with an outcome set Λ, the following implication holds:

$$\mathsf{A}_y \in \mathrm{conv}(\{\{\mathsf{B}_\pm^{(i)}\}_i, o, u\}) \; \forall y \in \Lambda \quad \Rightarrow \quad \mathsf{A} \in \mathfrak{sim}(\mathcal{B}).$$

Proof. Let A be an observable with outcome set Λ such that $\mathsf{A}_y \in \mathrm{conv}(\{\{\mathsf{B}_\pm^{(i)}\}_i, o, u\})$ for all $y \in \Lambda$ so that

$$\mathsf{A}_y = \sum_i (\lambda_+^{(i,y)} \mathsf{B}_+^{(i)} + \lambda_-^{(i,y)} \mathsf{B}_-^{(i)}) + \lambda^{(u,y)} u + \lambda^{(o,y)} o$$

$$= \sum_i (\omega_+^{(i,y)} \mathsf{B}_+^{(i)} + \omega_-^{(i,y)} \mathsf{B}_-^{(i)}), \tag{40}$$

where $\{\{\lambda_\pm^{(i,y)}\}_i, \lambda^{(u,y)}, \lambda^{(o,y)}\}$ is a probability distribution for all $y \in \Lambda$ and $\omega_\pm^{(i,y)} = \lambda_\pm^{(i,y)} + \frac{1}{m} \lambda^{(u,y)}$ for all $i = 1, \dots, m$ and $y \in \Lambda$. Here we have taken into account that $u = \frac{1}{m} \sum_i (\mathsf{B}_+^{(i)} + \mathsf{B}_-^{(i)})$.

Because of the normalization of A, we have that

$$u = \sum_y \mathsf{A}_y = \sum_i (\omega_+^{(i)} \mathsf{B}_+^{(i)} + \omega_-^{(i)} \mathsf{B}_-^{(i)})$$

$$= \left(\sum_i \omega_-^{(i)}\right) u + \sum_i (\omega_+^{(i)} - \omega_-^{(i)}) \mathsf{B}_+^{(i)}, \tag{41}$$

where $\omega_\pm^{(i)} = \sum_y \omega_\pm^{(i,y)} \geqslant 0$ for all $i = 1, \dots, m$.

Since effects u, $\mathsf{B}_+^{(1)}, \dots, \mathsf{B}_+^{(m)}$ are linearly independent, we conclude that $\omega_+^{(i)} = \omega_-^{(i)} =: p_i$ for all $i = 1, \dots, m$ and $\sum_i p_i = 1$. We can then define a postprocessing $\nu : \{1, \dots, m\} \times \{+, -\} \to \Lambda$ by setting

$$\nu_{(i,\pm)y} = \begin{cases} \frac{\omega_\pm^{(i,y)}}{p_i}, & \text{if } p_i \neq 0, \\ \frac{1}{m}, & \text{if } p_i = 0. \end{cases} \tag{42}$$

From Eq. (40), we can now confirm that

$$\mathsf{A}_y = \sum_i p_i (\nu_{(i,+)y} \mathsf{B}_+^{(i)} + \nu_{(i,-)y} \mathsf{B}_-^{(i)}) \tag{43}$$

for all $y \in \Lambda$ so that $\mathsf{A} \in \mathfrak{sim}(\mathcal{B})$. ∎

We note that if there is only one dichotomic simulator observable B, then $\{u, \mathsf{B}_+\}$ is linearly independent if and only if B is nontrivial. In the case of one simulator B, we can even have more outcomes for B provided that the effects of B are linearly independent.

Proposition 10. Let B be an observable with linearly independent effects and an outcome set Ω. For an observable A with an outcome set Λ, the following implication holds:

$$\mathsf{A}_y \in \mathrm{conv}(\{\{\mathsf{B}_x\}_x, o, u\}) \quad \forall y \in \Lambda \quad \Rightarrow \quad \mathsf{A} \in \mathfrak{sim}(\mathsf{B}).$$

Proof. Each effect A_y can be expressed as a convex decomposition into the effects B_x, o, and u so that

$$\mathsf{A}_y = \sum_x \lambda_x^{(y)} \mathsf{B}_x + \lambda_o^{(y)} o + \lambda_u^{(y)} u \tag{44}$$

for all $y \in \Lambda$ for some positive numbers $\lambda_x^{(y)}, \lambda_o^{(y)}$ and $\lambda_u^{(y)}$ such that $\sum_x \lambda_x^{(y)} + \lambda_o^{(y)} + \lambda_u^{(y)} = 1$ for all $y \in \Lambda$. Since

$u = \sum_x \mathsf{B}_x$, we have that

$$\mathsf{A}_y = \sum_x \left(\lambda_x^{(y)} + \lambda_u^{(y)}\right)\mathsf{B}_x \tag{45}$$

for all $y \in \Lambda$. From the normalization of observables A and B it follows that

$$\sum_x \mathsf{B}_x = u = \sum_y \mathsf{A}_y = \sum_x \left[\sum_y \left(\lambda_x^{(y)} + \lambda_u^{(y)}\right)\right]\mathsf{B}_x. \tag{46}$$

The linear independence of the effects B_x leads us to conclude that $\sum_y (\lambda_x^{(y)} + \lambda_u^{(y)}) = 1$ for all $x \in \Omega$. Thus, if we define a mapping $\nu : \Omega \to \Lambda$ by

$$\nu_{xy} = \lambda_x^{(y)} + \lambda_u^{(y)} \tag{47}$$

for all $x \in \Omega$ and $y \in \Lambda$, we see that now ν is a postprocessing and $\mathsf{A} = \nu \circ \mathsf{B}$. ∎

As an example of this, a simulation irreducible observable (or its minimally sufficient version) consists of linearly independent effects, and so in this case we have a sufficient condition for an observable to be simulated by it. However, we see that the condition is not a necessary one and also that if we try to increase the number of simulators then the proposition no longer holds. The converse of Proposition 8 is also seen to be false in general.

Example 8 (Simulation irreducible qubit observable). Let us consider a 4-outcome qubit observable B with effects

$$B(i) = \tfrac{1}{4}(\mathbb{1} + \vec{b}_i \cdot \vec{\sigma}), \quad i = 1,2,3,4,$$

where

$$\vec{b}_1 = \left(\frac{2\sqrt{2}}{3}, 0, -\frac{1}{3}\right), \quad \vec{b}_2 = \left(-\frac{\sqrt{2}}{3}, \sqrt{\frac{2}{3}}, -\frac{1}{3}\right),$$

$$\vec{b}_3 = \left(-\frac{\sqrt{2}}{3}, -\sqrt{\frac{2}{3}}, -\frac{1}{3}\right), \quad \vec{b}_4 = (0,0,1), \tag{48}$$

so that the four vectors form the vertices of a tetrahedron inside a unit ball. Clearly, any set of three of the four vectors form a linearly independent set and in fact the set of all four effects is linearly independent. Furthermore, since the effects $B(i)$ are rank 1 for all $i = 1,2,3,4$, we have that B is simulation irreducible.

To see that the converses of Propositions 8 and 10 do not hold, we define a dichotomic qubit observable A by setting

$$A(+) = B(1) + B(2) = \frac{1}{2}\left[\mathbb{1} + \left(\frac{\vec{b}_1 + \vec{b}_2}{2}\right) \cdot \vec{\sigma}\right],$$

$$A(-) = B(3) + B(4) = \frac{1}{2}\left[\mathbb{1} + \left(\frac{\vec{b}_3 + \vec{b}_4}{2}\right) \cdot \vec{\sigma}\right].$$

Clearly, $\mathsf{A} \in \mathfrak{sim}(\mathsf{B})$. We will show by contradiction that $A(+)$ does not belong to the convex set of effects $O, \mathbb{1}, B(1), B(2), B(3), B(4)$. Suppose that $A(+) \in \mathrm{conv}(\{B(1), B(2), B(3), B(4), O, \mathbb{1}\})$, i.e., for $A(+)$ there exists

a convex decomposition

$$A(+) = \sum_{i=1}^4 \lambda_i B(i) + \lambda_5 \mathbb{1} + \lambda_6 O$$

$$= \frac{1}{2}\left[\left(\frac{\sum_{i=1}^4 \lambda_i}{2} + 2\lambda_5\right)\mathbb{1} + \left(\frac{\sum_{i=1}^4 \lambda_i \vec{b}_i}{2}\right) \cdot \vec{\sigma}\right].$$

By comparing the coefficients of $\mathbb{1}$ and the Pauli matrices, we arrive at the following two equations:

$$\sum_{i=1}^4 \lambda_i + 4\lambda_5 = 2,$$

$$\sum_{i=1}^4 \lambda_i \vec{b}_i = \vec{b}_1 + \vec{b}_2.$$

By using the latter equation and the condition that $\sum_{i=1}^4 \vec{b}_i = \vec{0}$, we have that

$$(1 + \lambda_4 - \lambda_1)\vec{b}_1 + (1 + \lambda_4 - \lambda_2)\vec{b}_2 + (\lambda_4 - \lambda_3)\vec{b}_3 = \vec{0}. \tag{49}$$

Now the set $\{\vec{b}_1, \vec{b}_2, \vec{b}_3\}$ is linearly independent so that

$$\lambda_1 = \lambda_2 = 1 + \lambda_4 \quad \Rightarrow \quad \lambda_1 = \lambda_2 = 1, \tag{50}$$

which contradicts the fact that $\sum_{i=1}^6 \lambda_i = 1$. Thus, $A(+)$ cannot be contained in the convex hull of O, $\mathbb{1}$, and the effects of B. By similar arguments, we see that since, for example, the set $\{\vec{b}_2, \vec{b}_3, \vec{b}_4\}$ is linearly independent, then $A(-)$ also cannot be contained in the convex hull of O, $\mathbb{1}$, and the effects of B.

We also see that in Proposition 9 both the dichotomicity of observables $\mathsf{B}^{(i)}$ and the linear independence of the effects $\{u, \{\mathsf{B}_+^{(i)}\}_i\}$ is truly needed: Define dichotomic observables $\mathsf{C}^{(i)}$ by setting $C^{(i)}(+) = B(i)$ and $C^{(i)}(-) = \mathbb{1} - B_i$ for all $i = 1,2,3,4$. Clearly $B_i \in \mathrm{conv}(\{\{C^{(j)}(\pm)\}_{j=1}^4, O, \mathbb{1}\})$ for all $i = 1,2,3,4$ and even the effects $C^{(1)}(+), C^{(2)}(+), C^{(3)}(+)$, and $C^{(4)}(+)$ are linearly independent by themselves but not with the unit effect $\mathbb{1}$. If $\mathsf{B} \in \mathfrak{sim}(\{\mathsf{C}^{(i)}\}_{i=1}^4)$, then by the simulation irreducibility of B we would have that $\mathsf{B} \leftrightarrow \mathsf{C}^{(k)}$ for some $k \in \{1,2,3,4\}$ which is clearly not the case since when measuring $\mathsf{C}^{(k)}$ we only get information about the outcome k of the observable B and not the other outcomes. This also happens when we define two trichotomic observables $\mathsf{D}^{(1)}$ and $\mathsf{D}^{(2)}$ by setting

$$D^{(1)}(1) = B(1), \ D^{(1)}(2) = B(2), \ D^{(1)}(3) = B(3) + B(4),$$

$$D^{(2)}(1) = B(3), \ D^{(2)}(2) = B(4), \ D^{(2)}(3) = B(1) + B(2),$$

since then the effects $\mathbb{1}, D^{(1)}(1)$, and $D^{(2)}(1)$ are linearly independent and $B(i) \in \mathrm{conv}(\{\{D^{(j)}(k)\}_{j,k}, O, \mathbb{1}\})$ for all $i = 1,2,3,4$ but now by the same arguments as above we have that $\mathsf{B} \notin \mathfrak{sim}(\{\mathsf{D}^{(1)}, \mathsf{D}^{(2)}\})$. This also shows that Proposition 8 does not hold with simulated observables which have more than two outcomes.

If we restrict ourselves to sets of simulators composed of dichotomic observables, the converse of Proposition 8 is seen to hold even when allowing more outcomes for the simulated observables.

Proposition 11. Let $\mathcal{B} = \{\mathsf{B}^{(i)}\}_{i=1}^m$ be a collection of m dichotomic observables. For an observable A with an outcome set Λ, the following implication holds:

$$\mathsf{A} \in \mathfrak{sim}(\mathcal{B}) \quad \Rightarrow \quad \mathsf{A}_y \in \mathrm{conv}(\{\{\mathsf{B}_\pm^{(i)}\}_i, o, u\}) \quad \forall y \in \Lambda.$$

Proof. Denote $\mathcal{I}_m = \{1, \ldots, m\}$. Let $\mathsf{A} \in \mathfrak{sim}(\mathcal{B})$ so that

$$\mathsf{A}_y = \sum_{i \in \mathcal{I}_m} p_i(\nu_{(i,+)y} \mathsf{B}_+^{(i)} + \nu_{(i,-)y} \mathsf{B}_-^{(i)}) \tag{51}$$

for some probability distribution $(p_i)_{i=1}^m$ and a postprocessing $\nu : \mathcal{I}_m \times \{+,-\} \to \Lambda$.

For each $y \in \Lambda$, we denote $\mathcal{I}_y^+ = \{i \in \mathcal{I}_m \mid \nu_{(i,+)y} \geqslant \nu_{(i,-)y}\}$ and $\mathcal{I}_y^- = \mathcal{I}_m \setminus \mathcal{I}_y^+$. Now we may express each effect A_y as

$$\mathsf{A}_y = \sum_{i \in \mathcal{I}_y^+} p_i(\nu_{(i,+)y} \mathsf{B}_+^{(i)} + \nu_{(i,-)y} \mathsf{B}_-^{(i)})$$
$$+ \sum_{i \in \mathcal{I}_y^-} p_i(\nu_{(i,+)y} \mathsf{B}_+^{(i)} + \nu_{(i,-)y} \mathsf{B}_-^{(i)})$$
$$= \sum_{i \in \mathcal{I}_y^+} p_i[(\nu_{(i,+)y} - \nu_{(i,-)y}) \mathsf{B}_+^{(i)} + \nu_{(i,-)y} u]$$
$$+ \sum_{i \in \mathcal{I}_y^-} p_i[(\nu_{(i,-)y} - \nu_{(i,+)y}) \mathsf{B}_-^{(i)} + \nu_{(i,+)y} u]$$
$$= \sum_{i \in \mathcal{I}_y^+} p_i(\nu_{(i,+)y} - \nu_{(i,-)y}) \mathsf{B}_+^{(i)}$$
$$+ \sum_{i \in \mathcal{I}_y^-} p_i(\nu_{(i,-)y} - \nu_{(i,+)y}) \mathsf{B}_-^{(i)}$$
$$+ \left[\sum_{i \in \mathcal{I}_y^+} p_i \nu_{(i,-)y} + \sum_{i \in \mathcal{I}_y^-} p_i \nu_{(i,+)y}\right] u,$$

where we have used the fact that $\mathsf{B}_-^{(i)} = u - \mathsf{B}_+^{(i)}$ for all $i \in \mathcal{I}_m$. We see that now the coefficients of all the effects in the above expression are positive and for the total sum of the coefficients we have that

$$\sum_{i \in \mathcal{I}_y^+} p_i(\nu_{(i,+)y} - \nu_{(i,-)y}) + \sum_{i \in \mathcal{I}_y^-} p_i(\nu_{(i,-)y} - \nu_{(i,+)y})$$
$$+ \sum_{i \in \mathcal{I}_y^+} p_i \nu_{(i,-)y} + \sum_{i \in \mathcal{I}_y^-} p_i \nu_{(i,+)y}$$
$$= \sum_{i \in \mathcal{I}_y^+} p_i \nu_{(i,+)y} + \sum_{i \in \mathcal{I}_y^-} p_i \nu_{(i,-)y}$$
$$\leqslant \sum_{i \in \mathcal{I}_y^+} p_i + \sum_{i \in \mathcal{I}_y^-} p_i = \sum_{i \in \mathcal{I}_m} p_i = 1.$$

Thus, by adding the zero effect o in the last expression for A_y with a weight of $1 - \sum_{i \in \mathcal{I}_y^+} p_i \nu_{(i,+)y} - \sum_{i \in \mathcal{I}_y^-} p_i \nu_{(i,-)y}$, we get a convex decomposition for A_y so that

$$\mathsf{A}_y \in \mathrm{conv}(\{\{\mathsf{B}_\pm^{(i)}\}_i, o, u\}) \tag{52}$$

for all $y \in \Lambda$. $\blacksquare$

The previous proposition shows that if an observable is effectively dichotomic, all of its effects are contained in the convex hull of the zero effect, the unit effect, and the effects of the dichotomic simulator observables. That is, if for a given set of dichotomic observables corresponding to some measurement devices in a laboratory, we choose some postprocessing and a probability distribution such that we make a simulation with those measurement devices, the previous proposition can be used to extract the simulated observable's convex decomposition into the effects of the set of simulators and the zero and the unit effect, thereby giving us their mathematical expressions.

On the other hand, it gives a useful necessary condition for dichotomic simulability in an experimental setting. Let us say we have access to some fixed set of measurement devices that correspond to some dichotomic observables $\mathcal{B}$ and we want to know whether a given observable A can be simulated using the accessible measurements. If we find an effect of A that is not contained in the convex hull of o, u, and the effects of the observables in $\mathcal{B}$, we know that A cannot be simulated by $\mathcal{B}$.

In general, however, we note that if the set of simulators is not fixed, for any observable we can always find such dichotomic observables so that condition (52) is satisfied, namely the binarizations of the given observable.

From Propositions 9 and 11, we get the following corollary.

Corollary 2. Let $\mathcal{B} = \{\mathsf{B}^{(i)}\}_{i=1}^m$ be a collection of m dichotomic observables such that the set $\{u, \{\mathsf{B}_+^{(i)}\}_{i=1}^m\}$ is linearly independent. An observable A with an outcome set Λ is contained in $\mathfrak{sim}(\mathcal{B})$ if and only if $\mathsf{A}_y \in \mathrm{conv}(\{\{\mathsf{B}_\pm^{(i)}\}_i, o, u\})$ for all outcomes $y \in \Lambda$.

If the set of simulators as well as the simulated observable are all dichotomic, we get the following simple corollary from Propositions 8 and 11.

Corollary 3. Let $\mathcal{B} = \{\mathsf{B}^{(i)}\}_{i=1}^m$ be a collection of m dichotomic observables. A dichotomic observable A is contained in $\mathfrak{sim}(\mathcal{B})$ if and only if $\mathsf{A}_+ \in \mathrm{conv}(\{\{\mathsf{B}_\pm^{(i)}\}_i, o, u\})$.

From Propositions 10 and 11, we get a full characterization for the simulation set of a single simulation irreducible dichotomic observable.

Corollary 4. Let B be a simulation irreducible dichotomic observable. An observable A with an outcome set Λ is contained in $\mathfrak{sim}(\mathsf{B})$ if and only if $\mathsf{A}_y \in \mathrm{conv}(\{\mathsf{B}_+, \mathsf{B}_-, o, u\})$ for all $y \in \Lambda$.

Example 9. A qubit effect E can be written in the form $E = \frac{1}{2}[(1 + e_0)\mathbb{1} + \vec{e} \cdot \vec{\sigma}]$ for some $e_0 \in \mathbb{R}$ and $\vec{e} = (e_x, e_y, e_z) \in \mathbb{R}^3$ satisfying $|e_0| + \|\vec{e}\|_2 \leqslant 1$. The real number $e_0 \in [-1, 1]$ is called the *bias* of the effect E, with E being *unbiased* if $e_0 = 0$. We denote by X, Y, and Z the observables that have the effects $X(\pm) = \frac{1}{2}(\mathbb{1} \pm \sigma_x)$, $Y(\pm) = \frac{1}{2}(\mathbb{1} \pm \sigma_y)$, and $Z(\pm) = \frac{1}{2}(\mathbb{1} \pm \sigma_z)$, and consider the simulation set $\mathfrak{sim}(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ of those observables. We also denote by T the trivial observable with effects $T(+) = \mathbb{1}$ and $T(-) = O$.

Since the set of effects $\{\mathbb{1}, X(+), Y(+), Z(+)\}$ is linearly independent, it follows from Corollary 2 that a qubit observable E with an outcome set Ω is contained in $\mathfrak{sim}(\mathsf{X}, \mathsf{Y}, \mathsf{Z})$ if and only if the effects $E(j) = \frac{1}{2}[(1 + e_0^{(j)})\mathbb{1} + \vec{e}^{(j)} \cdot \vec{\sigma}] \in \mathrm{conv}(\{X(\pm), Y(\pm), Z(\pm), O, \mathbb{1}\})$ for all $j \in \Omega$. The set of effects $\{T(\pm), X(\pm), Y(\pm), Z(\pm)\}$ is convexly independent, so the set of extreme effects of $\mathrm{conv}(\{T(\pm), X(\pm), Y(\pm), Z(\pm)\})$ are exactly the effects $\{T(\pm), X(\pm), Y(\pm), Z(\pm)\}$. These effects correspond to vectors $\{(\pm 1, 0, 0, 0), (0, \pm 1, 0, 0), (0, 0,$
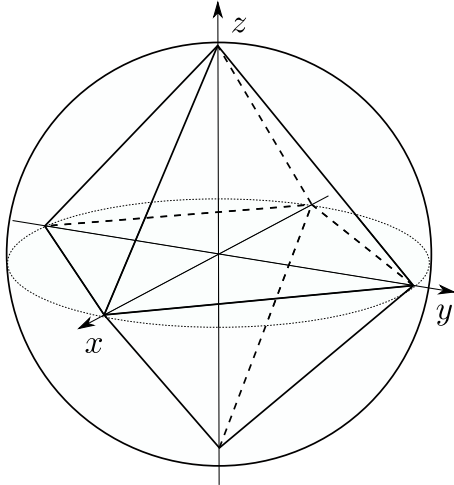
FIG. 6. The unbiased effects in $\mathfrak{sim}(\mathsf{X},\mathsf{Y},\mathsf{Z})$ form an octahedron.

$\pm 1, 0), (0, 0, 0, \pm 1)\}$ in $\mathbb{R}^4$, respectively, which in turn are the extreme points of the four-dimensional convex set

$$S^4 = \left\{ (r_0, r_1, r_2, r_3) \in \mathbb{R}^4 \mid \sum_{i=0}^{3} |r_i| \leqslant 1 \right\}. \quad (53)$$

Thus, there is a one-to-one correspondence with the effects in $\mathrm{conv}(\{X(\pm), Y(\pm), Z(\pm), O, \mathbb{1}\})$ and the points in S^4, and so the observable E with effects $E(j) = \frac{1}{2}[(1 + e_0^{(j)})\mathbb{1} + \vec{e}^{(j)} \cdot \vec{\sigma}]$ is in $\mathfrak{sim}(\mathsf{X},\mathsf{Y},\mathsf{Z})$ if and only if $(e_0^{(j)}, \vec{e}^{(j)}) \in S^4$ for all $j \in \Omega$, i.e.,

$$|e_0^{(j)}| + \|\vec{e}^{(j)}\|_1 \leqslant 1. \quad (54)$$

For the unbiased case, i.e., when $e_0^{(j)} = 0$ for all $j \in \Omega$, inequality (54) defines an octahedron in $\mathbb{R}^3$ which is depicted in Fig. 6. We also see that the set of unbiased effects in $\mathfrak{sim}(\mathsf{X},\mathsf{Y})$ forms a square in $\mathbb{R}^2$ [as do $\mathfrak{sim}(\mathsf{X},\mathsf{Z})$ and $\mathfrak{sim}(\mathsf{Y},\mathsf{Z})$ too].

VI. NONQUANTUM STATE SPACES

A. Classical state spaces

A state space $\mathcal{S}$ is *classical* if all pure states are distinguishable, or equivalently, $\mathcal{S}$ is simplex. Up to the labeling of outcomes, the observable that can distinguish all pure states is unique. It is clear that any classical state space $\mathcal{S}_{cl}$ has only one equivalence class of simulation irreducible observables: Let G be the observable on $\mathcal{S}_{cl}$ that distinguishes the pure states of $\mathcal{S}_{cl}$. For each observable A we define a postprocessing ν^A by setting $\nu_{xy}^\mathsf{A} = \mathsf{A}_y(s_x)$ for all outcomes y and pure states $s_x \in \mathcal{S}_{cl}^{ext}$. Since for any state $s = \sum_x \lambda_x s_x$ we have that $\mathsf{G}_x(s) = \lambda_x$, and so

$$\mathsf{A}_y(s) = \sum_x \lambda_x \mathsf{A}_y(s_x) = \sum_x \nu_{xy}^\mathsf{A} \mathsf{G}_x(s) = (\nu^\mathsf{A} \circ \mathsf{G})_y(s) \quad (55)$$

for all outcomes y and states $s \in \mathcal{S}_{cl}$, so $\mathsf{A} \in \mathfrak{sim}(\mathsf{G})$ and therefore $\mathcal{O} = \mathfrak{sim}(\mathsf{G})$. If G' is some other simulation irreducible observable, then $\mathsf{G}' \in \mathfrak{sim}(\mathsf{G})$, and from the fact that G' is postprocessing clean it follows that also $\mathsf{G} \in \mathfrak{sim}(\mathsf{G}')$ which yields $\mathsf{G}' \leftrightarrow \mathsf{G}$. Furthermore, the extreme simulation irreducible observable has the same number of outcomes as

the number of pure states in $\mathcal{S}_{cl}$. We conclude that the effective number of any observable in a classical state space is at most n, where n is the number of pure states.

On the other hand, if there exists only a single equivalence class of simulation irreducible observables on a state space $\mathcal{S}$, the state space must be classical; this follows from the result of Ref. [9]. In the following, we give an alternative proof of this fact, relying on the properties of simulation irreducible observables.

Let us denote $d = \dim(\mathrm{aff}(\mathcal{S}))$ so that as in Sec. II A we can consider $\mathcal{S}$ and $\mathcal{E}(\mathcal{S})$ to be embedded in $(d+1)$-dimensional ordered vector spaces $\mathcal{A}$ and $\mathcal{A}^*$ respectively. Denote by B the extreme simulation irreducible observable in the equivalence class and suppose it has n outcomes. From Proposition 6, it follows that every observable on $\mathcal{S}$ can be simulated with B. Now B consists of n linearly independent indecomposable effects B_i. For each indecomposable effect B_i there exists an extreme effect b_i and $\beta_i \in (0,1]$ such that $\mathsf{B}_i = \beta_i b_i$ for all $i = 1, \ldots, n$ [17]. Since the n dichotomic observables determined by the effects b_i must be simulable by B, there exist postprocessings $\nu^{(i)}$ such that $b_i = \sum_j \nu_{j+} \mathsf{B}_j = \sum_j \nu_{j+} \beta_j b_j$ for all $i = 1, \ldots, n$ and since the set $\{b_i\}_i$ is linearly independent, it follows that $\beta_i = 1$ for all $i = 1, \ldots, n$. Thus, the effects of B are actually extreme.

It is easy to see that for each extreme effect there exists an extreme state that gives probability one for the state [17]. Thus, for every effect B_i there exists a pure state s_i such that $\mathsf{B}_i(s_i) = 1$ for all $i = 1, \ldots, n$. Furthermore, due to the normalization of B, we have that

$$1 = u(s_i) = \sum_j \mathsf{B}_j(s_i) = \mathsf{B}_i(s_i) + \sum_{j \neq i} \mathsf{B}_j(s_i)$$
$$= 1 + \sum_{j \neq i} \mathsf{B}_j(s_i),$$

so that $\mathsf{B}_j(s_i) = 0$ and $s_j \neq s_i$ for all $j \neq i$ where $i = 1, \ldots, n$. Hence, B distinguishes the set of states $\{s_1, \ldots, s_n\}$.

We now note that the effects of B are the only indecomposable effects that lie on different extreme rays. Indeed, let e be any indecomposable effect and consider the dichotomic observable E with $\mathsf{E}_+ = e$. Since $\mathsf{E} \in \mathfrak{sim}(\mathsf{B})$, there exists a postprocessing μ such that $e = \sum_i \mu_{i+} \mathsf{B}_i$ so that from the indecomposability of e it follows that e is proportional to B_l for some $l \in \{1, \ldots, n\}$. Thus, there exist exactly n linearly independent extreme rays that define the generating positive cone in the $(d+1)$-dimensional effect space, and therefore we must have that $n = d + 1$.

It is straightforward to check that the states $\{s_1, \ldots, s_n\}$ are affinely independent so that $\dim(\mathrm{aff}(\{s_1, \ldots, s_n\})) = n - 1 = d = \dim(\mathrm{aff}(\mathcal{S}))$. Thus, every state $s \in \mathcal{S}$ can be expressed as an affine combination of the states $\{s_1, \ldots, s_n\}$, i.e., $s = \sum_i \gamma_i s_i$ for some $\{\gamma_i\}_i \subseteq \mathbb{R}$ such that $\sum_i \gamma_i = 1$. However, we see that

$$\gamma_j = \sum_i \gamma_i \mathsf{B}_j(s_i) = \mathsf{B}_j(s) \geqslant 0,$$

so the affine decomposition of s is actually convex, which shows that the only pure states are actually $s_1, \ldots, s_n$. Since $\mathcal{S}$ is then a convex hull of $d+1$ affinely independent (distinguishable) pure states, $\mathcal{S}$ must be a d simplex.
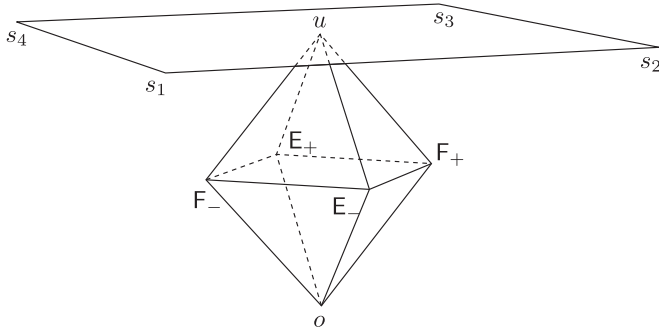
FIG. 7. Square state space (above) and the space of effects (below).

We can rephrase this result as follows.

Proposition 12. A state space is nonclassical if and only if there exist at least two inequivalent simulation irreducible observables.

B. Square bit state space

Consider a state space $\mathcal{S}_\square = \mathrm{conv}(\{s_1, s_2, s_3, s_4\})$ that is isomorphic to a square in $\mathbb{R}^2$, i.e., $s_1 + s_3 = s_2 + s_4$ (see Fig. 7). Such a state space is also referred to as the square bit state space or squit state space. The set of effects $\mathcal{E}(\mathcal{S}_\square)$ is an intersection of the positive dual cone $\mathcal{A}_+^*$ and the set $u - \mathcal{A}_+^*$, which is isomorphic to the octahedron in $\mathbb{R}^3$, Fig. 7.

In this section, we demonstrate that the set of all observables $\mathcal{O}$ on the square bit state space can be simulated from a set of two binary observables E and F defined as follows:

$$\mathsf{E}_+(s_1) = \mathsf{E}_+(s_2) = 0, \quad \mathsf{E}_+(s_3) = \mathsf{E}_+(s_4) = 1,$$
$$\mathsf{E}_-(s_1) = \mathsf{E}_-(s_2) = 1, \quad \mathsf{E}_-(s_3) = \mathsf{E}_-(s_4) = 0.$$
$$\mathsf{F}_+(s_1) = \mathsf{F}_+(s_4) = 0, \quad \mathsf{F}_+(s_2) = \mathsf{F}_+(s_3) = 1,$$
$$\mathsf{F}_-(s_1) = \mathsf{F}_-(s_4) = 1, \quad \mathsf{F}_-(s_2) = \mathsf{F}_-(s_3) = 0.$$

Since the set of effects $\{u, \mathsf{E}_+, \mathsf{F}_+\}$ is linearly independent, it follows from Corollary 2 that an observable A with outcome set Ω is contained in $\mathfrak{sim}(\{\mathsf{E}, \mathsf{F}\})$ if and only if $\mathsf{A}_x \in \mathrm{conv}(\{\mathsf{E}_+, \mathsf{E}_-, \mathsf{F}_+, \mathsf{F}_-, o, u\})$ for all $x \in \Omega$, which is always fulfilled because $\mathrm{conv}(\{\mathsf{E}_+, \mathsf{E}_-, \mathsf{F}_+, \mathsf{F}_-, o, u\}) = \mathcal{E}(\mathcal{S}_\square)$. Hence, $\mathfrak{sim}(\{\mathsf{E}, \mathsf{F}\}) = \mathcal{O}$.

The obtained result implies the following:

(1) The effective number of outcomes for any observable on the square bit state space is at most 2.

(2) Any simulation irreducible observable is postprocessing equivalent to either E or F.

(3) $\mathfrak{s}_{\min}(\mathcal{O}) = 2$.

It is known that the square bit state space possesses the feature of maximal incompatibility: There exists a pair of observables (which are actually exactly the observables E and F) such that the minimum amount of noise one has to mix them with to make their noisy versions compatible is enough to make any other pair of observables compatible in any theory [6]. In this sense, the square bit state space is even more nonclassical than any finite-dimensional quantum theory [15].

Since classical theories have only one equivalence class of simulation irreducible observables, we can argue that theories, such as square bit state space, having just two of such
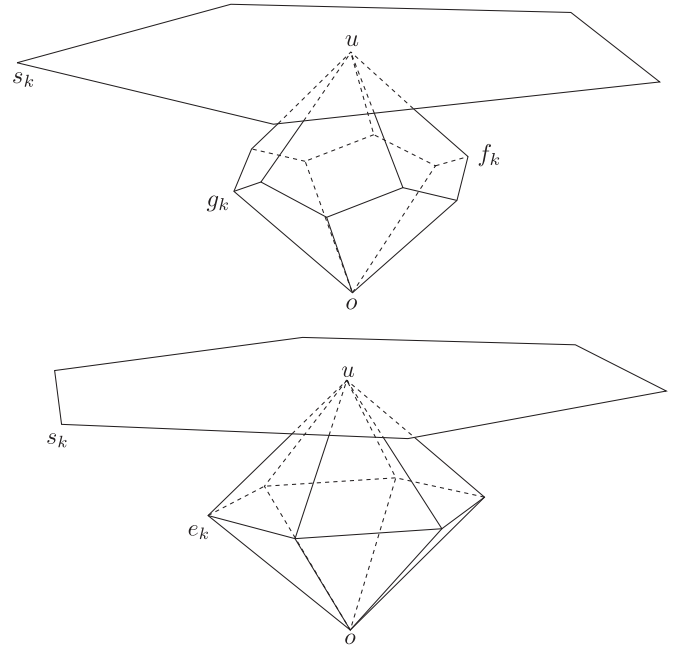

FIG. 8. Odd and even polygon state spaces $\mathcal{S}$ and corresponding sets of effects $\mathcal{E}(\mathcal{S})$.

equivalence classes are somewhat closest to classical theory. Furthermore, the effective number of all observables on this state space is the same as in the simplest and one of the most important classical theories, namely the bit. In this sense, the square state space is closest to classical theory amongst all nonclassical theories.

C. Polygon state spaces

We say that a convex set P_n is a regular n-sided polygon if there exist n vectors $\vec{p}_1, \ldots, \vec{p}_n$ in $\mathbb{R}^2$ such that $\|\vec{p}_1\| = \|\vec{p}_2\| = \cdots = \|\vec{p}_n\|$, and $\vec{p}_i \cdot \vec{p}_{i+1} = \|\vec{p}_i\|^2 \cos\left(\frac{2\pi}{n}\right)$ for all $i = 1, \ldots, n$ (where the addition is modulo n) such that P_n is isomorphic to $\mathrm{conv}(\{\vec{p}_1, \ldots, \vec{p}_n\})$. The extremal points of a polygon are its vertices, and faces are exactly the sides of the polygon; see Fig. 8.

As a state space, we consider polygons embedded in $\mathbb{R}^3$ lying on the $z = 1$ plane. A polygon state space $\mathcal{S}_n$ with n vertices is then given by the convex hull of n extremal states

$$\vec{s}_k = \begin{pmatrix} \sec\left(\frac{\pi}{n}\right) \cos\left(\frac{2k\pi}{n}\right) \\ \sec\left(\frac{\pi}{n}\right) \sin\left(\frac{2k\pi}{n}\right) \\ 1 \end{pmatrix}, \quad k = 1, \ldots, n. \quad (56)$$

As the polygons are two dimensional, the effects can also be represented as elements in $\mathbb{R}^3$. Hence, we can express each $e \in \mathcal{E}(\mathcal{S}_n)$ as $\vec{e} = (a_x, a_y, a_z)^T \in \mathbb{R}^3$. With this identification we have that $e(s) = \vec{e} \cdot \vec{s}$ for all $e \in \mathcal{E}(\mathcal{S}_n)$ and $s \in \mathcal{S}_n$, where now $\vec{e}, \vec{s} \in \mathbb{R}^3$ and $\cdot$ is the Euclidean dot product in $\mathbb{R}^3$. We omit the vector notation from here onward and simply denote the states and effects in $\mathbb{R}^3$ by s and e instead of $\vec{s}$ and $\vec{e}$. Clearly, we now have the zero effect $o = (0, 0, 0)^T$ and the unit effect $u = (0, 0, 1)^T$.

To find the positive dual cone $\mathcal{A}_+^* = \{e \mid e(s) \geqslant 0 \text{ for all } s \in \mathcal{S}_n\}$, it is enough to satisfy the requirement $e(s_k) \geqslant 0$ for

all extremal states (56). We have

$$e(s_k) = a_x \sec\left(\frac{\pi}{n}\right) \cos\left(\frac{2k\pi}{n}\right)$$
$$+ a_y \sec\left(\frac{\pi}{n}\right) \sin\left(\frac{2k\pi}{n}\right) + a_z \geqslant 0, \quad (57)$$

$k = 1, \ldots, n$. The extremal rays of the positive dual cone $\mathcal{A}_+^*$ correspond to the intersection of two adjacent planes $e(s_k) = 0$ and $e(s_{k-1}) = 0$ and have the form

$$e_k^+ = \begin{pmatrix} -a_z \cos\left(\frac{(2k-1)\pi}{n}\right) \\ -a_z \sin\left(\frac{(2k-1)\pi}{n}\right) \\ a_z \end{pmatrix}, \quad k = 1, \ldots, n, \quad a_z \geqslant 0. \quad (58)$$

Similarly, inequalities $e(s_k) \leqslant 1$, $k = 1, \ldots, n$ define the set $u - \mathcal{A}_+^*$ with extremal rays

$$e_k^- = \begin{pmatrix} b_z \cos\left(\frac{(2k-1)\pi}{n}\right) \\ b_z \sin\left(\frac{(2k-1)\pi}{n}\right) \\ 1 - b_z \end{pmatrix}, \quad k = 1, \ldots, n, \quad b_z \geqslant 0. \quad (59)$$

If n is even, then the extremal rays $e_{k+n/2}^+$ and e_k^- intersect, with the resulting nontrivial extremal effects being

$$e_k = \frac{1}{2} \begin{pmatrix} \cos\left(\frac{(2k-1)\pi}{n}\right) \\ \sin\left(\frac{(2k-1)\pi}{n}\right) \\ 1 \end{pmatrix}, \quad k = 1, \ldots, n. \quad (60)$$

If n is odd, then the rays e_k^+ and $e_{k'}^-$ do not intersect. In this case, the intersection of upward and downward cones results in two families of extremal effects. The first family corresponds to points at which $\frac{1}{2} e_{k+(n-1)/2}^+ + \frac{1}{2} e_{k+(n-1)/2-1}^+ = e_k^-$ and reads

$$f_k = \frac{1}{1 + \sec\left(\frac{\pi}{n}\right)} \begin{pmatrix} \cos\left(\frac{(2k-1)\pi}{n}\right) \\ \sin\left(\frac{(2k-1)\pi}{n}\right) \\ \sec\left(\frac{\pi}{n}\right) \end{pmatrix}, \quad k = 1, \ldots, n. \quad (61)$$

The second family corresponds to points at which $\frac{1}{2} e_{k+(n-1)/2}^- + \frac{1}{2} e_{k+(n-1)/2-1}^- = e_k^+$ and reads

$$g_k = \frac{1}{1 + \sec\left(\frac{\pi}{n}\right)} \begin{pmatrix} -\cos\left(\frac{(2k-1)\pi}{n}\right) \\ -\sin\left(\frac{(2k-1)\pi}{n}\right) \\ 1 \end{pmatrix}$$
$$= u - f_k, \quad k = 1, \ldots, n. \quad (62)$$

In this case of odd n, we note that the nontrivial extremal effects no longer lie in a single plane; see Fig. 8.

Thus, in the case of even polygon state spaces we have $\mathcal{E}(\mathcal{S}_n) = \text{conv}(\{e_1, \ldots, e_n, o, u\}) = \text{conv}(\{\mathsf{E}_\pm^{(1)}, \ldots, \mathsf{E}_\pm^{(n/2)}, o, u\})$, where we have defined the dichotomic observables $\mathsf{E}^{(i)}$ with effects $\mathsf{E}_+^{(i)} = e_i$ and $\mathsf{E}_-^{(i)} = e_{i+n/2}$, $i = 1, \ldots, \frac{n}{2}$. In the case of odd polygon state spaces, we have $\mathcal{E}(\mathcal{S}_n) = \text{conv}(\{f_1, \ldots, f_n, g_1, \ldots, g_n, o, u\}) = \text{conv}(\{\mathsf{F}_\pm^{(1)}, \ldots, \mathsf{F}_\pm^{(n)}, o, u\})$, where we have defined the dichotomic observables $\mathsf{F}^{(i)}$ with effects $\mathsf{F}_+^{(i)} = f_i$ and $\mathsf{F}_-^{(i)} = u - \mathsf{F}_+^{(i)} = g_i$, $i = 1, \ldots, n$.

The fundamental difference between the effect spaces for even and odd polygon state spaces is that, in the case of even n, to construct $\mathcal{E}(\mathcal{S}_n)$ one needs the effects of $\frac{n}{2}$ dichotomic observables (plus the zero and the unit effect), whereas in the case of odd n, one needs the effects of n dichotomic observables (plus the zero and the unit effect) to get the whole effect space $\mathcal{E}(\mathcal{S}_n)$.

However, we find that Proposition 8 has strong consequences in polygon state spaces in both even and odd cases. Namely, if A is a dichotomic observable on a polygon state space $\mathcal{S}_n$ with n vertices, then always

$$\mathsf{A}_+ \in \begin{cases} \text{conv}(\{\mathsf{E}_\pm^{(1)}, \ldots, \mathsf{E}_\pm^{(\frac{n}{2})}, o, u\}) & \text{if } n \text{ is even}, \\ \text{conv}(\{\mathsf{F}_\pm^{(1)}, \ldots, \mathsf{F}_\pm^{(n)}, o, u\}) & \text{if } n \text{ is odd}. \end{cases} \quad (63)$$

From Proposition 8, it follows that

$$\mathsf{A} \in \begin{cases} \mathfrak{sim}(\{\mathsf{E}^{(1)}, \ldots, \mathsf{E}^{(\frac{n}{2})}\}) & \text{if } n \text{ is even}, \\ \mathfrak{sim}(\{\mathsf{F}^{(1)}, \ldots, \mathsf{F}^{(n)}\}) & \text{if } n \text{ is odd}, \end{cases} \quad (64)$$

so that for the set $\mathcal{O}_\pm$ of all dichotomic observables on $\mathcal{S}_n$ we have

$$\mathfrak{s}_{\min}(\mathcal{O}_\pm) \leqslant \begin{cases} \frac{n}{2} & \text{if } n \text{ is even}, \\ n & \text{if } n \text{ is odd}. \end{cases} \quad (65)$$

Next, we will characterize the extreme simulation irreducible observables in polygon state spaces.

Proposition 13. The minimal simulation number for the set $\mathcal{O}$ of all observables on an even polygon state space $\mathcal{S}_{2m}$ equals $\mathfrak{s}_{\min}(\mathcal{O}) = m + \frac{1}{3} m(m-1)(m-2)$.

Proof. From Proposition 6, it follows that in order to find $\mathfrak{s}_{\min}(\mathcal{O})$ one merely needs to know the number of inequivalent simulation irreducible observables. By Corollary 1, it is enough to find the number of inequivalent observables A with linearly independent indecomposable effects. Since A is indecomposable, its effects belong the extreme rays of the positive effects cone, i.e., they are some positive scalar multiples of the nontrivial extremal effects e_k in (60). Furthermore, since the effects of A are linearly independent and contained in $\mathbb{R}^3$, A has at most three outcomes.

If A is dichotomic, then the only possibility is that $\mathsf{A}_+ = e_k$ and $\mathsf{A}_- = e_{k+m}$, $k = 1, \ldots, 2m$. Thus, there are $2m$ choices for the effects of A. Taking into account the bijective relabellings of outcomes, i.e., the permutations of the set $\{+, -\}$, we have $2m/2! = m$ inequivalent simulation irreducible dichotomic observables.

If A is trichotomic with effects A_1, A_2, and A_3, then $\mathsf{A}_j = c_j e_{k_j}$ for some $k_j \in \{1, \ldots, 2m\}$ and $0 < c_j \leqslant 1$ for all $j = 1, 2, 3$ such that $k_1 \neq k_2 \neq k_3 \neq k_1$. Denote $c \equiv \sum_{j=1}^3 c_j \neq 0$, and then from the normalization of A it follows that

$$\sum_{j=1}^3 \frac{c_j}{c} e_{k_j} = \frac{1}{c} u. \quad (66)$$

Since $\frac{1}{2} u$ is the only scalar multiple of u contained in the plane of nontrivial extreme effects, with necessity $c = 2$. Therefore, $\frac{1}{2} u$ must be contained in the convex hull of the extreme effects $\{e_{k_j}\}_{j=1}^3$, which limits the choices of the indices k_j. Moreover, since the convex hull of the three effects e_{k_1}, e_{k_2}, and e_{k_3} is always a simplex, the real numbers c_1, c_2, and c_3

are uniquely determined. By counting the possible indices k_j and reducing the bijective relabellings, we find that the number of inequivalent simulation irreducible trichotomic observables equals $\frac{1}{3}m(m-1)(m-2)$. For details of the combinatorics, we refer the interested reader to the appendix.

Combining the results for dichotomic and trichotomic observables concludes the proof. ∎

Proposition 14. The minimal simulation number for the set $\mathcal{O}$ of all observables on an odd polygon state space $\mathcal{S}_{2m+1}$ equals $\mathfrak{s}_{\min}(\mathcal{O}) = \frac{1}{6}m(m+1)(2m+1)$.

Proof. The proof follows from similar arguments as in the previous proposition. However, for odd polygon state spaces, there are no indecomposable dichotomic observables because the extreme rays (58) are aligned in such a way that no positive linear combination of two effects in the extreme rays can sum up to u. In other words, the complement of any indecomposable effect $c_j g_{k_j}$ does not belong to an extreme ray. For this reason, we focus on trichotomic simulation irreducible observables A with effects $\mathsf{A}_j = c_j g_{k_j}$, $j = 1, 2, 3$. Since we are interested in inequivalent observables A, the effects A_1, A_2, and A_3 are linear independent, which guarantees the uniqueness of the convex decomposition $\sum_{j=1}^{3} c_j g_{k_j} = u$. The number of such observables A is merely the number of ways to choose three points k_1, k_2, and k_3 among $2m+1$ vertices of a regular polygon with restriction that the center of the polygon belongs to the triangle $\triangle(k_1, k_2, k_3)$. The number of different ways equals $\frac{1}{6}m(m+1)(2m+1)$. For details of the combinatorics, we refer the interested reader to the appendix. ∎

Propositions 13 and 14 show that in any polygon state spaces with more than four vertices there always exists trichotomic simulation irreducible observables. Since any simulation irreducible observable can be simulated with its minimally sufficient representative, which has been shown to have at most three outcomes for polygon state spaces, we conclude that in any polygon state space with $n \geqslant 5$ vertices the effective number of outcomes for the whole space of observables $\mathcal{O}$ is exactly three.

Corollary 5. For any polygon state space $\mathcal{S}_n$ with $n \geqslant 5$, the set of all observables is effectively trichotomic, i.e., $\mathcal{O} = \mathcal{O}_3^{\text{eff}}$.

Finally, the following example illustrates the effect of noise on simulability of observables.

Example 10. Consider a hexagon state space $\mathcal{S}_6$ and a trichotomic simulation irreducible observable A with effects $\mathsf{A}_1 = \frac{2}{3} e_1$, $\mathsf{A}_2 = \frac{2}{3} e_3$, $\mathsf{A}_3 = \frac{2}{3} e_5$, where the effects e_k are given by formula (60). Obviously, A is effectively trichotomic as it is simulation irreducible. Let us show that the noisy observable A′ with effects $\mathsf{A}'_k = (1-\lambda)\mathsf{A}_k + \lambda \frac{1}{3} u$ becomes effectively dichotomic if $\frac{1}{4} \leqslant \lambda < 1$. In fact, if $\lambda = \frac{1}{4}$, then $\mathsf{A}'_1 = \frac{1}{3}(e_1 + \frac{1}{2} e_6 + \frac{1}{2} e_2)$, $\mathsf{A}'_2 = \frac{1}{3}(e_3 + \frac{1}{2} e_2 + \frac{1}{2} e_4)$, $\mathsf{A}'_3 = \frac{1}{3}(e_5 + \frac{1}{2} e_4 + \frac{1}{2} e_6)$. If this is the case, then $\mathsf{A}'_k = \frac{1}{3} \sum_{i=1}^{3} \sum_{x=\pm} \nu_{(i,x)k} \mathsf{B}^{(i)}_x$, where $\mathsf{B}^{(i)}$ is a dichotomic observable with effects $\mathsf{B}^{(i)}_+ = e_{2i-1}$ and $\mathsf{B}^{(i)}_- = e_{2i+2}$, $i = 1, 2, 3$ (addition in indices is modulo 6), $\nu_{(i,x)k}$ is the right stochastic matrix with elements $\nu_{(i,+)k} = 1$ and $\nu_{(i,-)k} = 0$ if $i = k$, $\nu_{(i,+)k} = 0$ and $\nu_{(i,-)k} = \frac{1}{2}$ if $i \neq k$. Clearly, for larger noise the observable A′ remains effectively dichotomic unless $\lambda = 1$, when the observable A′ becomes trivial.

The above example illustrates that sufficiently noisy observables can be simulated by dichotomic observables.

VII. CONCLUSIONS

Within the framework of generalized probabilistic theories, we have considered the fundamental properties of the set of observables $\mathfrak{sim}(\mathcal{B})$ that can be obtained from another set of observables $\mathcal{B}$ via mixing and postprocessing. Mathematically, the simulation map $\mathfrak{sim}(\cdot)$ is an algebraic closure operator on the set of observables. We introduced the concept of a simulation irreducible observable, which turned out to be useful in the analysis of simulability. In particular, we have shown that any observable can be simulated by a finite number of simulation irreducible ones.

The benefit of a simulation scheme is that a wide class of observables can be realized (experimentally) via a small number of simulators. We have discussed the minimal simulation number $\mathfrak{s}_{\min}(\mathcal{B})$ as an indicator of the incompatibility of a subset $\mathcal{B}$ of observables, and we pointed out its connection (in the case of quantum theory) to k compatibility of observables. Another way to benefit from a simulation scheme is that one can simulate observables with a larger number of outcomes as compared with the number of outcomes for simulators. This means that a class of observables with many outcomes can be achieved by using, e.g., dichotomic simulators, in which case we can regard those observables as effectively dichotomic.

We found that the effects of an effectively dichotomic observable have a simple geometric characterization in terms of the effects of the dichotomic simulator observables. This then serves as a useful necessary condition for dichotomic simulability when the set of available dichotomic measurement devices is fixed. We also showed that the condition becomes sufficient when we pose some additional restrictions on the simulator observables.

Finally, we have considered particular examples of non-quantum state spaces. The classical state spaces are the state spaces where there exists, up to equivalence, only one simulation irreducible observable. In general, the number of inequivalent simulation irreducible observables is a characteristic feature of a state space. We have considered even and odd polygon state spaces $\mathcal{S}_n$ in detail. In contrast to quantum theory, where there exists a continuum of inequivalent simulation irreducible observables, in any polygon state space the minimal simulation number for the set of all observables is finite. Also, we have shown that the set of all observables is effectively dichotomic for $n = 4$ and effectively trichotomic for $n \geqslant 5$. By a specific example, we have illustrated how an effectively trichotomic observable becomes effectively dichotomic under the addition of noise.

APPENDIX

1. Proof of Proposition 2

Proof. Suppose the nonzero effects $A_1, \ldots, A_n$ of an observable A are linearly dependent, i.e.,

$$\sum_{i=1}^{n} r_i A_i = 0 \tag{A1}$$

for real r_i such that $\sum_i |r_i| > 0$. This implies that

$$\sum_{i:\, r_i \geqslant 0} r_i A_i = \sum_{i:\, r_i < 0} |r_i| A_i. \tag{A2}$$

Denote $\lambda = \frac{1}{2\max_i |r_i|} > 0$ and consider two observables B and C defined as follows:

$$B_i = \begin{cases} (1 - \lambda r_i) A_i & \text{if } r_i \geqslant 0, \\ (1 + \lambda |r_i|) A_i & \text{if } r_i < 0, \end{cases} \tag{A3}$$

$$C_i = \begin{cases} (1 + \lambda r_i) A_i & \text{if } r_i \geqslant 0, \\ (1 - \lambda |r_i|) A_i & \text{if } r_i < 0. \end{cases} \tag{A4}$$

It is straightforward to see that B and C are indeed observables. Now it follows that

$$A = \tfrac{1}{2} B + \tfrac{1}{2} C. \tag{A5}$$

Therefore, A is not extreme. ∎

2. Proof of property (sim6)

Take $A, A' \in \mathfrak{sim}(\mathcal{B})$ so that there exists two finite sets of observables $\{B^{(i)}\}_{i=1}^{m}, \{B'^{(j)}\}_{j=1}^{m'} \subseteq \mathcal{B}$ with outcome sets X for $B^{(i)}$'s and X' for $B'^{(j)}$'s, probability distributions $\{p_i\}_{i=1}^{m}, \{p'_j\}_{j=1}^{m'} \subset [0,1]$, and postprocessings $\nu : \{1, \ldots, m\} \times X \to Y$ and $\nu' : \{1, \ldots, m'\} \times X' \to Y'$ for some outcome sets Y and Y' such that

$$A_y = \sum_{(i,x)} \nu_{(i,x)y} p_i B_x^{(i)}, \quad A'_{y'} = \sum_{(j,x')} \nu'_{(j,x')y'} p'_j B'^{(j)}_{x'} \tag{A6}$$

for all $y \in Y$ and $y' \in Y'$.

For any $0 \leqslant \lambda \leqslant 1$, we may form a mixture of A and A' with outcome set $Y_{\mathrm{mix}} \equiv Y \cup Y'$ so that

$$\lambda A_y + (1 - \lambda) A'_y$$
$$= \sum_{(i,x)} \nu_{(i,x)y} \lambda p_i B_x^{(i)} + \sum_{(j,x')} \nu'_{(j,x')y} (1 - \lambda) p'_j B'^{(j)}_{x'}, \tag{A7}$$

where we have also extended both postprocessings on Y_{mix} by setting $\nu_{(i,x)y} = 0$ if $y \notin Y$ and $\nu'_{(i,x)y} = 0$ if $y \notin Y'$.

We see now that we can use the observables $\{B^{(1)}, \ldots, B^{(m)}, B'^{(1)}, \ldots, B'^{(m')}\} \subseteq \mathcal{B}$ to simulate the mixture $\lambda A + (1 - \lambda) A'$. Namely, if we denote $B^{(m+i)} = B'^{(i)}$ for all $i = 1, \ldots, m'$ and consider the probability distribution $\{\tilde{p}_i\}_{i=1}^{m+m'} \equiv \{\lambda p_1, \ldots, \lambda p_m, (1 - \lambda) p'_1, \ldots, (1 - \lambda) p'_{m'}\} \subset [0,1]$, we may define the mixture observable $\tilde{B}$ with outcome set $\{1, \ldots, m + m'\} \times X_{\mathrm{mix}}$, where $X_{\mathrm{mix}} \equiv X \cup X'$, by

$$\tilde{B}_{(i,x)} = \tilde{p}_i B_x^{(i)} \tag{A8}$$

for all $i = 1, \ldots, m + m'$ that keeps track of the measured observable. Similarly, we can define a postprocessing $\mu :$ $\{1, \ldots, m + m'\} \times X_{\mathrm{mix}} \to Y_{\mathrm{mix}}$ by

$$\mu_{(i,x)y} = \chi_{\{1,\ldots,m\}}(i) \nu_{(i,x)y} + \chi_{\{m+1,\ldots,m+m'\}}(i) \nu'_{(i-m,x)y}, \tag{A9}$$

where χ_S is the characteristic function of a set $S \subset \mathbb{Z}$ so that $\chi_S(x) = 1$ if $x \in S$ and $\chi_S(x) = 0$ otherwise. Now

$$(\mu \circ \tilde{B})_y = \sum_{(i,x)} \mu_{(i,x)y} \tilde{B}_{(i,x)}$$
$$= \sum_{i=1}^{m} \sum_x \nu_{(i,x)y} \tilde{B}_{(i,x)} + \sum_{i=m+1}^{m+m'} \sum_x \nu'_{(i-m,x)y} \tilde{B}_{(i,x)}$$
$$= \sum_{i=1}^{m} \sum_x \nu_{(i,x)y} \lambda p_i B_x^{(i)}$$
$$\quad + \sum_{j=1}^{m'} \sum_x \nu'_{(j,x)y} (1 - \lambda) p'_j B'^{(j)}_x$$
$$= \lambda A_y + (1 - \lambda) A'_y$$

for all $y \in Y_{\mathrm{mix}}$ so that $\lambda A + (1 - \lambda) A' \in \mathfrak{sim}(\mathcal{B})$, which shows that $\mathfrak{sim}(\mathcal{B})$ is convex.

3. Proof of property (sim7)

Take $A \in \mathfrak{sim}(\mathcal{B})$ with an outcome set Y so that

$$A_y = \sum_{(i,x)} \nu_{(i,x)y} p_i B_x^{(i)} \tag{A10}$$

for all $y \in Y$, some finite set of observables $\{B^{(i)}\}_i \subseteq \mathcal{B}$ with outcome sets X, some probability distribution $\{p_i\}_i \subset [0,1]$, and some postprocessing $\nu : \cup_k \{k\} \times X \to Y$. If now $\mu : Y \to Z$ is a postprocessing from Y to some outcome set Z, then

$$(\mu \circ A)_z = \sum_y \mu_{yz} A_y$$
$$= \sum_y \mu_{yz} \left(\sum_{(i,x)} \nu_{(i,x)y} p_i B_x^{(i)} \right)$$
$$= \sum_{(i,x)} \left(\sum_y \nu_{(i,x)y} \mu_{yz} \right) p_i B_x^{(i)}$$
$$= \sum_{(i,x)} \eta_{(i,x)z} p_i B_x^{(i)},$$

where we have defined the postprocessing $\eta : \cup_k \{k\} \times X \to Z$ by $\eta_{(i,x)z} = \sum_y \nu_{(i,x)y} \mu_{yz}$ for all i, $x \in X$, and $z \in Z$. Thus, $\mu \circ A \in \mathfrak{sim}(\mathcal{B})$.

4. Combinatorics in proof of Proposition 13

When choosing effects $A_1 = c_1 e_{k_1}$, $A_2 = c_2 e_{k_2}$, $A_3 = c_3 e_{k_3}$, we cannot have $k_l = k_j + m$ for any $l \neq j$, since then from the decomposition $u = e_{k_j} + e_{k_j+m}$ it would follow that the remaining effect $A_i = c_i e_{k_i}$, $i \neq j \neq l \neq i$, is decomposable. Second, we cannot have $k_l = k_j \pm 1$ for any $l \neq j$ since this would force the remaining index k_i, $i \neq l \neq j \neq i$, to be either $k_i = k_j + m$ or $k_i = k_j \pm 1 + m$ in order for (66) to hold, which in turn would lead to a violation of the previous case. Thus, by considering possible cases for the indices k_j, $j = 1,2,3$, such that (66) holds, we see that the problem reduces to a simple problem of combinatorics:

(i) We can choose the effect A_1 to be proportional to any nontrivial extreme effect e_i, where $i \in \{1, \ldots, 2m\}$ so that A_1 has $2m$ possibilities.

(ii) For A_2 there are $2m - 4$ possibilities since A_2 cannot be proportional to e_{i-1}, e_i, e_{i+1}, or e_{i+m}. Thus, we have that A_2 is proportional to e_j, where either $j \in \{i+2, \ldots, i+m-1\}$ or $j \in \{i+m+1, \ldots, i+2m-2\}$ so that j has $m-2$ possibilities in both of these cases.

(iii) If $j \in \{i+2, \ldots, i+m-1\}$, the only possibility for A_3 is to be proportional to an effect e_k which is limited to be in some of the extreme rays between the complements of e_i and e_j since otherwise the convex hull of $\{e_i, e_j, e_k\}$ would not contain $u/2$. Thus, $k \in \{i+m+1, \ldots, j+m-1\}$ and since $j = i + l$ for some $l \in \{2, \ldots, m-1\}$ we have that k has a total of $l - 1$ possibilities. By the same argument, in the case when $j \in \{i+m+1, \ldots, i+2m-2\}$, we still have $l-1$ different possibilities, where again each l represents different j from (ii).

Now we can calculate the total number of different cases. As shown above, for A_1 we have $2m$ possibilities, and then for A_2 and A_3 there are

$$2 \sum_{l=2}^{m-1} (l-1) = 2 \sum_{l'=1}^{m-2} l'$$
$$= 2 \frac{(m-2)(m-1)}{2} = (m-1)(m-2) \quad \text{(A11)}$$

different possibilities, where the multiplier 2 came from two different sets of values for j in (ii). In order to not to include any bijective relabellings of the effects of A, we have to take into account the different permutations of the set $\{1, 2, 3\}$. Hence, the total number of inequivalent simulation irreducible trichotomic observable equals

$$\frac{2m(m-1)(m-2)}{3!} = \frac{m(m-1)(m-2)}{3}. \quad \text{(A12)}$$

5. Combinatorics in proof of Proposition 14

Effects $\mathsf{A}_1 = c_1 g_{k_1}$, $\mathsf{A}_2 = c_2 g_{k_2}$, $\mathsf{A}_3 = c_3 g_{k_3}$ can be chosen as follows:

(i) A_1 is proportional to one of the nontrivial extreme effects f_i, where $i \in \{1, \ldots, 2m+1\}$ so that for A_1 we have $2m+1$ possibilities.

(ii) For A_2 there are $2m$ possibilities since A_2 cannot be proportional to g_i. Thus, we have that A_2 is proportional to g_j, where either $j \in \{i+1, \ldots, i+m\}$ or $j \in \{i+m+1, \ldots, i+2m\}$ so that j has m possibilities in both of these cases.

(iii) If $j \in \{i+1, \ldots, i+m\}$, the only possibility for A_3 is to be proportional to an effect g_k with $k \in \{i+m+1, \ldots, j+m\}$ and since $j = i + l$ for some $l \in \{1, \ldots, m\}$ we have that k has a total of l possibilities. By the same argument, in the case when $j \in \{i+m+1, \ldots, i+2m\}$, we still have l different possibilities, where again each l represents different j from (ii).

From this, we can calculate the total number of different cases. As shown above, for A_1 we have $2m+1$ possibilities and then for A_2 and A_3, there are

$$2 \sum_{l=1}^{m} l = 2 \frac{m(m+1)}{2} = m(m+1) \quad \text{(A13)}$$

different possibilities, where the multiplier 2 came from two different sets of values for j in (ii). In order to not to include any bijective relabellings of the effects of A, we have to take into account the different permutations of the set $\{1, 2, 3\}$. Hence, the total number of inequivalent simulation irreducible trichotomic observable equals

$$\frac{(2m+1)m(m+1)}{3!}. \quad \text{(A14)}$$

[1] L. Guerini, J. Bavaresco, M. T. Cunha, and A. Acín, Operational framework for quantum measurement simulability, J. Math. Phys. **58**, 092102 (2017).

[2] M. Oszmaniec, L. Guerini, P. Wittek, and A. Acín, Simulating Positive-Operator-Valued Measures with Projective Measurements, Phys. Rev. Lett. **119**, 190501 (2017).

[3] F. Hirsch, M. T. Quintino, T. Vértesi, M. Navascués, and N. Brunner, Better local hidden variable models for two-qubit Werner states and an upper bound on the Grothendieck constant $K_G(3)$, Quantum **1**, 3 (2017).

[4] M. Kleinmann, T. Vértesi, and A. Cabello, Proposed experiment to test fundamentally binary theories, Phys. Rev. A **96**, 032104 (2017).

[5] M. Kleinmann and A. Cabello, Quantum Correlations are Stronger than all Nonsignaling Correlations Produced by n-Outcome Measurements, Phys. Rev. Lett. **117**, 150401 (2016).

[6] P. Busch, T. Heinosaari, J. Schultz, and N. Stevens, Comparing the degrees of incompatibility inherent in probabilistic physical theories, Europhys. Lett. **103**, 10002 (2013).

[7] N. Stevens and P. Busch, Steering, incompatibility, and Bell-inequality violations in a class of probabilistic theories, Phys. Rev. A **89**, 022123 (2014).

[8] M. Banik, Measurement incompatibility and Schrödinger-Einstein-Podolsky-Rosen steering in a class of probabilistic theories, J. Math. Phys. **56**, 052101 (2015).

[9] M. Plávala, All measurements in a probabilistic theory are compatible if and only if the state space is a simplex, Phys. Rev. A **94**, 042108 (2016).

[10] S. N. Filippov, T. Heinosaari, and L. Leppäjärvi, Necessary condition for incompatibility of observables in general probabilistic theories, Phys. Rev. A **95**, 032127 (2017).

[11] A. Jenčová and M. Plávala, Conditions on the existence of maximally incompatible two-outcome measurements in general probabilistic theory, Phys. Rev. A **96**, 022113 (2017).

[12] P. Janotta, C. Gogolin, J. Barrett, and N. Brunner, Limits on nonlocal correlations from the structure of the local state space, New J. Phys. **13**, 063024 (2011).

[13] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Found. Phys. **24**, 379 (1994).

[14] D. Gross, M. Müller, R. Colbeck, and O. C. O. Dahlsten, All Reversible Dynamics in Maximally Nonlocal Theories are Trivial, Phys. Rev. Lett. **104**, 080402 (2010).

[15] T. Heinosaari, J. Schultz, A. Toigo, and M. Ziman, Maximally incompatible quantum observables, Phys. Lett. A **378**, 1695 (2014).

[16] R. T. Rockafellar, *Convex Analysis* (Princeton University Press, Princeton, NJ, 1970).

[17] G. Kimura, K. Nuida, and H. Imai, Distinguishability measures and entropies for general probabilistic theories, Rep. Math. Phys. **66**, 175 (2010).

[18] H. Martens and W. M. de Muynck, Nonideal quantum measurements, Found. Phys. **20**, 255 (1990).

[19] F. Buscemi, G. M. D'Ariano, M. Keyl, P. Perinotti, and R. F. Werner, Clean positive operator valued measures, J. Math. Phys. **46**, 082109 (2005).

[20] Y. Kuramochi, Minimal sufficient positive-operator valued measure on a separable Hilbert space, J. Math. Phys. **56**, 102205 (2015).

[21] E. Haapasalo, T. Heinosaari, and J.-P. Pellonpää, Quantum measurements on finite dimensional systems: Relabeling and mixing, Quantum Inf. Process. **11**, 1751 (2012).

[22] K. R. Parthasarathy, Extremal decision rules in quantum hypothesis testing, Infin. Dimens. Anal. Quantum Probab. Relat. Top. **2**, 557 (1999).

[23] R. Kunjwal, C. Heunen, and T. Fritz, Quantum realization of arbitrary joint measurability structures, Phys. Rev. A **89**, 052126 (2014).

[24] T. Heinosaari, D. Reitzner, and P. Stano, Notes on joint measurability of quantum observables, Found. Phys. **38**, 1133 (2008).

[25] C. Carmeli, T. Heinosaari, D. Reitzner, J. Schultz, and A. Toigo, Quantum incompatibility in collective measurements, Mathematics **4**, 54 (2016).

[26] P. Busch, Informationally complete sets of physical quantities, Int. J. Theor. Phys. **30**, 1217 (1991).

[27] M. Singer and W. Stulpe, Phase-space representations of general statistical physical theories, J. Math. Phys. **33**, 131 (1992).