

FACTORIZATION OF MULTIMETERS

—A UNIFIED VIEW ON NONCLASSICAL QUANTUM PHENOMENA—

arxiv2504.19865

TIM ACHENBACH, ANDREAS BLUHM, LEEVI LEPPÄJÄRVI, ION NECHITA, AND MARTIN PLÁVALA

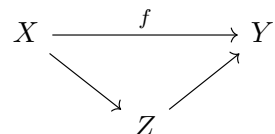
Main idea: Nonclassicality through factorization

Quantum theory's departure from classical physics is marked by a host of nonclassical features that can be classified as *local*, meaning that they involve only preparations, transformations, and measurements involving a single quantum system, or as *nonlocal*, meaning that they require at least two or more spatially separated quantum systems. Examples of local nonclassical features of quantum theory are the uncertainty relations [Hei27, Ken27, Rob29, Sch35], no-cloning and no-broadcasting theorems [WZ82, BCF⁺96, BBLW07], incompatibility of measurements [HMZ16, GHK⁺23], and contextuality as defined by Kochen and Specker [KS90, BCG⁺22] or as defined by Spekkens [Spe05]. Nonlocal nonclassical features of quantum theory are entanglement [GT09], steering [UCNG20], Bell nonlocality [BCP⁺14], and nonlocality without inputs in quantum networks [TPKLR22]. These phenomena are not only of foundational importance but also serve as essential resources for quantum technologies, from computation to cryptography. While traditionally studied as separate topics, they share deep interconnections. This work introduces a single, unifying mathematical framework that reveals many of these phenomena to be different manifestations of the same core principle: **factorizability**.

The central idea is to represent a collection of g quantum measurements each having k outputs as a single mathematical map, that we call a *multimeter*. This map takes a quantum state as input and produces a matrix of measurement outcome probabilities as output.

$$M : \varrho \mapsto M(\varrho) = (\text{Tr}[M_{a|x} \varrho])_{a \in [k], x \in [g]}.$$

With this representation, we can ask a simple yet powerful question: can this map be broken down into a sequence of simpler steps? Specifically, can a map $f : X \rightarrow Y$ be *factored* through an intermediate system Z , such that f is equivalent to first mapping $X \rightarrow Z$ and then $Z \rightarrow Y$?



The main appeal of the concept of factorization lies in the fact that the intermediate space Z can have a simpler structure than that of the source and target spaces X, Y . This fact has fundamental structural implications for the map f . By applying this framework to the maps representing quantum measurements, we transform profound physical questions about classical simulability and the nature of quantum resources into clean, structural questions about mathematical maps. This approach not only streamlines proofs of known results but also uncovers new connections and provides a powerful, intuitive toolkit for the quantum information community. The topic is of central importance to quantum information theory, as it provides a universal language to analyze the very resources that power the quantum advantage, making the results and techniques presented here relevant for physicists, computer scientists, and mathematicians alike.

Results: A new framework

The main conceptual contribution of our work is the realization that factorization is more than a mathematical curiosity; it is the natural language for describing simulation and resource reduction. When we ask if a complex quantum process can be simulated by a simpler one, we are fundamentally asking a question about factorization. We argue this fact by a series of results framing several fundamental incarnations of quantumness using the factorization toolkit we introduce.

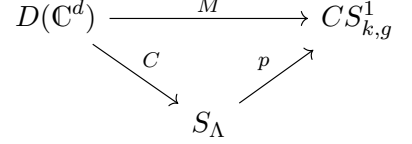
E-mail addresses: tim.n.achenbach@jyu.fi, andreas.bluhm@univ-grenoble-alpes.fr, leevi.i.leppajarvi@jyu.fi, nechita@irsamc.ups-tlse.fr, martin.plavala@itp.uni-hannover.de.

Measurement Incompatibility. The quintessential example is *measurement incompatibility*. Recall that a multimeter M is called *compatible* if there exists a single POVM C with Λ outcomes and a conditional probability distribution $p = (p_{\cdot|x,\lambda})$ on $[k]$ such that

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} p_{a|x,\lambda} C_{\lambda}.$$

We cast the compatibility question in our framework by establishing the equivalence of the following two statements:

- (1) The multimeter M consists of compatible measurements.
- (2) The channel $M : D(\mathbb{C}^d) \rightarrow CS_{k,g}^1$ can be factorized as $M = p \circ C$ for some single POVM $C : D(\mathbb{C}^d) \rightarrow S_{\Lambda}$ and a classical postprocessing $p : S_{\Lambda} \rightarrow CS_{k,g}^1$.



Above, $D(\mathbb{C}^d)$ denotes the set of density matrices, S_{Λ} is the probability simplex with Λ outcomes, while $CS_{k,g}^1$ is the *polysimplex*, the range of arbitrary multimeters consisting of g measurements each having k outcomes. Mathematically, $CS_{k,g}^1$ is the set of $k \times g$ column-stochastic matrices. With the help of the polysimplex, we can give an additional equivalent formulation of measurement compatibility using the language of tensor product of cones [Rya02, NB10, Pis20, vDdB20, Plá23]:

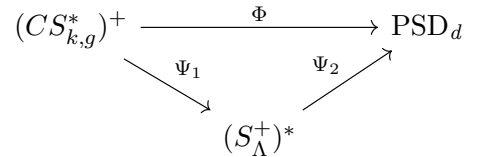
- (3) The multimeter $M \in \text{PSD}_d \hat{\otimes} CS_{k,g}^+$, seen as a tensor, is *separable*, i.e. $M \in \text{PSD}_d \dot{\otimes} CS_{k,g}^+$, where $\dot{\otimes}$ denotes the minimal, $\hat{\otimes}$ the maximal tensor product of cones and PSD_d the cone of positive semidefinite matrices.

The equivalent statements above rephrase a physical property (compatibility) into a clear structural property of the multimeter, seen either as a map or as a tensor. Going beyond (in-)compatibility, this perspective seamlessly extends to more general notions of *simulability* [OGWA17, GBTCA17, OMP19, FHL18, ISD⁺22, JUC⁺23, VPLU24, JEP24, BLN25], where one asks if a set of measurements can be simulated using fewer measurements, fewer outcomes, or measurements on a smaller-dimensional quantum system. On the left, the multimeter M having k measurements and g outcomes can be simulated as a multimeter N having l measurements and r outcomes followed by a classical post-processing p . On the right, M can be simulated by compression channels $\Phi_{\lambda} : D(\mathbb{C}^d) \rightarrow D(\mathbb{C}^{d'})$ followed by multimeters $N_{\lambda} : D(\mathbb{C}^{d'}) \rightarrow CS_{k,g}^1$, where λ is a label for classical information that can be used in the compression scheme.



Steering and Bell Nonlocality. The framework naturally extends from the single-system scenario of incompatibility to the bipartite settings of *steering* and *Bell nonlocality*. An assemblage $\sigma = (M_A \otimes \text{id}_B)(\rho_{AB}) \in CS_{k,g}^1 \hat{\otimes} D(\mathbb{C}^{d_B})$ admits a classical *Local Hidden State (LHS) model* if and only if the tensor describing it is separable $\sigma \in CS_{k,g}^1 \dot{\otimes} D(\mathbb{C}^{d_B})$.

Given an assemblage σ we can find an associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow \text{PSD}_d$ with $\Phi(\mathbb{1}_{CS_{k,g}^1}) = \sigma$. If σ admits an LHS model, then there are Ψ_1 and Ψ_2 , s.t. Φ factors through a simplex.



Due to the known relation between steering and measurement incompatibility [UBGP15] we can immediately transport all results about incompatibility to steering with the steering assemblage σ replacing the multimeter M . In quantum theory any steering assemblage can be obtained by measuring a bipartite state [Gis89, HJW93]. While the same is known to not hold in GPTs [BGW13, SB14], we do provide an explicit counterexample.

The same logic holds for Bell nonlocality: a set of correlations $(M \otimes N)(\varrho) \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ has a *Local Hidden Variable (LHV) model* if and only if it is separable $(M \otimes N)(\varrho) \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$. This demonstrates that the classicality of these distinctly nonlocal phenomena is governed by the same underlying mathematical structure.

Let $P \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ be some correlations, i.e a no-signaling distribution. We have an associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow CS_{l,r}^+$ with $\Phi(\mathbb{1}_{CS_{k,g}^1}) \in CS_{l,r}^1$, that factors through a simplex using maps Ψ_1 and Ψ_2 if and only if P admits an LHV model.

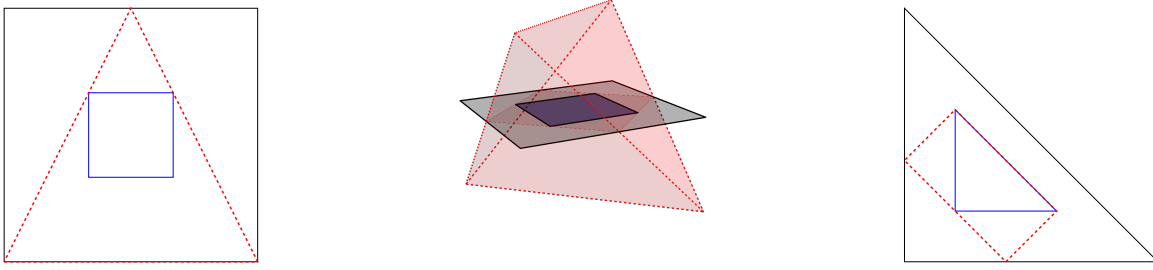
$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & CS_{l,r}^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & S_{\Lambda}^+ & \end{array}$$

A New Geometric Perspective on the CHSH Inequality. A key result of our work is a novel and intuitive perspective on the celebrated CHSH inequality. We use the formalism of multimeters to provide streamlined proofs of known results, such as that block-positive operators do not provide post-quantum behaviors [BBB⁺10], that in quantum theory any pair of dichotomic measurements violates the CHSH inequality if and only if it is incompatible [WPGF09], and that if one party measures a pair of dichotomic measurements while the other applies an arbitrary finite multimeter, then all Bell inequalities are just post-processings of the CHSH inequality [Pir14]. To prove the latter two results we will use another key idea also observed in [Jen18]: the CHSH inequalities can be identified with isomorphisms of a square via isomorphisms between tensors and linear maps. These results provide a powerful conceptual simplification of the structure of bipartite correlations.

Results: Symmetric extensions and Noise robustness

Our framework serves not just as a tool to further explore known facets of non-classicality, but also as an engine for the discovery of new manifestations and limitations of quantumness. We investigate *symmetric n -extensions* [DPS04, AMHP25] of multimeters $M \in \text{PSD}_d \hat{\otimes} CS_{k,g}^+$ and no-signaling behaviors $P \in CS_{k,g}^1 \hat{\otimes} CS_{k,g}^1$. These extensions are now tensors in $\text{PSD}_d \hat{\otimes} (CS_{k,g}^+)^{\hat{\otimes} n}$ and in $CS_{k,g}^1 \hat{\otimes} (CS_{k,g}^1)^{\hat{\otimes} n}$, respectively, such that by applying a symmetric reduction map $\gamma_n^\Phi : (CS_{k,g}^+)^{\hat{\otimes} n} \rightarrow CS_{k,g}^+$ the extension reduces to the original multimeter M and the original no-signaling behavior P , respectively. We find that for both M and P the existence of an g -extension is equivalent to the tensor being separable, in which case M is compatible and P has an LHV model. On the other hand, for $n < g$, the existence of an n -extension for M is equivalent to the fact that *every n -subset of measurements in M being compatible with the joint measurements satisfying particular no-signaling constraints*. Similarly for P having an n -extension is equivalent to P having restricted LHV models, i.e. LHV models for all the cases where the second party is restricted to only performing some n measurements. Again these restricted LHV models must satisfy some particular no-signaling constraints.

Our framework is naturally suited to analyzing the resilience of quantum resources [VT99, HKR15, DFK19]. We investigate the robustness of factorizability to noise: it is known that if a certain amount of white noise is added to the system, incompatible measurements become compatible, steering assemblages start having local hidden state models, and behaviors start having hidden variable models. We inspect factorization conditions for maps mixed with noise to get geometric criteria for when noisy multimeters can be simulated by multimeters with simpler measurement settings. This analysis is performed at the level of the classical polysimplices $CS_{k,g}^1$, making it suitable for applications to other non-classical phenomena. Mathematically, it amounts to finding embeddings of the simulating polysimplex between two scaled copies of the simulated polysimplex. Below are conjectured optimal embeddings for simulating 2 noisy binary measurements by a 3-outcome measurement (left) and a 4-outcome measurement (center), and that of a noisy 3-outcome measurement by a pair of binary measurements (right).



Conclusion: A new toolkit for non-classicality

This work presents a unified framework that recasts a wide range of nonclassical quantum phenomena as problems of map factorization. By translating physical questions into the natural language of maps and their compositional structure, our work provides a toolkit that is of broad interest to the entire quantum information theory community. We believe that the approach outlined in our work can be useful to further analyze known manifestations of non-classicality, their fundamental properties and relations, their robustness to various noise models, but also to uncover new such non-classical phenomena.

REFERENCES

- [AMHP25] Guillaume Aubrun, Alexander Müller-Hermes, and Martin Plávala. Monogamy of entanglement between cones. *Mathematische Annalen*, 391(1):1591–1609, 2025.
- [BBB⁺10] Howard Barnum, Salman Beigi, Sergio Boixo, Matthew B. Elliott, and Stephanie Wehner. Local quantum measurement and no-signaling imply quantum correlations. *Physical Review Letters*, 104(14), 2010.
- [BBLW07] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Generalized no-broadcasting theorem. *Physical Review Letters*, 99(24):240501, 2007.
- [BCF⁺96] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Non-commuting mixed states cannot be broadcast. *Physical Review Letters*, 76(15):2818, 1996.
- [BCG⁺22] Costantino Budroni, Adán Cabello, Otfried Gühne, Matthias Kleinmann, and Jan-Åke Larsson. Kochen-Specker contextuality. *Reviews of Modern Physics*, 94(4):045007, 2022.
- [BCP⁺14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of Modern Physics*, 86:419–478, 2014.
- [BGW13] Howard Barnum, Carl Philipp Gaebler, and Alexander Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories. *Foundations of Physics*, 43:1411–1427, 2013.
- [BLN25] Andreas Bluhm, Leevi Leppäjärvi, and Ion Nechita. On the simulation of quantum multimeters. *Quantum*, 9:1608, 2025.
- [DFK19] Sébastien Designolle, Máté Farkas, and Jędrzej Kaniewski. Incompatibility robustness of quantum measurements: a unified framework. *New Journal of Physics*, 21(11):113053, 2019.
- [DPS04] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2):022308, 2004.
- [FHL18] Sergey N. Filippov, Teiko Heinosaari, and Leevi Leppäjärvi. Simulability of observables in general probabilistic theories. *Physical Review A*, 97(6):062102, 2018.
- [GBTCA17] Leonardo Guerini, Jessica Bavaresco, Marcelo Terra Cunha, and Antonio Acín. Operational framework for quantum measurement simulability. *Journal of Mathematical Physics*, 58:092102, 2017.
- [GHK⁺23] Otfried Gühne, Erkkä Haapasalo, Tristan Kraft, Juha-Pekka Pellonpää, and Roope Uola. Colloquium: Incompatible measurements in quantum information science. *Reviews of Modern Physics*, 95(1):011003, 2023.
- [Gis89] Nicolas Gisin. Stochastic quantum dynamics and relativity. *Helvetica Physica Acta*, 62:363–371, 1989.
- [GT09] Otfried Gühne and Géza Tóth. Entanglement detection. *Physics Reports*, 474(1-6):1–75, 2009.
- [Hei27] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3):172–198, 1927.
- [HJW93] Lane P. Hughston, Richard Jozsa, and William K. Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183(1):14–18, 1993.
- [HKR15] Teiko Heinosaari, Jukka Kiukas, and Daniel Reitzner. Noise robustness of the incompatibility of quantum measurements. *Physical Review A*, 92:022115, 2015.
- [HMZ16] Teiko Heinosaari, Takayuki Miyadera, and Mário Ziman. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical*, 49(12):123001, 2016.
- [ISD⁺22] Marie Ioannou, Pavel Sekatski, Sébastien Designolle, Benjamin D. M. Jones, Roope Uola, and Nicolas Brunner. Simulability of high-dimensional quantum measurements. *Physical Review Letters*, 129:190401, 2022.
- [Jen18] Anna Jenčová. Incompatible measurements in a class of general probabilistic theories. *Physical Review A*, 98(1):012133, 2018.

- [JEPU24] Pauli Jokinen, Sophie Egelhaaf, Juha-Pekka Pellonpää, and Roope Uola. Compressing continuous variable quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 57:325302, 2024.
- [JUC⁺23] Benjamin D. M. Jones, Roope Uola, Thomas Cope, Marie Ioannou, Sébastien Designolle, Pavel Sekatski, and Nicolas Brunner. Equivalence between simulability of high-dimensional measurements and high-dimensional steering. *Physical Review A*, 107:052425, 2023.
- [Ken27] Earle H Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Zeitschrift für Physik*, 44(4):326–352, 1927.
- [KS90] Simon Kochen and Ernst P. Specker. The problem of hidden variables in quantum mechanics. *Ernst Specker Selecta*, pages 235–263, 1990.
- [NB10] L. Narici and E. Beckenstein. *Topological Vector Spaces*. CRC Press, 2010.
- [OGWA17] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín. Simulating positive-operator-valued measures with projective measurements. *Physical Review Letters*, 119:190501, 2017.
- [OMP19] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała. Simulating all quantum measurements using only projective measurements and postselection. *Physical Review A*, 100:012351, 2019.
- [Pir14] Stefano Pironio. All Clauser–Horne–Shimony–Holt polytopes. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424020, 2014.
- [Pis20] Gilles Pisier. *Tensor Products of C^* -Algebras and Operator Spaces: The Connes–Kirchberg Problem*. Cambridge University Press, 2020.
- [Plá23] Martín Plávala. General probabilistic theories: An introduction. *Physics Reports*, 1033:1–64, 2023.
- [Rob29] Howard Percy Robertson. The uncertainty principle. *Physical Review*, 34(1):163, 1929.
- [Rya02] Raymond A Ryan. *Introduction to tensor products of Banach spaces*. Springer, 2002.
- [SB14] Neil Stevens and Paul Busch. Steering, incompatibility, and Bell-inequality violations in a class of probabilistic theories. *Physical Review A*, 89(2):022123, 2014.
- [Sch35] Erwin Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Die Naturwissenschaften*, 48:23, 1935.
- [Spe05] Robert W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A—Atomic, Molecular, and Optical Physics*, 71(5):052108, 2005.
- [TPKLR22] Armin Tavakoli, Alejandro Pozas-Kerstjens, Ming-Xing Luo, and Marc-Olivier Renou. Bell nonlocality in networks. *Reports on Progress in Physics*, 85(5):056001, 2022.
- [UBGP15] Roope Uola, Costantino Budroni, Otfried Gühne, and Juha-Pekka Pellonpää. One-to-one mapping between steering and joint measurability problems. *Physical Review Letters*, 115(23):230402, 2015.
- [UCNG20] Roope Uola, Ana C. S. Costa, H. Chau Nguyen, and Otfried Gühne. Quantum steering. *Reviews of Modern Physics*, 92:015001, 2020.
- [vDdB20] Josse van Dobben de Bruyn. Tensor products of convex cones, part II: Closed cones in finite-dimensional spaces. *arXiv preprint arXiv:2009.11843*, 2020.
- [VPLU24] Isadora Veeren, Martín Plávala, Leevi Leppäjärvi, and Roope Uola. Semi-device-independent certification of the number of measurements. *Physical Review A*, 109(6):062203, 2024.
- [VT99] Guifré Vidal and Rolf Tarrach. Robustness of entanglement. *Physical Review A*, 59(1):141, 1999.
- [WPGF09] Michael M. Wolf, David Pérez-García, and Carlos Fernández. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Physical Review Letters*, 103:230402, 2009.
- [WZ82] William K. Wootters and Wojciech H. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982.

FACTORIZATION OF MULTIMETERS: A UNIFIED VIEW ON NONCLASSICAL QUANTUM PHENOMENA

TIM ACHENBACH, ANDREAS BLUHM, LEEVI LEPPÄJÄRVI, ION NECHITA, AND MARTIN PLÁVALA

ABSTRACT. Quantum theory exhibits various nonclassical features, such as measurement incompatibility, contextuality, steering, and Bell nonlocality, which distinguish it from classical physics. These phenomena are often studied separately, but they possess deep interconnections. This work introduces a unified mathematical framework based on commuting diagrams that unifies them. By representing collections of measurements (multimeters) as maps to the set of column-stochastic matrices, we show that measurement compatibility and simulability correspond to specific factorizations of these maps through intermediate systems. We apply this framework to put forward connections between different nonclassical notions and provide factorization-based characterizations for steering assemblages and Bell correlations, including a new perspective on the CHSH inequality witnessing measurement incompatibility. We also investigate the symmetric n -extensions of multimeters and no-signaling behaviors and connect these extensions to a notion of n -wise compatibility and to the existence of n -wise LHV models, respectively. Furthermore, we investigate robustness to noise of nonlocal features by examining factorization conditions for maps involving noisy state spaces, providing geometric criteria for when noisy multimeters can be simulated by simpler measurement settings.

CONTENTS

1. Introduction	2
2. Main results	3
3. Preliminaries	5
3.1. The formalism of general probabilistic theories	6
3.2. Positive maps and channels between GPTs	7
3.3. Meters and instruments in GPTs	8
3.4. State spaces of column stochastic matrices	10
3.5. Multimeters as channels	10
4. Compatibility and simulation of multimeters	11
4.1. Simulating multimeters on one state space by multimeters on another state space	11
4.2. Classical simulation of multimeters	13
5. Steering	19
5.1. Steering in quantum mechanics and general probabilistic theories	19
5.2. Assemblages as measurements on bipartite states	21
5.3. Factorizations of maps preparing ensembles	25
6. Bell nonlocality	30
6.1. Bell nonlocality in quantum theory	30
6.2. CHSH inequalities as incompatibility witnesses	33
6.3. Bell inequalities in general probabilistic theories	38
7. Separability and extendability	43
7.1. Extendability of measurements	45
7.2. Extendability of no-signaling behaviors	53
8. Robustness results	56
8.1. Compatibility	57
8.2. Two noisy binary measurements simulated by a three-outcome measurement	60
8.3. A noisy three-outcome measurement simulated by two dichotomic measurements	61

9. Concluding remarks	62
Acknowledgments	63
References	63
Appendix A. Cones, tensor products, and positive maps	66

1. INTRODUCTION

There are several nonclassical features of quantum theory that are of foundational importance since they illustrate the departure from classical physics and classical probability theory. These nonclassical features can be classified as local, meaning that they involve only preparations, transformations, and measurements involving a single quantum system, and as nonlocal, meaning that they require at least two or more spatially separated quantum systems. Examples of local nonclassical features of quantum theory are the uncertainty relations [Hei27, Ken27, Rob29, Sch35], no-cloning and no-broadcasting theorems [WZ82, BCF⁺96, BBLW07], incompatibility of measurements [HMZ16, GHK⁺23], and contextuality as defined by Kochen and Specker [KS90, BCG⁺22] or as defined by Spekkens [Spe05]. Nonlocal nonclassical features of quantum theory are entanglement [GT09], steering [UCNG20], Bell nonlocality [BCP⁺14], and nonlocality without inputs in quantum networks [TPKLR22].

The various nonclassical features also serve as the backbone of its applications. One can trace contextuality as the source of the quantum advantage present in quantum computers [HWVE14] and no-broadcasting is necessary for quantum cryptography. Steering and Bell nonlocality are necessary resources for semi-device-independent and device-independent quantum cryptography [Wol21], the latter of which was recently demonstrated experimentally [ZvLR⁺22, NDN⁺22]. In nonlocal phenomena as well as in communication tasks making use of quantum protocols, e.g., quantum random access codes, incompatibility is a necessary condition for observation of the phenomena or usefulness of the protocol.

Some of these concepts gave rise to further generalizations with their own operational interpretations. Spekkens contextuality was generalized to test whether a system is quantum [MG23]. Measurement incompatibility, which asks whether several measurements can be replaced by a single measurement, was generalized to simulability of measurements which concerns the question of whether one can obtain measurement statistics of a given set of measurements by performing a different set of measurement [OGWA17, GBTCA17, OMP19, FHL18, ISD⁺22, JUC⁺23, VPLU24, JEPU24, BLN25]. Moreover, many of these concepts are interconnected: Both measurement incompatibility and entanglement are necessary both for steering [UMG14] and Bell nonlocality [Fin82], while contextuality is a precondition for entanglement [PG24] and Bell nonlocality [WF23]. No-broadcasting is equivalent to contextuality [JWP⁺24], while measurement incompatibility and steering are equivalent to a weaker version of contextuality [TU20, Plá22].

In this work we show that most of these concepts are instances of the same mathematical problem: factorizability. We say that a map $f : X \rightarrow Y$ factorizes through Z if there are maps $g : X \rightarrow Z$ and $h : Z \rightarrow Y$ such that

$$f = h \circ g, \tag{1.1}$$

or, equivalently, if the following diagram commutes:

$$\begin{array}{ccc} X & \xrightarrow{f} & Y \\ & \searrow g \quad \nearrow h & \\ & Z & \end{array} \tag{1.2}$$

It was previously observed that Spekkens contextuality is equivalent to simplex-embeddability [SSW⁺21], which itself is an instance of factorizability. Since existence of a broadcasting map was

recently shown to be equivalent to contextuality [JWP⁺24], we immediately get that this problem reduces to factorizability as well.

Firstly, factorizability allows us to approach the problems of certification and quantification of the nonclassical features of quantum theory within a unified and consistent mathematical framework. Secondly, this allows us to use existing mathematical methods to offer solutions to the main problem in quantum information: finding certificates, the so-called witnesses, that yield computable criteria for the presence of the nonclassical features. Thirdly, this will allow us to find streamlined proofs of known results. The main mathematical tool that we will be using is the reformulation of all underlying concepts within the language of *general probabilistic theories* (GPTs) [Lam18, Mül21, Lep21, Plá23] and convex cones. While GPTs will play an important role in our proofs and constructions, we will now formulate all of our results within the language of quantum theory in order to showcase the power of our approach.

This paper is organized as follows. We start with a short section gathering some of our main results (Section 2). In Section 3 we introduce the necessary mathematical background on general probabilistic theories, convex cones and their tensor products, and quantum measurements. In Section 4 we develop our unified framework based on factorization of maps and show how it captures measurement incompatibility and simulability. Section 5 applies this framework to quantum steering, providing new characterizations of steering assemblages and their local hidden state models. In Section 6 we extend our analysis to Bell nonlocality, demonstrating how factorization conditions naturally describe local hidden variable models and providing a novel perspective on CHSH inequalities. In Section 7, we explore the symmetric extensions of multimeters and no-signaling behaviors, relating them to a constrained form of compatibility and a restricted local hidden variable (LHV) model, respectively. Finally, in Section 8 we investigate how noise affects these nonclassical features by examining factorization conditions for noisy state spaces and deriving geometric criteria for simulability of noisy measurements.

2. MAIN RESULTS

One of the main ideas in our framework is the interpretation of a collection of POVMs as maps from quantum states to outcome statistics which are arranged as matrices. A *multimeter* $M = \{M_{\cdot|x}\}_{x \in [g]}$ is a collection of POVMs that consists of g POVMs $M_{\cdot|x}$ each with k effects $\{M_{a|x}\}_{a \in [k]}$ and thus k outcomes. We will consider every multimeter as a map

$$M : \varrho \mapsto M(\varrho) = (\text{Tr}[M_{a|x} \varrho])_{a \in [k], x \in [g]}, \quad (2.1)$$

from the set of density matrices $D(\mathbb{C}^d)$ to the set of column-stochastic matrices, denoted by $CS_{k,g}^1$. In particular, we can think of this map as a channel (i.e., a positive map) between two GPT state spaces, namely from the quantum state space $D(\mathbb{C}^d)$ to the state space $CS_{k,g}^1$. Then we can look into the properties of the POVMs just by looking into the properties of this map. In particular, we look into what it means if this map can be factorized into a composition of two different maps. In the special case when $g = 1$, i.e., when we have just one POVM, we consider it as map to the probability simplex S_k which is isomorphic to $CS_{k,1}^1$.

Since GPT state spaces define cones, another way to study positive maps between state spaces is to consider them as maps between cones, in which case we can write the map as an element of a particular tensor product of the cones. Thus, since the set of density matrices $D(\mathbb{C}^d)$ generates the cone of $d \times d$ positive semi-definite matrices on $\mathbb{C}^d$, denoted by PSD_d , and $CS_{k,g}^1$ generates the cone of nonnegative matrices with equal column sums, denoted by $CS_{k,g}^+$, we have that

$$M \in \text{PSD}_d^* \hat{\otimes} CS_{k,g}^+ \simeq \text{PSD}_d \hat{\otimes} CS_{k,g}^+, \quad (2.2)$$

where $\hat{\otimes}$ denotes the *maximal tensor product* of the cones, see Appendix A.2. Thus, we can also look into the properties of the POVMs by looking into the properties of this tensor. In particular,

we look into what it means for this tensor to be separable, that is, to be an element of the *minimal tensor product* $\dot{\otimes}$.

Our motivating starting point are the different characterizations of measurement incompatibility. In contrast to classical physics, it is in general not possible to have a simultaneous readout of all measurement outcomes in quantum theory. To be more exact, a multimeter $M = \{M_{\cdot|x}\}_{x \in [g]}$ of g POVMs with k outcomes is called *compatible* if there exists a single POVM C with Λ outcomes and a conditional probability distribution $p = (p_{\cdot|x,\lambda})$ on $[k]$ such that

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} p_{a|x,\lambda} C_{\lambda} \quad (2.3)$$

for all $a \in [k]$ and $x \in [g]$. Our guiding example, which already appears in [Jen18] and from which we take inspiration, is the following known result that one can capture compatibility of multimeters either as a form of generalized separability, or via factorizability: for any multimeter $M = \{M_{\cdot|x}\}_{x \in [g]}$ of g POVMs with k outcomes on $\mathbb{C}^d$ the following are equivalent:

- (1) M consists of compatible measurements,
- (2) $M \in \text{PSD}_d \hat{\otimes} CS_{k,g}^+$ as a tensor is separable, i.e., $M \in \text{PSD}_d \dot{\otimes} CS_{k,g}^+$, where $\dot{\otimes}$ denotes the minimal tensor product of cones,
- (3) $M : D(\mathbb{C}^d) \rightarrow CS_{k,g}^1$ as a channel can be factorized as $M = \Phi \circ N$ for some single POVM $N : D(\mathbb{C}^d) \rightarrow S_{\Lambda}$ and some channel $\Phi : S_{\Lambda} \rightarrow CS_{k,g}^1$ for some $\Lambda \in \mathbb{N}$, i.e., the following diagram commutes:

$$\begin{array}{ccc} D(\mathbb{C}^d) & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow N \quad \nearrow \Phi & \\ & S_{\Lambda} & \end{array} \quad (2.4)$$

One can also formulate simulability of measurements in the same fashion: here given a multimeter $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^1$ on a d_A -dimensional Hilbert space we can ask whether M can be simulated (or compressed) to a multimeter on a d_B -dimensional Hilbert space. That is, we ask whether there are channels $\Phi_{\lambda} : D(\mathbb{C}^{d_A}) \rightarrow D(\mathbb{C}^{d_B})$ and multimeters $N_{\lambda} : D(\mathbb{C}^{d_B}) \rightarrow CS_{k,g}^1$, where λ is a label for classical information that can be used in the simulation scheme as well, such that the following diagram commutes:

$$\begin{array}{ccc} D(\mathbb{C}^{d_A}) & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow \Phi \quad \nearrow N & \\ & D(\mathbb{C}^{d_B}) \dot{\otimes} S_{\Lambda} & \end{array} \quad (2.5)$$

Here we have replaced all of the channels $\Phi_{\lambda} : D(\mathbb{C}^{d_A}) \rightarrow D(\mathbb{C}^{d_B})$ by a single channel $\Phi : D(\mathbb{C}^{d_A}) \rightarrow D(\mathbb{C}^{d_B}) \otimes S_{\Lambda}$, where S_{Λ} is a simplex, that is a classical state space, which stores the classical information λ , and similarly for N_{λ} . That is, Φ is an *instrument*. Given a multimeter $M : D(\mathbb{C}^d) \rightarrow CS_{k,g}^1$ we can also ask whether M can be performed using less measurements or outcomes, that is, whether there is a multimeter $N : D(\mathbb{C}^d) \rightarrow CS_{l,r}^1$ and a classical post-processing channel $\Phi : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ such that:

$$\begin{array}{ccc} D(\mathbb{C}^d) & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow N \quad \nearrow \Phi & \\ & CS_{l,r}^1 & \end{array} \quad (2.6)$$

Steering is a quantum phenomenon that may occur when one of the two parties sharing a bipartite state performs one out of several possible measurements. That is, in the steering scenario we have a bipartite state $\varrho \in D(\mathbb{C}^{d_A d_B})$ and a multimeter $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^1$ and we investigate the steering assemblage $\sigma = (M \otimes \text{id})(\varrho)$. In general $\sigma \in CS_{k,g}^1 \hat{\otimes} D(\mathbb{C}^{d_B})$ and we say that σ has local hidden state model if and only if σ is separable, $\sigma \in CS_{k,g}^1 \dot{\otimes} D(\mathbb{C}^{d_B})$. Due to the known relation between steering and measurement incompatibility [UBGP15] we can immediately transport all results about incompatibility to steering with the steering assemblage σ replacing the multimeter M . In quantum theory any steering assemblage can be obtained by measuring a bipartite state [Gis89, HJW93]. While the same is known to not hold in GPTs [BGW13, SB14], we do provide an explicit counterexample.

Similarly to steering, Bell nonlocality may occur when both parties sharing a bipartite state perform one out of several possible measurements. In this case we are working with a bipartite state $\varrho \in D(\mathbb{C}^{d_A d_B})$ and multimeters $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^1$, $N : D(\mathbb{C}^{d_B}) \rightarrow CS_{l,r}^1$ and we investigate the so-called behavior $(M \otimes N)(\varrho)$. In general we have $(M \otimes N)(\varrho) \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ and $(M \otimes N)(\varrho)$ has local hidden variable model if and only if it is separable, $(M \otimes N)(\varrho) \in CS_{k,g}^1 \dot{\otimes} CS_{l,r}^1$, which was recently exploited to find connection between measurement incompatibility and Bell nonlocality [PGQ24]. We will then use the formalism of multimeters to provide streamlined proofs of known results, such as that block-positive operators do not provide post-quantum behaviors [BBB⁺10], that in quantum theory any pair of dichotomic measurements violates the CHSH inequality if and only if it is incompatible [WPGF09], and that if one party measures a pair of dichotomic measurements while the other applies an arbitrary finite multimeter, then all Bell inequalities are just post-processings of the CHSH inequality [Pir14]. To prove the latter two results we will use another key idea also observed in [Jen18, Example 14]: the CHSH inequalities can be identified with isomorphisms of a square via isomorphisms between tensors and linear maps.

We also investigate symmetric n -extensions of multimeters $M \in \text{PSD}_d \hat{\otimes} CS_{k,g}^+$ and no-signaling behaviors $P \in CS_{k,g}^1 \hat{\otimes} CS_{k,g}^1$. These extensions are now tensors in $\text{PSD}_d (CS_{k,g}^+)^{\hat{\otimes} n}$ and in $CS_{k,g}^1 \hat{\otimes} (CS_{k,g}^1)^{\hat{\otimes} n}$, respectively, such that by applying a symmetric reduction map $\gamma_n^\Phi : (CS_{k,g}^+)^{\hat{\otimes} n} \rightarrow CS_{k,g}^+$ the extension reduces to the original multimeter M and the original no-signaling behavior P , respectively. We find that for both M and P the existence of an g -extension is equivalent to the tensor being separable, in which case M is compatible and P has an LHV model. On the other hand, for $n < g$, the existence of an n -extension for M is equivalent to every n -subset of measurements in M being compatible with the joint measurements satisfying particular no-signaling constraints. Similarly for P having an n -extension is equivalent to P having restricted LHV models, i.e., LHV models for all the cases where the second party is restricted to only performing some n measurements. Again these restricted LHV models must satisfy some particular no-signaling constraints.

Finally we investigate the robustness of factorizability to noise: it is known that if a certain amount of white noise is added to the system, incompatible measurements become compatible, steering assemblages start having local hidden state models, and behaviors start having hidden variable models. We inspect factorization conditions for maps mixed with noise to get geometric criteria for when noisy multimeters can be simulated by multimeters with simpler measurement settings.

3. PRELIMINARIES

In this section, we will review the formalism of GPTs and their transformations. We will start with a short review of the formalism in Section 3.1 and continue by describing positive maps and channels between such theories in Section 3.2. In Section 3.3, we focus on special classes of channels, namely measurements and instruments. Then, we will concentrate on the GPT of column stochastic

matrices in Section 3.4, which we will need to formalize multimeters as maps from state spaces to column stochastic matrices in Section 3.5.

3.1. The formalism of general probabilistic theories. In this section, we will review a general framework for describing operational theories. For a review of cones and the terminology used to describe them, see Appendix A.1. A GPT is described by a triple $(V, V^+, \mathbb{1}_V)$, where V is a real finite-dimensional vector space with a proper cone V^+ , and where $\mathbb{1}_V$ is an order unit in the interior of the dual cone $A^+ := (V^+)^* \subset V^* =: A$. Here, we have written V^* for the dual vector space of V , i.e., the vector space of linear functionals on V , and $(V^+)^*$ for the cone dual to V^+ . Then the *state space* of the theory is described by the compact convex subset $K \subset V$ defined as

$$K := \{v \in V^+ : \mathbb{1}_V(v) = 1\}, \quad (3.1)$$

which also forms a base for V^+ . On the other hand, if we have a compact convex set K , then if we set $A = A(K)$ (the set of affine functionals $f : K \rightarrow \mathbb{R}$), $A^+ = A(K)^+$ (the set of positive affine functionals on K) and take $\mathbb{1}_V$ to be the constant function $\mathbb{1}_K$ giving value 1 on K (which is an order unit in A). Then we can choose $V = A^*$ and $V^+ = (A^+)^*$ and it follows that K is isomorphic to a compact convex base of the cone V^+ determined by $\mathbb{1}_K$ as in Eq. (3.1). Thus, sometimes we might also write $(V(K), V(K)^+, \mathbb{1}_K)$ for a GPT which is determined by its state space K . We refer the reader to [Lam18, Plá23] for more background on GPTs. Before we continue, we will briefly describe what classical and quantum mechanics look like in the formalism of GPTs.

Example 3.1. Any classical system is described by the triple $\text{CM}_d := (\mathbb{R}^d, \mathbb{R}_+^d, \mathbb{1}_d)$, $d \in \mathbb{N}$, where $\mathbb{R}_+^d$ denotes the set of elements with nonnegative coordinates and $\mathbb{1}_d = (1, 1, \dots, 1)$. The classical state space is the probability simplex with d vertices S_d .

Example 3.2. Quantum theory corresponds to the triple $\text{QM}_d := (\mathcal{M}(\mathbb{C})_d^{\text{sa}}, \text{PSD}_d, \text{Tr})$, $d \in \mathbb{N}$, where PSD_d is the cone of $d \times d$ positive semi-definite complex matrices in the real vector space of self-adjoint matrices $\mathcal{M}(\mathbb{C})_d^{\text{sa}}$. The quantum state space is the set of density matrices $D(\mathbb{C}^d)$, i.e., the elements in PSD_d with trace equal to one.

Until now, we have only considered single systems. However, as in quantum mechanics, it makes sense to have systems that are multipartite, where each subsystem is a GPT. Let $(V_A, V_A^+, \mathbb{1}_{V_A})$ and $(V_B, V_B^+, \mathbb{1}_{V_B})$ be two GPTs. To construct a bipartite GPT out of them, we first need a joint vector space: For this, we consider simply the tensor product of the individual vector spaces, i.e., $V_{AB} := V_A \otimes V_B$. Second, we need an order unit, for which we can simply take the tensor product of order units, i.e., $\mathbb{1}_{V_{AB}} := \mathbb{1}_{V_A} \otimes \mathbb{1}_{V_B}$. Thus, it remains to choose a cone for the GPT on the bipartite system. Here, we are free to choose any tensor cone, i.e., any cone V_{AB}^+ such that

$$V_A^+ \dot{\otimes} V_B^+ \subseteq V_{AB}^+ \subseteq V_A^+ \hat{\otimes} V_B^+. \quad (3.2)$$

We refer the reader to Appendix A.2 for a recap of tensor products of cones. In fact, we can also define the *minimal tensor product of state spaces*

$$K_A \dot{\otimes} K_B := \text{conv}\{x_A \otimes x_B : x_A \in K_A, x_B \in K_B\} \quad (3.3)$$

and the *maximal tensor product of state spaces*

$$K_A \hat{\otimes} K_B := \{y \in V_{AB} : \langle \mathbb{1}_{K_A} \otimes \mathbb{1}_{K_B}, y \rangle = 1, \langle f_A \otimes f_B, y \rangle \geq 0 \ \forall f_A \in A(K_A)^+, \forall f_B \in A(K_B)^+\}. \quad (3.4)$$

Here, we have written for $x \in V$ an element in a vector space and $\alpha \in V^*$ a functional the expression $\langle \alpha, x \rangle$ to mean $\alpha(x)$. As $D(\mathbb{C}^{d_A}) \dot{\otimes} D(\mathbb{C}^{d_B})$ is the set of separable states, elements in $K_A \dot{\otimes} K_B$ are called *separable*. However, the set $D(\mathbb{C}^{d_A}) \hat{\otimes} D(\mathbb{C}^{d_B})$ encompasses more than $D(\mathbb{C}^{d_A d_B})$. The set is in fact equal to the set of block-positive matrices of unit trace and can

therefore be interpreted as the set of entanglement witnesses [CS14]. This can be seen from the fact that $V_A^+ \hat{\otimes} V_B^+ = ((V_A^*)^+ \hat{\otimes} (V_B^*)^+)^*$. With these definitions, it follows that

$$V(K_A \hat{\otimes} K_B)^+ = V(K_A)^+ \hat{\otimes} V(K_B)^+, \quad (3.5)$$

$$V(K_A \hat{\otimes} K_B)^+ = V(K_A)^+ \hat{\otimes} V(K_B)^+. \quad (3.6)$$

3.2. Positive maps and channels between GPTs. Let $(V_A, V_A^+, \mathbb{1}_{V_A})$ and $(V_B, V_B^+, \mathbb{1}_{V_B})$ be two GPTs, transformations between these two GPTs are described by *channels*, i.e., linear mappings $\Phi : V_A \rightarrow V_B$ that are positive, i.e., $\Phi(V_A^+) \subseteq V_B^+$, and normalization-preserving, i.e., $\mathbb{1}_{V_B}(\Phi(v)) = \mathbb{1}_{V_A}(v)$ for all $v \in V_A^+$. The latter property means that channels map elements of the state space K_A of $(V_A, V_A^+, \mathbb{1}_{V_A})$ to elements of the state space K_B of $(V_B, V_B^+, \mathbb{1}_{V_B})$. Regarding notation, we may also denote the channel $\Phi : V_A \rightarrow V_B$ simply as a map $\Phi : K_A \rightarrow K_B$ between the state spaces of the GPTs. If the map is merely positive but not a channel, we will often write $\Phi : V_A^+ \rightarrow V_B^+$ to emphasize that it is a map between cones. For the dual map $\Phi^* : V_B^* \rightarrow V_A^*$, it holds that Φ is positive if and only if Φ^* is. Additionally, Φ is a channel if and only if $\Phi^*(\mathbb{1}_{V_B}) = \mathbb{1}_{V_A}$, i.e., Φ^* is *unital*.

Remark 3.3. *Note that our notion of a positive map does not automatically mean completely positive as required in quantum theory. The problem with complete positivity in GPTs is that one needs to define it through an ancillary system and this requires specifying the ancillary GPT as well as the tensor product between the cones. We refer to [Plá23, Sec. 6.5] for a more in-depth discussion.*

Now, we can consider tensors that play the role of Choi matrices for GPTs. Let us consider two GPTs $(V(K_A), V(K_A)^+, \mathbb{1}_{K_A})$ and $(V(K_B), V(K_B)^+, \mathbb{1}_{K_B})$. Let $\{e_i\}_{i \in [d]}$ be a basis of $V(K_A)$ and let $\{a_j\}_{j \in [d]}$ be the corresponding dual basis of $A(K_A)$ for some $d \in \mathbb{N}$. Then, we can define a special tensor $\chi_{V(K_A)} \in A(K_A) \otimes V(K_A)$ as

$$\chi_{V(K_A)} = \sum_{i=1}^d a_i \otimes e_i. \quad (3.7)$$

It holds that $\chi_{V(K_A)} \in A(K_A) \hat{\otimes} V(K_A)$, an explicit proof can be found in [Jen18, Lemma A.1]. Also by [Jen18, Lemma A.1] (see also Appendix A.3), there is a one-to-one correspondence between $\xi_\Phi \in V(K_B)^+ \hat{\otimes} V(K_A)^+$ and $\Phi : A(K_A)^+ \rightarrow V(K_B)^+$ as $\xi_\Phi = (\Phi \otimes \text{id})(\chi_{V(K_A)})$ and

$$\langle \Phi(f_A), f_B \rangle = \langle \xi_\Phi, f_B \otimes f_A \rangle, \quad \forall f_A \in A(K_A), \forall f_B \in A(K_B). \quad (3.8)$$

In fact, the correspondence between maps and tensors is unique:

Lemma 3.4. *Let $\xi \in K_A \hat{\otimes} K_B$. Then, there is a unique map $\Phi : A(K_A)^+ \rightarrow V(K_B)^+$ such that $\xi = (\Phi \otimes \text{id})(\chi_{V(K_A)})$ and $\Phi(\mathbb{1}_{K_A}) \in K_B$.*

As PSD_d is self-dual, one can embed $D(\mathbb{C}^d)$ into both $V(D(\mathbb{C}^d))^+$ and $A(D(\mathbb{C}^d))^+$. However, GPTs are in general not self-dual and one cannot embed K_A into $A(K_A)^+$. Sometimes, it is nonetheless helpful to also define a state space in $A(K_A)^+$. Let $\varrho \in K_A$ and define

$$(K_A)_\varrho^* := \{f \in A(K_A)^+ : f(\varrho) = 1\}. \quad (3.9)$$

If ϱ is in the relative interior of K , then $(K_A)_\varrho^*$ is compact and $V((K_A)_\varrho^*)^+ = A(K_A)^+$, $A((K_A)_\varrho^*)^+ = V(K_A)^+$, see again [Jen18, Lemma A.1] for an explicit proof. Thus, $(K_A)_\varrho^*$ is a state space for the GPT $(A(K_A), A(K_A)^+, \varrho)$. If ϱ is not in the relative interior, there is $f \in A(K_A)^+$ such that $f \neq 0$ and $f(\varrho) = 0$, and $(K_A)_\varrho^*$ is not bounded.

To illustrate this construction, let us again consider classical and quantum mechanics:

Example 3.5. Let us consider classical system as a GPT $\text{CM}_d = (\mathbb{R}^d, \mathbb{R}_+^d, 1_d)$, $d \in \mathbb{N}$ with some classical state $q > 0$ (i.e., a probability distribution with positive definite entries). Then, we can readily verify that

$$(S_d)_q^* = \{p \in \mathbb{R}_+^d : \sum_{i=1}^d p_i q_i = 1\}. \quad (3.10)$$

Thus, we can identify $(S_d)_q^* \simeq S_d$, where the vertices of the new simplex are

$$\delta_i = (0, \dots, 0, \underbrace{q_i^{-1}}_{i\text{-th entry}}, 0, \dots, 0). \quad (3.11)$$

Example 3.6. Let us consider quantum theory as a GPT $\text{QM}_d = (\mathcal{M}(\mathbb{C})_d^{\text{sa}}, \text{PSD}_d, \text{Tr})$, $d \in \mathbb{N}$ with some quantum state $\varrho > 0$. Then, we can readily verify that

$$(D(\mathbb{C}^d))_\varrho^* = \{\varrho^{-\frac{1}{2}} \sigma \varrho^{-\frac{1}{2}} : \sigma \in D(\mathbb{C}^d)\}. \quad (3.12)$$

To conclude this subsection, we will characterize maps corresponding to separable tensors (also called *entanglement-breaking* in analogy to the case of quantum mechanics) as having a certain factorization:

Proposition 3.7. Let $\xi_\Phi \in K_A \hat{\otimes} K_B$ and $\Phi : A(K_A)^+ \rightarrow V(K_B)^+$ such that $\Phi(1_{K_A}) \in K_B$ is the unique associated map in Lemma 3.4. Then, $\xi_\Phi \in K_A \hat{\otimes} K_B$ if and only if there exist finitely many outcomes Λ , a positive map $\Psi_1 : A(K_A)^+ \rightarrow S_\Lambda^+$ such that $\Psi_1(1_{K_A}) \in S_\Lambda$, and a channel $\Psi_2 : S_\Lambda \rightarrow K_B$ such that the following diagram commutes:

$$\begin{array}{ccc} A(K_A)^+ & \xrightarrow{\Phi} & V(K_B)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & S_\Lambda^+ & \end{array} \quad (3.13)$$

If in addition Φ is a channel, then Ψ_1 can be taken to be a channel as well.

Proof. The factorization into positive Ψ_1 and Ψ_2 follows from [BJN22, Lemma 3.7] and [Jen18, Proposition A.1]. Thus, it remains to prove that indeed we can assume $\Psi_1(1_{K_A}) \in S_\Lambda$ and Ψ_2 to be a channel. Decomposing

$$\xi_\Phi = \sum_{\lambda=1}^{\Lambda} p_\lambda x_\lambda \otimes y_\lambda \quad (3.14)$$

with $x_\lambda \in K_A$ and $y_\lambda \in K_B$ for all $\lambda \in [\Lambda]$ and a probability distribution $(p_\lambda)_{\lambda \in [\Lambda]}$, we can explicitly choose

$$\Psi_1 : \alpha \mapsto \sum_{\lambda=1}^{\Lambda} p_\lambda \alpha(x_\lambda) \delta_\lambda, \quad \forall \alpha \in A(K_A), \quad (3.15)$$

$$\Psi_2 : \delta_\lambda \mapsto y_\lambda, \quad \forall \lambda \in [\Lambda]. \quad (3.16)$$

We can thus see directly that Ψ_2 is a channel and that $\Psi_1(1_{K_A}) \in S_\Lambda$, since $1_{K_A}(x_\lambda) = 1$ for all $\lambda \in [\Lambda]$. The fact that Ψ_1 can also be chosen to be a channel in case Φ is a channel follows from [Jen18, Proposition A.1]. $\square$

3.3. Meters and instruments in GPTs. Measurements are a special form of channels between GPTs, as we will see in this section. An *effect* corresponds to the most elementary ‘yes-no’ - measurement and it is described by an element $e \in A^+$ such that $0 \leq e(\varrho) \leq 1$ for all $\varrho \in K$. The set of effects on a state space K is denoted by $E(K)$ and it is called the effect space of K . The number $e(\varrho)$ is interpreted as the probability that the event corresponding to the effect e is detected in a measurement of a system in a state ϱ . We say that a nonzero effect e is *indecomposable* if

any decomposition of e into a sum of two other nonzero effects e_1, e_2 as $e = e_1 + e_2$ implies that $e = \alpha_1 e_1 = \alpha_2 e_2$ for some $\alpha_1, \alpha_2 > 0$ [KNI10]. It can be shown that an effect is indecomposable if and only if it lies on an extreme ray of the dual cone $A(K)^+$.

Definition 3.8. Let $(V, V^+, \mathbb{1}_V)$ be a GPT with a state space K . A measurement (or meter) on K with $k < \infty$ outcomes is described by a channel $f : K \rightarrow S_k$.

Let f be a k -outcome measurement on K . Then for all $\varrho \in K$ we have that $f(\varrho) \in S_k$ such that if we write $\{\delta_i\}_{i=1}^k$ for the vertices of the simplex S_k , then $f(\varrho)$ has a unique convex decomposition $f(\varrho) = \sum_{i=1}^k \alpha_i^{(f, \varrho)} \delta_i$ for some unique coefficients $\alpha_i^{(f, \varrho)} \geq 0$ for all $i \in [k]$ such that $\sum_{i=1}^k \alpha_i^{(f, \varrho)} = 1$. For all $i \in [k]$ we define a mapping $f_i : K \rightarrow [0, 1]$ by setting $f_i(\varrho) = \alpha_i^{(f, \varrho)}$. For all $i \in [k]$ we see that f_i is clearly well-defined and it is straightforward to see that f_i is affine. Furthermore, each f_i can be uniquely extended to an element (for which we use the same notation) f_i in $A(K)$. It is then clear that $f_i \in E(K)$ for all $i \in [k]$ and that $\sum_{i=1}^k f_i(\varrho) = 1$ for all $\varrho \in K$; hence $\sum_{i=1}^k f_i = \mathbb{1}_V$. Thus, each k -outcome measurement f on K is characterized by some k effects $f_1, \dots, f_k$ that satisfy $\sum_{i=1}^k f_i = \mathbb{1}_V$; hence

$$f(\varrho) = \sum_{i=1}^k f_i(\varrho) \delta_i \quad (3.17)$$

for all $\varrho \in K$. On the other hand, given k effects $f_1, \dots, f_k$ that satisfy $\sum_{i=1}^k f_i = \mathbb{1}_V$ one can use the above equation to define a channel from K to S_k . The interpretation is that when we measure a system which is in a state ϱ with a measurement/meter f , then $f_i(\varrho)$ is the probability that an outcome i is obtained.

Measurement devices which not only produce a classical measurement outcome but also output a post-measurement state are called *instruments*. Formally, we can define instruments as specific types of channels:

Definition 3.9. Let $(V_A, V_A^+, \mathbb{1}_{V_A})$ and $(V_B, V_B^+, \mathbb{1}_{V_B})$ be two GPTs with state spaces K_A and K_B , respectively. An instrument from K_A to K_B with k outcomes is described by a channel $\Phi : K_A \rightarrow K_B \dot{\otimes} S_k$.

Let $\Phi : K_A \rightarrow K_B \dot{\otimes} S_k$ be an instrument between state spaces K_A and K_B . Thus, we can write it as $\Phi(\varrho) = \sum_{i=1}^k \lambda_i^{\varrho} \xi_i^{\varrho} \otimes \delta_i$ for some $\xi_i^{\varrho} \in K_B$ and $\lambda_i^{\varrho} \geq 0$ for all $i \in [k]$ and $\varrho \in K_A$ such that $\sum_{i=1}^k \lambda_i^{\varrho} = 1$. Since $\{\delta_i\}_{i=1}^k$ is a basis of $\mathbb{R}^k$, this decomposition is unique. Thus, for all $i \in [k]$ we can define a map $\Phi_i : V_A \rightarrow V_B$ by setting $\Phi_i(\varrho) = \lambda_i^{\varrho} \xi_i^{\varrho}$ for all $\varrho \in K_A$, and it is straightforward to see that each Φ_i is positive, i.e., $\Phi_i(V_A^+) \subseteq V_B^+$ and that it is normalization-non-increasing, i.e., $\mathbb{1}_{K_B}(\Phi_i(\varrho)) \leq \mathbb{1}_{K_A}(\varrho)$ for all $\varrho \in V_A^+$. We refer to maps satisfying these two properties as *operations*. Furthermore, we see that $\sum_{i=1}^k \Phi_i$ is normalization-preserving, i.e., it is a channel between K_A and K_B . Thus, each k -outcome instrument Φ between K_A and K_B is characterized by a collection of operations $\Phi_1, \dots, \Phi_k$ such that $\sum_{i=1}^k \Phi_i$ is a channel from K_A to K_B which maps

$$\Phi(\varrho) = \sum_{i=1}^k \Phi_i(\varrho) \otimes \delta_i \quad (3.18)$$

for all $\varrho \in K_A$. On the other hand, given a collection of such maps $\Phi_1, \dots, \Phi_k$ one can construct an instrument $\Phi : K_A \rightarrow K_B \dot{\otimes} S_k$ by just setting $\Phi(\varrho) = \sum_{i=1}^k \Phi_i(\varrho) \otimes \delta_i$ for all $\varrho \in K_A$. The operational interpretation of an instrument is the following: when the system, which is in a state ϱ , is operated on by an instrument Φ , then a measurement outcome i is detected with probability $\mathbb{1}_{K_B}(\Phi_i(\varrho))$ after which the system can be found in the conditional output state $\Phi_i(\varrho) / \mathbb{1}_{K_B}(\Phi_i(\varrho))$. Note that we can without loss of generality assume that $\mathbb{1}_{K_B}(\Phi_i(\varrho)) \neq 0$, because otherwise outcome i never occurs and the post-measurement state is irrelevant.

3.4. State spaces of column stochastic matrices. It is now time to consider a GPT that will be essential for the definition of multimeters. Let us consider the GPT $(CS_{k,g}, CS_{k,g}^+, \mathbb{1}_{CS_{k,g}^+})$, where $CS_{k,g}$ is the real vector space of real $k \times g$ -matrices whose column sums are equal, $CS_{k,g}^+$ is the cone of nonnegative $k \times g$ -matrices whose column sums are equal and $\mathbb{1}_{CS_{k,g}^+} : CS_{k,g} \rightarrow \mathbb{R}$ is a functional defined as $\mathbb{1}_{CS_{k,g}^+}(M) := \frac{1}{g} \langle J_{k,g}, M \rangle$ for all $M \in CS_{k,g}$, where $J_{k,g}$ is the $k \times g$ -matrix of all ones and the inner product is the Hilbert-Schmidt inner product. The state space of this GPT is the set of column stochastic $k \times g$ -matrices (nonnegative $k \times g$ -matrices whose columns sum to one), which we denote by $CS_{k,g}^1$. This GPT is a special case of the polysimplex considered in [Jen18], where the number of vertices of all simplices involved is chosen to be equal for simplicity.

The extreme points of $CS_{k,g}^1$ are matrices whose columns have an entry 1 on all columns on some positions, i.e., matrices of the form

$$s_{i_1, \dots, i_g} = \sum_{j=1}^g E_{i_j, j} \quad (3.19)$$

for all $i_j \in [k]$ and $j \in [g]$, where $E_{i,j}$ is a matrix with entry 1 on the position (i, j) and zeros everywhere else. We can define measurements $m^{(j)}$ which project onto the j -th column and have effects $m_i^{(j)}$ defined as

$$m_i^{(j)}(s_{i_1, \dots, i_g}) := \langle E_{i,j}, s_{i_1, \dots, i_g} \rangle = \begin{cases} 1 & i_j = i \\ 0 & \text{otherwise} \end{cases}. \quad (3.20)$$

We note that $\sum_{i=1}^k m_i^{(j)}(M) = \mathbb{1}_{CS_{k,g}^1}(M)$ for all $M \in CS_{k,g}$ so that as functionals from $CS_{k,g}$ to $\mathbb{R}$ they are equal; $\sum_{i=1}^k m_i^{(j)} = \mathbb{1}_{CS_{k,g}^1}$ for all $j \in [g]$. Moreover, we define

$$e_i^{(j)} := E_{i,j} - E_{k,j}. \quad (3.21)$$

These definitions give us convenient bases for $CS_{k,g}$ and $CS_{k,g}^*$, respectively. Lemma 1 of [Jen18] states that the $m_i^{(j)}$, $i \in [k]$, $j \in [g]$ generate the extreme rays of $(CS_{k,g}^+)^*$. Furthermore,

$$\mathbb{1}_{CS_{k,g}^1}, m_1^{(1)}, \dots, m_{k-1}^{(1)}, m_1^{(2)}, \dots, m_{k-1}^{(g)} \quad (3.22)$$

form a basis of $CS_{k,g}^*$ and

$$s_{k, \dots, k}, e_1^{(1)}, \dots, e_{k-1}^{(1)}, e_1^{(2)}, \dots, e_{k-1}^{(g)} \quad (3.23)$$

the corresponding dual basis of $CS_{k,g}$.

3.5. Multimeters as channels. As we saw earlier in Section 3.3, a measurement or meter is characterized by a collection of effects. Next we focus on measurement devices that can implement the measurement of multiple meters; we call this a *multimeter*.

Definition 3.10. Let $(V, V^+, \mathbb{1}_V)$ be a GPT with a state space K . A multimeter on K with g measurements each with k outcomes is described by a channel $f : K \rightarrow CS_{k,g}^1$.

Measurements will be from now on seen as just multimeters with one measurement setting. For the same reason, in everything that follows we will simply talk about multimeters even if we are considering single measurements unless we want to specifically emphasize that we are talking about measurements.

Let f be a multimeter with g measurements and k outcomes on K . Then for all $\varrho \in K$ we have $f(\varrho) \in CS_{k,g}^1$. Let us define maps $M_{a|x} : K \rightarrow S_k$ by setting $M_{a|x} = m_a^{(x)} \circ f$ for all $a \in [k]$ and $x \in [g]$, where $m^{(x)} : CS_{k,g} \rightarrow S_k$ are the measurements defined in Eq. (3.20) which project onto the x -th column of the outcome probability matrix. As compositions of a channel and a measurement,

it follows that $M_{a|x}$ are effects such that $\sum_{a=1}^k M_{a|x} = \mathbb{1}_K$ so that the collection $\{M_{a|x}\}_{a=1}^k$ defines a measurement $M_{\cdot|x}$ on K for all $x \in [g]$.

In the basis of Eq. (3.23) we can write the multimeter $f : K \rightarrow CS_{k,g}^1$ now as

$$f(\varrho) = \mathbb{1}_K(\varrho) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} M_{a|x}(\varrho) e_a^{(x)} \quad (3.24)$$

for all $\varrho \in K$. One can readily check that indeed $m_a^{(x)}(f(\varrho)) = M_{a|x}(\varrho)$ for all $a \in [k]$, $x \in [g]$ and $\varrho \in K$ as defined before. Conversely, given a collection of g measurements $\{M_{\cdot|x}\}_{x \in [g]}$ on K each with k outcomes one can use the above equation to define a unique channel $f : K \rightarrow CS_{k,g}^1$.

Alternatively, we can also represent f more naturally as

$$f(\varrho) = \sum_{x=1}^g \sum_{a=1}^k M_{a|x}(\varrho) E_{a,x} = (M_{\cdot|1}(\varrho), \dots, M_{\cdot|g}(\varrho)) \quad (3.25)$$

for all $\varrho \in K$. In particular, then we see that a k -outcome multimeter f on K is characterized by g measurements $M_{\cdot|1}, \dots, M_{\cdot|g}$ with k outcomes respectively such that $f(\varrho) = (M_{\cdot|1}(\varrho), \dots, M_{\cdot|g}(\varrho))$ where the measurements define the columns of the outcome probability matrix. Thus, we may write $f = f_M$ for some collection of k -outcome measurements $M = \{M_{\cdot|x}\}_{x \in [g]}$. In this representation, f manifestly maps states to states. For multimeters the interpretation is that when we measure a system which is in a state ϱ with a multimeter M by using a measurement setting $x \in [g]$, i.e., by using a measurement $M_{\cdot|x}$, then $M_{a|x}(\varrho)$ is the probability that an outcome $a \in [k]$ is obtained.

An important property of measurements is compatibility.

Definition 3.11. *A multimeter $M = \{M_{\cdot|x}\}_{x \in [g]}$ of g measurements with k outcomes on K is compatible if there exists a single l -outcome measurement N on K and some conditional probability distribution $\nu = (\nu_{\cdot|b,x})_{b \in [l], x \in [g]}$ on $[k]$ such that*

$$M_{a|x} = \sum_{b=1}^l \nu_{a|b,x} N_b \quad (3.26)$$

for all $a \in [k]$ and $x \in [g]$.

We refer the reader to [HMZ16, GHK⁺23] for an introduction to incompatible measurements in quantum mechanics and other operational theories. In the next section, we will characterize compatibility in terms of the corresponding map $f : K \rightarrow CS_{k,g}^1$.

4. COMPATIBILITY AND SIMULATION OF MULTIMETERS

4.1. Simulating multimeters on one state space by multimeters on another state space.

We continue by generalizing to GPTs the notion of simulability of quantum measurements considered in the recent works [ISD⁺22, JUC⁺23, JEP24].

Definition 4.1. *Let $M = \{M_{\cdot|x}\}_{x \in [g]}$ be a multimeter of g measurements with k outcomes on a state space K_A . It is K_B -simulable if there exists a finite number of outcomes Λ , an instrument $\Phi : K_A \rightarrow K_B \otimes S_\Lambda$ with operations $\Phi_\lambda : V(K_A)^+ \rightarrow V(K_B)^+$, and a multimeter $N = \{N_{\cdot|x,\lambda}\}_{x \in [g], \lambda \in [\Lambda]}$ of $g \cdot \Lambda$ measurements with k outcomes on K_B such that*

$$M_{a|x} = \sum_{\lambda=1}^\Lambda \Phi_\lambda^*(N_{a|x,\lambda}) \quad (4.1)$$

for all $a \in [k]$ and $x \in [g]$.

Here the recipe of simulation is as follows: a measurement with a label x is obtained first by measuring the input state $\varrho \in K_A$ by the instrument Φ , obtaining an outcome λ and resulting the system in a (unnormalized) conditional output state $\Phi_\lambda(\varrho)$, which is then measured by the POVM $N_{\cdot|x,\lambda}$ from which an outcome a is obtained and then reported as the final outcome of the transformed measurement with a label x .

We will show now that K_B -simulability is equivalent to a factorization of the associated multi-meter, seen as a channel into column stochastic matrices.

Theorem 4.2. *Let $M = \{M_{\cdot|x}\}_{x \in [g]}$ be a multimeter of g measurements with k outcomes on a state space K_A . Then M is K_B -simulable if and only if there exists a finite number of outcomes Λ , an instrument $\Phi : K_A \rightarrow K_B \dot{\otimes} S_\Lambda$ and a multimeter $\tilde{N} : K_B \dot{\otimes} S_\Lambda \rightarrow CS_{k,g}^1$ such that the following diagram commutes:*

$$\begin{array}{ccc} K_A & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow \Phi \quad \nearrow \tilde{N} & \\ & K_B \dot{\otimes} S_\Lambda & \end{array} \quad (4.2)$$

We start by proving the following lemma.

Lemma 4.3. *There is a one-to-one correspondence between multimeters $N : K_B \rightarrow CS_{k,g,\Lambda}^1$ and channels $\tilde{N} : K_B \dot{\otimes} S_\Lambda \rightarrow CS_{k,g}^1$.*

Proof. Informally, the multimeter N maps states in K_B to $g \cdot \Lambda$ tuples of probability vectors of size k . If we see the range of this map as Λ -tuples of elements of $CS_{k,g}^1$, the data of the channel N is equivalent to mapping $K_B \dot{\otimes} S_\Lambda$ to $CS_{k,g}^1$. Next, we make this intuition formal.

First, given the measurements $\{N_{\cdot|x,\lambda}\}_{x \in [g], \lambda \in [\Lambda]}$ of the multimeter $N : K_B \rightarrow CS_{k,g,\Lambda}^1$ we can define a linear map $\tilde{N} : V(K_B) \otimes \mathbb{R}^\Lambda \rightarrow CS_{k,g}^1$ by setting

$$\tilde{N}(X) = \mathbb{1}_{K_B \dot{\otimes} S_\Lambda}(X)_{s_{k,\dots,k}} + \sum_{\lambda=1}^{\Lambda} \sum_{x=1}^g \sum_{a=1}^{k-1} (N_{a|x,\lambda} \otimes b_\lambda)(X) e_a^{(x)}, \quad (4.3)$$

for all $X = \sum_{\lambda=1}^{\Lambda} X_\lambda \otimes \delta_\lambda \in V(K_B) \otimes \mathbb{R}^\Lambda$, where $\{b_\lambda\}_{\lambda \in [\Lambda]}$ is the dual basis of $\{\delta_\lambda\}_{\lambda \in [\Lambda]}$. Clearly now

$$\tilde{N}(\varrho \otimes \delta_\lambda) = \mathbb{1}_{K_B}(\varrho)_{s_{k,\dots,k}} + \sum_{x=1}^g \sum_{a=1}^{k-1} N_{a|x,\lambda}(\varrho) e_a^{(x)} = \sum_{x=1}^g \sum_{a=1}^k N_{a|x,\lambda}(\varrho) E_{a,x} \quad (4.4)$$

for all $\varrho \in V(K_B)^+$ and $\lambda \in [\Lambda]$. Thus, by linearity $\tilde{N}$ maps states to states so that it is a channel.

On the other hand, if we have a channel $\tilde{N} : K_B \dot{\otimes} S_\Lambda \rightarrow CS_{k,g}^1$, then it defines a multimeter $N : K_B \rightarrow CS_{k,g,\Lambda}^1$ as follows. Namely, as $\tilde{N}$ is linear and maps states to states, it can be written as

$$\tilde{N}(X) = \mathbb{1}_{K_B \dot{\otimes} S_\Lambda}(X)_{s_{k,\dots,k}} + \sum_{x=1}^g \sum_{a=1}^{k-1} A_{a|x}(X) e_a^{(x)}, \quad (4.5)$$

where $A_{a|x} : V(K_B) \otimes \mathbb{R}^\Lambda \rightarrow \mathbb{R}$ are some linear functionals. We see immediately that

$$m_a^{(x)}(\tilde{N}(X)) = A_{a|x}(X) \quad \forall a \in [k-1], x \in [g], \quad (4.6)$$

which is positive for all $X \in V(K_B)^+ \dot{\otimes} S_\Lambda^+$, because $\tilde{N}$ is positive and the $m_a^{(x)}$ generate the extreme rays of $(CS_{k,g}^+)^*$. Thus, $A_{a|x} \in (V(K_B)^+ \dot{\otimes} S_\Lambda^+)^*$ for all $a \in [k-1], x \in [g]$ and also

$A_{k|x} := \mathbb{1}_{K_B \dot{\otimes} S_\Lambda} - \sum_{a=1}^{k-1} A_{a|x} \in (V(K_B)^+ \dot{\otimes} S_\Lambda^+)^*$ since

$$m_k^{(x)}(\tilde{N}(X)) = \mathbb{1}_{K_B \dot{\otimes} S_\Lambda}(X) - \sum_{a=1}^{k-1} A_{a|x}(X). \quad (4.7)$$

Hence, $\{A_{\cdot|x}\}_{x \in [g]}$ is a set of g measurements with k outcomes. Defining $N_{a|x,\lambda}(\varrho) := A_{a|x}(\varrho \otimes \delta_\lambda)$ for all $\varrho \in K_B$, $a \in [k]$, $x \in [g]$ and $\lambda \in [\Lambda]$, we find for $X = \sum_{\lambda=1}^\Lambda X_\lambda \otimes \delta_\lambda$ that

$$\tilde{N}(X) = \mathbb{1}_{K_B \dot{\otimes} S_\Lambda}(X) s_{k,\dots,k} + \sum_{\lambda=1}^\Lambda \sum_{x=1}^g \sum_{a=1}^{k-1} N_{a|x,\lambda}(X_\lambda) e_a^{(x)}. \quad (4.8)$$

Clearly $N = \{N_{\cdot|x,\lambda}\}_{x \in [g], \lambda \in [\Lambda]}$ now defines a multimeter $N : K_B \rightarrow CS_{k,g,\Lambda}^1$. $\square$

By using the previous lemma we prove Theorem 4.2.

Proof of Theorem 4.2. Let us now first assume that the multimeter M on K_A is K_B -simulable. Thus, there exists some finite number of outcomes Λ , an instrument $\Phi : K_A \rightarrow K_B \dot{\otimes} S_\Lambda$ and a multimeter $N : K_B \rightarrow CS_{k,g,\Lambda}^1$ such that $M_{a|x} = \sum_{\lambda=1}^\Lambda \Phi_\lambda^*(N_{a|x,\lambda})$ for all $a \in [k]$ and $x \in [g]$. As was explained in Lemma 4.3, we can use the multimeter $N : K_B \rightarrow CS_{k,g,\Lambda}^1$ to define a multimeter $\tilde{N} : K_B \dot{\otimes} S_\Lambda \rightarrow CS_{k,g}^1$. What remains to show is just that $M = \tilde{N} \circ \Phi$. First, as we explained in Section 3.5, we have that

$$\mathbb{1}_{CS_{k,g}^1}(M(\varrho)) = \mathbb{1}_{K_A}(\varrho), \quad m_a^{(x)}(M(\varrho)) = M_{a|x}(\varrho) \quad \forall a \in [k-1], x \in [g], \varrho \in K_A. \quad (4.9)$$

By using Eq. (4.4) it is easy to see that $\mathbb{1}_{CS_{k,g}^1}(\tilde{N}(\Phi(\varrho))) = \mathbb{1}_{K_A}(\varrho)$. Finally, for all $a \in [k-1]$, $x \in [g]$, $\varrho \in K_A$,

$$m_a^{(x)}(\tilde{N}(\Phi(\varrho))) = \sum_{\lambda=1}^\Lambda \tilde{N}_{a|x,\lambda}(\Phi_\lambda(\varrho)) = \sum_{\lambda=1}^\Lambda (\Phi_\lambda^*(N_{a|x,\lambda}))(\varrho) = \left(\sum_{\lambda=1}^\Lambda \Phi_\lambda^*(\tilde{N}_{a|x,\lambda}) \right)(\varrho) = M_{a|x}(\varrho), \quad (4.10)$$

which shows that indeed $M = \tilde{N} \circ \Phi$.

On the other hand, if $M = \tilde{N} \circ \Phi$ with the channels $\tilde{N}$ and Φ described in the diagram, then we can use Lemma 4.3 again to see that M is K_B -simulable because by using Eq. (4.8) we have that

$$M_{a|x}(\varrho) = m_a^{(x)}(M(\varrho)) = m_a^{(x)}(\tilde{N}(\Phi(\varrho))) = \sum_{\lambda=1}^\Lambda N_{a|x,\lambda}(\Phi_\lambda(\varrho)) = \left(\sum_{\lambda=1}^\Lambda \Phi_\lambda^*(N_{a|x,\lambda}) \right)(\varrho) \quad (4.11)$$

for all $a \in [k]$, $x \in [g]$ and $\varrho \in K_A$. $\square$

4.2. Classical simulation of multimeters. In this section, we consider a different notion of simulation of measurements, purely in terms of classical mixing and post-processing of measurements [OGWA17, GBTC17, FHL18].

Definition 4.4. Let $M = \{M_{\cdot|x}\}_{x \in [g]}$ be a multimeter of g measurements with k outcomes on a state space K . We say that M can be classically simulated (or is classically simulable) with a multimeter $N = \{N_{\cdot|y}\}_{y \in [r]}$ of r measurements with l outcomes on K if there exist conditional probability distributions $\pi = (\pi_{\cdot|x})_{x \in [g]}$ on $[r]$ and $\nu = (\nu_{\cdot|b,x,y})_{b \in [l], x \in [g], y \in [r]}$ on $[k]$ such that

$$M_{a|x} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} N_{b|y} \quad (4.12)$$

for all $a \in [k]$ and $x \in [g]$.

The operational interpretation of classical simulation is the following: given a measurement label x , we measure the input state $\varrho \in K$ with the measurement $N_{\cdot|y}$ with probability $\pi_{y|x}$ from which we obtain a measurement outcome b . Instead of registering this outcome we report an outcome a with probability $\nu_{a|b,x,y}$ as the final outcome of the simulated measurement with the label x .

It turns out that we can also express classical simulability as factorization of the multimeter through a CS of different size. For this we need to first characterize the channels between two sets of column stochastic matrices of different size.

Theorem 4.5. *A linear map $\Phi : CS_{l,r}^1 \rightarrow CS_{k,g}$ is a channel if and only if there exist conditional probability distributions $\pi = (\pi_{\cdot|x})_{x \in [g]}$ on $[r]$ and $\nu = (\nu_{\cdot|b,x,y})_{b \in [l], x \in [g], y \in [r]}$ on $[k]$ such that*

$$\Phi(X) = \mathbb{1}_{CS_{l,r}^1}(X) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} \left[\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)}(X) \right] e_a^{(x)} \quad (4.13)$$

for all $X \in CS_{l,r}$.

Proof. As was explained in Section 3.5, for a state space K , a linear map $\Phi : V(K) \rightarrow CS_{k,g}$ is a channel, i.e., a multimeter, if and only if there exist g measurements $\{A_{\cdot|x}\}_{x \in [g]}$ with k outcomes such that

$$\Phi(\varrho) = \mathbb{1}_K(\varrho) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} A_{a|x}(\varrho) e_a^{(x)} \quad (4.14)$$

for all $\varrho \in V(K)$.

Let us now take $K = CS_{l,r}^1$. Since now $A_{\cdot|x}$ is a measurement on $CS_{l,r}^1$ for each $x \in [g]$ we can write it as

$$A_{\cdot|x}(\varrho) = \sum_{a=1}^k A_{a|x}(\varrho) \delta_a \quad (4.15)$$

for all $\varrho \in CS_{l,r}^1$, where $A_{1|x}, \dots, A_{k|x}$ are the effects of $A_{\cdot|x}$. Since $A_{a|x}$ is an effect, we know that in particular $A_{a|x} \in (CS_{l,r}^+)^*$. Recall that the $m_b^{(y)}$ generate the extreme rays of $(CS_{l,r}^+)^*$. Therefore, since $(CS_{l,r}^+)^*$ is a convex cone, it follows that we can write $A_{a|x}$ as

$$\forall x \in [g], \forall a \in [k] \quad A_{a|x} = \sum_{y=1}^r \sum_{b=1}^l \gamma_{a,b,x,y} m_b^{(y)} \quad (4.16)$$

for some $\gamma_{a,b,x,y} \geq 0$. Positivity of Φ imposes the following constraints:

$$\forall x \in [g], \forall a \in [k], \forall \vec{b} \in [l]^y \quad A_{a|x}(s_{b_1,\dots,b_r}) = \sum_{y=1}^r \gamma_{a,b_y,x,y} \geq 0 \quad (4.17)$$

and the fact that Φ maps states to states:

$$\forall x \in [g], \forall \vec{b} \in [l]^y \quad \sum_{a=1}^k A_{a|x}(s_{b_1,\dots,b_r}) = \sum_{a=1}^k \sum_{y=1}^r \gamma_{a,b_y,x,y} = 1. \quad (4.18)$$

Now fix some $y_0 \in [r]$ and let us take $b_y = l$ for all $y \neq y_0$ in the above equation. Then

$$\forall x \in [g], \forall b \in [l] \quad \sum_{a=1}^k \gamma_{a,b,x,y_0} = 1 - \sum_{a=1}^k \sum_{y \in [r] \setminus \{y_0\}} \gamma_{a,l,x,y}. \quad (4.19)$$

As the right hand side is independent of b , it follows that we can define

$$\pi_{y_0|x} := \sum_{a=1}^k \gamma_{a,b,x,y_0} \quad (4.20)$$

for some $b \in [l]$ and for all $x \in [g]$ and $y_0 \in [r]$. Since $y_0 \in [r]$ was arbitrary, it follows from Eq. (4.18) that $\pi := (\pi_{\cdot|x})_{x \in [g]}$ is a conditional probability distribution on $[r]$. Setting

$$\forall x \in [g], \forall a \in [k], \forall y \in [r], \forall b \in [l] \quad \nu_{a|b,x,y} := \begin{cases} \gamma_{a,b,x,y}/\pi_{y|x}, & \pi_{y|x} \neq 0 \\ 1/k, & \pi_{y|x} = 0, \end{cases} \quad (4.21)$$

it follows that $\nu = (\nu_{\cdot|b,x,y})_{b \in [l], x \in [g], y \in [r]}$ is a conditional probability distribution on $[k]$. Hence, we may rewrite $A_{a|x}$ as

$$A_{a|x} = \sum_{y=1}^r \sum_{b=1}^l \gamma_{a,b,x,y} m_b^{(y)} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)} \quad (4.22)$$

for all $a \in [k]$ and $x \in [g]$ so that by Eq. (4.14) we have that

$$\Phi(X) = \mathbb{1}_{CS_{l,r}^1}(X) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} \left[\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)}(X) \right] e_a^{(x)} \quad (4.23)$$

for all $X \in CS_{l,r}$. For the converse, it is straightforward to verify that any map as in Eq. (4.13) is a channel. $\square$

We can use the previous result to conclude that every factorization through a state space of column stochastic matrices is given by some simulation scheme.

Corollary 4.6. *A multimeter $M = \{M_{\cdot|x}\}_{x \in [g]}$ of g measurements with k outcomes on a state space K can be classically simulated with a multimeter $N = \{N_{\cdot|y}\}_{y \in [r]}$ of r measurements with l outcomes on K if and only if there exists a channel $\Phi : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ such that the following diagram commutes:*

$$\begin{array}{ccc} K & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow N & \nearrow \Phi \\ & CS_{l,r}^1 & \end{array} \quad (4.24)$$

Proof. If M can be classically simulated with N , then the conditional probability distributions π and ν in Eq. (4.12) can be used to define a channel $\Phi : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ by using Eq. (4.13) of Theorem 4.5. Then

$$\Phi(N(\varrho)) = \mathbb{1}_{CS_{l,r}^1}(N(\varrho)) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} \left[\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)}(N(\varrho)) \right] e_a^{(x)} \quad (4.25)$$

$$= \mathbb{1}_K(\varrho) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} \left[\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} N_{y|b}(\varrho) \right] e_a^{(x)} \quad (4.26)$$

$$= \mathbb{1}_K(\varrho) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} M_{a|x}(\varrho) e_a^{(x)} \quad (4.27)$$

$$= M(\varrho) \quad (4.28)$$

for all $\varrho \in V(K)$.

On the other hand, if there exists a channel $\Phi : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ such that $M = \Phi \circ N$, then there exists conditional probability distributions π and ν as in Theorem 4.5 such that Eq. (4.25)-Eq. (4.28) hold. In particular, comparing the basis decompositions in Eq. (4.26) and Eq. (4.27) one immediately sees that $M_{a|x} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} N_{y|b}$ for all $a \in [k]$ and $x \in [g]$ so that M can be classically simulated with N . $\square$

As a special case of Corollary 4.6 we get the following known characterization of measurement incompatibility [Jen18].

Corollary 4.7. *Let $M = \{M_{\cdot|x}\}_{x \in [g]}$ be a multimeter of g measurements with k outcomes on a state space K . Then, the following are equivalent:*

- (1) $(\text{id} \otimes M)(\chi_{V(K)}) \in K_{\varrho}^* \dot{\otimes} CS_{k,g}^1$ for some ϱ in the relative interior of K
- (2) M consists of compatible measurements
- (3) There exist a finite number of outcomes Λ , a measurement $N : K \rightarrow S_{\Lambda}$, and a channel $\Phi : S_{\Lambda} \rightarrow CS_{k,g}^1$ such that the following diagram commutes:

$$\begin{array}{ccc} K & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow N & \nearrow \Phi \\ & S_{\Lambda} & \end{array} \quad (4.29)$$

Proof. The equivalence between (1) and (3) follows from Proposition 3.7. The equivalence between (2) and (3) follows directly from Corollary 4.6 with $r = 1$, as the factorization is equivalent by the corollary to

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} \nu_{a|\lambda,x} N_{\lambda} \quad \forall x \in [g], \forall a \in [k]. \quad (4.30)$$

□

If we add an additional classical system S_{Λ} into the factorization of Corollary 4.6 then we get the following characterization:

Corollary 4.8. *For a multimeter $M = \{M_{\cdot|x}\}_{x \in [g]}$ of g measurements with k outcomes on a state space K there exist channels $N : K \rightarrow CS_{l,r}^1 \dot{\otimes} S_{\Lambda}$ and $\tilde{\Phi} : CS_{l,r}^1 \dot{\otimes} S_{\Lambda} \rightarrow CS_{k,g}^1$ such that the following diagram commutes*

$$\begin{array}{ccc} K & \xrightarrow{M} & CS_{k,g}^1 \\ & \searrow N & \nearrow \tilde{\Phi} \\ & CS_{l,r}^1 \dot{\otimes} S_{\Lambda} & \end{array} \quad (4.31)$$

if and only if there exist conditional probability distributions $\pi = (\pi_{\cdot|x,\lambda})_{x \in [g], \lambda \in [\Lambda]}$ on $[r]$ and $\nu = (\nu_{\cdot|b,x,\lambda,y})_{b \in [l], x \in [g], \lambda \in [\Lambda], y \in [r]}$ on $[k]$ and a multimeter $N = \{N_{\cdot,|y}\}_{y \in [r]}$ consisting of r measurements with $l \cdot \Lambda$ outcomes on K satisfying

$$\sum_{b=1}^l N_{b,\lambda|y} = \sum_{b=1}^l N_{b,\lambda|y'} \quad (4.32)$$

for all $\lambda \in [\Lambda]$ and $y, y' \in [r]$ such that

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} \sum_{y=1}^r \pi_{y|x,\lambda} \sum_{b=1}^l \nu_{a|b,x,y,\lambda} N_{b,\lambda|y} \quad (4.33)$$

for all $a \in [k]$ and $x \in [g]$.

Proof. Let us first assume that the diagram commutes. Per Lemma 4.3 we can use $\tilde{\Phi} : CS_{l,r}^1 \dot{\otimes} S_{\Lambda} \rightarrow CS_{k,g}^1$ to define a channel $\Phi : CS_{l,r}^1 \rightarrow CS_{k,g,\Lambda}^1$ so that by Theorem 4.5 we can write it as

$$\Phi(X) = \mathbb{1}_{CS_{l,r}^1}(X) s_{k,\dots,k} + \sum_{\lambda=1}^{\Lambda} \sum_{x=1}^g \sum_{a=1}^{k-1} \left[\sum_{y=1}^r \pi_{y|x,\lambda} \sum_{b=1}^l \nu_{a|b,x,y,\lambda} m_b^{(y)}(X) \right] e_a^{(x,\lambda)} \quad (4.34)$$

for all $X \in CS_{l,r}$. By Theorem 4.2 together with the previous expression we have that

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} N_{\lambda}^*(\Phi_{a|x,\lambda}) = \sum_{\lambda=1}^{\Lambda} \Phi_{a|x,\lambda} \circ N_{\lambda} = \sum_{\lambda=1}^{\Lambda} \left[\sum_{y=1}^r \pi_{y|x,\lambda} \sum_{b=1}^l \nu_{a|b,x,\lambda,y} m_b^{(y)} \right] \circ N_{\lambda}, \quad (4.35)$$

where $N_{\lambda} : V(K)^+ \rightarrow CS_{l,r}^+$ are the operations of the instrument N , which we can write as

$$N_{\lambda}(X) = \mathbb{1}_{CS_{l,r}^1}(N_{\lambda}(X))_{s_{l,\dots,l}} + \sum_{y=1}^r \sum_{b=1}^{l-1} N_{b,\lambda|y}(X) e_b^{(y)} \quad (4.36)$$

for all $\lambda \in [\Lambda]$ and $X \in V(K)$. Here, $N_{b,\lambda|y} \in (V(K)^+)^*$ are positive functionals on K for all $b \in [l-1]$, $\lambda \in [\Lambda]$ and $y \in [r]$. Let us define $N_{l,\lambda|y} := \mathbb{1}_{CS_{l,r}^1} \circ N_{\lambda} - \sum_{b=1}^{l-1} N_{b,\lambda|y} \in (V(K)^+)^*$ for all $\lambda \in [\Lambda]$ and $y \in [r]$. Now we have that

$$\forall y \in [r], \forall \lambda \in [\Lambda] : \sum_{b=1}^l N_{b,\lambda|y} = \mathbb{1}_{CS_{l,r}^1} \circ N_{\lambda}, \quad (4.37)$$

$$\forall y \in [r] : \sum_{\lambda=1}^{\Lambda} \sum_{b=1}^l N_{b,\lambda|y} = \mathbb{1}_K, \quad (4.38)$$

where the second equation follows from the fact the the operations N_{λ} form an instrument. Thus, the effects $N_{b,\lambda|y}$ form a multimeter $N : K \rightarrow CS_{l,\Lambda,r}^1$ such that $\sum_{b \in [l]} N_{b,\lambda|y} = \sum_{b \in [l]} N_{b,\lambda|y'}$ for all $\lambda \in [\Lambda]$ and $y, y' \in [r]$. It follows from Eq. (4.35) and Eq. (4.36) that

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} \sum_{y=1}^r \pi_{y|x,\lambda} \sum_{b=1}^l \nu_{a|b,x,\lambda,y} N_{b,\lambda|y} \quad (4.39)$$

for all $a \in [k]$ and $x \in [g]$.

On the other hand if Eq. (4.33) holds for some conditional probability distributions π and ν and some multimeter $N : K \rightarrow CS_{l,\Lambda,r}^1$ such that $\sum_{b \in [l]} N_{b,\lambda|y} = \sum_{b \in [l]} N_{b,\lambda|y'}$ for all $\lambda \in [\Lambda]$ and $y, y' \in [r]$, then we can use Eq. (4.34) and Eq. (4.36) to define a channel $\Phi : CS_{l,r}^1 \rightarrow CS_{k,l,\Lambda}^1$ and an instrument $N : K \rightarrow CS_{l,r}^1 \dot{\otimes} S_{\Lambda}$, respectively. By Lemma 4.3 we can use Φ to define a channel $\tilde{\Phi} : CS_{l,r}^1 \dot{\otimes} S_{\Lambda} \rightarrow CS_{k,g}^1$ so that then by Eq. (4.33) we can see that Eq. (4.35) must hold so that by Theorem 4.2 the diagram commutes. $\square$

We will next explore classical simulability a bit more with the aim of providing more operational intuition behind Theorem 4.5. In particular, we will give an operational proof for that result solely based on classical simulation showing that it is no coincidence that a channel between two spaces of column stochastic matrices is characterized by classical simulation. First we need to explore some concepts related to classical simulability.

Naturally every measurement can be used to trivially classically simulate itself. Following [FHL18] we next look into measurements for which this is the only way to simulate them (up to postprocessing equivalence).

Definition 4.9. A measurement A is simulation irreducible if for any multimeter M that can be used to simulate A there exists some measurement in M that is postprocessing equivalent with A .

Remark 4.10. Here by postprocessing equivalent measurements we mean the case when two measurements A and B can be postprocessed from each other via some conditional probability distributions ν and μ so that $A_i = \sum_j \nu_{i|j} B_j$ and $B_j = \sum_i \mu_{j|i} A_i$. In fact, then it follows that the set of measurements can be partitioned into equivalence classes of measurements and in many cases it is

convenient to consider properties of measurements only between different postprocessing equivalence classes.

As already the name suggests, it turns out that every measurement can be reduced to a classical simulation of some set of simulation irreducible measurements [FHL18].

Proposition 4.11 ([FHL18]). *For every multimeter M there exists a multimeter consisting only of simulation irreducible measurements which can classically simulate M .*

Thus, in terms of understanding classical simulability studying the simulation irreducible measurements is extremely useful. It can be shown that in the postprocessing equivalence class of each simulation irreducible measurement there exists a unique extremal measurement which we take as representative of the equivalence class (and which is the same as the *minimally sufficient representative* in [Kur15]). In [FHL18] the following characterization was given:

Proposition 4.12 ([FHL18]). *A measurement is simulation irreducible and extremal if and only if it consists of linearly independent indecomposable effects.*

Now we can use the above result to characterize all the extremal simulation irreducible measurements on the state space of column stochastic matrices.

Proposition 4.13. *The k -outcome measurements $m^{(1)}, \dots, m^{(g)}$ defined in Eq. (3.20) are the only extremal simulation irreducible measurements on $CS_{k,g}^1$.*

Proof. First, let us note that each $m^{(x)}$ consists of linearly independent indecomposable effects so by Proposition 4.12 it is simulation irreducible and extremal for all $x \in [g]$.

On the other hand, let us look into the structure of (other possible) simulation irreducible measurements on $CS_{k,g}^1$. Since simulation irreducible measurements consist of indecomposable effects and since the effects of the measurements $m^{(1)}, \dots, m^{(g)}$ are exactly all the extremal indecomposable effects, without loss of generality we can represent an arbitrary extremal indecomposable measurement $\tilde{f}$ as having the outcome set $[g] \times [k]$ such that $\tilde{f}_{(x,a)} = \alpha_a^{(x)} m_a^{(x)}$ for some $\alpha_a^{(x)} \in [0, 1]$ for all $a \in [k]$ for all $x \in [g]$. Since $\tilde{f}$ and each $m^{(x)}$ are measurements the normalization $\sum_{x \in [g]} \sum_{a \in [k]} \tilde{f}_{(x,a)} = \mathbb{1}_{CS_{k,g}^1} = \sum_{a \in [k]} m_a^{(x)}$ holds for all $x \in [g]$ so that we have

$$0 = \mathbb{1}_{CS_{k,g}^1} - \sum_{x=1}^g \sum_{a=1}^k \alpha_a^{(x)} m_a^{(x)} = \left(1 - \sum_{x=1}^g \alpha_k^{(x)}\right) \mathbb{1}_{CS_{k,g}^1} + \sum_{x=1}^g \sum_{a=1}^{k-1} (\alpha_k^{(x)} - \alpha_a^{(x)}) m_a^{(x)}. \quad (4.40)$$

Now, since the set $\{\mathbb{1}_{CS_{k,g}^1}, m_1^{(1)}, \dots, m_{k_g-1}^{(1)}, \dots, m_1^{(g)}, \dots, m_{k_g-1}^{(g)}\}$ forms a basis of $CS_{k,g}^*$ so that in particular the set is linearly independent, we have that $\alpha_a^{(x)} = \alpha_{a'}^{(x)} =: \alpha^{(x)}$ for all $a, a' \in [k]$ for all $x \in [g]$ and that $\sum_{x \in [g]} \alpha^{(x)} = 1$.

Since $\tilde{f}$ is extremal it consist of linearly independent effects. Thus, if we denote $G_{\neq 0} = \{x \in [g] \mid \alpha^{(x)} \neq 0\}$, which is clearly non-empty since $\tilde{f}$ is a measurement, then $\{\tilde{f}_{(x,a)}\}_{a \in [k], x \in G_{\neq 0}}$ is a set of linearly independent elements. Let us now fix $x \in G_{\neq 0}$. Now we see that

$$\frac{1}{\alpha^{(x)}} \sum_{a=1}^k \tilde{f}_{(x,a)} = \frac{1}{\alpha^{(x)}} \sum_{a=1}^k \alpha_a^{(x)} m_a^{(x)} = \mathbb{1}_{CS_{k,g}^1} = \sum_{y=1}^g \alpha^{(y)} \mathbb{1}_{CS_{k,g}^1} = \sum_{y=1}^g \alpha^{(y)} \sum_{a=1}^k m_a^{(y)} = \sum_{y \in G_{\neq 0}} \sum_{a=1}^k \tilde{f}_{(y,a)}. \quad (4.41)$$

By rearranging the terms we have that

$$0 = \alpha^{(x)} \sum_{y \in G_{\neq 0} \setminus \{x\}} \sum_{a=1}^k \tilde{f}_{(y,a)} + (\alpha^{(x)} - 1) \sum_{a=1}^k \tilde{f}_{(x,a)}. \quad (4.42)$$

Since $\{\tilde{f}_{(x,a_x)}\}_{a_x \in [k_x], x \in G_{\neq 0}}$ is a set of linearly independent elements and since we have that $\alpha^{(x)} \neq 0$ we must have that $G_{\neq 0} \setminus \{x\} = \emptyset$, i.e., $G_{\neq 0} = \{x\}$, so that by the fact that $\sum_{y \in [g]} \alpha^{(y)} = 1$ we in fact have that $\alpha^{(x)} = 1$ and thus the only nonzero effects of $\tilde{f}$ are $\tilde{f}_{(x,a)} = m_a^{(x)}$ for all $a \in [k]$. This concludes the proof. $\square$

Together with Proposition 4.11 we can use the above result for a more operational proof of Theorem 4.5.

Alternative proof for Theorem 4.5. As we have already argued before, a linear map $\Phi : CS_{l,r} \rightarrow CS_{k,g}$ is a channel (or a multimeter) if and only if there exist g measurements $\{A_{\cdot|x}\}_{x \in [g]}$ with k outcomes such that

$$\Phi(\varrho) = \mathbb{1}_{CS_{l,r}^1}(\varrho) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} A_{a|x}(\varrho) e_a^{(x)} \quad (4.43)$$

for all $\varrho \in CS_{l,r}$. Since now $A_{\cdot|x}$ is a measurement on $CS_{l,r}^1$ for each $x \in [g]$ we can write it as

$$A_{\cdot|x}(\varrho) = \sum_{a=1}^k A_{a|x}(\varrho) \delta_a \quad (4.44)$$

for all $\varrho \in CS_{l,r}^1$, where $A_{1|x}, \dots, A_{k|x}$ are the effects of $A_{\cdot|x}$.

Since $A_{\cdot|x}$ is a measurement on $CS_{l,r}^1$ for each $x \in [g]$, by Proposition 4.11 there exists some multimeter $M^{(x)}$ consisting of simulation irreducible measurements on $CS_{l,r}^1$ which can be used to simulate $A_{\cdot|x}$ for all $x \in [g]$. Since every simulation irreducible measurement is postprocessing equivalent with an extremal simulation irreducible measurement we can take $M^{(x)}$ to contain only extremal simulation irreducible measurements. Furthermore, since $M^{(x)}$ can be used to simulate $A_{\cdot|x}$ also the multimeter M which consists of all of the extremal simulation irreducible measurements on $CS_{l,r}^1$ can be used to simulate $A_{\cdot|x}$ for all $x \in [g]$. By Proposition 4.13 the multimeter M thus consists exactly of the measurements $m^{(1)}, \dots, m^{(r)}$ defined in Eq. (3.20). By the definition of classical simulability there thus exists some conditional probability distributions $\pi = (\pi_{\cdot|x})_{x \in [g]}$ on $[r]$ and $\nu = (\nu_{\cdot|b,x,y})_{b \in [l], x \in [g], y \in [r]}$ on $[k]$ such that

$$A_{a|x} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)} \quad (4.45)$$

for all $a \in [k]$ and $x \in [g]$. Then by Eq. (4.43) we have that

$$\Phi(X) = \mathbb{1}_{CS_{l,r}^1}(X) s_{k,\dots,k} + \sum_{x=1}^g \sum_{a=1}^{k-1} \left[\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)}(X) \right] e_a^{(x)} \quad (4.46)$$

for all $X \in CS_{l,r}$. Again, the converse is straightforward to verify. $\square$

5. STEERING

5.1. Steering in quantum mechanics and general probabilistic theories. One form of quantum nonlocality that is closely related to measurement incompatibility is quantum steering. Therefore, we will now show how a steering assemblage can be understood in the formalism used so far. We will first stick with steering in quantum mechanics and then generalize it to GPTs. Proofs will come at the end of the section, as it is convenient to prove everything in the GPT formalism and to recover quantum mechanics as a special case.

Let us consider a multimeter consisting of g POVMs acting on a d -dimensional system with k outcomes each and a bipartite state $\varrho \in D(\mathbb{C}^{d_A d_B})$. They enable us to construct the following object:

$$[(M \otimes \text{id})(\varrho)]_{a|x} = \text{Tr}_A[(M_{a|x} \otimes \mathbb{1}_{K_B})\varrho] = \sigma_{a|x} \quad (5.1)$$

This object is called an *assemblage*. Assemblages in quantum mechanics are defined as sets of positive operators $\sigma := \{\sigma_{a|x}\}_{a \in [k], x \in [g]} \subset \text{PSD}_{d_B}$ with the property that there exists a $\bar{\sigma} \in D(\mathbb{C}^{d_B})$ such that

$$\bar{\sigma} = \sum_{a=1}^k \sigma_{a|x} \quad \forall x \in [g]. \quad (5.2)$$

In Eq. (5.1), we can easily verify that $\bar{\sigma} = \text{Tr}_A(\varrho)$ and that indeed $\sigma_{a|x} \geq 0$ for all $a \in [k]$, $x \in [g]$ as required.

The interpretation of an assemblage is that it corresponds to a probabilistic state preparator such that given a classical input x the state $\hat{\sigma}_{a|x} := \sigma_{a|x}/p_{a|x}$ is prepared with probability $p_{a|x} = \text{Tr}[\sigma_{a|x}]$. The condition of Eq. (5.2) then means that without knowing the classical output a , i.e., when averaging over all the possible states, we can have no knowledge of the classical input x . Now Eq. (5.1) is a particular realization of the assemblage in the two-party setting so that the states are prepared by performing a measurement on one part of a joint state, and in this case the condition of Eq. (5.2) corresponds to a no-signaling condition so that without having access to the measurement outcome the party with the prepared states cannot know which measurement setting was chosen on the other part of the shared state.

As for multimeters, we can now consider assemblages as tensors:

Lemma 5.1. *Given a multimeter $M : D(\mathbb{C}^{d_A}) \rightarrow \text{CS}_{k,g}^1$ and a bipartite state $\varrho \in D(\mathbb{C}^{d_A d_B})$, the assemblage obtained as $\sigma = (M \otimes \text{id})(\varrho)$ in Eq. (5.1) are elements of the set $\text{CS}_{k,g}^1 \hat{\otimes} D(\mathbb{C}^{d_B})$.*

The statement follows from the fact that $\varrho \in D(\mathbb{C}^{d_A d_B}) \subset D(\mathbb{C}^{d_A}) \hat{\otimes} D(\mathbb{C}^{d_B})$. In fact, it is known that the converse also holds. From the GHJW-theorem [Gis89, HJW93], it follows that for each assemblage, there is a set of measurements and a bipartite state such that the assemblage is obtained as in Eq. (5.1).

Lemma 5.2. *Given an element $\sigma \in \text{CS}_{k,g}^1 \hat{\otimes} D(\mathbb{C}^{d_B})$, for $d_A \geq d_B$ there exists a multimeter $M : D(\mathbb{C}^{d_A}) \rightarrow \text{CS}_{k,g}^1$ and a bipartite state $\varrho \in D(\mathbb{C}^{d_A d_B})$ such that $\sigma = (M \otimes \text{id})(\varrho)$.*

Since the assemblages are in one-to-one correspondence with elements in the maximal tensor product of quantum states and column-stochastic matrices, the question arises what happens if a certain assemblage is separable, i.e., part of the minimal tensor product of these state spaces. This question lies at the core of this work and the answer will follow from our results at the end of this section. We recover the result that separable elements are in one-to-one correspondence with assemblages admitting a local hidden state (LHS) model [Jen18, Theorem 4].

Proposition 5.3. *For an assemblage σ , it holds that $\sigma \in \text{CS}_{k,g}^1 \dot{\otimes} D(\mathbb{C}^d)$ if and only if there exists an LHS model that reproduces it. By LHS model we mean that there exists a $\Lambda \in \mathbb{N}$, operators $\{B_\lambda\}_{\lambda \in [\Lambda]} \subset \text{PSD}_d$ and a conditional probability distribution $p = (p_{\cdot|\lambda,x})_{x \in [g], \lambda \in [\Lambda]}$ on $[k]$ such that*

$$\sigma_{a|x} = \sum_{\lambda=1}^{\Lambda} p_{a|\lambda,x} B_\lambda \quad (5.3)$$

for all $a \in [k]$ and $x \in [g]$.

The proposition follows in particular from Corollary 5.21. For a recent review of quantum steering, we refer the reader to [UCNG20].

With this preparation, we can define steering in the GPT setting. Let us choose a GPT $(V(K), V(K)^+, \mathbb{1}_K)$ for what follows. We start with the definition of an assemblage.

Definition 5.4. Let $(V(K), V(K)^+, \mathbb{1}_K)$ be a GPT. An assemblage is a set $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ of states $\varrho_{a|x} \in K$ and conditional probability distributions $(q_{a|x})_{a \in [k]}$ for all $x \in [g]$ together with an average state $\bar{\varrho} \in K$ such that

$$\bar{\varrho} = \sum_{a \in [k]} q_{a|x} \varrho_{a|x} \quad \forall x \in [g]. \quad (5.4)$$

As was explained earlier in the quantum setting, the interpretation of an assemblage is that it corresponds to a probabilistic state preparator where given a classical input x a state $\varrho_{a|x} \in K$ is prepared with probability $q_{a|x}$. For an assemblage $\varrho = \{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ we may also write $\varrho = \{\tilde{\varrho}_{a|x}\}_{a \in [k], x \in [g]}$ where $\tilde{\varrho}_{a|x} = q_{a|x} \varrho_{a|x} \in V(K)^+$. Since the state space K forms a base for the cone $V(K)^+$ and $\mathbb{1}(\tau) = 1$ for all $\tau \in K$, we can recover the conditional probabilities q simply as $q_{a|x} = \mathbb{1}(\tilde{\varrho}_{a|x})$. Thus, we may go back and forth between expressing the assemblage as probabilities together with the states and expressing the assemblages just with the unnormalized states. We use both notations interchangeably.

The definition for assemblages having an LHS model can be straightforwardly generalized:

Definition 5.5. Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage. Then, the assemblage admits an LHS model if there exist $\Lambda \in \mathbb{N}$, a set of states $\{\sigma_\lambda\}_{\lambda \in [\Lambda]}$, a probability distribution $(p_\lambda)_{\lambda \in [\Lambda]}$ and conditional probability distributions $(\nu_{a|\lambda, x})_{a \in [k]}$ for all $\lambda \in [\Lambda]$, $x \in [g]$, such that

$$q_{a|x} \varrho_{a|x} = \sum_{\lambda=1}^{\Lambda} \nu_{a|\lambda, x} p_\lambda \sigma_\lambda \quad \forall a \in [k], \forall x \in [g]. \quad (5.5)$$

Theorem 4 of [Jen18] states that for any $\xi \in CS_{k,g}^1 \hat{\otimes} K$, there is an assemblage $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ such that

$$\xi = s_{k, \dots, k} \otimes \bar{\varrho} + \sum_{x=1}^g \sum_{a=1}^{k-1} q_{a|x} e_a^{(x)} \otimes \varrho_{a|x}. \quad (5.6)$$

Conversely, any element of the form in Eq. (5.6) is in $CS_{k,g}^1 \hat{\otimes} K$, such that there is a one-to-one correspondence between assemblages and tensors in $CS_{k,g}^1 \hat{\otimes} K$. This implies in particular Lemma 5.1.

5.2. Assemblages as measurements on bipartite states. The considerations at the end of the previous section, however, do not give us a counterpart to Lemma 5.2, because Definition 5.4 does not make any reference to applying a multimeter to any bipartite state. We will start with a purely mathematical construction to work around this. As discussed in Section 3.2, any $\xi \in CS_{k,g}^1 \hat{\otimes} K$ can be written as $\xi = (\text{id} \otimes \Phi)(\chi_{CS_{k,g}^*}) = (\Phi^* \otimes \text{id})(\chi_{V(K)})$, where $\Phi : (CS_{k,g}^*)^+ \rightarrow V(K)^+$ such that (see Eq. (5.6))

$$\Phi(\mathbb{1}_{CS_{k,g}^1}) = \bar{\varrho} \quad (5.7)$$

$$\Phi(m_a^{(x)}) = q_{a|x} \varrho_{a|x} \quad \forall x \in [g], \forall a \in [k]. \quad (5.8)$$

The dual map $\Phi^* : A(K)^+ \rightarrow CS_{k,g}^+$ is such that for any $\alpha \in A(K)^+$,

$$\Phi^* : \alpha \mapsto \alpha(\bar{\varrho}) s_{k, \dots, k} + \sum_{x=1}^g \sum_{a=1}^{k-1} q_{a|x} \alpha(\varrho_{a|x}) e_a^{(x)}. \quad (5.9)$$

Now, we would like to see Φ^* as a multimeter, i.e., we need a state space in $A(K)^+$ which Φ^* maps to $CS_{k,g}^1$ and hence makes Φ^* a channel. Let us assume that $\bar{\varrho}$ is in the relative interior of K . Then, we consider $K_{\bar{\varrho}}^*$ defined in Eq. (3.9). It follows from the definition of $K_{\bar{\varrho}}^*$ that for all $\alpha \in K_{\bar{\varrho}}^*$

$$\mathbb{1}_{CS_{k,g}^1}(\Phi^*(\alpha)) = \alpha(\bar{\varrho}) = 1, \quad (5.10)$$

thus indeed $\Phi^* : K_{\bar{\varrho}}^* \rightarrow CS_{k,g}^1$ is a channel and hence a multimeter.

Remark 5.6. *The assumption that $\bar{\varrho}$ is in the relative interior of K is no restriction, since otherwise K is in a sense too large. More precisely, if $\bar{\varrho}$ is not in the relative interior of K , then it is in a face of the convex set K [Roc70, Theorem 18.2]. Let F be the smallest face that contains $\bar{\varrho}$. This implies that $\bar{\varrho}$ is in the relative interior of F , since otherwise there would be a smaller face containing $\bar{\varrho}$. F is compact because K is. Therefore, we can define a GPT $(V(F), V(F)^+, \mathbb{1}_F)$ from F . As $\bar{\varrho} \in F$ and $\bar{\varrho} = \sum_{a=1}^k q_{a|x} \varrho_{a|x}$ for all $x \in [g]$, it follows that $\varrho_{a|x} \in F$ for all $a \in [k]$, $x \in [g]$. Indeed, for $0 < q_{a|x} < 1$ this is a consequence of F being a face. For $q_{a|x} = 1$, it follows that $\varrho_{a|x} = \bar{\varrho}$ and for $q_{a|x} = 0$, we can choose $\varrho_{a|x}$ any way we like. Thus, $\Phi((CS_{k,g}^*)^+) \subseteq V(F)^+$ and we can consider Φ as a map $\Phi : (CS_{k,g}^*)^+ \rightarrow V(F)^+$ instead.*

We can now more generally raise the question if there always exists a realization as in Eq. (5.1) for a given steering assemblage as it is the case in quantum theory. To be more precise, given a state space K_A we consider an assemblage $\sigma = \{\sigma_{a|x}\}_{a \in [k], x \in [g]} \subset V(K_A)^+$ with the property that $\sum_{a \in [k]} \sigma_{a|x} = \bar{\sigma} \in K_A$ for all $x \in [g]$. Then, there are two questions:

- (1) Is there a state space K_B , a tensor product $\otimes$, a state $z \in K_A \otimes K_B$ and a multimeter $M : K_B \rightarrow CS_{k,g}^1$ such that $\sigma_{a|x} = (\text{id} \otimes M_{a|x})(z)$?
- (2) Is there a tensor product $\otimes$, a state $w \in K_A \otimes K_A$ and a multimeter $N : K_A \rightarrow CS_{k,g}^1$ such that $\sigma_{a|x} = (\text{id} \otimes N_{a|x})(w)$?

The answer to these questions will show that the existence of a realization cannot be taken for granted and depends on the second state space.

If we may choose K_B freely, we can give a suitable choice and thus a positive answer to question (1) in the following proposition. This already follows from the considerations at the beginning of this section, but we will give a simpler proof below.

Proposition 5.7. *An assemblage $\sigma = \{\sigma_{a|x}\}_{a \in [k], x \in [g]} \subset V(K_A)^+$, with an average state $\sum_a \sigma_{a|x} = \bar{\sigma}$, always finds a realization, i.e., $\sigma_{a|x} = (\text{id} \otimes M_{a|x})(z)$ with a state $z \in K_A \otimes K_B$ and a multimeter $M : K_B \rightarrow CS_{k,g}^1$ if K_B is chosen as $K_B := (K_A)_{\bar{\sigma}}^*$ and $\bar{\sigma}$ is in the relative interior of K_A .*

Proof. Let $\sigma = \{\sigma_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with $\sum_a \sigma_{a|x} = \bar{\sigma}$. We consider a trivial map $\varphi_{\text{id}} : A(K_A)^+ \rightarrow A(K_A)^+$, that is, as a tensor $\xi_{\varphi_{\text{id}}} = \chi_{A(K_A)} \in V(K_A)^+ \hat{\otimes} A(K_A)^+$ (see Eq. (3.7)). If we could choose $M_{a|x} = \sigma_{a|x}$, we would be done since then by Eq. (3.8)

$$(\text{id} \otimes M_{a|x})(\xi_{\varphi_{\text{id}}}) = \sigma_{a|x}. \quad (5.11)$$

Thus, we set $K_B = (K_A)_{\bar{\sigma}}^*$ and the assemblages become functionals on the states in K_B . We recall from Section 3.2 that $\bar{\sigma}$ is the order unit for the GPT constructed from the state space $(K_A)_{\bar{\sigma}}^*$. Moreover, for all $a \in [k]$, $x \in [g]$ it holds that $\sigma_{a|x} \in A((K_A)_{\bar{\sigma}}^*)^+$, as $A((K_A)_{\bar{\sigma}}^*)^+ = V(K_A)^+$. Thus, setting $M_{a|x} := \sigma_{a|x}$ indeed gives a valid multimeter acting on $(K_A)_{\bar{\sigma}}^*$. It remains for us to verify that $\xi_{\varphi_{\text{id}}} \in K_A \hat{\otimes} K_B$: Using again that the order unit on $(K_A)_{\bar{\sigma}}^*$ is $\bar{\sigma}$, we can confirm that

$$\mathbb{1}_{K_A} \otimes \mathbb{1}_{K_B}(\xi_{\varphi_{\text{id}}}) = \mathbb{1}_{K_A}(\bar{\sigma}) = 1, \quad (5.12)$$

which is all we needed to check. $\square$

However, if K_B is fixed, a realization does not always exist. While this was known previously [Jen18], explicit counterexamples are lacking in the literature. Necessary conditions on state spaces such that every assemblage admits a realization can be found in [BGW13, SB14]. We will find an explicit counterexample by considering a cube and an octahedron as shown in Figure 1. One can choose a state space represented by an octahedron. The corresponding dual state space is then represented by the cube. Note that this also works the other way round. For the measurements on the respective bodies we consider the ones that pick certain faces of the body (see also Figure 1), which generate all other measurements on that state space.

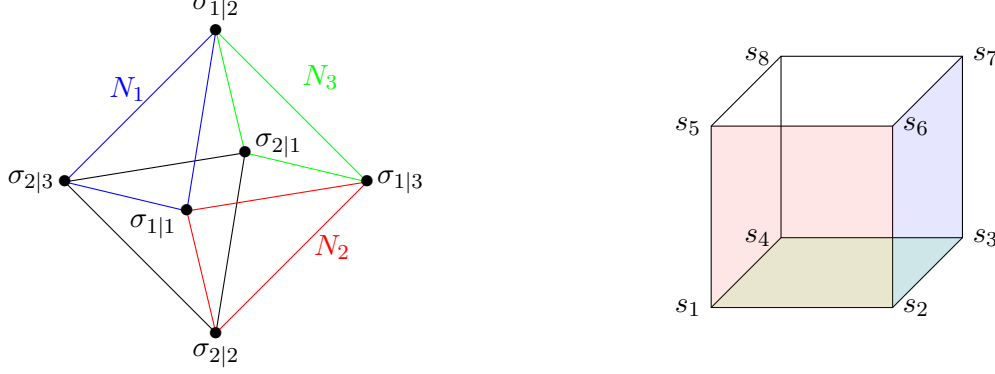


FIGURE 1. The two state spaces represented by an octahedron (left) and a cube (right) are dual to each other. For the measurements on the octahedron we choose the measurements N_i that pick out a face (colored). The opposing face is then picked by $\mathbb{1}_{K_{\text{octa}}} - N_i$ and the fourth measurements follows from normalization. For the cube we choose measurements that pick front (red, M_3), bottom (green, M_2) and right side (blue, M_1) of the cube.

Before we can proceed, we need to prove some facts about the measurement incompatibility in the cube and the octahedron: For a multimeter $M : K \rightarrow CS_{k,g}^1$, we define its compatibility robustness as

$$R_m(M) = \max \left\{ \lambda \in [0, 1] : \exists (p_{\cdot|x})_{x \in [g]} \text{ such that } \{ \lambda M_{\cdot|x} + (1 - \lambda) p_{\cdot|x} \mathbb{1}_K \}_{x \in [g]} \text{ is compatible} \right\}. \quad (5.13)$$

Lemma 5.8. *Let $M = \{M_1, M_2, M_3\}$ be a multimeter containing the measurements that pick out the front (M_3), bottom (M_2) and right (M_1) side of the cube as in Figure 1. Then, $R_m(M) = \frac{1}{3}$.*

Proof. Analogously to an argument later in the proof of Theorem 6.11 which shows that $CS_{2,2}^1$ is isomorphic to a square, one can show that $CS_{2,3}^1$ is isomorphic to a cube (see also [Jen18, Proposition 3] for a more general statement). Thus, we can consider the multimeter M as a channel $M : CS_{2,3}^1 \rightarrow CS_{2,3}^1$. Then, as shown in [Jen18, Example 12], the dichotomic measurements of the multimeter M are maximally incompatible, that is, $R_m(M) = \frac{1}{3}$. $\square$

Let us now shift the focus from measurements on the cube to measurements on the octahedron. The aim is to find distinct separability constraints, i.e., different thresholds for robustness.

Lemma 5.9. *Let K_{octa} be the state space represented by an octahedron and let $N = \{N_1, N_2, N_3, N_4\}$ be the multimeter containing the face-picking measurements on K_{octa} from Figure 1. Then every dichotomic measurement can be classically simulated by N .*

Proof. In [FHL18, Corollary 3] it was shown that a dichotomic measurement can be classically simulated by a multimeter containing dichotomic measurements if and only if the effects of the measurement are contained in the convex hull of the effects of the measurements contained in the simulating multimeter together with the zero effect 0 and the unit effect $\mathbb{1}_{K_{\text{octa}}}$. The effects of the four measurements N_1, N_2, N_3, N_4 generate the extreme rays of the effect cone and together with the zero and the unit effect, their convex hull is the whole effect space $E(K_{\text{octa}})$. Thus, any dichotomic measurement can be simulated by N . $\square$

Lemma 5.10. *Let K be a state space. Then for any multimeter $M : K \rightarrow CS_{k,g}^1$ and any classical simulation $\Phi : CS_{k,g}^1 \rightarrow CS_{l,r}^1$ we have that $R_m(\Phi \circ M) \geq R_m(M)$.*

Proof. Let $R_m(M) = \hat{\lambda}$. Thus, there exists a conditional probability distribution $p = (p_{\cdot|x})_{x \in [g]}$ on $[k]$ such that the multimeter $\hat{M} = \{\hat{\lambda}M_{\cdot|x} + (1 - \hat{\lambda})p_{\cdot|x} \mathbb{1}_K\}_{x \in [g]}$ is compatible. By Corollary 4.7 this means that there exists some measurement $G : K \rightarrow S_n$ and a channel $\mu : S_n \rightarrow CS_{k,g}^1$ such that $\hat{M} = \mu \circ G$. Now clearly

$$(\Phi \circ \mu) \circ G = \Phi \circ (\mu \circ G) = \Phi \circ \hat{M} \quad (5.14)$$

so that again by Corollary 4.7 the multimeter $\Phi \circ \hat{M} : K \rightarrow CS_{l,r}^1$ is compatible with a joint measurement $G : K \rightarrow S_n$ and a postprocessing channel $\Phi \circ \mu : S_n \rightarrow CS_{l,r}^1$. Now since $p \in CS_{k,g}^1$ we can set $q = \Phi(p)$ and we have that

$$(\Phi \circ \hat{M})_{b|y} = \hat{\lambda}(\Phi \circ M)_{b|y} + (1 - \hat{\lambda})q_{b|y} \mathbb{1}_K \quad (5.15)$$

for all $b \in [l]$ and $y \in [r]$. Since $\Phi \circ \hat{M}$ is compatible, it follows that $R_m(\Phi \circ M) \geq \hat{\lambda}$. $\square$

Lemma 5.11. *For any multimeter E on K_{octa} consisting of dichotomic measurements it holds that $R_m(E) \geq \frac{1}{2}$.*

Proof. Consider the dichotomic measurements N_1, N_2, N_3, N_4 on the octahedron as visualized in Figure 1 and let us denote the defining effects for them as e_1, e_2, e_3, e_4 . A joint measurement for the noisy versions of these measurements can be given by the measurement

$$G = \left\{ \frac{e_1}{2}, \frac{e_2}{2}, \frac{e_3}{2}, \frac{e_4}{2}, \mathbb{1}_{K_{\text{octa}}} - \sum_{i=1}^4 \frac{e_i}{2} \right\}, \quad (5.16)$$

as it can be used to obtain the noisy measurements $\{\frac{e_i}{2}, \mathbb{1}_{K_{\text{octa}}} - \frac{e_i}{2}\}$ for all $i \in [4]$. We rewrite each of these measurements as $\{\frac{1}{2}e_i + \frac{1}{2} \cdot 0 \cdot \mathbb{1}_{K_{\text{octa}}}, \frac{1}{2}(\mathbb{1}_{K_{\text{octa}}} - e_i) + \frac{1}{2} \cdot 1 \cdot \mathbb{1}_{K_{\text{octa}}}\}$ so that they correspond to the noisy measurements $\{\lambda N_i + (1 - \lambda)T_i\}$, where we have $\lambda = \frac{1}{2}, T_{+|i} = 0, T_{-|i} = \mathbb{1}_{K_{\text{octa}}}$ for all $i \in [4]$. Thus, for the multimeter $N = \{N_1, N_2, N_3, N_4\}$ we have that $R_m(N) \geq \frac{1}{2}$. Since by Lemma 5.9 every dichotomic multimeter E can be classically simulated by N , and since according to Lemma 5.10 classical simulation cannot decrease the robustness, we have that $R_m(E) \geq R_m(N) \geq \frac{1}{2}$ for any dichotomic multimeter E . $\square$

Now we can finally provide our counterexample.

Proposition 5.12. *Let K_{octa} be the state space of an octahedron. There exists a steering assemblage $\sigma = \{\sigma_{a|x}\}_{a,x} \subset V(K_{\text{octa}})^+$ which does not have a realization in terms of a multimeter on K_{octa} and a state in $K_{\text{octa}} \otimes K_{\text{octa}}$ for any tensor product $\otimes$.*

Proof. It is a well-known fact that the dual of a cube is an octahedron and vice versa. Consequently, $(V(K_{\text{cube}})^+)^* = V(K_{\text{octa}})^+$ and Lemma A.6 implies $(V(K_{\text{octa}})^+)^* = V(K_{\text{cube}})^+$.

We start by defining the robustness for assemblages $\sigma = \{\sigma_{a|x}\}_{a \in [k], x \in [g]}, \sum_{a=1}^k \sigma_{a|x} = \bar{\sigma}$, as

$$R_s(\sigma) = \max \left\{ \mu \in [0, 1] : \exists (p_{\cdot|x})_{x \in [g]} \text{ s.t. } \{\mu \sigma_{a|x} + (1 - \mu)p_{a|x} \bar{\sigma}\}_{a,x} \text{ has an LHS model} \right\}. \quad (5.17)$$

Let us consider the assemblage on the octahedron given by the (unnormalized) opposing extreme points embedded in a hyperplane in a 4-dimensional space, i.e.,

$$\sigma_{1|1} = \frac{1}{2}(1, 0, 0, 1), \quad \sigma_{1|2} = \frac{1}{2}(0, 0, 1, 1), \quad \sigma_{1|3} = \frac{1}{2}(0, 1, 0, 1), \quad (5.18)$$

$$\sigma_{2|1} = \frac{1}{2}(-1, 0, 0, 1), \quad \sigma_{2|2} = \frac{1}{2}(0, 0, -1, 1), \quad \sigma_{2|3} = \frac{1}{2}(0, -1, 0, 1). \quad (5.19)$$

Here, the notation for the extreme points is as in Figure 1. We can identify this assemblage on the octahedron with measurements on the cube: if we consider the cube embedded on a hyperplane of a 4-dimensional space as $K_{\text{cube}} = \{(x, y, z, 1) : -1 \leq x, y, z \leq 1\}$ it follows that $\mathbb{1}_{K_{\text{cube}}} =$

$\bar{\sigma} = (0, 0, 0, 1)$, and also that now $\{\sigma_{a|x}\}_{a \in [2], x \in [3]}$ are the only non-trivial ray-extremal effects on K_{cube} and that $E(K_{\text{cube}}) = \text{conv}(\{\{\sigma_{a|x}\}_{a \in [2], x \in [3]}, 0, \mathbb{1}_{K_{\text{cube}}}\})$. As the ray-extremal effects (in polytope state spaces) are in one-to-one correspondence with the facets of the state space [HLP19, Proposition 16], then considering σ as a multimeter we can identify $\sigma_{\cdot|1}$ with M_1 , $\sigma_{\cdot|2}$ with M_2 and $\sigma_{\cdot|3}$ with M_3 from Figure 1. Then from Lemma 5.8 it follows that $R_m(\sigma) = \frac{1}{3}$.

Now, if $R_s(\sigma) = \lambda$, this means that there exists a conditional probability distribution $p = (p_{\cdot|x})_{x \in [3]}$ such that $\{\lambda\sigma_{\cdot|x} + (1 - \lambda)p_{\cdot|x}\bar{\sigma}\}_{x \in [3]}$ has an LHS model. By Definition 5.5 there exists $n \in \mathbb{N}$, an ensemble $\varrho = \{\varrho_c\}_{c \in [n]} \subset V(K_{\text{octa}})^+$ with $\bar{\varrho} := \sum_{c=1}^n \varrho_c = \bar{\sigma}$ and a conditional probability distribution $\nu = (\nu_{\cdot|c,x})_{c \in [n], x \in [3]}$ on $[2]$ such that $\lambda\sigma_{a|x} + (1 - \lambda)p_{a|x}\bar{\sigma} = \sum_{c=1}^n \nu_{a|c,x}\varrho_c$ for all $a \in [2]$ and $x \in [3]$. Now, since $\{\varrho_c\}_{c \in [n]} \subset V(K_{\text{octa}})^+ = (V(K_{\text{cube}})^+)^*$ and $\bar{\varrho} = \sum_c \varrho_c = \mathbb{1}_{K_{\text{cube}}}$, it follows that ϱ is a measurement on K_{cube} . Thus, the multimeter $\{\lambda\sigma_{\cdot|x} + (1 - \lambda)p_{\cdot|x}\mathbb{1}_{K_{\text{cube}}}\}_{x \in [3]}$ is compatible, meaning that $R_m(\sigma) \geq \lambda$ so that $R_s(\sigma) \leq R_m(M) = \frac{1}{3}$.

Suppose now that $E : K_{\text{octa}} \rightarrow \text{CS}_{2,3}^1$ is a multimeter and $w \in K_{\text{octa}} \otimes K_{\text{octa}}$ such that

$$\sigma_{a|x} = (E_{a|x} \otimes \text{id})(w) \quad \forall a \in [2], x \in [3]. \quad (5.20)$$

Then clearly $\bar{\sigma} = (\mathbb{1}_{K_{\text{octa}}} \otimes \text{id})(w)$. Let $R_m(E) = \mu$ so that there exists a measurement $F = \{F_d\}_{d=1}^m$ on K_{octa} and conditional probability distributions $q = (q_{\cdot|x})_{x \in [3]}$ and $\eta = \{\eta_{\cdot|d,x}\}_{d \in [m], x \in [3]}$ on $[2]$ such that

$$\mu E_{a|x} + (1 - \mu)q_{a|x}\mathbb{1}_{K_{\text{octa}}} = \sum_{d=1}^m \eta_{a|d,x} F_d \quad (5.21)$$

for all $a \in [2]$ and $x \in [3]$. Now

$$\mu\sigma_{a|x} + (1 - \mu)q_{a|x}\bar{\sigma} = ((\mu E_{a|x} + (1 - \mu)q_{a|x}\mathbb{1}_{K_{\text{octa}}}) \otimes \text{id})(w) = \sum_{d=1}^m \eta_{a|d,x} (F_d \otimes \text{id})(w) \quad (5.22)$$

for all $a \in [2]$ and $x \in [3]$. If we define $\kappa_d = (F_d \otimes \text{id})(w) \in V(K_{\text{octa}})^+$, we see that $\bar{\kappa} := \sum_d \kappa_d = \bar{\sigma}$ and thus the assemblage $\{\mu\sigma_{a|x} + (1 - \mu)q_{a|x}\bar{\sigma}\}_{a \in [2], x \in [3]}$ has an LHS model. This shows that $R_s(\sigma) \geq \mu$. Finally, by Lemma 5.11 we have that $R_m(E) = \mu \geq \frac{1}{2}$ so that also $R_s(\sigma) \geq \frac{1}{2}$. This contradicts our earlier observation that $R_s(\sigma) \leq \frac{1}{3}$.

Thus, the presented assemblage σ on K_{octa} cannot have an realization in terms of a multimeter on K_{octa} and a state in $K_{\text{octa}} \otimes K_{\text{octa}}$ for any tensor product $\otimes$. $\square$

5.3. Factorizations of maps preparing ensembles. For multimeters, we considered different factorizations of the associated map and related them to the compatibility of measurements, K_B -simulability and classical simulability. In this section, our aim is similar, but this time we consider a map that prepares the assemblage and then factorize it.

Now assume that you are given a bipartite state space $K_A \otimes K_B$ with an appropriate tensor product $\otimes$ and a state $\varrho_{AB} \in K_A \otimes K_B$. We recall that one way to prepare an assemblage from a bipartite state is to apply a multimeter M to one of the systems, e.g., the system K_A . This yields

$$q_{a|x}\varrho_{a|x} = (M_{a|x} \otimes \text{id})(\varrho_{AB}) \quad (5.23)$$

with $\varrho_{a|x} \in K_B$ and $\sum_{a=1}^k q_{a|x}\varrho_{a|x} = \bar{\varrho}$ for all $x \in [g]$. Let's assume now that the multimeter M is in addition K'_A -simulable, i.e., there exists an instrument Φ with operations $\Phi_\lambda : V(K_A)^+ \rightarrow V(K'_A)^+$, and a multimeter $N = \{N_{\cdot|x,\lambda}\}_{x \in [g], \lambda \in [\Lambda]}$ on K'_A such that

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} \Phi_\lambda^*(N_{a|x,\lambda}) \quad (5.24)$$

for all $a \in [k]$ and $x \in [g]$. If we have moreover a family of channels $\Psi_\lambda : K_B \rightarrow K'_B$ that can access the outcome λ of the instrument Φ , then

$$q_{a|x} \varrho_{a|x} = \sum_{\lambda=1}^{\Lambda} \Psi_\lambda(\sigma_{a,\lambda|x}) \quad \forall a \in [k], x \in [g], \quad (5.25)$$

where the $\sigma_{a,\lambda|x}$ are subnormalized states defined as $\sigma_{a,\lambda|x} = ((N_{a|x,\lambda} \circ \Phi_\lambda) \otimes \text{id})(\varrho_{AB})$. Hence, using the properties of the multimeter and the instrument, $\sum_{a=1}^k \sigma_{a,\lambda|x} = \bar{\sigma}_\lambda$ for all $x \in [g]$, where $\bar{\sigma}_\lambda$ is again a subnormalized state, and $\sum_{\lambda=1}^{\Lambda} \bar{\sigma}_\lambda = \bar{\sigma} \in K'_B$. This motivates the following definition:

Definition 5.13. Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K_A$. The assemblage is K_B -simulable if there exists a finite number of outcomes Λ , a family of channels $\Psi_\lambda : K_B \rightarrow K_A$, and a set of subnormalized states $\{\sigma_{a,\lambda|x}\}_{a \in [k], \lambda \in [\Lambda], x \in [g]} \subset V(K_B)^+$ with $\sum_{a=1}^k \sigma_{a,\lambda|x} = \bar{\sigma}_\lambda \in V(K_B)^+$ for all $x \in [g]$, and $\sum_{\lambda=1}^{\Lambda} \bar{\sigma}_\lambda = \bar{\sigma} \in K_B$, such that

$$q_{a|x} \varrho_{a|x} = \sum_{\lambda=1}^{\Lambda} \Psi_\lambda(\sigma_{a,\lambda|x}) \quad \forall a \in [k], x \in [g]. \quad (5.26)$$

As in the case of K_B -simulability of multimeters, we can also characterize K_B -simulability of assemblages in terms of a particular factorization of the associated map.

Theorem 5.14. Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K_A$ and associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow V(K_A)^+$ such that $\Phi(\mathbb{1}_{CS_{k,g}^1}) = \bar{\varrho}$. Then this assemblage is K_B -simulable if and only if there exists a finite number of outcomes Λ , a map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow V(K_B \dot{\otimes} S_\Lambda)^+$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) \in K_B \dot{\otimes} S_\Lambda$ and a channel $\Psi_2 : K_B \dot{\otimes} S_\Lambda \rightarrow K_A$ such that the following diagram commutes:

$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & V(K_A)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & V(K_B \dot{\otimes} S_\Lambda)^+ & \end{array} \quad (5.27)$$

Proof. Let us assume the existence of the maps Ψ_1, Ψ_2 as in the commuting diagram. Let us consider the map Ψ_1 first and let us define $\Psi_1(m_a^{(x)}) =: \tilde{\sigma}_{a|x}$ for all $a \in [k]$ and $x \in [g]$. Since $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) =: \hat{\sigma} \in K_B \dot{\otimes} S_\Lambda$, it follows that

$$\hat{\sigma} = \sum_{a=1}^k \tilde{\sigma}_{a|x} \quad \forall x \in [g]. \quad (5.28)$$

We can furthermore decompose uniquely

$$\tilde{\sigma}_{a|x} = \sum_{\lambda=1}^{\Lambda} \sigma_{a,\lambda|x} \otimes \delta_\lambda \quad \forall a \in [k], \forall x \in [g]. \quad (5.29)$$

Here, $\sigma_{a,\lambda|x} \in V(K_B)^+$ for all $a \in [k]$, $x \in [g]$, and $\lambda \in [\Lambda]$. We can likewise decompose uniquely $\hat{\sigma} = \sum_{\lambda=1}^{\Lambda} \bar{\sigma}_\lambda \otimes \delta_\lambda$ with $\bar{\sigma}_\lambda \in V(K_B)^+$ for all $\lambda \in [\Lambda]$. Since both decompositions are unique, we conclude

$$\forall x \in [g] \quad \sum_{a=1}^k \sigma_{a,\lambda|x} = \bar{\sigma}_\lambda, \quad \sum_{\lambda=1}^{\Lambda} \bar{\sigma}_\lambda = (\text{id} \otimes \mathbb{1}_{S_\Lambda})(\hat{\sigma}) =: \bar{\sigma} \in K_B. \quad (5.30)$$

As Ψ_2 is a channel, so is $\Psi_\lambda(\cdot) := \Psi_2(\cdot \otimes \delta_\lambda)$ for all $\lambda \in [\Lambda]$. Combining Eq. (5.7) and Eq. (5.8) with $\Phi = \Psi_2 \circ \Psi_1$, we obtain that indeed

$$q_{a|x} \varrho_{a|x} = \sum_{\lambda=1}^{\Lambda} \Psi_\lambda(\sigma_{a,\lambda|x}) \quad \forall a \in [k], x \in [g], \quad (5.31)$$

which proves the converse. The other direction follows straightforwardly along the same lines, verifying the properties of Ψ_1 and Ψ_2 . Since $\{\mathbb{1}_{CS_{k,g}^1}, m_a^{(x)}\}_{a \in [k-1], x \in [g]}$ is a basis for $CS_{k,g}^*$, the $\tilde{\sigma}_{a|x}$ and $\hat{\sigma}$ uniquely define a positive map Ψ_1 with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) \in K_B \otimes S_\Lambda$. Finally, we can set

$$\Psi_2(x) = \sum_{\lambda=1}^{\Lambda} q_\lambda \Psi_\lambda(x_\lambda) \quad \forall x \in K_B \dot{\otimes} S_\Lambda, \quad (5.32)$$

since any such x has a unique decomposition as $x = \sum_{\lambda=1}^{\Lambda} q_\lambda x_\lambda \otimes \delta_\lambda$ for $(q_\lambda)_{\lambda \in [\Lambda]}$ a probability distribution and $x_\lambda \in K_B$ for all $\lambda \in [\Lambda]$. Thus, Ψ_2 defines a channel, which concludes the proof. $\square$

Now by looking at the dual map Φ^* , which we were able to interpret as a multimeter, we can prove a very similar factorization:

Theorem 5.15. *Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K_A$ and associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow V(K_A)^+$ such that $\Phi(\mathbb{1}_{CS_{k,g}^1}) = \bar{\varrho}$. Let $\bar{\varrho} \in K_A$ be in the relative interior of K_A . Then, there exists a finite number of outcomes $\Lambda \in \mathbb{N}$, a map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow V(K_B \dot{\otimes} S_\Lambda)^+$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \bar{\sigma} \otimes q$ for a probability distribution $(q_\lambda)_{\lambda \in [\Lambda]}$ such that $q_\lambda > 0$ for all $\lambda \in [\Lambda]$ and $\bar{\sigma}$ in the relative interior of K_B , and a map $\Psi_2 : V(K_B \dot{\otimes} S_\Lambda)^+ \rightarrow V(K_A)^+$ with $\Psi_2(\bar{\sigma} \otimes q) = \bar{\varrho}$ such that the following diagram commutes:*

$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & V(K_A)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & V(K_B \dot{\otimes} S_\Lambda)^+ & \end{array} \quad (5.33)$$

if and only if the multimeter $\Phi^* : (K_A)_{\bar{\varrho}}^* \rightarrow CS_{k,g}^1$ is $(K_B)_{\bar{\sigma}}^*$ -simulable.

Proof. As $\bar{\sigma}$ and q are both in the relative interior of K_B and S_Λ , respectively, we can define the state spaces $(K_B)_{\bar{\sigma}}^*$ and $(S_\Lambda)_q^*$. By Example 3.5, we can identify $(S_\Lambda)_q^* \simeq S_\Lambda$. We observe that $\Psi_1^* : (K_B)_{\bar{\sigma}}^* \dot{\otimes} (S_\Lambda)_q^* \rightarrow CS_{k,g}^1$ is a channel, since for all probability distributions $(p_i)_i$ and $\alpha_i \in (K_B)_{\bar{\sigma}}^*$, $s_i \in (S_\Lambda)_q^*$,

$$\left\langle \mathbb{1}_{CS_{k,g}^1}, \Psi_1^* \left(\sum_i p_i \alpha_i \otimes s_i \right) \right\rangle = \sum_i \left\langle \Psi_1(\mathbb{1}_{CS_{k,g}^1}), p_i \alpha_i \otimes s_i \right\rangle = \sum_i p_i \alpha_i(\bar{\sigma}) s_i(q) = 1. \quad (5.34)$$

Moreover, as $\Psi_2(\hat{\sigma}) = \bar{\varrho}$ with $\hat{\sigma} = \bar{\sigma} \otimes q$, we have that for all $\alpha \in (K_A)_{\bar{\varrho}}^*$,

$$\langle \Psi_2^*(\alpha), \bar{\sigma} \otimes q \rangle = \alpha(\Psi_2(\hat{\sigma})) = \alpha(\bar{\varrho}) = 1, \quad (5.35)$$

hence $\Psi_2^* : (K_A)_{\bar{\varrho}}^* \rightarrow (K_B)_{\bar{\sigma}}^* \dot{\otimes} (S_\Lambda)_q^*$ is a channel. The assertion then follows by Theorem 4.2. It thus follows that the multimeter Φ^* is $(K_B)_{\bar{\sigma}}^*$ -simulable.

Conversely, for $(K_B)_{\bar{\sigma}}^*$ to be a state space $\bar{\sigma}$ has to be in the relative interior of K_B . Theorem 4.2 guarantees the factorization $\Phi^* = \Psi_1^* \circ \Psi_2^*$ with channels $\Psi_1^* : (K_B)_{\bar{\sigma}}^* \dot{\otimes} S_\Lambda \rightarrow CS_{k,g}^1$ and $\Psi_2^* : (K_A)_{\bar{\varrho}}^* \rightarrow (K_B)_{\bar{\sigma}}^* \dot{\otimes} S_\Lambda$. Thus, Ψ_1 and Ψ_2 are both unital. Picking any q in the relative interior of S_Λ , we can define $S_\Lambda \simeq (S_\Lambda)_q^*$ such that $\mathbb{1}_{S_\Lambda} \simeq q$ (see Example 3.5). Note also that $\mathbb{1}_{(K_B)_{\bar{\sigma}}^*} = \bar{\sigma}$. We recall that $A((K_A)_{\bar{\varrho}}^*)^+ = V(K_A)^+$ and $A((K_B)_{\bar{\sigma}}^*)^+ = V(K_B)^+$. Hence, $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow$

$V(K_B)^+ \dot{\otimes} V((S_\Lambda)_q^*)^+$ fulfills $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \bar{\sigma} \otimes q \in K_B \dot{\otimes} (S_\Lambda)_q^*$ and $\Psi_2 : V(K_B)^+ \dot{\otimes} V((S_\Lambda)_q^*)^+ \rightarrow V(K_A)^+$ fulfills $\Psi_2(\bar{\sigma} \otimes q) = \bar{\varrho}$. $\square$

Remark 5.16. *The main differences between the factorizations in Theorem 5.14 and Theorem 5.15 is that in Theorem 5.14 the map Ψ_2 is a channel, whereas we only require $\Psi_2(\bar{\sigma} \otimes q) = \bar{\varrho}$ in Theorem 5.15. On the other hand, $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \bar{\sigma} \otimes q$ in Theorem 5.15, whereas in Theorem 5.14 $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) \in K_B \dot{\otimes} S_\Lambda$ needs not be a pure tensor product.*

Remark 5.17. *Let us take a closer look at Theorem 5.15 for the case of quantum theory. Seeing $\Phi^* = \{M_{\cdot|x}\}_{x \in [g]}$ as a multimeter on $(D(\mathbb{C}^d))_{\bar{\varrho}}^*$, the factorization of Φ in Theorem 5.15 is equivalent to the existence of a finite number of outcomes Λ , of an instrument $\Psi : (D(\mathbb{C}^d))_{\bar{\varrho}}^* \rightarrow (D(\mathbb{C}^{d'}))_{\bar{\sigma}}^* \dot{\otimes} S_\Lambda$ with operations $\Psi_\lambda : \text{PSD}_d \rightarrow \text{PSD}_{d'}$, and of a multimeter $N = \{N_{\cdot|x,\lambda}\}_{x \in [g], \lambda \in [\Lambda]}$ of $g \cdot \Lambda$ measurements with k outcomes on $(D(\mathbb{C}^{d'}))_{\bar{\sigma}}^*$ such that*

$$M_{a|x} = \sum_{\lambda=1}^{\Lambda} \Psi_\lambda^*(N_{a|x,\lambda}) \quad (5.36)$$

for all $a \in [k]$ and $x \in [g]$. We will now relate this to the existence of ordinary quantum multimeters and instruments. Example 3.6 tells us the shape of $(D(\mathbb{C}^d))_{\bar{\varrho}}^*$ and $(D(\mathbb{C}^{d'}))_{\bar{\sigma}}^*$. Thus, there exists a multimeter $\{\tilde{M}_{\cdot|x}\}_{x \in [g]}$ on $D(\mathbb{C}^d)$ such that

$$\tilde{M}_{a|x}(\tau) = M_{a|x}(\bar{\varrho}^{-\frac{1}{2}} \tau \bar{\varrho}^{-\frac{1}{2}}) \quad \forall a \in [k], \forall x \in [g], \forall \tau \in D(\mathbb{C}^d). \quad (5.37)$$

Moreover, there exist positive maps $\tilde{\Psi}_\lambda^* : \text{PSD}_d \rightarrow \text{PSD}_{d'}$ such that $\sum_{\lambda=1}^{\Lambda} \tilde{\Psi}_\lambda^* : D(\mathbb{C}^d) \rightarrow D(\mathbb{C}^{d'})$ is a channel (although it is not a quantum channel, since it is not necessarily completely positive) and such that

$$\tilde{\Psi}_\lambda^*(\tau) = \bar{\sigma}^{-\frac{1}{2}} \Psi_\lambda^*(\bar{\varrho}^{-\frac{1}{2}} \tau \bar{\varrho}^{-\frac{1}{2}}) \bar{\sigma}^{\frac{1}{2}} \quad \forall \lambda \in [\Lambda], \forall \tau \in D(\mathbb{C}^d). \quad (5.38)$$

Finally, there exists a multimeter $\{\tilde{N}_{\cdot|x,\lambda}\}_{x \in [g], \lambda \in [\Lambda]}$ on $D(\mathbb{C}^{d'})$ such that

$$\tilde{N}_{a|x,\lambda}(\tau) = N_{a|x,\lambda}(\bar{\sigma}^{-\frac{1}{2}} \tau \bar{\sigma}^{-\frac{1}{2}}) \quad \forall a \in [k], \forall x \in [g], \forall \lambda \in [\Lambda], \forall \tau \in D(\mathbb{C}^{d'}). \quad (5.39)$$

Hence, since $\bar{\varrho}$ and $\bar{\sigma}$ are invertible, the factorization of Φ is equivalent to

$$\tilde{M}_{a|x} = \sum_{\lambda=1}^{\Lambda} \tilde{\Psi}_\lambda^*(\tilde{N}_{a|x,\lambda}) \quad (5.40)$$

for all $a \in [k]$ and $x \in [g]$.

Similarly to the discussion that motivated the definition of K_B -simulable assemblages, we can instead assume the multimeter preparing the assemblage to be classically simulable. This motivates the following definition:

Definition 5.18. *Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K$. We say that the assemblage can be classically simulated (or is classically simulable) by an assemblage $\{v_{b|y}, \sigma_{b|y}\}_{b \in [l], y \in [r]}$ with $\bar{\sigma} = \bar{\varrho} \in K$ if there exist conditional probability distributions $\pi = (\pi_{\cdot|x})_{x \in [g]}$ on $[r]$ and $\nu = (\nu_{\cdot|b,x,y})_{b \in [l], x \in [g], y \in [r]}$ on $[k]$ such that*

$$q_{a|x} \varrho_{a|x} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} v_{b|y} \sigma_{b|y} \quad (5.41)$$

for all $a \in [k]$ and $x \in [g]$.

We can again find a characterization of classically simulable assemblages in terms of factorizations.

Theorem 5.19. *Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K$ and associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow V(K)^+$ such that $\Phi(\mathbb{1}_{CS_{k,g}^1}) = \bar{\varrho}$. Then this assemblage is classically simulable by some assemblage $\{v_{b|y}, \sigma_{b|y}\}_{b \in [l], y \in [r]}$ with $\bar{\sigma} = \bar{\varrho} \in K$ if and only if there exist a map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow (CS_{l,r}^*)^+$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \mathbb{1}_{CS_{l,r}^1}$ and a map $\Psi_2 : (CS_{l,r}^*)^+ \rightarrow V(K)^+$ with $\Psi_2(\mathbb{1}_{CS_{l,r}^1}) \in K$ such that the following diagram commutes:*

$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & V(K)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & (CS_{l,r}^*)^+ & \end{array} \quad (5.42)$$

Proof. Using Theorem 4.5 and duality of maps, every map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow (CS_{l,r}^*)^+$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \mathbb{1}_{CS_{l,r}^1}$ can be written as

$$\Psi_1(\alpha) = \alpha(s_{k,\dots,k}) \mathbb{1}_{CS_{l,r}^1} + \sum_{x=1}^g \sum_{a=1}^{k-1} \alpha(e_a^{(x)}) \left[\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} m_b^{(y)} \right], \quad (5.43)$$

where $\alpha \in (CS_{k,g}^*)^+$. Now given $\Psi_2 : (CS_{l,r}^*)^+ \rightarrow V(K)^+$, we can define $\Psi_2(m_b^{(y)}) =: v_{b|y} \sigma_{b|y}$ with $\sigma_{b|y} \in K$ and $v_{b|y} := \mathbb{1}_{CS_{l,r}^1}(\Psi_2(m_b^{(y)}))$. Setting moreover

$$\sum_{b=1}^l v_{b|y} \sigma_{b|y} =: \bar{\sigma} \in K \quad \forall y \in [r], \quad (5.44)$$

it follows that $(v_{b|y})_{b \in [l]}$ is a probability distribution for all $y \in [r]$. As $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \mathbb{1}_{CS_{l,r}^1}$, it follows that $\bar{\sigma} = \bar{\varrho}$. Using $\Phi = \Psi_2 \circ \Psi_1$, we find

$$q_{a|x} \varrho_{a|x} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} v_{b|y} \sigma_{b|y} \quad (5.45)$$

for all $a \in [k-1]$ and $x \in [g]$. For $a = k$, we infer

$$q_{k|x} \varrho_{k|x} = \bar{\varrho} - \underbrace{\sum_{a=1}^{k-1} \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{a|b,x,y} v_{b|y} \sigma_{b|y}}_{\sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l (1 - \nu_{k|b,x,y}) v_{b|y} \sigma_{b|y}} = \sum_{y=1}^r \pi_{y|x} \sum_{b=1}^l \nu_{k|b,x,y} v_{b|y} \sigma_{b|y}. \quad (5.46)$$

The converse is straightforward defining the map Ψ_2 using the assemblage $\{v_{b|y}, \sigma_{b|y}\}_{b \in [l], y \in [r]}$. $\square$

Instead of seeing classical simulability as a property of the assemblage, we can consider it as a property of the associated multimeter, as the next corollary shows.

Corollary 5.20. *Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K$ and associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow V(K)^+$ such that $\Phi(\mathbb{1}_{CS_{k,g}^1}) = \bar{\varrho}$. Let $\bar{\varrho}$ be in the relative interior of K . Then this assemblage is classically simulable if and only if the multimeter $\Phi^* : K_{\bar{\varrho}}^* \rightarrow CS_{k,g}^1$ is classically simulable.*

Proof. By Theorem 5.19, the assemblage is classically simulable if and only if there exist a map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow (CS_{l,r}^*)^+$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \mathbb{1}_{CS_{l,r}^1}$ and a map $\Psi_2 : (CS_{l,r}^*)^+ \rightarrow V(K)^+$ with $\Psi_2(\mathbb{1}_{CS_{l,r}^1}) \in K$ such that $\Phi = \Psi_2 \circ \Psi_1$. Therefore, $\Psi_2^* : K_{\bar{\varrho}}^* \rightarrow CS_{l,r}^1$ is a multimeter, and as Ψ_1 is unital, $\Psi_1^* : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ is a channel as well. As $\Phi = \Psi_2 \circ \Psi_1$, we have $\Phi^* = \Psi_1^* \circ \Psi_2^*$. The assertion follows from Corollary 4.6 and the converse follows in the same manner. $\square$

Finally, we can recover from our results a characterization of assemblages admitting an LHS model, see also Proposition 5.3.

Corollary 5.21. *Let $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ be an assemblage with average state $\bar{\varrho} \in K$ and associated map $\Phi : (CS_{k,g}^*)^+ \rightarrow V(K)^+$ such that $\Phi(\mathbb{1}_{CS_{k,g}^1}) = \bar{\varrho}$. Then, the following are equivalent:*

- (1) $(\text{id} \otimes \Phi)(\chi_{CS_{k,g}^*}) \in CS_{k,g}^1 \hat{\otimes} K$
- (2) $\{q_{a|x}, \varrho_{a|x}\}_{a \in [k], x \in [g]}$ has an LHS model
- (3) *There exist a finite number Λ , a map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow (S_\Lambda^+)^*$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \mathbb{1}_{S_\Lambda}$ and a map $\Psi_2 : (S_\Lambda^+)^* \rightarrow V(K)^+$ with $\Psi_2(\mathbb{1}_{S_\Lambda}) \in K$ such that the following diagram commutes:*

$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & V(K)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & (S_\Lambda^+)^* & \end{array} \quad (5.47)$$

Proof. The equivalence of (1) and (3) follows from Proposition 3.7. The equivalence of (2) and (3) follows from Theorem 5.19 with $r = 1$ and $l = \Lambda$, because a classical simulation of this kind has the form

$$q_{a|x} \varrho_{a|x} = \sum_{\lambda=1}^{\Lambda} \nu_{a|\lambda, x} v_\lambda \sigma_\lambda. \quad (5.48)$$

for all $a \in [k]$ and $x \in [g]$. □

6. BELL NONLOCALITY

Another form of quantum nonlocality that is closely related to measurement incompatibility is Bell nonlocality, a phenomenon that questions the assumption of locality in an experiment. A no-signaling probability distribution (Bell behavior) is said to be local if it has a local realization (local hidden variable (LHV) model). Then Bell nonlocality can be detected through a violation of a Bell inequality. We can use our formalism to show how no-signaling probability distributions can be seen as tensors and that separability of that tensor corresponds to having a local realization through an LHV model. Again, this will correspond to a factorization of the associated map through a simplex and we see how different factorizations give different simulation processes for the no-signaling distributions. We will also see how Bell inequalities can be formulated in this framework and we will be able to recover some important known results from [WPGF09, BBB⁺10, Pir14].

6.1. Bell nonlocality in quantum theory. We begin investigating Bell nonlocality by considering two multimeters $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^+$, $N : D(\mathbb{C}^{d_B}) \rightarrow CS_{l,r}^+$ and $\varrho \in D(\mathbb{C}^{d_A d_B})$. Both multimeters together yield a map

$$M \otimes N : \varrho \mapsto (M \otimes N)(\varrho), \quad (6.1)$$

which forms a bipartite Bell behavior with the outcome statistic

$$\text{Tr}[(M_{a|x} \otimes N_{b|y}) \varrho] = p_{a,b|x,y} \geq 0 \quad \text{with } a \in [k], b \in [l], x \in [g], y \in [r]. \quad (6.2)$$

The following is straightforward to confirm.

Lemma 6.1. *The image $(M \otimes N)(\varrho)$ is in the maximal tensor product $CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$.*

As a classical explanation to nonlocality one could search for a model that explains nonlocal correlations by local hidden variables. Knowledge of these hidden variables would allow for a complete determination of the measurement outcomes and hence leave the inherent randomness of quantum theory behind. Such models are consequently called local hidden variable models and will be abbreviated with LHV models from now on.

Definition 6.2. A Bell behavior $(p_{a_1, \dots, a_N | x_1, \dots, x_N})_{a_j \in [k_j], x_j \in [g_j], j \in [N]}$ with N parties, each party $j \in [N]$ performing g_j measurements with k_j outcomes, can be described by an LHV model if there is a locally hidden variable $\Lambda \in \mathbb{N}$ and probability distributions q on $[\Lambda]$ and $(p_{\cdot | \lambda, x_j})_{\lambda \in [\Lambda], x_j \in [g_j]}$ on $[k_j]$ for all $j \in [N]$ such that the outcome statistics defined likewise as in Eq. (6.2) can be written as

$$p_{a_1, \dots, a_N | x_1, \dots, x_N} = \sum_{\lambda \in [\Lambda]} q_\lambda \prod_{j=1}^N p_{a_j | \lambda, x_j} \quad (6.3)$$

for all $a_j \in [k_j]$, $x_j \in [g_j]$ and $j \in [N]$.

The idea is partially similar to LHS models, and as with LHS models, there is a quite similar connection between separability and LHV models to be found. However, in contrast to steering, where every bipartite quantum assemblage has a realization (Lemma 5.2), not all outcome statistics can be achieved by quantum theory. Formally, this means $\exists P \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ st. $P \neq (M \otimes N)(\varrho)$, where M, N are quantum multimeters and ϱ is a density matrix. An example of such objects P are Popescu-Rohrlich nonlocal boxes [PR94].

Typically, in scenarios where more than one party performs sets of measurements on a shared system, the parties are considered spatially separated. For example, the parties are in distinct laboratories performing their operations and measurements. Thus, it is fair to assume that the outcomes of one party are not affected by the measurement choice of another party, since that would require instantaneous or superluminal signal transmission. Hence, the assumption is well motivated by special relativity and is widely known as no-signaling.

Definition 6.3. A conditional probability distribution $(p_{\cdot, \cdot | xy})_{x \in [g], y \in [r]}$ on $[k] \times [l]$ is said to be no-signaling if it satisfies:

- (1) $\sum_{a=1}^k p_{a,b|x,y} = \sum_{a=1}^k p_{a,b|x',y} \quad \forall x, x' \in [g], b \in [l]$
- (2) $\sum_{b=1}^l p_{a,b|x,y} = \sum_{b=1}^l p_{a,b|x,y'} \quad \forall y, y' \in [r], a \in [k]$

In the context of experiments, these probability distributions are also called no-signaling behaviors.

The concept of no-signaling is more general and weaker than local hidden variables. As the next lemma shows, all physical statistics are covered in the set $CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ (see Proposition 6.14 for proof, originally shown in [Jen18]).

Lemma 6.4. The set of all no-signaling behaviors is in one-to-one correspondence with $CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$.

As we saw with the steering assemblages, the separable assemblages correspond to those that are classically explainable with LHS models. We can make a similar conclusion here with the separable no-signaling behaviors being classically explainable with LHV models (see Corollary 6.19 for proof, originally shown in [Pl17, Jen18]):

Lemma 6.5. Let $P \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$. Then $P \in CS_{k,g}^1 \dot{\otimes} CS_{l,r}^1$ if and only if P has an LHV model.

In particular, if a behavior P has a realization as $P = (M \otimes N)(\varrho)$ for some multimeters M and N and some shared state ϱ , we get the following:

Corollary 6.6. Consider $M \otimes N : D(\mathbb{C}^{d_A d_B}) \rightarrow CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ and $\varrho \in D(\mathbb{C}^{d_A d_B})$. Then $(M \otimes N)(\varrho)$ has a LHV model if and only if $(M \otimes N)(\varrho) \in CS_{k,g}^1 \dot{\otimes} CS_{l,r}^1$.

It is possible to detect nonlocality in Bell behaviors with Bell inequalities. Statistics from quantum theory are able to exceed bounds to statistics stemming from local realist behaviors, that is, statistics that have an LHV model. There are also bounds to the statistics of quantum theory, for instance the Tsirelson bound, which in turn is exceeded by the already mentioned Popescu-Rohrlich boxes. Those are consequently also nonlocal. A necessary condition for violating a Bell inequality is the inseparability of the quantum states and incompatibility of the measurements in the Bell

experiment [Fin82]. As we saw earlier, incompatibility can also be seen as a form of entanglement. Therefore, the Bell inequality acts as an entanglement witness in a generalized sense.

Definition 6.7. *A Bell inequality is a witness-type map defined as $W : CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1 \rightarrow \mathbb{R}$ such that $W(P) \geq 0$ for all $P \in CS_{k,g}^1 \dot{\otimes} CS_{l,r}^1$.*

This definition can easily be extended to every finite number of parties. We call functionals witnesses if they are positive on separable elements. Moreover, we will also often make use of the fact that for every inseparable (entangled) element, there is always a witness that detects it, i.e., maps it to a negative number. For more details, the reader is asked to see Appendix A.2. The claim from above that Bell inequalities are generalized entanglement witnesses shall now be formalized for two parties.

Lemma 6.8. *Every Bell inequality is an element of the set $(CS_{k,g}^*)^+ \hat{\otimes} (CS_{l,r}^*)^+$ and vice versa, every functional in this set is a Bell inequality.*

The statement follows directly from the fact that $(CS_{k,g}^*)^+ \hat{\otimes} (CS_{l,r}^*)^+ = (CS_{k,g}^+ \dot{\otimes} CS_{l,r}^+)^*$ (see Appendix A.2).

As mentioned before, it is a well-known fact that measurement incompatibility in the Bell setup is a necessity for violating Bell inequalities. In fact, no Bell inequality with N parties can be violated if $N - 1$ parties perform compatible measurements [Fin82]. In the bipartite case, this reduces to one set of compatible POVMs. Both entanglement and incompatibility are a form of inseparability, which can be witnessed by a Bell inequality.

Popescu-Rohrlich nonlocal boxes are able to achieve the algebraic maximum of Bell inequalities, succeeding the Tsirelson bounds of quantum theory. A question arises whether normalized elements of $D(\mathbb{C}^{d_A}) \hat{\otimes} D(\mathbb{C}^{d_B})$ lead to statistics unachievable by quantum states. That is, we could relax the condition on the joint system and instead of considering density matrices, let it be a more general object, the elements of which get mapped to probabilities by POVMs. In this way, we might get more than quantum correlations in $CS_{k,g}^+ \dot{\otimes} CS_{l,r}^+$. However, in [BBB⁺10] this possibility was already ruled out. Here we are able to write down a different proof and also give another point of view on this result.

Theorem 6.9. *Let $d_B \geq d_A$. For all multimeters $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^1$, $N : D(\mathbb{C}^{d_B}) \rightarrow CS_{l,r}^1$ and $\Gamma \in D(\mathbb{C}^{d_A}) \hat{\otimes} D(\mathbb{C}^{d_B})$ there exist multimeter $\tilde{N} : D(\mathbb{C}^{d_B}) \rightarrow CS_{l,r}^1$ and $\varrho \in D(\mathbb{C}^{d_A d_B})$ such that*

$$(M \otimes N)(\Gamma) = (M \otimes \tilde{N})(\varrho). \quad (6.4)$$

Proof. For this Γ we can write

$$(\text{id} \otimes N)(\Gamma) \in D(\mathbb{C}^{d_A}) \hat{\otimes} CS_{k,g}^1, \quad (M \otimes N)(\Gamma) \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1. \quad (6.5)$$

These objects are still positive and we get an assemblage σ from Γ . According to Lemma 5.2 the obtained assemblage σ finds a quantum realization

$$(\text{id} \otimes N)(\Gamma) = \sigma = (\text{id} \otimes \tilde{N})(\varrho). \quad (6.6)$$

Now we are already done since for all $a \in [k]$, $b \in [l]$, $x \in [g]$, $y \in [r]$ it will hold that

$$\text{Tr}[(M_{a|x} \otimes N_{b|y}) \Gamma] = \text{Tr}[(M_{a|x} \otimes \tilde{N}_{b|y}) \varrho]. \quad (6.7)$$

□

It follows that there cannot be a Bell inequality to detect if the statistics in $CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ stemmed from a density matrix in $D(\mathbb{C}^{d_A d_B})$ or a block-positive, unit trace matrix $\Gamma \in D(\mathbb{C}^{d_A}) \hat{\otimes} D(\mathbb{C}^{d_B})$.

Corollary 6.10. *For every $\Gamma \in D(\mathbb{C}^{d_A}) \hat{\otimes} D(\mathbb{C}^{d_B})$ and multimeters $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^1$, $N : D(\mathbb{C}^{d_B}) \rightarrow CS_{l,r}^1$ there is no Bell inequality $W \in (CS_{k,g}^*)^+ \hat{\otimes} (CS_{l,r}^*)^+$ with $W((M \otimes N)(\Gamma)) < 0$ and $W((\tilde{M} \otimes \tilde{N})(\varrho)) \geq 0$ for all multimeters $\tilde{M} : D(\mathbb{C}^{d_A}) \rightarrow CS_{k,g}^1$, $\tilde{N} : D(\mathbb{C}^{d_B}) \rightarrow CS_{l,r}^1$ and density matrices $\varrho \in D(\mathbb{C}^{d_A d_B})$.*

Note that the correctness of this theorem is not clear anymore when extending to the general multipartite case. The technique of the given proof breaks down since there is no quantum realization for every steering assemblage in case of more than two parties as it was shown in [SBC⁺15] by considering PR-boxes again.

6.2. CHSH inequalities as incompatibility witnesses. This section is about proving a theorem linking CHSH inequalities and isomorphisms between cones with square basis, and about applying it in order to retrieve interesting results in quantum theory. The CHSH scenario, in which two parties perform two dichotomic (also binary or two-outcome) measurements, therefore plays a special role among Bell experiments. Most commonly, a CHSH inequality is represented by quantum correlations, which are expectation values of the products of outcomes ± 1 . A lower and an upper bound together with 4 positions of a negative sign lead to 8 CHSH inequalities.

$$\begin{aligned} -2 &\leq E_{1,1} + E_{1,2} + E_{2,1} - E_{2,2} \leq 2 \\ -2 &\leq E_{1,1} + E_{1,2} - E_{2,1} + E_{2,2} \leq 2 \\ -2 &\leq E_{1,1} - E_{1,2} + E_{2,1} + E_{2,2} \leq 2 \\ -2 &\leq -E_{1,1} + E_{1,2} + E_{2,1} + E_{2,2} \leq 2 \end{aligned} \tag{6.8}$$

where $E_{x,y} = p_{1,1|x,y} - p_{1,2|x,y} - p_{2,1|x,y} + p_{2,2|x,y}$. Here, the first index corresponds to the first party and the second index to the second party, whereas the value indicates which measurement is being used. Those inequalities hold for all classical correlations and can be violated by nonlocal theories such as quantum mechanics or theories containing PR-boxes.

Theorem 6.11. *The tensors $\xi_L \in (CS_{2,2}^*)^+ \hat{\otimes} (CS_{2,2}^*)^+$ corresponding to linear isomorphisms $L : CS_{2,2}^+ \rightarrow (CS_{2,2}^*)^+$ are in one-to-one correspondence with CHSH inequalities.*

This was also observed in [Jen18, Example 14]. Here, we give a more elementary proof. We first provide an outline of the proof.

- (1) We show that the cone $CS_{2,2}^+$ is generated by matrices $\{X_i\}_{i \in [4]}$ satisfying the square condition and we map them to vectors $\{x_i\}_{i \in [4]}$ in the Euclidean space $\mathbb{R}^3$.
- (2) For the vectors $\{x_i\}_{i \in [4]} \subset \mathbb{R}^3$ we construct dual vectors $\{\varphi_j\}_{j \in [4]} \subset (\mathbb{R}^3)^* \cong \mathbb{R}^3$ and we make sure that both sets of vectors lie in a square and span the corresponding vector space.
- (3) A linear isomorphism L maps vectors $\{x_i\}_{i \in [4]}$ to dual vectors $\{\varphi_j\}_{j \in [4]}$ in $(\mathbb{R}^3)^*$. Using the one-to-one correspondence of linear maps and maximal tensor products of cones, we can decompose the tensor ξ_L .
- (4) We find the dual matrices $\{F_j\}_{j \in [4]}$ corresponding to the set of dual vectors and specify ξ_L .
- (5) The decomposition enables us to show that ξ_L is a CHSH inequality.
- (6) We show that all CHSH inequalities correspond to such a tensor ξ_L , or equivalently, to such an isomorphism L .

Proof. Any matrix from $CS_{2,2}^1$ can be written in the following form, with $a, b \in [0, 1]$:

$$\begin{bmatrix} a & b \\ 1-a & 1-b \end{bmatrix} = a \begin{bmatrix} 1 & b \\ 0 & 1-b \end{bmatrix} + (1-a) \begin{bmatrix} 0 & b \\ 1 & 1-b \end{bmatrix} \tag{6.9}$$

$$= a \left(b \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} + (1-b) \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \right) + (1-a) \left(b \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} + (1-b) \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix} \right) \tag{6.10}$$

So we can define the generating matrices

$$X_1 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}, X_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, X_3 = \begin{bmatrix} 0 & 0 \\ 1 & 1 \end{bmatrix}, X_4 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \quad (6.11)$$

which form the basis of the cone $CS_{2,2}^+$. Note that:

$$X_1 + X_3 = X_2 + X_4, \quad (6.12)$$

which will be referred to as the “square condition” because it says that two diagonals intersect in the center of the shape the 4 points form. Consequently the cone $CS_{2,2}^+$ is generated by a square. We now use an isomorphism $\iota : CS_{2,2}^+ \rightarrow \mathbb{R}^3$, $X_i \mapsto x_i$ to map the four matrices X_i to vectors in Euclidean space $\mathbb{R}^3$ that, satisfying Eq. (6.12), lie in a square.

$$x_1 = \begin{pmatrix} 1 \\ -1 \\ 1 \end{pmatrix}, x_2 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}, x_3 = \begin{pmatrix} 0 \\ 1 \\ 0 \end{pmatrix}, x_4 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad (6.13)$$

The choice was arbitrary except that the vectors $\{x_i\}_{i \in [4]}$ have to satisfy Eq. (6.12) and as such they can be seen as the extremal points of $CS_{2,2}^+$. From here we want to find dual vectors $\{\varphi_j\}_{j \in [4]}$ that fulfill the constraint

$$\varphi_j(x_i) \geq 0. \quad (6.14)$$

To find the dual vectors, fix one $\varphi_j = (a \ b \ c)$ and with Eq. (6.14) one can solve for the components of φ_j .

$$\varphi_1 = (1 \ 1 \ 0), \varphi_2 = (0 \ 1 \ 1), \varphi_3 = (0 \ 0 \ 1), \varphi_4 = \varphi_1 - \varphi_2 + \varphi_3 = (1 \ 0 \ 0) \quad (6.15)$$

Also these dual vectors form a square and they span $(\mathbb{R}^3)^*$. As both cones are generated by a square, they are isomorphic to each other $CS_{2,2}^+ \cong (CS_{2,2}^*)^+$. The linear isomorphism is $L : CS_{2,2}^+ \rightarrow (CS_{2,2}^*)^+$. We can map a square to a square by mapping a vertex to a vertex. Since the isomorphism must be linear, neighboring vertices must remain being neighbors. Therefore we are left with $4 \cdot 2 = 8$ isomorphisms L . For instance, one could choose $\hat{L}$ such that:

$$x_1 \xrightarrow{h} \varphi_2, x_2 \xrightarrow{h} \varphi_3, x_3 \xrightarrow{h} \varphi_4, x_4 \xrightarrow{h} \varphi_1, \quad (6.16)$$

where $h = (\iota^*)^{-1} \circ \hat{L} \circ \iota^{-1} : \mathbb{R}^3 \rightarrow (\mathbb{R}^3)^*$, but up to the isomorphisms this can be thought of as just $\hat{L}$. For an isomorphism L we define the tensor $\xi_L \in (CS_{2,2}^*)^+ \hat{\otimes} (CS_{2,2}^*)^+$:

$$\varphi_{jL}(x_i) = \langle L(x_j), x_i \rangle = \langle L(\iota^{-1}(X_j)), \iota^{-1}(X_i) \rangle = \langle \xi_L, X_j \otimes X_i \rangle, \quad (6.17)$$

Let $\{e_i\}_{i \in [3]}$ and $\{f_j\}_{j \in [3]}$ be the standard bases of $\mathbb{R}^3$ and $\mathbb{R}^{3*}$ respectively, such that $f_j(e_i) = \delta_{ji}$ with the usual Kronecker delta. We write down the action of ξ_L using its basis decomposition $\xi_L = \sum_{k,l=1}^3 \alpha_{kl} \mu_k \otimes \mu_l$, where $\mu_k = \iota^*(f_k)$, so that

$$\langle \xi_L, X_i \otimes X_j \rangle = \left\langle \sum_{k,l=1}^3 \alpha_{kl} \mu_k \otimes \mu_l, \iota^{-1}(x_i) \otimes \iota^{-1}(x_j) \right\rangle \quad (6.18)$$

$$= \sum_{k,l=1}^3 \alpha_{kl} \langle \mu_k, \iota^{-1}(x_i) \rangle \langle \mu_l, \iota^{-1}(x_j) \rangle \quad (6.19)$$

$$= \sum_{k,l=1}^3 \alpha_{kl} \langle (\iota^{-1})^*(\mu_k), x_i \rangle \langle (\iota^{-1})^*(\mu_l), x_j \rangle \quad (6.20)$$

$$= \left\langle \sum_{k,l=1}^3 \alpha_{kl} f_k \otimes f_l, x_i \otimes x_j \right\rangle \quad (6.21)$$

and the coefficients are found via $\langle \xi_L, \iota^{-1}(e_k) \otimes \iota^{-1}(e_l) \rangle = (L(e_k))(e_l) = \alpha_{kl}$. Since the set of dual vectors $\{\varphi_j\}_{j \in [4]}$ span $(\mathbb{R}^3)^*$, the standard basis $\{f_j\}_{j \in [3]}$ can be expressed through them and vice versa. Expressing ξ_L with the φ 's yields for the above example:

$$\langle \xi_L, \cdot \rangle = \langle \varphi_3 \otimes \varphi_2 + \varphi_2 \otimes \varphi_4 + (\varphi_4 - \varphi_3) \otimes \varphi_3, \iota \otimes \iota(\cdot) \rangle \quad (6.22)$$

Due to the isomorphisms ι and ξ_L there must be a correspondence between the functionals $\{\varphi_j\}_{j \in [4]}$ and dual matrices $\{F_j\}_{j \in [4]}$ via:

$$\text{Tr}(F_j^\top X_i) = \langle F_j, X_i \rangle = \langle F_j \circ \iota^{-1} \circ \iota, X_i \rangle = \langle \varphi_j, \iota(X_i) \rangle = \varphi_j(x_i) \quad (6.23)$$

Recall that the matrices X contain conditional probabilities $X_{a|x} = p_{a|x}$ for $a, x \in [2]$. So, the result of Eq. (6.23) has to be a combination of (conditional) probabilities. Let q be an arbitrary vector in convex hull of $\{x_1, x_2, x_3, x_4\}$ in $\mathbb{R}^3$ corresponding to some column stochastic matrix

$$\begin{bmatrix} p_{1|1} & p_{1|2} \\ p_{2|1} & p_{2|2} \end{bmatrix}. \quad (6.24)$$

Then, there is a convex combination of matrices $\{X_i\}_{i \in [4]}$, such that:

$$\varphi_j(q) = \left\langle \varphi_j, \iota \left(\sum_{i=1}^4 \lambda_i X_i \right) \right\rangle = \text{Tr} \left(F_j^\top \begin{bmatrix} p_{1|1} & p_{1|2} \\ p_{2|1} & p_{2|2} \end{bmatrix} \right) \quad (6.25)$$

This can be written down explicitly for every q by calculating $\varphi_j(x_i)$ because $\{x_i\}_{i \in [4]}$ span $\mathbb{R}^3$. This is explained best by showing one instance, that is, we fix one index j , say φ_1 . We have

$$\varphi_1(x_1) = \varphi_1(x_4) = 0, \varphi_1(x_2) = \varphi_1(x_3) = 1 \quad (6.26)$$

and the matrices $\{X_i\}_{i \in [4]}$ were given above. By solving Eq. (6.23) for every i we can deduce that $F_1^\top = \begin{bmatrix} 1 & 1 \\ -1 & 0 \end{bmatrix}$ and so

$$\varphi_1(q) = \text{Tr} \left(F_1^\top \begin{bmatrix} p_{1|1} & p_{1|2} \\ p_{2|1} & p_{2|2} \end{bmatrix} \right) = p_{1|1} + p_{2|1} - p_{1|2} = 1 - p_{1|2} = p_{2|2}. \quad (6.27)$$

Note that the construction of F_j is not unique although matrices, vectors and dual vectors are fixed. We repeat the same procedure for the remaining φ_j to compute $\varphi_j(q)$ and find:

$$\varphi_1(q) = p_{2|2}, \varphi_2(q) = p_{2|1}, \varphi_3(q) = p_{1|2}, \varphi_4(q) = p_{1|1} \quad (6.28)$$

The results we obtained again satisfy the square condition Eq. (6.12). By construction (recall Eq. (6.14)) $\langle \xi_L, P \rangle \geq 0$ for any P that is a convex combination of product elements, i.e., for any separable $P \in CS_{2,2}^+ \dot{\otimes} CS_{2,2}^+$ and hence $\xi_L \in (CS_{2,2}^*)^+ \dot{\otimes} (CS_{2,2}^*)^+$ for any L . It follows from Lemma 6.8 that ξ_L is a Bell inequality.

For $P = \sum_k \lambda_k Q_k \otimes \tilde{Q}_k \in CS_{2,2}^+ \dot{\otimes} CS_{2,2}^+$, we use the above positivity and plug the obtained expressions of $\varphi_j(q_k)$ and $\varphi_j(\tilde{q}_k)$ for the corresponding vectors q_k and $\tilde{q}_k$ of the matrices Q_k and $\tilde{Q}_k$ into the decomposition Eq. (6.22) to find that

$$\langle \xi_L, P \rangle = \langle (\iota^* \otimes \iota^*)(\varphi_3 \otimes \varphi_2 + \varphi_2 \otimes \varphi_4 + (\varphi_4 - \varphi_3) \otimes \varphi_3), P \rangle \quad (6.29)$$

$$= \sum_{k=1}^n \lambda_k \langle F_3 \otimes F_2 + F_2 \otimes F_4 + (F_4 - F_3) \otimes F_3, Q_k \otimes \tilde{Q}_k \rangle \quad (6.30)$$

$$= \sum_{k=1}^n \lambda_k \left(p_{1|2}^{(k)} \cdot \tilde{p}_{2|1}^{(k)} + p_{2|1}^{(k)} \cdot \tilde{p}_{1|1}^{(k)} + p_{1|1}^{(k)} \cdot \tilde{p}_{1|2}^{(k)} - p_{1|2}^{(k)} \cdot \tilde{p}_{1|2}^{(k)} \right) \quad (6.31)$$

$$= p_{1,2|2,1} + p_{2,1|1,1} + p_{1,1|1,2} - p_{1,1|2,2} \geq 0, \quad (6.32)$$

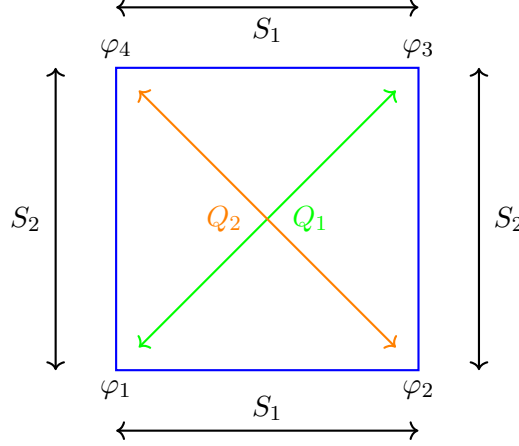


FIGURE 2. The isomorphisms S_1 , S_2 , Q_1 and Q_2 are reflections of the square and can also be used to get counterclockwise rotations by $\frac{\pi}{2}$, π and $\frac{3\pi}{2}$. The arrows indicate the effects of the maps, that is, they reflect across the symmetry axes of the square. As a consequence, measurement choices and outcomes get swapped.

where we have set $p_{a,b|x,y} = \sum_{k=1}^n \lambda_k p_{a|x}^{(k)} \tilde{p}_{b|y}^{(k)}$, which is a LHV model for P . We want to rewrite this in terms of expectation values in order to show that this is indeed a CHSH inequality. Using the normalization of the probabilities and the marginal probabilities as well as standard correlations

$$E_{x,y} = p_{1,1|x,y} - p_{1,2|x,y} - p_{2,1|x,y} + p_{2,2|x,y}, \quad (6.33)$$

the inequality obtained from Eq. (6.32) is equivalent to

$$E_{1,1} + E_{2,1} + E_{2,2} - E_{1,2} \leq 2. \quad (6.34)$$

This is a CHSH inequality. Constructing the remaining CHSH inequalities now becomes easier as we can obtain them by applying the symmetries of a square to $\hat{L}$.

By symmetries of a square we mean reflections along the symmetry axes and rotations as indicated in Fig. 2, i.e. the elements of the dihedral group D_4 , which is known to have 8 of them. By comparing the effects of the symmetries in Fig. 2 with the explicit expressions in Eq. (6.28), it can readily be verified that the symmetries have physical implications, e.g. $\varphi_1(q) = p_{2|2} \xrightarrow{S_1} \varphi_2(q) = p_{2|1}$.

- (1) id: identity
- (2) S_1 : reflection along vertical axis (flip measurement choice)
- (3) S_2 : reflection along horizontal axis (flip measurement and outcome)
- (4) Q_1 : reflection along diagonal $\varphi_1 - \varphi_3$ axis (flip outcomes of second measurement)
- (5) Q_2 : reflection along diagonal $\varphi_2 - \varphi_4$ axis (flip outcomes of first measurement)
- (6) $R_{\frac{\pi}{2}} = S_1 \circ Q_2$: rotation by $\frac{\pi}{2}$
- (7) $R_{\pi} = S_1 \circ S_2$: rotation by π
- (8) $R_{\frac{3\pi}{2}} = S_2 \circ Q_2$: rotation by $\frac{3\pi}{2}$

A key step in obtaining the inequality corresponding to $\hat{L}$ was to find the decomposition of $\xi_{\hat{L}}$ in terms of $\{\varphi_j\}_{j \in [4]}$ (see Eq. (6.22)). By concatenating $\hat{L}$ with an element from D_4 , that is, another isomorphism between squares, we get a decomposition of a new tensor ξ_L . This allows for recycling the previously calculated α_{kl} , e.g.

$$L(e_1) = L(x_2) = R_{\pi} \circ \hat{L}(x_2) = R_{\pi}(\varphi_3) = \varphi_1 = \hat{L}(x_4) = \hat{L}(e_3) \Rightarrow \alpha_{1l} = \alpha_{3l}^{(\hat{L})} \forall l \in [3]. \quad (6.35)$$

This leads to another CHSH inequality by repeating the same procedure as before and in the end, to all CHSH inequalities.

$$\langle \xi_{R_{\frac{\pi}{2}} \circ \hat{L}}, P \rangle = p_{1,2|2,1} + p_{1,1|1,1} - p_{1,2|1,2} + p_{2,2|2,2} \geq 0 \Leftrightarrow -2 \leq E_{1,1} - E_{2,1} + E_{2,2} + E_{1,2} \quad (6.36)$$

$$\langle \xi_{S_1 \circ \hat{L}}, P \rangle = p_{2,2|1,1} + p_{1,2|1,2} - p_{2,2|2,2} + p_{2,1|2,1} \geq 0 \Leftrightarrow -E_{1,1} + E_{2,1} + E_{2,2} + E_{1,2} \leq 2 \quad (6.37)$$

$$\langle \xi_{R_{\pi \circ \hat{L}}, P \rangle = p_{1,1|1,1} + p_{2,2|2,1} - p_{2,1|2,2} + p_{2,1|1,2} \geq 0 \Leftrightarrow -2 \leq E_{1,1} + E_{2,1} + E_{2,2} - E_{1,2} \quad (6.38)$$

$$\langle \xi_{S_2 \circ \hat{L}}, P \rangle = p_{1,2|1,1} + p_{1,2|1,2} - p_{1,2|2,2} + p_{1,1|2,1} \geq 0 \Leftrightarrow -2 \leq -E_{1,1} + E_{2,1} + E_{2,2} + E_{1,2} \quad (6.39)$$

$$\langle \xi_{R_{\frac{3\pi}{2}} \circ \hat{L}}, P \rangle = p_{1,2|1,1} + p_{2,1|1,2} - p_{1,1|2,2} + p_{1,1|2,1} \geq 0 \Leftrightarrow E_{1,1} - E_{2,1} + E_{2,2} + E_{1,2} \leq 2 \quad (6.40)$$

$$\langle \xi_{Q_1 \circ \hat{L}}, P \rangle = p_{1,2|1,2} + p_{2,1|1,1} - p_{1,1|2,1} + p_{1,1|2,2} \geq 0 \Leftrightarrow E_{1,1} + E_{2,1} - E_{2,2} + E_{1,2} \leq 2 \quad (6.41)$$

$$\langle \xi_{Q_2 \circ \hat{L}}, P \rangle = p_{1,1|1,2} + p_{2,2|1,1} - p_{1,2|2,1} + p_{1,2|2,2} \geq 0 \Leftrightarrow -2 \leq E_{1,1} + E_{2,1} - E_{2,2} + E_{1,2} \quad (6.42)$$

This concludes that all 8 CHSH inequalities can be obtained by such an isomorphism L from above. $\square$

Of course, there are easier ways to derive the CHSH inequality and the above result was already discovered. So, let us immediately make use of the obtained isomorphism. In [WPGF09] the authors showed the equivalence of dichotomic POVMs being incompatible and them allowing for violation of the CHSH inequality. We reformulate this result in our terms and prove it using our established formalism.

Theorem 6.12. *A multimeter $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{2,2}^1$ is incompatible if and only if there exists another multimeter $N : D(\mathbb{C}^{d_B}) \rightarrow CS_{2,2}^1$ and a bipartite entangled state $\varrho \in D(\mathbb{C}^{d_A d_B})$ such that the resulting behavior violates the CHSH inequality.*

Proof. To begin with, let $M : D(\mathbb{C}^{d_A}) \rightarrow CS_{2,2}^1$ be a multimeter and let us consider the corresponding tensor $\xi_M \in \text{PSD}_{d_A}^* \hat{\otimes} CS_{2,2}^+$. From Lemma A.10 we get a witness-type functional W_M such that $\langle W_M, \xi_M \rangle < 0$ if ξ_M is not separable, i.e., M is incompatible.

$$W_M \in \text{PSD}_{d_A} \hat{\otimes} (CS_{2,2}^*)^+ = (\text{id} \otimes L)(\text{PSD}_{d_A} \hat{\otimes} CS_{2,2}^+), \quad (6.43)$$

where the linear isomorphism $L : CS_{2,2}^+ \rightarrow (CS_{2,2}^*)^+$ was used. We are free to choose a normalization here, so we can take $W_M \in (\text{id} \otimes L)(D(\mathbb{C}^{d_A}) \hat{\otimes} CS_{2,2}^1)$. The set $D(\mathbb{C}^{d_A}) \hat{\otimes} CS_{2,2}^1$ is the already known set of assemblages σ . Recall that for each assemblage $\sigma \in D(\mathbb{C}^{d_A}) \hat{\otimes} CS_{2,2}^1$ there is a multimeter $N : D(\mathbb{C}^{d_B}) \rightarrow CS_{2,2}^1$ and a bipartite state $\varrho \in D(\mathbb{C}^{d_A d_B})$ that reproduce it as $\sigma = (\text{id} \otimes N)(\varrho)$. Thus, there exists such σ such that $W_M = (\text{id} \otimes L)(\sigma)$. By expressing $\varrho = \sum_a \alpha_a \otimes \beta_a$ we have that

$$0 > \langle W_M, \xi_M \rangle = \langle (\text{id} \otimes L)(\sigma), \xi_M \rangle = \sum_a \langle \alpha_a \otimes L(N(\beta_a)), \xi_M \rangle \quad (6.44)$$

$$= \sum_a \langle M(\alpha_a), L(N(\beta_a)) \rangle = \sum_a \langle \xi_L, M(\alpha_a) \otimes N(\beta_a) \rangle \quad (6.45)$$

$$= \langle \xi_L, (M \otimes N)(\varrho) \rangle, \quad (6.46)$$

where $\xi_L \in (CS_{2,2}^*)^+ \hat{\otimes} (CS_{2,2}^*)^+$ is the tensor corresponding to the map L which by Theorem 6.11 can be interpreted as a CHSH inequality.

For the converse direction, one just needs to read the previous equations backward and conclude $\langle (\text{id} \otimes L)(\sigma), \xi_M \rangle < 0$. Now $W_M = (\text{id} \otimes L)(\sigma)$ is an incompatibility witness for M . As an image of the assemblage $\sigma \in D(\mathbb{C}^{d_A}) \hat{\otimes} CS_{2,2}^1$ under the positive map $\text{id} \otimes L : \text{PSD}_{d_A} \hat{\otimes} CS_{2,2}^+ \rightarrow \text{PSD}_{d_A} \hat{\otimes} (CS_{2,2}^*)^+$ we have that $W_M \in \text{PSD}_{d_A} \hat{\otimes} (CS_{2,2}^*)^+$. By the definition of the maximal tensor product we then have that $\langle W_M, \xi \rangle \geq 0$ for all $\xi \in \text{PSD}_{d_A}^* \hat{\otimes} CS_{2,2}^+$. Thus, W_M detects the incompatibility of M . $\square$

The more well-known direction is the implication that if the set of measurements M enables the violation of a Bell inequality, then it must be incompatible. Usually, this is shown via contraposition: If in a bipartite scenario a Bell inequality is violated, then all measurements must have been incompatible because the measurement incompatibility on both parties is necessary. As soon as one party has more than two measurements, incompatibility does not imply Bell nonlocality anymore [HQB18, BV18].

We go a bit further and relax the conditions on one party. So now suppose that the second party has an arbitrary number of measurements g as well as arbitrary many outcomes k , whereas without loss of generality k can be set to an equal constant for every measurement by adding zero outcomes. This gives some measurement statistics, which could arise from any (nonlocal) model. Then the same witness-type argument can be made as before. Again, all nontrivial facets of the Bell polytope can be constructed from CHSH, which was found in [Pir14].

Theorem 6.13. *For any Bell inequality $W \in (CS_{2,2}^*)^+ \hat{\otimes} (CS_{k,g}^*)^+$ there is a positive map $\Xi^* : CS_{k,g}^+ \rightarrow CS_{2,2}^+$ such that W can be expressed through the CHSH inequality as $W = (\text{id} \otimes \Xi^*)(\xi_L)$, where $\xi_L \in (CS_{2,2}^*)^+ \hat{\otimes} (CS_{2,2}^*)^+$ is the tensor corresponding to L in Theorem 6.11.*

The channel Ξ^* can be understood as a post-processing, enabling the reduction of the witness W to a CHSH inequality.

Proof. We can rewrite $W \in (CS_{2,2}^*)^+ \hat{\otimes} (CS_{k,g}^*)^+$ by using the isomorphism $L : CS_{2,2}^+ \rightarrow (CS_{2,2}^*)^+$, which is the already familiar CHSH inequality, simply as $W = (L \otimes \text{id})(L^{-1} \otimes \text{id})(W) = (L \otimes \text{id})(\xi_\Xi)$, where we have defined $\xi_\Xi = (L^{-1} \otimes \text{id})(W) \in CS_{2,2}^+ \hat{\otimes} (CS_{k,g}^*)^+$. Now ξ_Ξ corresponds to a positive map $\Xi : (CS_{2,2}^*)^+ \rightarrow (CS_{k,g}^*)^+$.

Now if we write $P = \sum_i A_i \otimes B_i$ we see that

$$0 > \langle W, P \rangle = \sum_i \langle (L \otimes \text{id})(\xi_\Xi), P \rangle = \sum_i \langle \xi_\Xi, L^*(A_i) \otimes B_i \rangle = \sum_i \langle L^*(A_i), \Xi^*(B_i) \rangle \quad (6.47)$$

$$= \sum_i \langle \xi_L, A_i \otimes \Xi^*(B_i) \rangle = \langle \xi_L, (\text{id} \otimes \Xi^*)(P) \rangle, \quad (6.48)$$

where $\Xi^* : CS_{k,g}^+ \rightarrow CS_{2,2}^+$. The map Ξ^* is positive since its dual map Ξ is. □

Up to normalization, we see that Ξ^* is exactly the type of map that was characterized in Theorem 4.5 as classical simulations. The way Ξ^* acts on P is by reducing the $k \times g$ column stochastic matrix to a 2×2 column stochastic matrix.

6.3. Bell inequalities in general probabilistic theories. In this section, we will again look at Bell inequalities from a factorization point of view. The following in particular proves Lemma 6.4.

Proposition 6.14. $\xi \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$ if and only if there exist no-signaling probability distributions $(p_{a,b|x,y})_{a \in [k], b \in [l], x \in [g], y \in [r]}$ such that

$$\xi = \sum_{x=1}^g \sum_{y=1}^r \sum_{a=1}^{k-1} \sum_{b=1}^{l-1} p_{a,b|x,y} e_a^{(x)} \otimes e_b^{(y)} + \sum_{x=1}^g \sum_{a=1}^{k-1} \left(\sum_{b=1}^l p_{a,b|x,1} \right) e_a^{(x)} \otimes s_{l,\dots,l} \quad (6.49)$$

$$+ \sum_{y=1}^r \sum_{b=1}^{l-1} \left(\sum_{a=1}^k p_{a,b|1,y} \right) s_{k,\dots,k} \otimes e_b^{(y)} + s_{k,\dots,k} \otimes s_{l,\dots,l}. \quad (6.50)$$

Proof. Let $\xi \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$. Then, we can use the basis of the polysimplex to write

$$\xi = \sum_{j=1}^g \sum_{q=1}^r \sum_{i=1}^{k-1} \sum_{p=1}^{l-1} \lambda_{i,p|j,q} e_i^{(j)} \otimes e_p^{(q)} + \sum_{j=1}^g \sum_{i=1}^{k-1} \mu_{i|j} e_i^{(j)} \otimes s_{l,\dots,l} + \sum_{q=1}^r \sum_{p=1}^{l-1} \mu'_{p|q} s_{k,\dots,k} \otimes e_p^{(q)} + \eta s_{k,\dots,k} \otimes s_{l,\dots,l}. \quad (6.51)$$

We infer $\langle \mathbb{1}_{CS_{k,g}^1} \otimes \mathbb{1}_{CS_{l,r}^1}, \xi \rangle = \eta$, thus $\eta = 1$. Moreover, using that the $m_i^{(j)}$ are positive functionals for all i and j , we obtain

$$\langle m_i^{(j)} \otimes m_p^{(q)}, \xi \rangle = \lambda_{i,p|j,q} \geq 0 \quad \forall j \in [g], q \in [r], i \in [k-1], p \in [l-1], \quad (6.52)$$

$$\langle m_i^{(j)} \otimes m_l^{(q)}, \xi \rangle = \mu_{i|j} - \sum_{p=1}^{l-1} \lambda_{i,p|j,q} \geq 0 \quad \forall j \in [g], q \in [r], i \in [k-1], \quad (6.53)$$

$$\langle m_k^{(j)} \otimes m_p^{(q)}, \xi \rangle = \mu'_{p|q} - \sum_{i=1}^{k-1} \lambda_{i,p|j,q} \geq 0 \quad \forall j \in [g], q \in [r], p \in [l-1], \quad (6.54)$$

$$\langle m_i^{(j)} \otimes \mathbb{1}_{CS_{l,r}^1}, \xi \rangle = \mu_{i|j} \geq 0 \quad \forall j \in [g], i \in [k-1], \quad (6.55)$$

$$\langle \mathbb{1}_{CS_{g,k}^1} \otimes m_p^{(q)}, \xi \rangle = \mu'_{p|q} \geq 0 \quad \forall q \in [r], p \in [l-1] \quad (6.56)$$

and

$$\langle m_k^{(j)} \otimes m_l^{(q)}, \xi \rangle = 1 + \sum_{i=1}^{k-1} \sum_{p=1}^{l-1} \lambda_{i,p|j,q} - \sum_{i=1}^{k-1} \mu_{i|j} - \sum_{p=1}^{l-1} \mu'_{p|q} \geq 0. \quad (6.57)$$

We can therefore define $p_{a,b|x,y} = \langle m_a^{(x)} \otimes m_b^{(y)}, \xi \rangle$ for all $x \in [g], y \in [r], a \in [k], b \in [l]$, i.e.,

$$p_{a,b|x,y} = \begin{cases} \lambda_{a,b|x,y} & x \in [g], y \in [r], a \in [k-1], b \in [l-1] \\ \mu'_{b|y} - \sum_{a=1}^{k-1} \lambda_{a,b|x,y} & x \in [k], y \in [r], a = k, b \in [l-1] \\ \mu_{a|x} - \sum_{b=1}^{l-1} \lambda_{a,b|x,y} & x \in [k], y \in [r], b = l, a \in [k-1] \\ 1 + \sum_{a=1}^{k-1} \sum_{b=1}^{l-1} \lambda_{a,b|x,y} - \sum_{a=1}^{k-1} \mu_{a|x} - \sum_{b=1}^{l-1} \mu'_{b|y} & x \in [k], y \in [r], a = k, b = l \end{cases}$$

From the inequalities above, we can verify that the $(p_{a,b|x,y})_{a \in [k], b \in [l]}$ defined in this way form valid probability distributions. As $\sum_i m_i^{(j)} = \mathbb{1}_{CS_{g,k}^1}$ for all j , they satisfy the no-signaling condition, as

$$\sum_{b=1}^l p_{a,b|x,y} = \mu_{a|x} \quad \forall y \in [r], \quad (6.58)$$

$$\sum_{a=1}^k p_{a,b|x,y} = \mu'_{b|y} \quad \forall x \in [g]. \quad (6.59)$$

The converse follows as the $m_i^{(j)}$ form the extreme rays of $(CS_{g,k}^+)^*$ and $(CS_{l,r}^+)^*$. $\square$

By the general correspondence between tensors and positive maps (see Section 3.2) we can also equate the no-signaling probability distributions with some positive maps.

Corollary 6.15. $\Phi : (CS_{k,g}^+)^* \rightarrow CS_{l,r}^+$ with $\Phi(\mathbb{1}_{CS_{k,g}^1}) \in CS_{l,r}^1$ if and only if and there exist no-signaling probability distributions $(p_{a,b|x,y})_{a \in [k], b \in [l]}$ such that

$$\langle m_b^{(y)}, \Phi(m_a^{(x)}) \rangle = p_{a,b|x,y} \quad \forall a \in [k], \forall b \in [l], \forall x \in [g], \forall y \in [r] \quad (6.60)$$

$$\langle \mathbb{1}_{CS_{l,r}^1}, \Phi(m_a^{(x)}) \rangle = \sum_{b=1}^l p_{a,b|x,1} \quad \forall a \in [k], \forall x \in [g] \quad (6.61)$$

$$\langle m_b^{(y)}, \Phi(\mathbb{1}_{CS_{k,g}^1}) \rangle = \sum_{a=1}^k p_{a,b|1,y} \quad \forall b \in [l], \forall y \in [r] \quad (6.62)$$

$$\langle \mathbb{1}_{CS_{l,r}^1}, \Phi(\mathbb{1}_{CS_{k,g}^1}) \rangle = 1 \quad (6.63)$$

Proof. This follows directly from Lemma 3.4 and Proposition 6.14. $\square$

Analogously to K -simulability of multimeters and assemblages, we can look into K -simulability of no-signaling distributions. It turns out this relates to how the no-signaling distribution can be realized in terms of multimeters and state assemblages on the simulating state space.

Theorem 6.16. *Let $\Phi : (CS_{k,g}^+)^* \rightarrow CS_{l,r}^+$ with $\Phi(\mathbb{1}_{CS_{k,g}^1}) \in CS_{l,r}^1$ and let $(V(K), V(K)^+, \mathbb{1}_K)$ be a GPT. Furthermore, let $\Psi : (CS_{k,g}^+)^* \rightarrow V(K)^+ \dot{\otimes} S_\Lambda^+$ such that $\Psi(\mathbb{1}_{CS_{k,g}^1}) \in K \dot{\otimes} S_\Lambda$ and let $F : K \dot{\otimes} S_\Lambda \rightarrow CS_{l,r}^1$. Then, the following diagram commutes*

$$\begin{array}{ccc} (CS_{k,g}^+)^* & \xrightarrow{\Phi} & CS_{l,r}^+ \\ & \searrow \Psi & \nearrow F \\ & V(K)^+ \dot{\otimes} S_\Lambda^+ & \end{array} \quad (6.64)$$

if and only if the no-signaling probability distributions defining Φ (see Corollary 6.15) satisfy

$$p_{a,b|x,y} = \sum_{\lambda=1}^\Lambda w_\lambda \langle F_{b|y,\lambda}, q_{a|x,\lambda} \varrho_{a|x,\lambda} \rangle \quad \forall a \in [k], \forall b \in [l], \forall x \in [g], \forall y \in [r]. \quad (6.65)$$

Here, $(w_\lambda)_{\lambda \in [\Lambda]}$ is a probability distribution, $(q_{a|x,\lambda})_{a \in [k]}$ are some conditional probability distributions, $F_{b|y,\lambda} \in A(K)^+$ are such that $\sum_{b=1}^l F_{b|y,\lambda} = \mathbb{1}_K$ for all $y \in [r]$ and $\lambda \in [\Lambda]$; in other words, F is a multimeter. The $\varrho_{a|x,\lambda} \in K$ are such that $\sum_{a=1}^k q_{a|x,\lambda} \varrho_{a|x,\lambda} = \bar{\varrho}_\lambda$ for all $x \in [g]$ and $\lambda \in [\Lambda]$ with $\bar{\varrho}_\lambda \in K$. Finally, $\sum_{\lambda \in [\Lambda]} w_\lambda \bar{\varrho}_\lambda = \bar{\varrho} \in K$, i.e., $\{q_{a|x,\lambda}, \varrho_{a|x,\lambda}\}_{a \in [k], x \in [g]}$ is an assemblage for all $\lambda \in [\Lambda]$.

Proof. We start by realizing that any map $F : K \dot{\otimes} S_\Lambda \rightarrow CS_{l,r}^1$ is a multimeter and can therefore be written as $F = \{F_{b|y,\lambda}\}_{y \in [r], \lambda \in [\Lambda]}$ (see also Lemma 4.3). In particular, $m_b^{(y)} \circ F(\cdot \otimes \delta_\lambda) = F_{b|y,\lambda}$. Moreover, we find that

$$\Psi(m_a^{(x)}) =: \tilde{\sigma}_{a|x} \in V(K)^+ \dot{\otimes} S_\Lambda^+ \quad \forall a \in [k], \forall x \in [g] \quad (6.66)$$

$$\Psi(\mathbb{1}_{CS_{k,g}^1}) =: \bar{\sigma} \in K \dot{\otimes} S_\Lambda \quad (6.67)$$

by definition and $\sum_{a=1}^k \tilde{\sigma}_{a|x} = \bar{\sigma}$ as $\sum_{a=1}^k m_a^{(x)} = \mathbb{1}_{CS_{k,g}^1}$ for all $x \in [g]$. We can furthermore decompose

$$\tilde{\sigma}_{a|x} = \sum_{\lambda \in [\Lambda]} \tilde{\varrho}_{a,\lambda|x} \otimes \delta_\lambda \quad (6.68)$$

and

$$\bar{\sigma} = \sum_{\lambda=1}^\Lambda w_\lambda \varrho_\lambda \otimes \delta_\lambda \quad (6.69)$$

We can set $\bar{\varrho} := (\text{id} \otimes \mathbb{1}_{S_\Lambda}) \bar{\sigma}$ and $\tilde{q}_{a,\lambda|x} = \mathbb{1}_K(\tilde{\varrho}_{a,\lambda|x})$ such that $\tilde{\varrho}_{a,\lambda|x} = \tilde{q}_{a,\lambda|x} \varrho_{a|x,\lambda}$ with $\varrho_{a|x,\lambda} \in K$ for all $a \in [k]$, $\lambda \in [\Lambda]$ and $x \in [g]$. The equation $\sum_{a=1}^k \tilde{q}_{a,\lambda|x} \varrho_{a|x,\lambda} = w_\lambda \bar{\varrho}_\lambda$ can be easily inferred

from a comparison of Eq. (6.68) and Eq. (6.69), which also implies that the $(\tilde{q}_{a,\lambda|x})_{a \in [k], \lambda \in [\Lambda]}$ are conditional probability distributions. We can define a probability distribution

$$q_{a|x,\lambda} = \begin{cases} q_{a,\lambda|x}/w_\lambda & w_\lambda \neq 0 \\ 1/k & w_\lambda = 0 \end{cases}. \quad (6.70)$$

Note that for $w_\lambda = 0$ we can just choose $\bar{\varrho}_\lambda$ as we please in Eq. (6.69), thus enforcing the condition $\sum_{a=1}^k q_{a|x,\lambda} \varrho_{a|x,\lambda} = \bar{\varrho}_\lambda$ also in this case. The assertion follows then from the relations in Corollary 6.15, as Ψ can easily be defined from an assemblage in the way shown above. $\square$

Remark 6.17. *The roles of $CS_{l,r}$ and $CS_{k,g}$ in Theorem 6.16 can be reversed. In quantum theory, we can find bipartite states $\varrho_{AB}^{(\lambda)}$ and a multimeter such that $q_{a|x,\lambda} \varrho_{a|x,\lambda} = \text{Tr}_A[(E_{a|x,\lambda} \otimes \mathbb{1}_B)\varrho_{AB}^{(\lambda)}]$. Thus, Eq. (6.65) can be written as*

$$p_{a,b|x,y} = \sum_{\lambda=1}^{\Lambda} w_\lambda \text{Tr}[(E_{a|x,\lambda} \otimes F_{b|y,\lambda})\varrho_{AB}^{(\lambda)}] \quad \forall a \in [k], \forall b \in [l], \forall x \in [g], \forall y \in [r]. \quad (6.71)$$

Similarly we can consider classical simulability of no-signaling probability distributions and characterize this in terms of a particular factorization. We note that since we want to consider classically simulating no-signaling distributions with no-signaling distributions, we cannot simply consider any joint classical simulations $\Phi : CS_{k_1,l_1,g_1,h_1}^1 \rightarrow CS_{k_2,l_2,g_2,h_2}^1$ since they do not necessarily preserve the no-signaling property. Instead, we consider a separate classical simulation process in which we classically simulate the first input-output pair and the second input-output pair separately. It turns out that this kind of classical simulation results in more of a sequential factorization process. This is formalized in the following result:

Theorem 6.18. *Let $\Phi : (CS_{k,g}^+)^* \rightarrow CS_{l,r}^+$ with $\Phi(\mathbb{1}_{CS_{k,g}^1}) \in CS_{l,r}^1$. Then, the no-signaling probability distributions defining Φ (see Corollary 6.15) satisfy*

$$p_{a,b|x,y} = \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} q_{c,d|w,z} \right], \quad (6.72)$$

where $(\pi_{w|x})_{w \in [u]}$, $(\theta_{z|y})_{z \in [v]}$, $(\nu_{a|c,x,w})_{a \in [k]}$ and $(\mu_{b|d,y,z})_{b \in [l]}$ are conditional probability distributions and $(q_{c,d|w,z})_{c \in [s], d \in [t]}$ are no-signaling probability distributions if and only if there exist maps $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow (CS_{s,u}^*)^+$ with $\Psi_1(\mathbb{1}_{CS_{k,g}^1}) = \mathbb{1}_{CS_{s,u}^1}$, $\Psi_2 : (CS_{s,u}^*)^+ \rightarrow CS_{t,v}^+$ with $\Psi_2(\mathbb{1}_{CS_{s,u}^1}) \in CS_{t,v}^1$, and a channel $\Psi_3 : CS_{t,v}^1 \rightarrow CS_{l,r}^1$ such that the following diagram commutes:

$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & CS_{l,r}^+ \\ & \searrow \Psi_1 & \nearrow \Psi_3 \\ & (CS_{s,u}^*)^+ & \xrightarrow{\Psi_2} CS_{t,v}^+ \end{array} \quad (6.73)$$

Proof. Let us define $\langle m_d^{(z)}, \Psi_2(m_c^{(w)}) \rangle =: q_{c,d|w,z}$ for all $c \in [s]$, $d \in [t]$, $w \in [u]$, $z \in [v]$ using Corollary 6.15. We will use Theorem 4.5 again to obtain for any $\alpha \in (CS_{k,g}^*)^+$

$$\Psi_1(\alpha) = \alpha(s_{k,\dots,k}) \mathbb{1}_{CS_{s,u}^1} + \sum_{x=1}^g \sum_{a=1}^{k-1} \alpha(e_a^{(x)}) \left[\sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} m_c^{(w)} \right] \quad (6.74)$$

and for any $X \in CS_{t,v}$

$$\Psi_3(X) = \mathbb{1}_{CS_{t,v}^1}(X) s_{l,\dots,l} + \sum_{y=1}^r \sum_{b=1}^{l-1} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} m_d^{(z)}(X) \right] e_b^{(y)}. \quad (6.75)$$

Hence, we infer for $a \in [k-1]$, using Corollary 6.15,

$$\Psi_3(\Psi_2(\Psi_1(m_a^{(x)}))) = \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \mathbb{1}_{CS_{t,v}^1}(\Psi_2(m_c^{(w)})) s_{l,\dots,l} \quad (6.76)$$

$$+ \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \sum_{y=1}^r \sum_{b=1}^{l-1} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} m_d^{(z)}(\Psi_2(m_c^{(w)})) \right] e_b^{(y)} \quad (6.77)$$

$$= \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \sum_{d=1}^t q_{c,d|w,1} s_{l,\dots,l} \quad (6.78)$$

$$+ \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \sum_{y=1}^r \sum_{b=1}^{l-1} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} q_{c,d|w,z} \right] e_b^{(y)} \quad (6.79)$$

Moreover,

$$\Psi_3(\Psi_2(\Psi_1(\mathbb{1}_{CS_{k,g}^1}))) = \mathbb{1}_{CS_{t,v}^1}(\Psi_2(\mathbb{1}_{CS_{s,u}^1})) s_{l,\dots,l} + \sum_{y=1}^r \sum_{b=1}^{l-1} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} m_d^{(z)}(\Psi_2(\mathbb{1}_{CS_{s,u}^1})) \right] e_b^{(y)} \quad (6.80)$$

$$= s_{l,\dots,l} + \sum_{y=1}^r \sum_{b=1}^{l-1} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} \sum_{c=1}^s q_{c,d|1,z} \right] e_b^{(y)}. \quad (6.81)$$

Thus, for $b \in [l-1]$ and $a \in [k-1]$,

$$p_{a,b|x,y} = \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} q_{c,d|w,z} \right]. \quad (6.82)$$

For $b = l$ and $a \in [k-1]$, we have using the no-signaling of $q_{c,d|w,z}$

$$p_{a,l|x,y} = \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \sum_{d=1}^t q_{c,d|w,1} \quad (6.83)$$

$$- \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \sum_{b=1}^{l-1} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} q_{c,d|w,z} \right] \quad (6.84)$$

$$= \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t q_{c,d|w,z} \underbrace{\left(1 - \sum_{b=1}^{l-1} \mu_{b|d,y,z} \right)}_{\mu_{l|d,y,z}}. \quad (6.85)$$

For $b \in [l-1]$ and $a = k$, it follows that, using the no-signaling of $q_{c,d|w,z}$,

$$p_{k,b|x,y} = m_b^{(y)}[\Psi_3(\Psi_2(\Psi_1(\mathbb{1}_{CS_{k,g}^1})))] - \sum_{a=1}^{k-1} m_b^{(y)}[\Psi_3(\Psi_2(\Psi_1(m_a^{(x)})))] \quad (6.86)$$

$$= \sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} \sum_{c=1}^s q_{c,d|1,z} - \sum_{a=1}^{k-1} \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \left[\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} q_{c,d|w,z} \right] \quad (6.87)$$

$$= \sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s q_{c,d|w,z} \underbrace{\left(1 - \sum_{a=1}^{k-1} \nu_{a|c,x,w} \right)}_{\nu_{k|c,x,w}}. \quad (6.88)$$

Finally, we find, using again the no-signaling of $q_{c,d|w,z}$,

$$p_{k,l|x,y} = \left(\mathbb{1}_{CS_{l,r}^1} - \sum_{b=1}^{h-1} m_b^{(y)} \right) \left[\Psi_3(\Psi_2(\Psi_1(\mathbb{1}_{CS_{k,g}^1}))) - \sum_{a=1}^{k-1} \Psi_3(\Psi_2(\Psi_1(m_a^{(x)}))) \right] \quad (6.89)$$

$$= 1 - \sum_{b=1}^{h-1} \sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{b|d,y,z} \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s q_{c,d|w,z} \nu_{k|c,x,w} \quad (6.90)$$

$$- \sum_{a=1}^{k-1} \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s \nu_{a|c,x,w} \underbrace{\sum_{d=1}^t q_{c,d|w,1}}_{\sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t q_{c,d|w,z}} \underbrace{1}_{\sum_{b=1}^l \mu_{b|d,y,z}} \quad (6.91)$$

$$= \sum_{z=1}^v \theta_{z|y} \sum_{d=1}^t \mu_{l|d,y,z} \sum_{w=1}^u \pi_{w|x} \sum_{c=1}^s q_{c,d|w,z} \nu_{k|c,x,w}. \quad (6.92)$$

The converse is straightforward as we can define the maps Ψ_1 , Ψ_2 and Ψ_3 via the conditional and no-signaling probabilities. $\square$

Finally, as with multimeters and assemblages, for no-signaling probability distributions we can equate the special case of factoring through a simplex and being separable with having a classical realization, i.e., having an LHV model.

Corollary 6.19. *Let $\Phi : (CS_{k,g}^*)^+ \rightarrow CS_{l,r}^+$ with $\Phi(\mathbb{1}_{CS_{k,g}^1}) \in CS_{l,r}^1$. Then, the following are equivalent:*

- (1) $(\Phi \otimes \text{id})(\chi_{CS_{k,g}^1}) \in CS_{l,r}^1 \dot{\otimes} CS_{k,g}^1$.
- (2) *The associated no-signaling probability distributions $(p_{a,b|x,y})_{a \in [k], b \in [l]}$ by Corollary 6.15 admit an LHV model.*
- (3) *There exist a finite number of outcomes Λ , a map $\Psi_1 : (CS_{k,g}^*)^+ \rightarrow S_\Lambda^+$ such that $\Psi(\mathbb{1}_{CS_{k,g}^1}) \in S_\Lambda$, and a channel $\Psi_2 : S_\Lambda \rightarrow CS_{l,r}^1$ such that the following diagram commutes:*

$$\begin{array}{ccc} (CS_{k,g}^*)^+ & \xrightarrow{\Phi} & CS_{l,r}^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & S_\Lambda^+ & \end{array} \quad (6.93)$$

Proof. The equivalence of (1) and (3) follows from Proposition 3.7. The equivalence of (2) and (3) follows from Theorem 6.16 with $K = S_1$ as follows: We obtain that (3) is equivalent to

$$p_{a,b|x,y} = \sum_{\lambda=1}^{\Lambda} w_\lambda q_{a|x,\lambda} f_{b|y,\lambda} \quad \forall a \in [k], \forall b \in [l], \forall x \in [g], \forall y \in [r], \quad (6.94)$$

where $(f_{b|y,\lambda})_{b \in [l]}$ is a conditional probability distribution and $q_{a|x,\lambda} = 1$ for all $a \in [k]$, $x \in [g]$, $\lambda \in [\Lambda]$ in this case. This gives an LHV model for $(p_{a,b|x,y})_{a \in [k], b \in [l], x \in [g], y \in [r]}$. $\square$

7. SEPARABILITY AND EXTENDABILITY

The notion of *extendability* has been recognized to be central in the context of quantum entanglement [DPS04]. In this section relate the extendability of relevant tensors to intermediate notion of classicality in the setting of measurement compatibility (Section 7.1) and existence of LHV models for correlations (Section 7.2). In order to introduce the concept of extendability, we recall a known instance from quantum theory. Consider a separable bipartite quantum state

$\varrho_{AB} = \sum_{\lambda} p(\lambda) \varrho_{\lambda}^{(A)} \otimes \varrho_{\lambda}^{(B)}$. We can construct a possible extension of this state with finitely many parties as follows:

$$\sigma_{AB_1 B_2 \dots B_n} = \sum_{\lambda} p(\lambda) \varrho_{\lambda}^{(A)} \otimes \left(\varrho_{\lambda}^{(B)} \right)^{\otimes n}, \quad (7.1)$$

such that

- (1) $\text{Tr}_{B_2 \dots B_n} [\sigma_{AB_1 B_2 \dots B_n}] = \varrho_{AB}$
- (2) $\pi_{B_1 B_2 \dots B_n} (\sigma_{AB_1 B_2 \dots B_n}) = \sigma_{AB_1 B_2 \dots B_n}$, where π permutes $B_1, \dots, B_n$

In [DPS04] it was found that if a quantum state $\sigma_{AB_1 B_2 \dots B_n}$ having properties (1) and (2) exists for all $n \geq 2$, then ϱ_{AB} is separable. On the other hand, if there is some $n \geq 2$ with no state $\sigma_{AB_1 B_2 \dots B_n}$ satisfying above properties, then ϱ_{AB} must be entangled. This result was also recently generalized to arbitrary convex cones [AMHP25, ZKSP26], where one obtains the very same hierarchy, but trace is replaced by a strictly positive functional.

We will first of all argue that factorizability can be solved using the generalization of the aforementioned methods of symmetric extensions. Consider the case where we are given state spaces K_A, K_B and a map $\Phi : K_A \rightarrow K_B$. We want to determine whether given the state space K_C there are maps $\Psi_1 : K_A \rightarrow K_C \dot{\otimes} S_{\Lambda}$ and $\Psi_2 : K_C \dot{\otimes} S_{\Lambda} \rightarrow K_B$ such that the following diagram commutes:

$$\begin{array}{ccc} V(K_A)^+ & \xrightarrow{\Phi} & V(K_B)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & V(K_C \dot{\otimes} S_{\Lambda})^+ & \end{array} \quad (7.2)$$

Here, as before, S_{Λ} represents a classical side information. For every value $\lambda \in \Lambda$ we define the partial maps $\Psi_{1,\lambda} : V(K_A)^+ \rightarrow V(K_C)^+$ and $\Psi_{2,\lambda} : V(K_C)^+ \rightarrow V(K_B)^+$ by fixing the value of λ in the output or input respectively. Then, commutativity of the diagram (7.2) is equivalent to

$$\Phi = \sum_{\lambda} \Psi_{2,\lambda} \circ \Psi_{1,\lambda}. \quad (7.3)$$

Defining the evaluation functional $\chi_{V(K_C)}$ on $V(K_C) \otimes A(K_C)$ via

$$\chi_{V(K_C)}(x \otimes f) = f(x) \quad (7.4)$$

for $x \in K_C, f \in A(K_C)$, we get that the commutativity of (7.2) is equivalent to

$$\Phi = (\text{id} \otimes \chi_{V(K_C)} \otimes \text{id}) \left(\sum_{\lambda} \Psi_{1,\lambda} \otimes \Psi_{2,\lambda} \right). \quad (7.5)$$

It thus follows that commutativity of (7.2) is equivalent to the existence of a separable tensor $\sum_{\lambda} \Psi_{2,\lambda} \otimes \Psi_{1,\lambda}$ such that (7.5) holds. Thus one can use the hierarchy of symmetric extensions [AMHP25] that generalizes the idea of (7.1) to rule out the existence of suitable maps $\Psi_{1,\lambda}$ and $\Psi_{2,\lambda}$. We will investigate this in detail in the case of joint measurability bellow, but before that we have to address the case without classical side information.

Consider again state spaces K_A, K_B and a map $\Phi : K_A \rightarrow K_B$. We want to determine whether given the state space K_C there are maps $\Psi_1 : K_A \rightarrow K_C$ and $\Psi_2 : K_C \rightarrow K_B$ such that the following diagram commutes:

$$\begin{array}{ccc} V(K_A)^+ & \xrightarrow{\Phi} & V(K_B)^+ \\ & \searrow \Psi_1 \quad \nearrow \Psi_2 & \\ & V(K_C)^+ & \end{array} \quad (7.6)$$

Repeating the steps as above, we get that this is equivalent to

$$\Phi = (\text{id} \otimes \chi_{V(K_C)} \otimes \text{id})(\Psi_1 \otimes \Psi_2). \quad (7.7)$$

In this case $\Psi_1 \otimes \Psi_2$ is a product tensor rather than a separable tensor and one clearly cannot just replace $\Psi_1 \otimes \Psi_2$ with an arbitrary separable tensor $\sum_\lambda \Psi_{2,\lambda} \otimes \Psi_{1,\lambda}$ as this would change the definition of the problem. Nevertheless even in this case one can construct a hierarchy of symmetric extensions that could be used to rule out the existence of $\Psi_1 : K_A \rightarrow K_C$ and $\Psi_2 : K_C \rightarrow K_B$ such that (7.6) commutes using the methods outlined in [PLG25].

7.1. Extendability of measurements. Let us now investigate the case of extendability and joint measurability in depth. We will start by recalling the first two theorems from [AMHP25]. Let us consider proper cones V_A^+ and V_B^+ in vector spaces V_A and V_B . First we define a reduction map of n -th order:

$$\gamma_n^\Phi : V_B^{\otimes n} \rightarrow V_B, \quad \gamma_n^\Phi = \frac{1}{n} \sum_{i=1}^n \Phi^{\otimes i-1} \otimes \text{id} \otimes \Phi^{\otimes n-i}, \quad (7.8)$$

where Φ is a linear form in the interior of $(V_B^+)^*$, e.g., the trace function Tr in the case of quantum theory. We make the following definition:

Definition 7.1. *Let Φ be a linear form in the interior of $(V_B^+)^*$. A tensor $\xi \in V_A^+ \hat{\otimes} V_B^+$ is n -extendable with respect to Φ if there exists $\xi^{(n)} \in V_A^+ \hat{\otimes} (V_B^+)^{\otimes n}$ such that $(\text{id} \otimes \gamma_n^\Phi)(\xi^{(n)}) = \xi$.*

As shown in [AMHP25], if a tensor is n -extendable then it is also $n-1$ extendable so that we get a hierarchy. In particular, Theorem 1 of [AMHP25] states the following:

Lemma 7.2 ([AMHP25]). *A tensor $\xi \in V_A^+ \hat{\otimes} V_B^+$ is separable, i.e., $\xi \in V_A^+ \dot{\otimes} V_B^+$, if and only if it is n -extendable for all $n \in \mathbb{N}$ with respect to some linear form Φ in the interior of $(V_B^+)^*$, i.e.,*

$$V_A^+ \dot{\otimes} V_B^+ = \bigcap_{n \geq 1} (\text{id} \otimes \gamma_n^\Phi) (V_A^+ \hat{\otimes} (V_B^+)^{\otimes n}) \quad (7.9)$$

Thus, the consequence is similar as in quantum theory, we get a separable vector via reducing all finite extensions of it. The procedure in Eq. (7.9) can be thought as cutting the set of all vectors closer to the subset of separable vectors with increasing n , eventually converging. Although in practice this is not very applicable unless one wants to check a small set of n -extensions where the reduction fails in order to find an entangled vector; otherwise one would need to check infinitely many n which is naturally not feasible.

In special cases, however, [AMHP25, Theorem 2] provides a remedy. Under a certain condition there is a threshold, i.e., one must only check finitely many extensions up to some fixed n .

Lemma 7.3 ([AMHP25]). *For $n \geq 1$ the following are equivalent:*

- (i) $V_A^+ \dot{\otimes} V_B^+ = (\text{id} \otimes \gamma_n^\Phi) (V_A^+ \hat{\otimes} (V_B^+)^{\otimes n})$
- (ii) *The base $K_\Phi = V_B^+ \cap \Phi^{-1}(1)$ is affinely equivalent to a Cartesian product of at most n simplices, that is, a polysimplex with at most n simplices.*

We want to apply this result to multimeters in order to find a criterion for compatibility. To do this, let $M = \{M_{\cdot|x}\}_{x \in [g]}$ be a multimeter that we identify with the tensor $\xi_M \in A(K)^+ \hat{\otimes} CS_{k,g}^+$. From now on, we fix the reduction map Φ as the equal sums of the columns, that is, $\Phi(C) = \mathbb{1}_{CS_{k,g}^1}(C) = \sum_{i=1}^k C_{ij}$ for $j \in [g]$ and $C \in CS_{k,g}^+$, and when we consider n -extendability we mean n -extendability with respect to this choice of Φ . Now the base $K_\Phi = CS_{k,g}^+ \cap \Phi^{-1}(1)$ is the set of column stochastic matrices $CS_{k,g}^1$. As already discussed, the state space $CS_{k,g}^1$ is a special case of a polysimplex and in particular it is affinely isomorphic to a Cartesian product of g simplices S_k .

Thus, by combining Proposition 3.7 and Lemma 7.3 we get the following characterization for compatibility:

Corollary 7.4. *A multimeter $M : K \rightarrow CS_{k,g}^1$ is compatible if and only if $\xi_M \in A(K)^+ \hat{\otimes} CS_{k,g}^+$ is g -extendable.*

The next question that arises is can we conclude something about M by considering its n -extensions for $n < g$? It turns out that we can and that it is related to the concept of n -wise compatibility.

Definition 7.5. A multimeter $M = \{M_{\cdot|x}\}_{x \in [g]}$ is said to be n -wise compatible with $n \leq g$ if for all $\{x_1, \dots, x_n\} \subset [g]$ the measurements $\{M_{\cdot|x_i}\}_{i \in [n]}$ are compatible.

Here, the choice of n measurements from M is arbitrary, i.e., if $n < g$, every subset of n measurements from M is compatible. Of course, if $n = g$ then M is (fully) compatible. The key to connecting n -extendability to n -wise compatibility is to see that applying the reduction map γ_n^Φ to an n -extension $\xi^{(n)}$ of ξ_M corresponds to obtaining M as a marginal of the map corresponding to $\xi^{(n)}$.

In order to show the connection, we first need some structural results on the extensions. The extensions of $\xi_M \in A(K)^+ \hat{\otimes} CS_{k,g}^+$ are tensors $A(K)^+ \hat{\otimes} (CS_{k,g}^+)^{\hat{\otimes} n}$. Equivalently we can consider the extensions as positive maps from $V(K)^+$ to $(CS_{k,g}^+)^{\hat{\otimes} n}$. In fact, if $M : K \rightarrow CS_{k,g}^1$ is a multimeter, and $\xi^{(n)} \in A(K)^+ \hat{\otimes} (CS_{k,g}^+)^{\hat{\otimes} n}$ is its n -extension, then actually the corresponding map $N : V(K)^+ \rightarrow (CS_{k,g}^+)^{\hat{\otimes} n}$ is a channel (we will prove this in Theorem 7.8). Thus, we will look into properties of normalized elements in $(CS_{k,g}^+)^{\hat{\otimes} n}$ similarly as we did in Proposition 6.14.

Recall that the symmetric group $\mathfrak{S}_n$ is the set of permutations of the set $[n]$, i.e., the set of bijective functions $\pi : [n] \rightarrow [n]$. For a permutation $\pi \in \mathfrak{S}_n$, we denote the *descent set* of π by $\text{desc}(\pi)$ as the set of indices $i \in [n-1]$ such that $\pi(i) > \pi(i+1)$. Note that for the identity permutation e we have $\text{desc}(e) = \emptyset$ and it is the only permutation with this property. For $m \in [n-1]$, we denote $\Pi_m := \{\pi \in \mathfrak{S}_n : \text{desc}(\pi) \subseteq \{m\}\}$, i.e., it is the set of permutations with at most one descent at index m . Equivalently, $\Pi_m = \{e\} \cup \{\pi \in \mathfrak{S}_n : \text{desc}(\pi) = \{m\}\}$. We note that $|\Pi_m| = \binom{n}{m}$ as it characterizes in how many ways we can pick the elements $\pi(1) < \dots < \pi(m)$ from the set $[n]$. We will also make use of the fact that for all $g \in \mathbb{N}$ the symmetric group $\mathfrak{S}_n$ acts naturally on $[g]^n$ by the linear map $\alpha_\pi : [g]^n \rightarrow [g]^n$ defined as $\alpha_\pi(y_1, \dots, y_n) = (y_{\pi(1)}, \dots, y_{\pi(n)})$ for all $(y_1, \dots, y_n) \in [g]^n$ for all $\pi \in \mathfrak{S}_n$. Similarly $\mathfrak{S}_n$ also acts on $(CS_{k,g})^{\hat{\otimes} n}$ by the linear map $\mathcal{U}_\pi : (CS_{k,g})^{\hat{\otimes} n} \rightarrow (CS_{k,g})^{\hat{\otimes} n}$ defined as $\mathcal{U}_\pi(X_1 \otimes \dots \otimes X_n) = X_{\pi(1)} \otimes \dots \otimes X_{\pi(n)}$ for all $X_i \in CS_{k,g}$ for all $\pi \in \mathfrak{S}_n$.

Proposition 7.6. $\xi \in (CS_{k,g}^1)^{\hat{\otimes} n}$ with $2 \leq n \in \mathbb{N}$ if and only if there exists a no-signalling conditional probability distribution $p = (p_{\dots, \cdot | x_1, \dots, x_n})_{x_1, \dots, x_n \in [g]}$ on $[k]^n$ such that

$$\begin{aligned} \xi &= (s_{k, \dots, k})^{\otimes n} + \sum_{x_1=1}^g \dots \sum_{x_n=1}^g \sum_{a_1=1}^{k-1} \dots \sum_{a_n=1}^{k-1} p_{a_1, \dots, a_n | x_1, \dots, x_n} \bigotimes_{i=1}^n e_{a_i}^{(x_i)} \\ &+ \sum_{m=1}^{n-1} \sum_{\pi \in \Pi_m} \sum_{x_{\pi(1)}=1}^g \dots \sum_{x_{\pi(m)}=1}^g \sum_{a_{\pi(1)}=1}^{k-1} \dots \sum_{a_{\pi(m)}=1}^{k-1} \sum_{a_{\pi(m+1)}=1}^k \dots \sum_{a_{\pi(n)}=1}^k p_{a_1, \dots, a_n | \alpha_\pi(x_{\pi(1)}, \dots, x_{\pi(m)}, 1, \dots, 1)} \\ &\quad \times \mathcal{U}_\pi \left(e_{a_{\pi(1)}}^{(x_{\pi(1)})} \otimes \dots \otimes e_{a_{\pi(m)}}^{(x_{\pi(m)})} \otimes (s_{k, \dots, k})^{\otimes n-m} \right). \end{aligned} \tag{7.10}$$

Proof. We start by writing a given tensor $\xi \in (CS_{k,g}^1)^{\otimes n}$ in the tensor basis of $CS_{k,g}$:

$$\begin{aligned} \xi = & \eta(s_{k,\dots,k})^{\otimes n} + \sum_{x_1=1}^g \cdots \sum_{x_n=1}^g \sum_{a_1=1}^{k-1} \cdots \sum_{a_n=1}^{k-1} N_{a_1,\dots,a_n|x_1,\dots,x_n} \bigotimes_{i=1}^n e_{a_i}^{(x_i)} + \sum_{m=1}^{n-1} \sum_{\pi \in \Pi_m} \sum_{x_{\pi(1)}=1}^g \cdots \sum_{x_{\pi(m)}=1}^g \\ & \sum_{a_{\pi(1)}=1}^{k-1} \cdots \sum_{a_{\pi(m)}=1}^{k-1} N_{a_{\pi(1)},\dots,a_{\pi(m)}|x_{\pi(1)},\dots,x_{\pi(m)}}^{(\pi)} \mathcal{U}_{\pi} \left(e_{a_{\pi(1)}}^{(x_{\pi(1)})} \otimes \cdots \otimes e_{a_{\pi(m)}}^{(x_{\pi(m)})} \otimes (s_{k,\dots,k})^{\otimes n-m} \right). \end{aligned} \quad (7.11)$$

To see that we get all the basis elements exactly once by applying the permutations in Π_m for all $m \in [n-1]$ we argue the following: Let us fix $m \in [n-1]$. Let us consider the tensor product as a list of length n and let us consider filling it with the m terms of $e_{a_1}^{(x_1)}, \dots, e_{a_m}^{(x_m)}$ and the $n-m$ copies of $s_{k,\dots,k}$. (Note that in Eq. (7.11) we also relabel the indices i of x_i and a_i but just to get all the basis elements we do not need to consider that). In general there are exactly $n!$ ways to do this, but in exactly $m!$ ways the $e_{a_i}^{(x_i)}$ terms are in the same m positions, just differently permuted. We only need one of these $m!$ products and so we select the products where the relative order of the terms $e_{a_i}^{(x_i)}$ remains the same, i.e., where $e_{a_j}^{(x_j)}$ appears before $e_{a_{j+1}}^{(x_{j+1})}$. Thus, there are now $\frac{n!}{m!}$ of these products. To actually get these products, let us start from the product $e_{a_1}^{(x_1)} \otimes \cdots \otimes e_{a_m}^{(x_m)} \otimes (s_{k,\dots,k})^{\otimes n-m}$. Since we want to preserve the relative order of the $e_{a_i}^{(x_i)}$'s we can only apply permutations $\pi \in \mathfrak{S}_n$ which have an ascending run of $[m]$, i.e., the set of indices i where $\pi(i) > \pi(i+1)$ is $[m]$. By taking all such permutations we get all $\frac{n!}{m!}$ products that we wanted. However, in these products there are still multiplicities of the basis elements where the $s_{k,\dots,k}$ terms are permuted among themselves. Since there are $(n-m)!$ ways to permute the $s_{k,\dots,k}$ terms, in the end we want to end up with $\binom{n}{m} = \frac{n!}{(n-m)!m!}$ products corresponding to the basis elements. One way to achieve this is to take only those permutations which also preserve the relative ordering of the $s_{k,\dots,k}$, i.e., those permutations which have an ascending run of $\{m+1, \dots, n\}$. Thus, in the end we are left with permutations with one ascending run $[n]$, which corresponds to the identity permutation, or two ascending runs $[m]$ and $[n] \setminus [m]$. This is equivalent to saying that we have permutations with either zero descends (the identity permutation) or exactly one descent at position m . But this is exactly how we defined Π_m so in the end we thus get all the basis elements exactly once for all $m \in [n-1]$.

Since ξ is a state, it is clear that $\langle \mathbb{1}_{CS_{k,g}^1}^{\otimes n}, \xi \rangle = 1 = \eta$. We make use of the positivity of the measurements $m_a^{(x)}$ for all $a \in [k]$ and $x \in [g]$ to define the conditional probability distribution $p = (p_{\cdot,\dots,\cdot|x_1,\dots,x_n})_{x_1,\dots,x_n \in [g]}$ on $[k]^n$ as follows:

$$p_{a_1,\dots,a_n|x_1,\dots,x_n} = \langle \bigotimes_{i=1}^n m_{a_i}^{(x_i)}, \xi \rangle \quad \forall a_i \in [k], x_i \in [g], i \in [n] \quad (7.12)$$

Clearly, since $\sum_{a=1}^k m_a^{(x)} = \mathbb{1}_{CS_{k,g}^1}$ for all $x \in [g]$, we have:

$$\sum_{a_1=1}^k \cdots \sum_{a_n=1}^k p_{a_1,\dots,a_n|x_1,\dots,x_n} = \langle \sum_{a_1=1}^k \cdots \sum_{a_n=1}^k m_{a_1}^{(x_1)} \otimes \cdots \otimes m_{a_n}^{(x_n)}, \xi \rangle = \langle \mathbb{1}_{CS_{k,g}^1}^{\otimes n}, \xi \rangle = 1 \quad (7.13)$$

so that $p = (p_{\cdot,\dots,\cdot|x_1,\dots,x_n})_{x_1,\dots,x_n \in [g]}$ is indeed a conditional probability distribution on $[k]^n$ for all $x_1, \dots, x_n \in [g]$.

What remains to see is that p is no-signaling and that the decompositions in Eq. (7.10) and Eq. (7.11) match. To see that p is no-signaling, we fix some $\{i_1, \dots, i_v\} \subset [n]$ for some $v \in [n-1]$. Let $\sigma \in \mathfrak{S}_n$ be the unique permutation with at most one descent at index l such that $\sigma(j) = i_j$ for

all $j \in [l]$. Thus, $\sigma \in \Pi_l$. We fix some $a_{i_j} \in [k]$ and $x_{i_j} \in [g]$ for all $j \in [l]$. Now for all $x_i \in [g]$, $i \in [n] \setminus \{i_1, \dots, i_l\} = \sigma([n] \setminus [l])$ we have that

$$\sum_{a_{\sigma(l+1)}=1}^k \cdots \sum_{a_{\sigma(n)}=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_n} \quad (7.14)$$

$$= \sum_{i \in \sigma([n] \setminus [l])} \sum_{a_i=1}^k \langle m_{a_1}^{(x_1)} \otimes \cdots \otimes m_{a_{i_1}}^{(x_{i_1})} \otimes \cdots \otimes m_{a_{i_l}}^{(x_{i_l})} \otimes \cdots \otimes m_{a_n}^{(x_n)}, \xi \rangle \quad (7.15)$$

$$= \langle \mathbb{1}_{CS_{k,g}^1} \otimes \cdots \otimes \mathbb{1}_{CS_{k,g}^1} \otimes m_{a_{\sigma(1)}}^{(x_{\sigma(1)})} \otimes \mathbb{1}_{CS_{k,g}^1} \otimes \cdots \otimes \mathbb{1}_{CS_{k,g}^1} \otimes m_{a_{\sigma(l)}}^{(x_{\sigma(l)})} \otimes \mathbb{1}_{CS_{k,g}^1} \otimes \cdots \otimes \mathbb{1}_{CS_{k,g}^1}, \xi \rangle \quad (7.16)$$

$$= \left\langle \mathcal{U}_{\sigma^{-1}}^* \left(m_{a_{\sigma(1)}}^{(x_{\sigma(1)})} \otimes \cdots \otimes m_{a_{\sigma(l)}}^{(x_{\sigma(l)})} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes n-l} \right), \xi \right\rangle, \quad (7.17)$$

where $\mathcal{U}_{\sigma^{-1}}^* : (CS_{k,g}^*)^{\otimes n} \rightarrow (CS_{k,g}^*)^{\otimes n}$ is the adjoint of $\mathcal{U}_{\sigma^{-1}}$ and arranges the tensor product of the vectors in $CS_{k,g}^*$ in the same order as $\mathcal{U}_{\sigma}$ arranges the tensor product of the vectors in $CS_{k,g}$. Clearly the result is independent of what we choose as $x_i \in [g]$ for $i \in \sigma([n] \setminus [l])$ so that the probability distribution is no-signaling. Thus, in this case we may choose $x_i = 1$ for all $i \in \sigma([n] \setminus [l])$. Furthermore, if $a_{i_j} \in [k-1]$ for all $j \in [l]$, then we have that

$$\sum_{a_{\sigma(l+1)}=1}^k \cdots \sum_{a_{\sigma(n)}=1}^k p_{a_1, \dots, a_n | \alpha_{\sigma}(x_{\sigma(1)}, \dots, x_{\sigma(l)}, 1, \dots, 1)} = \left\langle \mathcal{U}_{\sigma^{-1}}^* \left(m_{a_{\sigma(1)}}^{(x_{\sigma(1)})} \otimes \cdots \otimes m_{a_{\sigma(l)}}^{(x_{\sigma(l)})} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes n-l} \right), \xi \right\rangle \quad (7.18)$$

$$= N_{a_{\sigma(1)}, \dots, a_{\sigma(l)} | x_{\sigma(1)}, \dots, x_{\sigma(l)}}^{(\sigma)}, \quad (7.19)$$

which gives us exactly the decomposition in Eq. (7.10).

The converse follows as the $m_a^{(x)}$ form the extreme rays of $(CS_{g,k}^+)^*$. $\square$

Remark 7.7. *With the procedure in the above proof we also could have shown the same statement with $\xi \in CS_{k_1, g_1}^1 \hat{\otimes} CS_{k_2, g_2}^1 \hat{\otimes} \cdots \hat{\otimes} CS_{k_n, g_n}^1$, i.e., extending with arbitrary state cones. One just needs to label the indices accordingly and make sure the measurements act on the correct spaces. The challenge is that the notation becomes even more challenging, see Proposition 6.14 for the $n=2$ case. However, if we want to make use of the extension as described in the beginning of this section, it is important to have copies of the same state cone.*

At this point we are able to link the extensions with joint measurements.

Theorem 7.8. *Let $M : K \rightarrow CS_{k,g}^1$ be a multimeter and let $\xi_M \in A(K)^+ \hat{\otimes} CS_{k,g}^+$ be the corresponding tensor. Then M is n -extendable if and only if it is n -wise compatible such that for all measurements $\{x_1, \dots, x_n\} \subset [g]$ there exists joint measurements $G^{(x_1, \dots, x_n)} = \{G_{a_1, \dots, a_n}^{(x_1, \dots, x_n)}\}_{a_1, \dots, a_n=1}^k$ which satisfy the following no-signaling constraints*

$$\sum_{a_i=1}^k G_{a_1, \dots, a_n}^{(x_1, \dots, x_{i-1}, x_i, x_{i+1}, \dots, x_n)} = \sum_{a_i=1}^k G_{a_1, \dots, a_n}^{(x_1, \dots, x_{i-1}, \tilde{x}_i, x_{i+1}, \dots, x_n)} \quad (7.20)$$

for all $a_j \in [k]$, $x_j \in [g]$, $j \in [n] \setminus \{i\}$ and $x_i, \tilde{x}_i \in [g]$ for all $i \in [n]$.

Proof. Let us start with an n -extendable ξ_M . Then there exists $N^{(n)} : V(K)^+ \rightarrow (CS_{k,g}^+)^{\otimes n}$, that is, $\xi_{N^{(n)}} \in A(K)^+ \hat{\otimes} (CS_{k,g}^+)^{\otimes n}$ such that $(\text{id} \otimes \gamma_n^\Phi)(\xi_{N^{(n)}}) = \xi_M$ for $\Phi = \mathbb{1}_{CS_{k,g}^1}$. Let us first see how we can get a given measurement $M_{\cdot|x}$ from the extension:

$$M_{a|x}(\varrho) = \langle M(\varrho), m_a^{(x)} \rangle = \langle \xi_M, \varrho \otimes m_a^{(x)} \rangle \quad (7.21)$$

$$= \langle (\text{id} \otimes \gamma_n^\Phi)(\xi_{N^{(n)}}), \varrho \otimes m_a^{(x)} \rangle = \langle \xi_{N^{(n)}}, \varrho \otimes (\gamma_n^\Phi)^*(m_a^{(x)}) \rangle \quad (7.22)$$

$$= \langle N^{(n)}(\varrho), (\gamma_n^\Phi)^*(m_a^{(x)}) \rangle = \langle \gamma_n^\Phi(N^{(n)}(\varrho)), m_a^{(x)} \rangle \quad (7.23)$$

for all $a \in [k]$, $x \in [g]$ and $\varrho \in K$. In particular we see that $M = \gamma_n^\Phi \circ N^{(n)}$.

Now, for $\varrho \in K$ we see that $N^{(n)}(\varrho) \in (CS_{k,g}^1)^{\hat{\otimes} n}$ so that $N^{(n)}$ is a channel. This follows as M is a channel for all $\varrho \in K$:

$$\langle \mathbb{1}_{CS_{k,g}^1}^{\otimes n}, N^{(n)}(\varrho) \rangle = \langle (\gamma_n^\Phi)^*(\mathbb{1}_{CS_{k,g}^1}), N^{(n)}(\varrho) \rangle = \langle \mathbb{1}_{CS_{k,g}^1}, \gamma_n^\Phi(N^{(n)}(\varrho)) \rangle = \langle \mathbb{1}_{CS_{k,g}^1}, M(\varrho) \rangle = \langle \mathbb{1}_K, \varrho \rangle = 1, \quad (7.24)$$

where according to the results of [AMHP25] we can write $(\gamma_n^\Phi)^*(X) = P_{\text{Sym}_n(CS_{k,g}^*)}(X \otimes \Phi^{\otimes(n-1)})$ for all $X \in CS_{k,g}^*$, where

$$P_{\text{Sym}_n(CS_{k,g}^*)} = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} \mathcal{U}_\sigma^* \quad (7.25)$$

is the symmetric projection onto the symmetric subspace of $(CS_{k,g}^*)^{\hat{\otimes} n}$. Clearly now for $\Phi = \mathbb{1}_{CS_{k,g}^1}$ we have $(\gamma_n^\Phi)^*(\mathbb{1}_{CS_{k,g}^1}) = P_{\text{Sym}_n(CS_{k,g}^*)}(\mathbb{1}_{CS_{k,g}^1}^{\otimes n}) = \mathbb{1}_{CS_{k,g}^1}^{\otimes n}$ and the normalization of $N^{(n)}$ follows.

On the other hand, by using the same expression for $(\gamma_n^\Phi)^*$, we see that

$$M_{a|x}(\varrho) = \langle N^{(n)}(\varrho), (\gamma_u^\Phi)^*(m_a^{(x)}) \rangle = \langle N^{(n)}(\varrho), P_{\text{Sym}_n(CS_{k,g}^*)}(m_a^{(x)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes(n-1)}) \rangle \quad (7.26)$$

for all $a \in [k]$, $x \in [g]$ and $\varrho \in K$. For any $\{x_1, \dots, x_n\} \subset [g]$, we can define the joint measurement $G^{(x_1, \dots, x_n)}$ for $\{M_{\cdot|x_1}, \dots, M_{\cdot|x_n}\}$ as

$$G_{a_1, \dots, a_n}^{(x_1, \dots, x_n)}(\varrho) = \langle N^{(n)}(\varrho), P_{\text{Sym}_n(CS_{k,g}^*)}(m_{a_1}^{(x_1)} \otimes \dots \otimes m_{a_n}^{(x_n)}) \rangle \quad (7.27)$$

for all $a_1, \dots, a_n \in [k]$ and $\varrho \in K$. Clearly $G^{(x_1, \dots, x_n)}$ is a positive map and it correctly gives the measurements $\{M_{\cdot|x_1}, \dots, M_{\cdot|x_n}\}$ as marginals:

$$\sum_{\substack{j=1 \\ j \neq i}}^n \sum_{a_j=1}^k G_{a_1, \dots, a_n}^{(x_1, \dots, x_n)}(\varrho) = \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{a_j=1}^k \langle N^{(n)}(\varrho), P_{\text{Sym}_n(CS_{k,g}^*)}(m_{a_1}^{(x_1)} \otimes \dots \otimes m_{a_n}^{(x_n)}) \rangle \quad (7.28)$$

$$= \langle N^{(n)}(\varrho), P_{\text{Sym}_n(CS_{k,g}^*)}(m_{a_i}^{(x_i)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes(n-1)}) \rangle \quad (7.29)$$

$$= M_{a_i|x_i}(\varrho) \quad (7.30)$$

for all $a_i \in [k]$, $x_i \in [g]$, $i \in [n]$ and $\varrho \in K$. Thus, M is n -wise compatible. Finally, by Proposition 7.6 for all $\varrho \in K$ we have that $G_{a_1, \dots, a_n}^{(x_1, \dots, x_n)}(\varrho)$ is of the form

$$G_{a_1, \dots, a_n}^{(x_1, \dots, x_n)}(\varrho) = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} p_{a_{\sigma(1)}, \dots, a_{\sigma(n)}|x_{\sigma(1)}, \dots, x_{\sigma(n)}}^{N^{(n)}(\varrho)}, \quad (7.31)$$

where $p^{N^{(n)}(\varrho)}$ is the no-signaling probability distribution related to $N^{(n)}(\varrho) \in (CS_{k,g}^1)^{\hat{\otimes} n}$. Alternatively, one sees that $G_{a_1, \dots, a_n}^{(x_1, \dots, x_n)}(\varrho) = p_{a_1, \dots, a_n|x_1, \dots, x_n}^{\tilde{N}^{(n)}(\varrho)}$, where $\tilde{N}^{(n)} = P_{\text{Sym}_n(CS_{k,g})} \circ N^{(n)}$. One then directly sees that the joint measurements inherit the no-signaling constraints from the extension so that Eq. (7.20) is satisfied. This also means that the map $\tilde{N}^{(n)} = P_{\text{Sym}_n(CS_{k,g})} \circ N^{(n)}$ is also a channel.

On the other hand, if M is n -wise compatible and we are given the joint measurements which satisfy the no-signaling conditions in Eq. (7.20), we can explicitly construct the n -extension $\xi_{N^{(n)}}$ of ξ_M . We will do this by using Proposition 7.6 by defining the no-signaling probability distributions

$p^{N^{(n)}(\varrho)}$ for the elements $N^{(n)}(\varrho) \in (CS_{k,g}^1)^{\hat{\otimes} n}$ of the channel $N : K \rightarrow (CS_{k,g}^1)^{\hat{\otimes} n}$ corresponding to the tensor $\xi_{N^{(n)}} \in A(K)^+ \hat{\otimes} (CS_{k,g}^+)^{\hat{\otimes} n}$.

Let us start with the joint measurements $G^{(x_1, \dots, x_n)}$ for all measurement choices $\{x_1, \dots, x_n\} \subset [g]$. Note that for any $m < n$ we can also construct a joint measurement of the measurements $\{M_{|x_1}, \dots, M_{|x_m}\}$ from the joint measurement $G^{(x_1, \dots, x_n)}$ of any set of n measurements $\{M_{|x_1}, \dots, M_{|x_n}\}$ containing $\{M_{|x_1}, \dots, M_{|x_m}\}$ as

$$\tilde{G}_{a_1, \dots, a_m}^{(x_1, \dots, x_m, x_{m+1}, \dots, x_n)} := \sum_{a_{m+1}=1}^k \dots \sum_{a_n=1}^k G_{a_1, \dots, a_m, a_{m+1}, \dots, a_n}^{(x_1, \dots, x_m, x_{m+1}, \dots, x_n)} \quad (7.32)$$

for all $a_1, \dots, a_m \in [k]$. Furthermore, since the joint measurements satisfy the no-signaling conditions in Eq. (7.20), we see that if we choose some different measurement settings $\{\tilde{x}_{m+1}, \dots, \tilde{x}_n\}$ for the remaining $n-m$ measurements and use a joint measurement $G^{(x_1, \dots, x_m, \tilde{x}_{m+1}, \dots, \tilde{x}_n)}$ for the settings $\{x_1, \dots, x_m, \tilde{x}_{m+1}, \dots, \tilde{x}_n\}$ to define the joint measurement of the measurements $\{M_{|x_1}, \dots, M_{|x_m}\}$, we have that

$$\tilde{G}_{a_1, \dots, a_m}^{(x_1, \dots, x_m, \tilde{x}_{m+1}, \dots, \tilde{x}_n)} := \sum_{a_{m+1}=1}^k \dots \sum_{a_n=1}^k G_{a_1, \dots, a_m, a_{m+1}, \dots, a_n}^{(x_1, \dots, x_m, \tilde{x}_{m+1}, \dots, \tilde{x}_n)} \quad (7.33)$$

$$= \sum_{a_{m+1}=1}^k \dots \sum_{a_n=1}^k G_{a_1, \dots, a_m, a_{m+1}, \dots, a_n}^{(x_1, \dots, x_m, x_{m+1}, \dots, x_n)} \quad (7.34)$$

$$= \tilde{G}_{a_1, \dots, a_m}^{(x_1, \dots, x_m, x_{m+1}, \dots, x_n)} \quad (7.35)$$

for all $a_1, \dots, a_m \in [k]$. Thus, the two joint measurements for the measurements $\{M_{|x_1}, \dots, M_{|x_m}\}$ are actually the same so that it only depends on the measurement settings $\{x_1, \dots, x_m\}$.

In the joint measurements the measurement choices are always different from each other but to construct the n -extensions we need to define it also in the cases when some of the measurement choices are identical. We will do this by giving the equal measurements always the same outcome. Next we will do this more formally.

Let us take $x_i \in [g]$ for all $i \in [n]$. Let m be the number of different measurement settings in $\{x_1, \dots, x_n\}$ and let $y_1, \dots, y_m$ be those different measurement settings. Let l_j be the number of measurement settings in $\{x_1, \dots, x_n\}$ which are equal to y_j . Let us partition $[n]$ into intervals of length l_i such that $[n] = \{1, \dots, l_1, l_1 + 1, \dots, l_1 + l_2, l_1 + l_2 + 1, \dots, \sum_{j=1}^m l_j\}$. Let us define $i_r = \sum_{j=1}^r l_j \in [n]$ for all $r \in [m]$ so that now we have $[n] = \{1, \dots, i_1, i_1 + 1, \dots, i_2, \dots, i_m\}$, where $i_m = n$. Let us define a permutation $\sigma \in \mathfrak{S}_n$ as the permutation which orders the measurement settings such that

$$x_{\sigma^{-1}(1)} = \dots = x_{\sigma^{-1}(i_1)} = y_1, \quad (7.36)$$

$$x_{\sigma^{-1}(i_1+1)} = \dots = x_{\sigma^{-1}(i_2)} = y_2, \quad (7.37)$$

$$\vdots \quad (7.38)$$

$$x_{\sigma^{-1}(i_{m-1}+1)} = \dots = x_{\sigma^{-1}(i_m)} = y_m, \quad (7.39)$$

and $\sigma^{-1}(i) < \sigma^{-1}(i')$ for all $i < i'$, $i, i' \in \{i_{r-1} + 1, \dots, i_r\}$ for all $r \in [m]$, where we set $i_0 = 0$. Now we can define the conditional probability distribution $(p_{\cdot, \dots, \cdot | x_1, \dots, x_n}^{N^{(n)}(\varrho)})$ as

$$p_{a_1, \dots, a_n | x_1, \dots, x_n}^{N^{(n)}(\varrho)} = \sum_{b_1=1}^k \dots \sum_{b_m=1}^k \prod_{z=1}^m \delta_{a_{\sigma^{-1}(i_{z-1}+1)}, \dots, a_{\sigma^{-1}(i_z)}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho) \quad (7.40)$$

for all $a_1, \dots, a_n \in [k]$ and $\varrho \in K$, where we are using the Kronecker delta as

$$\delta_{a_1, \dots, a_u} = \begin{cases} 1, & a_1 = \dots = a_u, \\ 0, & \text{else,} \end{cases} \quad \text{for some } u \in \mathbb{N}. \quad (7.41)$$

Clearly $\sum_{i=1}^n \sum_{a_i=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_n}^{N^{(n)}(\varrho)} = 1$ so that it is a valid probability distribution for all $x_1, \dots, x_n \in [g]$ and $\varrho \in K$. As an example, to see how the construction looks like explicitly, say we want to define $p_{a_1, a_2, a_3 | x_1, x_1, x_3}^{N^{(3)}(\varrho)}$ where the first and second measurement choices are equal. Then, we have $y_1 = x_1$ and $y_2 = x_3$ and we pick the joint measurement between the first and the third party:

$$p_{a_1, a_2, a_3 | x_1, x_1, x_3}^{N^{(3)}(\varrho)} = \sum_{b_1=1}^k \sum_{b_2=1}^k \delta_{a_1, a_2, b_1} \delta_{a_3, b_2} G_{b_1, b_2}^{(y_1, y_2)}(\varrho) = \delta_{a_1, a_2} G_{a_1, a_3}^{(x_1, x_3)}(\varrho) \quad (7.42)$$

for all $a_1, a_3 \in [k]$ and $\varrho \in K$. Thus, we have just set the measurement outcomes the same for the equal measurements.

We now observe that if we fix $i \in [n]$ so that $x_i = y_{j_i}$ for some $j_i \in [m]$ then we have that

$$\sum_{\substack{t=1 \\ t \neq i}}^n \sum_{a_t=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_n}^{N^{(n)}(\varrho)} = \sum_{\substack{s=1 \\ s \neq j_i}}^m \sum_{b_s=1}^k G_{b_1, \dots, b_{j_i-1}, a_i, b_{j_i+1}, \dots, b_m}^{(y_1, \dots, y_{j_i}, \dots, y_n)}(\varrho) = M_{a_i | y_{j_i}}(\varrho) = M_{a_i | x_i}(\varrho) \quad (7.43)$$

for all $a_i \in [k]$ and $\varrho \in K$.

According to Proposition 7.6, in order to define the channel $N^{(n)} : K \rightarrow (CS_{k,g}^1)^{\hat{\otimes} n}$ via the conditional probabilities $p^{N^{(n)}(\varrho)} = (p_{\cdot, \dots, \cdot | x_1, \dots, x_n}^{N^{(n)}(\varrho)})_{x_1, \dots, x_n \in [g]}$ for all $\varrho \in K$ we still need to check that the no-signaling constraints are satisfied. Thus, let $r \in [n]$. Now $x_r = y_{j_r}$ for some $j_r \in [m]$ and thus $r \in \sigma^{-1}(\{i_{j_r-1} + 1, \dots, i_{j_r}\})$. Then

$$\sum_{a_r=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_r, \dots, x_n}^{N^{(n)}(\varrho)} = \sum_{a_r=1}^k \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \prod_{z=1}^m \delta_{a_{\sigma^{-1}(i_{z-1}+1)}, \dots, a_{\sigma^{-1}(i_z)}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho) \quad (7.44)$$

$$= \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \left(\sum_{a_r=1}^k \delta_{a_{\sigma^{-1}(i_{j_r-1}+1)}, \dots, a_{\sigma^{-1}(i_{j_r})}, b_{j_r}} \right) \prod_{\substack{z=1 \\ z \neq j_r}}^m \delta_{a_{\sigma^{-1}(i_{z-1}+1)}, \dots, a_{\sigma^{-1}(i_z)}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho) \quad (7.45)$$

On the other hand, let us now change the measurement setting x_r to $\tilde{x}_r$ for some $\tilde{x}_r \in [g]$, $x_r \neq \tilde{x}_r$. Thus, we have the total measurement settings $\{x_1, \dots, x_{r-1}, \tilde{x}_r, x_{r+1}, \dots, x_n\}$. Let first $\tilde{x}_r \in \{y_1, \dots, y_m\}$ so that there exists $\tilde{j}_r \in [m]$ such that $\tilde{x}_r = y_{\tilde{j}_r}$. Just as before, there is a permutation $\tilde{\sigma}$ which groups the same measurement settings together. The positions where the measurement settings change are now denoted by $\tilde{i}_1, \dots, \tilde{i}_m$. Thus, now $r \in \tilde{\sigma}^{-1}(\{\tilde{i}_{\tilde{j}_r-1} + 1, \dots, \tilde{i}_{\tilde{j}_r}\})$. Let now $\pi \in \mathfrak{S}_n$ be a permutation such that $\pi\sigma = \tilde{\sigma}$. We note that now

$$\pi(\{i_{z-1} + 1, \dots, i_z\}) = \{\tilde{i}_{z-1} + 1, \dots, \tilde{i}_z\} \quad (7.46)$$

for all $z \in [m]$. Then we have that

$$\sum_{a_r=1}^k p_{a_1, \dots, a_n | x_1, \dots, \tilde{x}_r, \dots, x_n}^{N^{(n)}(\varrho)} = \sum_{a_r=1}^k \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \prod_{z=1}^m \delta_{a_{\tilde{\sigma}^{-1}(\tilde{i}_{z-1}+1)}, \dots, a_{\tilde{\sigma}^{-1}(\tilde{i}_z)}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho) \quad (7.47)$$

$$= \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \left(\sum_{a_r=1}^k \delta_{a_{\tilde{\sigma}^{-1}(\tilde{i}_{\tilde{j}_r-1}+1)}, \dots, a_{\tilde{\sigma}^{-1}(\tilde{i}_{\tilde{j}_r})}, b_{\tilde{j}_r}} \right) \prod_{\substack{z=1 \\ z \neq \tilde{j}_r}}^m \delta_{a_{\tilde{\sigma}^{-1}(\tilde{i}_{z-1}+1)}, \dots, a_{\tilde{\sigma}^{-1}(\tilde{i}_z)}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho) \quad (7.48)$$

$$\begin{aligned}
&= \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \left(\sum_{a_r=1}^k \delta_{a_{\sigma^{-1}(\pi^{-1}(\tilde{i}_{j_r-1}+1)), \dots, a_{\sigma^{-1}(\pi^{-1}(\tilde{i}_{j_r}))}, b_{j_r}} \right) \\
&\quad \times \prod_{\substack{z=1 \\ z \neq j_r}}^m \delta_{a_{\sigma^{-1}(\pi^{-1}(\tilde{i}_{z-1}+1)), \dots, a_{\sigma^{-1}(\pi^{-1}(\tilde{i}_z))}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho)
\end{aligned} \tag{7.49}$$

$$\begin{aligned}
&= \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \left(\sum_{a_r=1}^k \delta_{a_{\sigma^{-1}(\tilde{i}_{j_r-1}+1), \dots, a_{\sigma^{-1}(\tilde{i}_{j_r})}, b_{j_r}} \right) \prod_{\substack{z=1 \\ z \neq j_r}}^m \delta_{a_{\sigma^{-1}(\tilde{i}_{z-1}+1), \dots, a_{\sigma^{-1}(\tilde{i}_z)}}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho)
\end{aligned} \tag{7.50}$$

$$= \sum_{a_r=1}^k \sum_{b_1=1}^k \cdots \sum_{b_m=1}^k \prod_{z=1}^m \delta_{a_{\sigma^{-1}(\tilde{i}_{z-1}+1), \dots, a_{\sigma^{-1}(\tilde{i}_z)}}, b_z} G_{b_1, \dots, b_m}^{(y_1, \dots, y_m)}(\varrho) \tag{7.51}$$

$$= \sum_{a_r=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_r, \dots, x_n}^{N^{(n)}(\varrho)} \tag{7.52}$$

so that the no-signaling constraints are satisfied.

Above we considered the case when $\tilde{x}_r \in \{y_1, \dots, y_m\}$. However, if $\tilde{x}_r \notin \{y_1, \dots, y_m\}$ we can just repeat the whole procedure by defining $p_{a_1, \dots, a_n | x_1, \dots, x_r, \dots, x_n}^{N^{(n)}(\varrho)}$ and $p_{a_1, \dots, a_n | x_1, \dots, \tilde{x}_r, \dots, x_n}^{N^{(n)}(\varrho)}$ through the joint measurement of the measurement settings $\{y_1, \dots, y_m, y_{m+1}\}$, where now $y_{m+1} = \tilde{x}_r$. Then in Eq. (7.40) and Eq. (7.47) if some setting in $\{y_1, \dots, y_m, y_{m+1}\}$ is not used in the settings $\{x_1, \dots, x_r, \dots, x_n\}$ or $\{x_1, \dots, \tilde{x}_r, \dots, x_n\}$, we just sum over its outcomes in the joint measurement $G^{(y_1, \dots, y_{m+1})}$. By the same procedure we see that the no-signaling constraints hold also in this case.

Thus, the conditional probability distributions $(p_{a_1, \dots, a_n | x_1, \dots, x_n}^{N^{(n)}(\varrho)})$ satisfy the no-signaling conditions for all $x_1, \dots, x_n \in [g]$ and $\varrho \in K$. Then according to Proposition 7.6 we have that the tensors $N^{(n)}(\varrho) \in (CS_{k,g}^+)^{\hat{\otimes} n}$ defined by the no-signaling probability distributions $p_{a_1, \dots, a_n | x_1, \dots, x_n}^{N^{(n)}(\varrho)}$ are actually normalized, i.e., $N^{(n)}(\varrho) \in (CS_{k,g}^1)^{\hat{\otimes} n}$ for all $\varrho \in K$. Thus, if we define a map $N^{(n)} : V(K)^+ \rightarrow (CS_{k,g}^+)^{\hat{\otimes} n}$ by $\varrho \mapsto N^{(n)}(\varrho)$ for all $\varrho \in K$, we see that it is a channel, i.e., $N^{(n)} : K \rightarrow (CS_{k,g}^1)^{\hat{\otimes} n}$. As was observed before, if $N^{(n)} : K \rightarrow (CS_{k,g}^1)^{\hat{\otimes} n}$ is a channel, so is $\tilde{N}^{(n)} := P_{\text{Sym}_n(CS_{k,g})} \circ N^{(n)}$. In fact, we have that

$$p_{a_1, \dots, a_n | x_1, \dots, x_n}^{\tilde{N}^{(n)}(\varrho)} = \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} p_{a_{\sigma(1)}, \dots, a_{\sigma(n)} | x_{\sigma(1)}, \dots, x_{\sigma(n)}}^{N^{(n)}(\varrho)} \tag{7.53}$$

for all $a_1, \dots, a_n \in [k]$ and $x_1, \dots, x_n \in [g]$. Using Eq. (7.43) we see that

$$\sum_{\substack{t=1 \\ t \neq i}}^n \sum_{a_t=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_n}^{\tilde{N}^{(n)}(\varrho)} = M_{a_i | x_i}(\varrho) \tag{7.54}$$

for all $a_i \in [k]$, $x_i \in [g]$, $i \in [n]$ and $\varrho \in K$.

Finally we have that

$$\langle \xi_M, \varrho \otimes m_{a_i}^{(x_i)} \rangle = \langle M(\varrho), m_{a_i}^{(x_i)} \rangle = M_{a_i | x_i}(\varrho) = \sum_{\substack{t=1 \\ t \neq i}}^n \sum_{a_t=1}^k p_{a_1, \dots, a_n | x_1, \dots, x_n}^{\tilde{N}^{(n)}(\varrho)} \tag{7.55}$$

$$= \langle \tilde{N}^{(n)}(\varrho), \mathbf{1}_{CS_{k,g}^1}^{\otimes(i-1)} \otimes m_{a_i}^{(x_i)} \otimes \mathbf{1}_{CS_{k,g}^1}^{\otimes(n-i)} \rangle \tag{7.56}$$

$$= \langle P_{\text{Sym}_n(CS_{k,g})}(N^{(n)}(\varrho)), \mathbb{1}_{CS_{k,g}^1}^{\otimes(i-1)} \otimes m_{a_i}^{(x_i)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes(n-i)} \rangle \quad (7.57)$$

$$= \langle N^{(n)}(\varrho), P_{\text{Sym}_n(CS_{k,g}^*)}(\mathbb{1}_{CS_{k,g}^1}^{\otimes(i-1)} \otimes m_{a_i}^{(x_i)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes(n-i)}) \rangle \quad (7.58)$$

$$= \langle N^{(n)}(\varrho), P_{\text{Sym}_n(CS_{k,g}^*)}(m_{a_i}^{(x_i)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes(n-1)}) \rangle \quad (7.59)$$

$$= \langle N^{(n)}(\varrho), (\gamma_n^\Phi)^*(m_{a_i}^{(x_i)}) \rangle \quad (7.60)$$

$$= \langle \xi_{N^{(n)}}, \varrho \otimes (\gamma_n^\Phi)^*(m_{a_i}^{(x_i)}) \rangle \quad (7.61)$$

$$= \langle (\text{id} \otimes \gamma_n^\Phi)(\xi_{N^{(n)}}), \varrho \otimes m_{a_i}^{(x_i)} \rangle \quad (7.62)$$

for all $a_i \in [k]$, $x_i \in [g]$, $i \in [n]$ and $\varrho \in K$. Since the states in K and the vectors $m_a^{(x)}$ span the corresponding vector spaces, we have that $\xi_M = (\text{id} \otimes \gamma_n^\Phi)(\xi_{N^{(n)}})$ which shows that ξ_M is n -extendable. $\square$

7.2. Extendability of no-signaling behaviors. Going back to the setting of Section 6, we can also consider the concept of extendability of the general no-signaling statistics $P \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$, where separability corresponds to the corresponding no-signaling distribution having an LHV-model. As n -wise compatibility can be seen as a restricted version of compatibility, we can also consider a no-signaling behavior to have a restricted LHV-model.

Definition 7.9. Let $g, r, k, l \in \mathbb{N}$ and $m \in [g]$ and $n \in [r]$. A no-signaling probability distribution $p = (p_{\cdot, \cdot | x, y})_{x \in [g], y \in [r]}$ on $[k] \times [l]$ has an (m, n) -LHV model if for all $(x_1, \dots, x_m) \subseteq [g]$ and $(y_1, \dots, y_n) \subseteq [r]$ the restricted no-signaling probability distribution $p^{(x_1, \dots, x_m, y_1, \dots, y_n)} = (p_{\cdot, \cdot | x, y})_{x \in \{x_1, \dots, x_m\}, y \in \{y_1, \dots, y_n\}}$ on $[k] \times [l]$ has an LHV model.

We note that $p = (p_{\cdot, \cdot | x, y})_{x \in [g], y \in [r]}$ has an LHV model if and only if it has an (g, r) -LHV model. Before connecting n -extendability to these restricted LHV models, let us first take a closer look on the set of LHV probability distributions. Let us denote by $\mathcal{P}_{(k,g),(l,r)}^{NS}$ the set of no-signaling probability distributions $p = (p_{\cdot, \cdot | x, y})_{x \in [g], y \in [r]}$ on $[k] \times [l]$; by Proposition 6.14 they are exactly the conditional probability distributions related to tensors $P \in CS_{k,g}^1 \hat{\otimes} CS_{l,r}^1$. We denote by $\mathcal{P}_{(k,g),(l,r)}^{LHV}$ the subset of $\mathcal{P}_{(k,g),(l,r)}^{NS}$ that have an LHV-model. Thus, each $p \in \mathcal{P}_{(k,g),(l,r)}^{LHV}$ can be written as

$$p_{a,b|x,y} = \sum_{\lambda \in [\Lambda]} q_\lambda p_{a|\lambda,x}^A p_{b|\lambda,y}^B \quad (7.63)$$

for all $a \in [k]$, $b \in [l]$, $x \in [g]$ and $y \in [r]$ for some probability distributions q on $[\Lambda]$, $p^A = (p_{\cdot|\lambda,x}^A)_{\lambda \in [\Lambda], x \in [g]}$ on $[k]$ and $p^B = (p_{\cdot|\lambda,y}^B)_{\lambda \in [\Lambda], y \in [r]}$ on $[l]$ for some $\Lambda \in \mathbb{N}$. It is clear that $\mathcal{P}_{(k,g),(l,r)}^{LHV}$ is a polytope so that it can be expressed as the convex hull of its finite number of extreme points. The extreme points are then probability distributions $(p_{\cdot|\lambda,x}^A p_{\cdot|\lambda,y}^B)_{x \in [g], y \in [r]}$ on $[k] \times [l]$ such that $p_{a|x}^A$ is non-zero only for one outcome $a_x \in [k]$ for all $x \in [g]$ and $p_{b|y}^B$ is non-zero only for one outcome $b_y \in [l]$ for all $y \in [r]$. We can thus label the extreme points $p^{\vec{a}, \vec{b}}$ by the outcome vectors $\vec{a} = (a_1, \dots, a_g) \subset [k]^g$ and $\vec{b} = (b_1, \dots, b_r) \subset [l]^r$ such that

$$p_{a,b|x,y}^{\vec{a}, \vec{b}} = \delta_{a, a_x} \delta_{b, b_y} \quad (7.64)$$

for all $a \in [k]$, $b \in [l]$, $x \in [g]$ and $y \in [r]$. Now if we can express a probability distribution $p \in \mathcal{P}_{(k,g),(l,r)}^{NS}$ as a mixture of these extreme points, i.e.,

$$p_{a,b|x,y} = \sum_{\vec{a} \in [k]^g} \sum_{\vec{b} \in [l]^r} q_{\vec{a}, \vec{b}} p_{a,b|x,y}^{\vec{a}, \vec{b}} \quad (7.65)$$

for some probability distribution q on $[k]^g \times [l]^r$, then we know that $p \in \mathcal{P}_{(k,g),(l,r)}^{LHV}$. Also, vice versa, if $p \in \mathcal{P}_{(k,g),(l,r)}^{LHV}$ then it has a convex decomposition as in Eq. (7.65). We call a decomposition, or simply the probability distribution q in Eq. (7.65) an *LHV-decomposition* of p . We note that such a decomposition is not always unique.

Similarly for the restricted LHV models, we see that $p = (p_{\cdot, \cdot | x, y})_{x \in [g], y \in [r]}$ has an (n, m) -LHV model if and only if an LHV-decomposition exists for all restricted distributions $p^{(x_1, \dots, x_m, y_1, \dots, y_n)} = (p_{\cdot, \cdot | x, y})_{x \in \{x_1, \dots, x_m\}, y \in \{y_1, \dots, y_n\}}$ on $[k] \times [l]$ for all $(x_1, \dots, x_m) \subseteq [g]$ and $(y_1, \dots, y_n) \subseteq [r]$.

In the next result we will take $l = k$ and $r = g$ since we will apply Proposition 7.6 where the cones match. As was noted in Remark 7.7, the same result could be obtained also if the cones are different, but we will not show it explicitly in this case. This restriction can also be avoided by simply embedding the smaller cone inside the larger one. In the proof we draw inspiration from [TDS03], where the authors showed that in a bipartite setting where the parties are both measuring g measurements on a shared quantum state, then if the quantum state has a symmetric g -quasiextension (the extension need not be a quantum state but only in the maximum tensor product of PSD _{d} cones), then it can be only used to generate behaviors which have an LHV model.

Theorem 7.10. *Let $P \in CS_{k,g}^1 \hat{\otimes} CS_{k,g}^1$ with the corresponding no-signaling probability distribution $p = (p_{\cdot, \cdot | x, y})_{x, y \in [g]}$ on $[k] \times [k]$ and let $n \in [g]$. Then P is n -extendable if and only if p has an (g, n) -LHV model such that for all settings $(y_1, \dots, y_n) \subseteq [g]$ there exists some (g, n) -LHV decompositions $q^{(y_1, \dots, y_n)}$ satisfying the following no-signaling constraints*

$$\sum_{\substack{z=1 \\ z \neq i}}^g \sum_{a_z=1}^k \sum_{b_j=1}^k q_{a_1, \dots, a_i, \dots, a_g, b_1, \dots, b_j, \dots, b_n}^{(y_1, \dots, y_j, \dots, y_n)} = \sum_{\substack{z=1 \\ z \neq i}}^g \sum_{a_z=1}^k \sum_{b_j=1}^k q_{a_1, \dots, a_i, \dots, a_g, b_1, \dots, b_j, \dots, b_n}^{(y_1, \dots, \tilde{y}_j, \dots, y_n)} \quad (7.66)$$

for all $b_1, \dots, b_n \in [k]$, $y_j, \tilde{y}_j \in [g]$ and $i \in [g]$, $j \in [n]$. If $n = g$, then $P \in CS_{k,g}^1 \dot{\otimes} CS_{k,g}^1$.

Proof. Let first P be n -extendable. Thus, there exists $P^{(n)} \in CS_{k,g}^1 \hat{\otimes} (CS_{k,g}^1)^{\hat{\otimes} n}$ s.t. $(\text{id} \otimes \gamma_n^\Phi)(P^{(n)}) = P$ for $\Phi = \mathbb{1}_{CS_{k,g}^1}$. By Proposition 7.6 we have that for all $a, b_i \in [k]$ and $x, y_i \in [g]$

$$p_{a, b_i | x, y_i} = \langle P, m_a^{(x)} \otimes m_{b_i}^{(y_i)} \rangle \quad (7.67)$$

$$= \langle (\text{id} \otimes \gamma_n^\Phi)(P^{(n)}), m_a^{(x)} \otimes m_{b_i}^{(y_i)} \rangle \quad (7.68)$$

$$= \langle P^{(n)}, m_a^{(x)} \otimes P_{\text{Sym}_n}(CS_{k,g}^*) (m_{b_i}^{(y_i)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes n}) \rangle \quad (7.69)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} p_{a, \alpha_\sigma(b_1, \dots, b_n) | x, \alpha_\sigma(1, \dots, 1, y_i, 1, \dots, 1)}^{(n)} \quad (7.70)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k \tilde{p}_{a, b_1, \dots, b_n | x, 1, \dots, 1, y_i, 1, \dots, 1}^{(n)}, \quad (7.71)$$

where $\tilde{p}^{(n)} = (\tilde{p}_{\cdot, \dots, \cdot | z_1, \dots, z_{n+1}}^{(n)})_{z_1, \dots, z_{n+1} \in [g]}$ is the no-signaling probability distribution on $[k]^{n+1}$ related to the tensor $(\text{id} \otimes \gamma_n^\Phi)(P^{(n)}) \in (CS_{k,g}^1)^{\hat{\otimes} (n+1)}$.

Let us now fix $\{y_1, \dots, y_n\} \subset [g]$. We set $r_{b_1, \dots, b_n}^{(y_1, \dots, y_n)} = \sum_{a=1}^k \tilde{p}_{a, b_1, \dots, b_n | 1, y_1, \dots, y_n}^{(n)}$ for all $b_i \in [k]$ for all $i \in [n]$ and define a probability distribution $q^{(y_1, \dots, y_n)}$ on $[k]^{gn}$ by setting

$$q_{a_1, \dots, a_g, b_1, \dots, b_n}^{(y_1, \dots, y_n)} = \frac{1}{(r_{b_1, \dots, b_n}^{(y_1, \dots, y_n)})^{g-1}} \prod_{x=1}^g \tilde{p}_{a_x, b_1, \dots, b_n | x, y_1, \dots, y_n}^{(n)} \quad (7.72)$$

for all $a_1, \dots, a_g, b_1, \dots, b_n \in [k]$. Clearly, since $\tilde{p}^{(n)}$ is also a no-signaling distribution, and in Eq. (7.66) we are summing over all a_i 's except for one, we see that the no-signaling constraints of Eq. (7.66) are satisfied for $q^{(y_1, \dots, y_n)}$.

Now for all $a, b \in [k]$, $x \in [g]$ and $i \in [n]$ we have that

$$\sum_{z=1}^g \sum_{a_z=1}^k \sum_{j=1}^n \sum_{b_j=1}^k q_{a_1, \dots, a_g, b_1, \dots, b_n}^{(y_1, \dots, y_n)} p_{a, b | x, y_i}^{(a_1, \dots, a_g), (b_1, \dots, b_n)} \quad (7.73)$$

$$= \sum_{z=1}^g \sum_{a_z=1}^k \sum_{j=1}^n \sum_{b_j=1}^k q_{a_1, \dots, a_g, b_1, \dots, b_n}^{(y_1, \dots, y_n)} \delta_{a, a_x} \delta_{b, b_i} \quad (7.74)$$

$$= \sum_{\substack{z=1 \\ z \neq x}}^g \sum_{a_z=1}^k \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k q_{a_1, \dots, a_{x-1}, a, a_{x+1}, \dots, a_g, b_1, \dots, b_{i-1}, b, b_{i+1}, \dots, b_n}^{(y_1, \dots, y_n)} \quad (7.75)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k \tilde{p}_{a, b_1, \dots, b_{i-1}, b, b_{i+1}, \dots, b_n | x, y_1, \dots, y_n}^{(n)} \quad (7.76)$$

$$= p_{a, b | x, y_i}. \quad (7.77)$$

Hence, the restricted no-signaling distribution $p^{(y_1, \dots, y_n)} = (p_{\cdot, \cdot | x, y})_{x \in [g], y \in \{y_1, \dots, y_n\}}$ on $[k] \times [k]$ for all $\{y_1, \dots, y_n\} \subset [g]$ has a LHV-model which means that p has a (g, n) -LHV model.

On the other hand, let now p have a (g, n) -LHV model s.t. for all settings $(y_1, \dots, y_n) \subseteq [g]$ there exists some (g, n) -LHV decompositions $q^{(y_1, \dots, y_n)}$ satisfying the no-signaling constraints in Eq. (7.66). We can define a conditional probability distribution $p'^{(n)} = (p'_{\cdot, \dots, \cdot | x, y_1, \dots, y_n})_{x, y_1, \dots, y_n \in [g]}$ on $[k]^{n+1}$ by setting

$$p'_{a, b_1, \dots, b_n | x, y_1, \dots, y_n}^{(n)} = \sum_{\substack{i=1 \\ i \neq x}}^g \sum_{a_i=1}^k q_{a_1, \dots, a_{i-1}, a, a_{i+1}, \dots, a_g, b_1, \dots, b_n}^{(y_1, \dots, y_n)}. \quad (7.78)$$

Clearly

$$\sum_{a=1}^k p'_{a, b_1, \dots, b_n | x, y_1, \dots, y_n}^{(n)} = \sum_{a=1}^k p'_{a, b_1, \dots, b_n | \tilde{x}, y_1, \dots, y_n}^{(n)} \quad (7.79)$$

for all $b_1, \dots, b_n \in [k]$, $x, \tilde{x}, y_1, \dots, y_n \in [g]$. Also, since the no-signaling constraints in Eq. (7.66) are satisfied, we also have that

$$\sum_{b_j=1}^k p'_{a, b_1, \dots, b_n | x, y_1, \dots, y_j, \dots, y_n}^{(n)} = \sum_{b_j=1}^k p'_{a, b_1, \dots, b_n | x, y_1, \dots, \tilde{y}_j, \dots, y_n}^{(n)} \quad (7.80)$$

for all $a, b_1, \dots, b_n \in [k]$, $x, y_1, \dots, y_n, \tilde{y}_j \in [g]$ for all $j \in [n]$. Hence, $p'^{(n)}$ is a no-signaling probability distribution and thus by Proposition 7.6 it corresponds to a tensor $P'^{(n)} \in (CS_{k,g}^1)^{\hat{\otimes}(n+1)}$. We see that now for all $a, b \in [k]$ and $x, y_i \in [g]$:

$$\langle P, m_a^{(x)} \otimes m_b^{(y_i)} \rangle = p_{a, b | x, y_i} = \sum_{z=1}^g \sum_{a_z=1}^k \sum_{j=1}^n \sum_{b_j=1}^k q_{a_1, \dots, a_g, b_1, \dots, b_n}^{(y_1, \dots, y_n)} p_{a, b | x, y_i}^{(a_1, \dots, a_g), (b_1, \dots, b_n)} \quad (7.81)$$

$$= \sum_{\substack{z=1 \\ z \neq x}}^g \sum_{a_z=1}^k \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k q_{a_1, \dots, a_{x-1}, a, a_{x+1}, \dots, a_g, b_1, \dots, b_{i-1}, b, b_{i+1}, \dots, b_n}^{(y_1, \dots, y_n)} \quad (7.82)$$

$$= \sum_{\substack{z=1 \\ z \neq x}}^g \sum_{a_z=1}^k \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k q_{a_1, \dots, a_{x-1}, a, a_{x+1}, \dots, a_g, b_1, \dots, b_{i-1}, b, b_{i+1}, \dots, b_n}^{(1, \dots, 1, y_i, 1, \dots, 1)} \quad (7.83)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k p_{a, b_1, \dots, b_n | x, 1, \dots, 1, y_i, 1, \dots, 1}^{(n)} \quad (7.84)$$

$$= \sum_{\substack{j=1 \\ j \neq i}}^n \sum_{b_j=1}^k \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} p_{a, \alpha_\sigma(b_1, \dots, b_n) | x, \alpha_\sigma(1, \dots, 1, y_i, 1, \dots, 1)}^{(n)} \quad (7.85)$$

$$= \langle P^{(n)}, m_a^{(x)} \otimes P_{\text{Sym}_n}(CS_{k,g}^*)(m_b^{(y_i)} \otimes \mathbb{1}_{CS_{k,g}^1}^{\otimes(n-1)}) \rangle \quad (7.86)$$

$$= \langle (\text{id} \otimes \gamma_n^\Phi)(P^{(n)}), m_a^{(x)} \otimes m_b^{(y_i)} \rangle, \quad (7.87)$$

where the equality between Eq. (7.84) and Eq. (7.85) follows from the no-signaling constraints in Eq. (7.80). Since the vectors $m_a^{(x)}$ span $CS_{k,g}^*$ we have that $P = (\text{id} \otimes \gamma_n^\Phi)(P^{(n)})$ so that P is n -extendable. $\square$

8. ROBUSTNESS RESULTS

In the previous sections, we have seen that whether a positive linear map between two cones factorizes through a third cone as a composition of positive maps depends heavily on the structure of the three cones and on the types of tensor product one uses in constructing the cones. In this section we would like to understand what type of limitations on factorization one can derive directly from the structure of the state space of column stochastic matrices.

As a first case study, let us suppose that the identity channel $\text{id} : CS_{l,r}^1 \rightarrow CS_{l,r}^1$ can be factorized through $CS_{k,g}^1$. That means that there are channels $\Psi : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ and $\Phi : CS_{k,g}^1 \rightarrow CS_{l,r}^1$ such that $\text{id} = \Phi \circ \Psi$. Operationally this means that the action of a particular simulation could always be reversed. Namely, if $M : K \rightarrow CS_{l,r}^1$ is any multimeter on a state space K , then from the factorization of id we have that $M = \Phi \circ \Psi \circ M$ so that the simulation of M by the channel Ψ can be reversed by applying the channel Φ . Naturally, there are cases when such factorization cannot be feasible: if $g = 1$ so that $\Psi \circ M$ is a single measurement meaning that $M = \Phi \circ \Psi \circ M$ implies the compatibility of all of the measurements in M even when M would contain incompatible measurements. On the other hand, in some cases this kind of reversibility is natural, for example when $k \geq l$ and $g \geq r$ so that $CS_{l,r}^1$ can be trivially embedded in $CS_{k,g}^1$. We can use Proposition 4.13 to show that actually this is the only case when the identity channel can be factorized like that.

Proposition 8.1. *Let $l, r \in \mathbb{N}$. The identity channel $\text{id} : CS_{l,r}^1 \rightarrow CS_{l,r}^1$ can be factorized through $CS_{k,g}^1$ for some $k, g \in \mathbb{N}$ if and only if $k \geq l$ and $g \geq r$.*

Proof. If $k \geq l$ and $g \geq r$, we can trivially consider a column stochastic matrix of size $l \times r$ as a column stochastic matrix of size $k \times g$ by adding zero rows and adding extra columns with element 1 on some row (this would correspond to adding trivial measurement $\mathbb{1}_K$ to your multimeter).

We first note that the identity channel $\text{id} : CS_{l,r}^1 \rightarrow CS_{l,r}^1$ can be seen as a multimeter on $CS_{l,r}^1$. Let us write this multimeter explicitly in the basis of $CS_{l,r}^1$. In fact, one can confirm that actually

$$\text{id}(X) = \mathbb{1}_{CS_{l,r}^1}(X) s_{l,\dots,l} + \sum_{y=1}^r \sum_{b=1}^{l-1} m_b^{(y)}(X) e_b^{(y)} \quad (8.1)$$

for all $X \in CS_{l,r}^1$. Thus, id corresponds to a multimeter consisting of the measurements $\{m^{(y)}\}_{y \in [r]}$ on $CS_{l,r}^1$.

If now the identity channel $\text{id} : CS_{l,r}^1 \rightarrow CS_{l,r}^1$ is factorizable through $CS_{k,g}^1$ then there exists channels $\Psi : CS_{l,r}^1 \rightarrow CS_{k,g}^1$ and $\Phi : CS_{k,g}^1 \rightarrow CS_{l,r}^1$ such that $\text{id} = \Phi \circ \Psi$. We can now interpret $\Psi = \{\Psi_{\cdot|x}\}_{x \in [g]}$ as a multimeter on $CS_{l,r}^1$ which thus consists of g measurements with k outcomes. Thus, by Corollary 4.6 the factorization $\text{id} = \Phi \circ \Psi$ can be interpreted as classically simulating the multimeter id by the multimeter Ψ . However, since by Proposition 4.13 the multimeter id consists only of simulation irreducible measurements $\{m^{(y)}\}_{y \in [r]}$, it means that for all $y \in [r]$, there exists $x_y \in [g]$ such that $\Psi_{\cdot|x_y}$ is post-processing equivalent with $m^{(y)}$. Also, because none of the measurements $m^{(1)}, \dots, m^{(r)}$ are post-processing equivalent, we must have that $x_y \neq x_{y'}$ for all $y, y' \in [r]$ such that $y \neq y'$. This can only hold if $g \geq r$. Furthermore, since $m^{(y)}$ is the unique minimally sufficient representative of its equivalence class $[[m^{(y)}]]$, to which $\Psi_{\cdot|x_y}$ also belongs, we must have that $m^{(y)}$ can be obtained from $\Psi_{\cdot|x_y}$ by joining all of its pairwise linearly dependent effects as was explained in Proposition 4.12. This means that for the outcomes we must have that $k \geq l$. $\square$

Thus, the only time an identity channel on a state space of column stochastic matrices can be factorized through a state space of column stochastic matrices of different size is exactly when column stochastic matrix is a submatrix of a larger column stochastic matrix. This means that no pathologies appear in the classical simulability of a multimeter M , since otherwise we could decrease the size of the column stochastic matrices via factorization of the identity channel.

8.1. Compatibility. Recall that a multimeter $M_{a|x}$ acting on a GPT state space K can be viewed as a channel $M : K \rightarrow CS_{k,g}^1$, where k denotes the number of outcomes for each of the g measurement settings. The state space $CS_{k,g}^1$ is a polysimplex.

Let us prove the following well-known result using the factorization characterization of compatibility.

Proposition 8.2. *Let $M_{a|x}$ be an arbitrary multimeter, $P_{a|x} = p_{a|x} \mathbb{1}_K$ a trivial multimeter, and q_x a probability measure on g . Then, the noisy multimeter*

$$N_{a|x} = q_x M_{a|x} + (1 - q_x) p_{a|x} \mathbb{1}_K \quad (8.2)$$

is compatible. In particular,

$$\tilde{M}_{a|x} = \frac{1}{g} M_{a|x} + \left(1 - \frac{1}{g}\right) \frac{\mathbb{1}_K}{k} \quad (8.3)$$

is compatible.

Proof. For any $z \in [g]$, consider the following multimeter:

$$M_{a|x}^{(z)} = \begin{cases} M_{a|x} & \text{if } x = z \\ p_{a|x} \mathbb{1}_K & \text{if } x \neq z. \end{cases} \quad (8.4)$$

The intuition from quantum mechanics is that this multimeter should be compatible, since the effects of different measurements commute. Indeed, we have

$$\left[K \xrightarrow{M^{(z)}} CS_{k,g}^1 \right] = \left[K \xrightarrow{M} CS_{k,g}^1 \xrightarrow{\text{marg}_z} S_k \xrightarrow{\text{triv}_z} CS_{k,g}^1 \right]. \quad (8.5)$$

where marg_z is the channel that discards all the indices but z :

$$\text{marg}_z[(q_{a|x})_{a \in [k], x \in [g]}] = q_{\cdot|z} \quad (8.6)$$

and triv_z is the channel that embeds the z -th copy of the simplex S_k into the polysimplex $CS_{k,g}^1$ in the following way:

$$\text{triv}_z[(q_a)] = (r_{a|x}) \quad \text{with } r_{a|x} = \begin{cases} q_a & \text{if } x = z \\ p_{a|x} \mathbb{1}_K & \text{if } x \neq z. \end{cases} \quad (8.7)$$

Note that the multimeter $M^{(z)}$ is compatible, since it factors through the z -th copy of S_k . Finally, we have

$$N = \sum_{z \in [g]} q_z M^{(z)} \quad (8.8)$$

proving the claim via convexity. $\square$

Let us consider now noisy multimeters as GPT channels between some GPT with state space K and some noisy version of the polysimplex, seen as a GPT, that depends on the amount and the type of noise considered.

We have

$$K \xrightarrow{M} CS_{k,g;\text{noisy}}^1. \quad (8.9)$$

Let us consider two examples. Firstly, let us consider adding white noise to k -outcome measurements: The noisy effects are

$$\tilde{A}_i := tA_i + (1-t)\frac{\mathbb{1}_K}{k}, \quad (8.10)$$

where $t \in [0, 1]$ is the *noise parameter*. The value $t = 1$ corresponds to having no noise (the original POVM), while the value $t = 0$ corresponds to the uniform trivial measurement having effects $\mathbb{1}_K/k$. Note that the range of the map

$$K \xrightarrow{A} S_k, \quad (8.11)$$

when restricted to states in K , is necessarily *smaller* than the full probability simplex on k vertices. Indeed, if the outcome probabilities for the original measurement are p , then the probabilities corresponding to the noisy measurement are

$$\tilde{p}_i = tp_i + \frac{1-t}{k}. \quad (8.12)$$

Hence, the noisy measurement has range

$$K \xrightarrow{\tilde{A}} S_{k;t}, \quad (8.13)$$

where $S_{k;t}$ is the state space of the GPT having $\mathbb{R}^k$ as a vector space, an order unit identical to the one of S_k , i.e., for $v \in \mathbb{R}^k$

$$\mathbb{1}_{S_{k;t}}(v) = \sum_{i=1}^k v_i, \quad (8.14)$$

but a *thinner* cone given by

$$(\mathbb{R}_+^k)_t := \{v \in \mathbb{R}^k : \forall i \in [k], v_i \geq (1-t)\bar{v}, \text{ where } \bar{v} := k^{-1} \sum_{j=1}^k v_j\}. \quad (8.15)$$

Geometrically, the set $S_{k;t}$ is obtained by scaling the usual simplex S_k by a factor of t , around its “central” point $(1/k, \dots, 1/k)$.

As a second example, let us consider a multimeter M with two 2-outcome measurements, to which we apply uniform noise, with parameters t_1 and t_2 respectively:

$$\tilde{M}_{a|x} = t_x M_{a|x} + (1 - t_x) \frac{\mathbb{1}_K}{2}. \quad (8.16)$$

The situation is similar to the one before: the noisy multimeter has range

$$K \xrightarrow{\tilde{M}} CS_{2,2;(t_1,t_2)}^1, \quad (8.17)$$

where the noisy polysimplex $CS_{2,2;(t_1,t_2)}^1$ has vector space

$$\mathbb{R}^{2+2} \cap E, \quad (8.18)$$

where

$$E = \{(v_{a|x}) \in \mathbb{R}^4 : \sum_a v_{a|1} = \sum_{a'} v_{a'|2}\}, \quad (8.19)$$

order unit

$$\mathbb{1}_{CS_{2,2;(t_1,t_2)}^1}(v) = \sum_a v_{a|1} = \sum_{a'} v_{a'|2}, \quad (8.20)$$

and cone

$$[(\mathbb{R}_+^2)_{t_1} \times \mathbb{R}_+^2] \cap E, \quad (8.21)$$

where the thin cones $(\mathbb{R}_+^2)_{t_1}$ and $(\mathbb{R}_+^2)_{t_2}$ have been defined above. We display the noisy (poly)simplices in Figure 3.

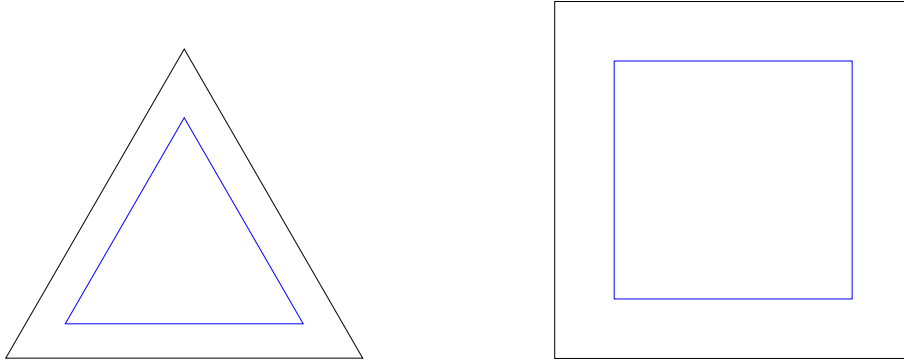


FIGURE 3. State spaces of the probability simplex GPT S_3 (left) and the polysimplex $CS_{2,2}^1$, along with their noisy versions in blue, for noise parameters $t = 2/3$, respectively $(t_1, t_2) = (2/3, 2/3)$.

We shall be interested in what follows in the following question: is it possible to factorize the identity map between a noisy version of some (poly)simplex and its noiseless version through a different (poly)simplex:

$$CS_{k,g;\text{noisy}}^1 \xrightarrow{\text{id}} CS_{k,g}^1 = CS_{k,g;\text{noisy}}^1 \xrightarrow{\varphi} CS_{l,g}^1 \xrightarrow{\psi} CS_{k,g}^1? \quad (8.22)$$

Above, the maps φ, ψ are GPT channels, i.e. positive and unit-preserving linear maps.

This type of factorization implies that a noisy multimeter with outcome counts $\underline{k}$ can be simulated by a multimeter with outcome counts $\underline{l}$, and that *independently of the GPT G* that the multimeter acts on. Importantly, the factorization above depends only on the respective polysimplices, no other physical GPTs being involved.

In practice, we shall tackle the existence of such factorizations from a geometric perspective, as follows: First, choose coordinates for the vector space space of the GPT $CS_{k,g}^1$ such that

$$V(CS_{k,g}^1) \ni v = (s, v_1, \dots, v_n) \quad \text{and} \quad \mathbb{1}_{CS_{k,g}^1}(v) = s, \quad (8.23)$$

where $n := g(k-1)$. Proceed similarly for the GPT $CS_{l,g}^1$ and its associated vector space $\mathbb{R}^m$, where $m = g(l-1)$. Note that $m \geq n$ is a necessary condition for the existence of such a factorization, for obvious rank reasons. We have the following result.

Proposition 8.3. *A factorization as in Eq. (8.22) exists if and only if there exists T , an affine image of $CS_{l,g}^1$ in $\mathbb{R}^m$, such that:*

- $CS_{k,g}^1 \oplus 0_{m-n} \subseteq T$;
- *there exists a projection $\Pi : \mathbb{R}^m \rightarrow \mathbb{R}^n$, not necessarily orthogonal, such that $\Pi(T) \subseteq CS_{k,g}^1$.*

Proof. Consider the linear extensions of the affine maps φ and ψ from Eq. (8.22). Let $\tilde{\varphi} : \mathbb{R}^n \rightarrow \mathbb{R}^m$ and $\tilde{\psi} : \mathbb{R}^m \rightarrow \mathbb{R}^n$ denote their unique linear extensions to the corresponding vector spaces. Let $\iota : \mathbb{R}^n \rightarrow \mathbb{R}^m$ be the standard inclusion $x \mapsto (x, 0)$ and $\pi : \mathbb{R}^m \rightarrow \mathbb{R}^n$ be the standard projection $(x, y) \mapsto x$. Note $\pi \circ \iota = \text{id}_{\mathbb{R}^n}$.

Define the affine map $A : \mathbb{R}^m \rightarrow \mathbb{R}^m$ by $A = \iota \circ \tilde{\psi}$. Let $T := A(CS_{l,g}^1) = \iota(\tilde{\psi}(CS_{l,g}^1))$. Since ι is injective and $\tilde{\psi}$ maps the m -dimensional affine space $\text{aff}(CS_{l,g}^1)$ to the n -dimensional affine space $\text{aff}(CS_{k,g}^1)$, T is an affine image of $CS_{l,g}^1$. It is obvious that $\Pi := \pi$ verifies the second point in the statement. $\square$

8.2. Two noisy binary measurements simulated by a three-outcome measurement. It is a standard result that any pair of dichotomic measurements are compatible, provided that they have noise parameters $t_1 = t_2 = t \leq 1/2$. Moreover, the two measurements can be post-processed from a joint 4-outcome measurement. We shall now investigate the same question, with the difference that we ask about pairs of dichotomic measurements that can be post-processed from a 3-outcome measurement.

Proposition 8.4. *For all $t \leq 1/3$, the identity map*

$$\text{id} : CS_{2,2;(t,t)}^1 \rightarrow CS_{2,2}^1 \quad (8.24)$$

factorizes through S_3 . In particular, any pair of noisy binary measurements with noise parameter $t \leq 1/3$ admit a joint measurement with 3 outcomes.

Proof. The construction is depicted in Figure 4, left panel. More precisely, noisy measurements

$$\tilde{X}_i = \frac{1}{3}X_i + \frac{1}{3}\mathbb{1}_K \quad i = 1, 2 \quad (8.25)$$

$$\tilde{Y}_j = \frac{1}{3}Y_j + \frac{1}{3}\mathbb{1}_K \quad j = 1, 2 \quad (8.26)$$

can be post-processed from the 3-outcome measurement (A, B, C) with

$$A = \tilde{Y}_1 = \frac{1}{3}Y_1 + \frac{1}{3}\mathbb{1}_K \quad (8.27)$$

$$B = \tilde{X}_1 - \frac{1}{2}\tilde{Y}_1 = \frac{1}{3}X_1 + \frac{1}{6}Y_2 \quad (8.28)$$

$$C = \mathbb{1}_K - \tilde{X}_1 - \frac{1}{2}\tilde{Y}_1 = \frac{1}{3}X_2 + \frac{1}{6}Y_2. \quad (8.29)$$

$\square$

Conjecture 8.5. *The value of t in the result above is optimal.*

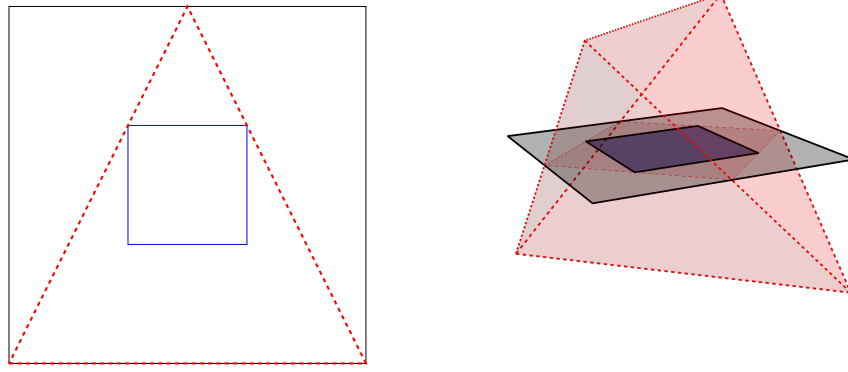


FIGURE 4. Any pair of two noisy dichotomic measurements with noise parameter $t = 1/3$ can be post-processed from a 3-outcome measurement (left). Any pair of two noisy dichotomic measurements with noise parameter $t = 1/2$ can be post-processed from a 4-outcome measurement (right). The state space of $CS_{2,2}^1$ is the outer square (black), its noisy version is the inner square (blue), and the affine image of the probability simplex S_3 (resp. S_4) is the triangle (resp. tetrahedron), in red. The inclusions of the state spaces correspond to those from Proposition 8.3.

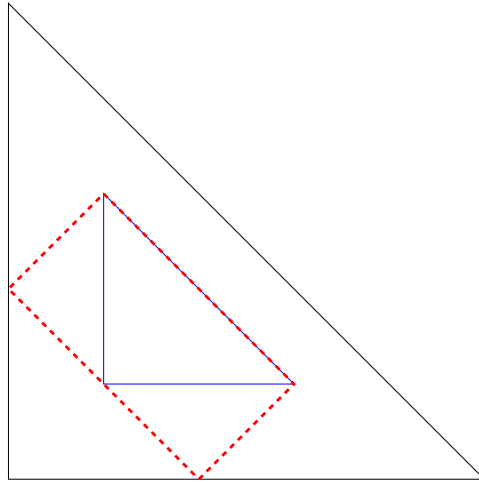


FIGURE 5. Any noisy measurement with noise parameter $t = 2/5$ can be post-processed from a pair of dichotomic measurements. The state space of S_3 is the outer triangle (black), its noisy version is the inner triangle (blue), and the affine image of the polysimplex $CS_{2,2}^1$ is the red rectangle. The inclusions of the state spaces correspond to those from Proposition 8.3; note that here the state space dimensions are equal.

Note that if we allow measurements with four outcomes as the middle polysimplex, then one can achieve $t = 1/2$; this is standard result about compatibility of two noisy dichotomic measurements, see Figure 4, right panel.

8.3. A noisy three-outcome measurement simulated by two dichotomic measurements.

We ask now the reverse question from the previous section: when can one noisy 3-outcome measurement can be simulated by two dichotomic measurements?

One can think that this is always possible: given a three outcome measurement (A, B, C) , simply define

$$X_1 = A \quad X_2 = B + C \quad (8.30)$$

$$Y_1 = B \quad Y_2 = A + C. \quad (8.31)$$

However, the original three outcome measurement cannot be post-processed from X, Y because of the ambiguity in the case where the two measurements yield the first result.

Proposition 8.6. *For all $t \leq 2/5$, the identity map*

$$\text{id} : S_{3,t} \rightarrow S_3 \quad (8.32)$$

factorizes through $CS_{2,2}^1$. In particular, any noisy ternary measurement with noise parameter $t \leq 2/5$ can be post-processed from a pair of dichotomic measurements.

Proof. The construction is depicted in Figure 5. More precisely, the noisy measurement

$$\tilde{A} = \frac{2}{5}A + \frac{1}{5}\mathbb{1}_K \quad (8.33)$$

$$\tilde{B} = \frac{2}{5}B + \frac{1}{5}\mathbb{1}_K \quad (8.34)$$

$$\tilde{C} = \frac{2}{5}C + \frac{1}{5}\mathbb{1}_K \quad (8.35)$$

can be post-processed from the two dichotomic measurements $(X_1, \mathbb{1}_K - X_1)$, $(Y_1, \mathbb{1}_K - Y_1)$ with

$$X_1 = \frac{5}{2}\tilde{A} + \frac{5}{2}\tilde{B} - \mathbb{1}_K \quad (8.36)$$

$$Y_1 = -\frac{5}{4}\tilde{A} + \frac{5}{4}\tilde{B} + \frac{1}{2}\mathbb{1}_K. \quad (8.37)$$

One can easily check that X_1 and Y_1 above define valid measurements. $\square$

Conjecture 8.7. *The value of t in the result above is optimal.*

9. CONCLUDING REMARKS

In this work we have considered several nonlocal and local non-classical phenomena arising in a probabilistic theory. Even though they are qualitatively different in their nature, we traced back those phenomena to one common mathematical origin: the factorization of a corresponding map, e.g. a multimeter. If the map could be factored through classical space, that is, a simplex, the non-classicality was absent, i.e. the phenomenon could be described by classical theory. The phenomena we considered were measurement incompatibility, steering and Bell-nonlocality. Even though not explicitly discussed, (Spekkens) contextuality is also naturally connected to our results as it can be expressed through simplex-embeddability.

Considering a generalized Choi-tensor and the state space of column stochastic matrices, which can be understood as a polysimplex, allowed us to find a unique separable tensor corresponding to the simplex-factoring map. This provides the recipe for the theme governing this work: Formulate a non-classical phenomenon in the GPT framework in terms of the appropriate objects (measurements, assemblages, statistics), find the corresponding map and tensor and check for factorizability and separability. In this way we also get conditions on non-classicality.

Towards the end of the paper, we focused on a more practical side of our framework. We provided a criterion for incompatibility and Bell-nonlocality by making use of generalized symmetric extensions. Here the advantage is, that any finite number of extension already allows for conclusions about the multimeter or Bell statistic in question. Lastly, we gave conditions on factorizability through the space of column stochastic matrices. In the trivial case of only one column, those

reduce to a probability simplex. We examined the noise robustness of the factorizing maps by pushing the noise into a noisy state space.

As for future work, our framework might be useful to give criteria on steerability. Due to the equivalence of incompatibility and steering problems in the bipartite case, it is clear that our results on extendability apply. We saw that realization of a steering is assemblage cannot be taken for granted in a GPT. It would be interesting to see the conditions on the GPT and the joint space in order have a realization of any assemblage from the theory.

ACKNOWLEDGMENTS

TA acknowledges support from the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation, project numbers 447948357 and 440958198), the Sino-German Center for Research Promotion (Project M-0294), the German Ministry of Education and Research (Project QuKuK, BMBF Grant No. 16KIS1618K), the DAAD, and the Alexander von Humboldt Foundation. TA and LL are supported by the Business Finland project BEQAH (Between Quantum Algorithms and Hardware). AB was supported by the ANR project PraQPv, grant number ANR-24-CE47-3023. IN was supported by the ANR project [ESQuisses](#), grant number ANR-20-CE47-0014-01. MP acknowledges support from the Niedersächsisches Ministerium für Wissenschaft und Kultur.

REFERENCES

- [ALPP21] Guillaume Aubrun, Ludovico Lami, Carlos Palazuelos, and Martin Plávala. Entangleability of cones. *Geometric and Functional Analysis*, 31(2):181–205, 2021. [67](#)
- [AMHP25] Guillaume Aubrun, Alexander Müller-Hermes, and Martin Plávala. Monogamy of entanglement between cones. *Mathematische Annalen*, 391(1):1591–1609, 2025. [44](#), [45](#), [49](#)
- [BBB⁺10] Howard Barnum, Salman Beigi, Sergio Boixo, Matthew B. Elliott, and Stephanie Wehner. Local quantum measurement and no-signaling imply quantum correlations. *Physical Review Letters*, 104(14), 2010. [5](#), [30](#), [32](#)
- [BBLW07] Howard Barnum, Jonathan Barrett, Matthew Leifer, and Alexander Wilce. Generalized no-broadcasting theorem. *Physical Review Letters*, 99(24):240501, 2007. [2](#)
- [BCF⁺96] Howard Barnum, Carlton M. Caves, Christopher A. Fuchs, Richard Jozsa, and Benjamin Schumacher. Noncommuting mixed states cannot be broadcast. *Physical Review Letters*, 76(15):2818, 1996. [2](#)
- [BCG⁺22] Costantino Budroni, Adán Cabello, Otfried Gühne, Matthias Kleinmann, and Jan-Åke Larsson. Kochen-Specker contextuality. *Reviews of Modern Physics*, 94(4):045007, 2022. [2](#)
- [BCP⁺14] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell non-locality. *Reviews of Modern Physics*, 86:419–478, 2014. [2](#)
- [BGW13] Howard Barnum, Carl Philipp Gaebler, and Alexander Wilce. Ensemble steering, weak self-duality, and the structure of probabilistic theories. *Foundations of Physics*, 43:1411–1427, 2013. [5](#), [22](#)
- [BJN22] Andreas Bluhm, Anna Jenčová, and Ion Nechita. Incompatibility in general probabilistic theories, generalized spectrahedra, and tensor norms. *Communications in Mathematical Physics*, 393(3):1125–1198, 2022. [8](#)
- [BLN25] Andreas Bluhm, Leevi Leppäjärvi, and Ion Nechita. On the simulation of quantum multimeters. *Quantum*, 9:1608, 2025. [2](#)
- [BV18] Erika Bene and Tamás Vértesi. Measurement incompatibility does not give rise to Bell violation in general. *New Journal of Physics*, 20(1):013021, 2018. [38](#)
- [CS14] Dariusz Chruściński and Gniewomir Sarbicki. Entanglement witnesses: construction, analysis and classification. *Journal of Physics A: Mathematical and Theoretical*, 47(48):483001, 2014. [7](#)
- [DPS04] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Physical Review A*, 69(2):022308, 2004. [43](#), [44](#)
- [FHL18] Sergey N. Filippov, Teiko Heinosaari, and Leevi Leppäjärvi. Simulability of observables in general probabilistic theories. *Physical Review A*, 97(6):062102, 2018. [2](#), [13](#), [17](#), [18](#), [23](#)
- [Fin82] Arthur Fine. Hidden variables, joint probability, and the Bell inequalities. *Physical Review Letters*, 48:291–295, 1982. [2](#), [32](#)
- [GBTCA17] Leonardo Guerini, Jessica Bavaresco, Marcelo Terra Cunha, and Antonio Acín. Operational framework for quantum measurement simulability. *Journal of Mathematical Physics*, 58:092102, 2017. [2](#), [13](#)

- [GHK⁺23] Otfried Gühne, Erkki Haapasalo, Tristan Kraft, Juha-Pekka Pellonpää, and Roope Uola. Colloquium: Incompatible measurements in quantum information science. *Reviews of Modern Physics*, 95(1):011003, 2023. [2](#), [11](#)
- [Gis89] Nicolas Gisin. Stochastic quantum dynamics and relativity. *Helvetica Physica Acta*, 62:363–371, 1989. [5](#), [20](#)
- [GT09] Otfried Gühne and Géza Tóth. Entanglement detection. *Physics Reports*, 474(1-6):1–75, 2009. [2](#)
- [Hei27] Werner Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43(3):172–198, 1927. [2](#)
- [HHH96] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Separability of mixed states: necessary and sufficient conditions. *Physics Letters A*, 223(1):1–8, 1996. [68](#)
- [HJW93] Lane P. Hughston, Richard Jozsa, and William K. Wootters. A complete classification of quantum ensembles having a given density matrix. *Physics Letters A*, 183(1):14–18, 1993. [5](#), [20](#)
- [HLP19] Teiko Heinosaari, Leevi Leppäjärvi, and Martin Plávala. No-free-information principle in general probabilistic theories. *Quantum*, 3:157, 2019. [25](#)
- [HMZ16] Teiko Heinosaari, Takayuki Miyadera, and Mário Ziman. An invitation to quantum incompatibility. *Journal of Physics A: Mathematical and Theoretical*, 49(12):123001, 2016. [2](#), [11](#)
- [HQB18] Flavien Hirsch, Marco Túlio Quintino, and Nicolas Brunner. Quantum measurement incompatibility does not imply Bell nonlocality. *Physical Review A*, 97(1), 2018. [38](#)
- [HWVE14] Mark Howard, Joel Wallman, Victor Veitch, and Joseph Emerson. Contextuality supplies the ‘magic’ for quantum computation. *Nature*, 510(7505):351–355, 2014. [2](#)
- [ISD⁺22] Marie Ioannou, Pavel Sekatski, Sébastien Designolle, Benjamin D. M. Jones, Roope Uola, and Nicolas Brunner. Simulability of high-dimensional quantum measurements. *Physical Review Letters*, 129:190401, 2022. [2](#), [11](#)
- [Jen18] Anna Jenčová. Incompatible measurements in a class of general probabilistic theories. *Physical Review A*, 98(1):012133, 2018. [4](#), [5](#), [7](#), [8](#), [10](#), [16](#), [20](#), [21](#), [22](#), [23](#), [31](#), [33](#)
- [JEPU24] Pauli Jokinen, Sophie Egelhaaf, Juha-Pekka Pellonpää, and Roope Uola. Compressing continuous variable quantum measurements. *Journal of Physics A: Mathematical and Theoretical*, 57:325302, 2024. [2](#), [11](#)
- [JUC⁺23] Benjamin D. M. Jones, Roope Uola, Thomas Cope, Marie Ioannou, Sébastien Designolle, Pavel Sekatski, and Nicolas Brunner. Equivalence between simulability of high-dimensional measurements and high-dimensional steering. *Physical Review A*, 107:052425, 2023. [2](#), [11](#)
- [JWP⁺24] Pauli Jokinen, Mirjam Weilenmann, Martin Plávala, Juha-Pekka Pellonpää, Jukka Kiukas, and Roope Uola. No-broadcasting characterizes operational contextuality. *Physical Review Letters*, 133(24):240201, 2024. [2](#), [3](#)
- [Ken27] Earle H Kennard. Zur Quantenmechanik einfacher Bewegungstypen. *Zeitschrift für Physik*, 44(4):326–352, 1927. [2](#)
- [KNI10] Gen Kimura, Koji Nuida, and Hideki Imai. Distinguishability measures and entropies for general probabilistic theories. *Reports on Mathematical Physics*, 66(2):175–206, 2010. [9](#)
- [KS90] Simon Kochen and Ernst P. Specker. The problem of hidden variables in quantum mechanics. *Ernst Specker Selecta*, pages 235–263, 1990. [2](#)
- [Kur15] Yui Kuramochi. Minimal sufficient positive-operator valued measure on a separable Hilbert space. *Journal of Mathematical Physics*, 56:102205, 2015. [18](#)
- [Lam18] Ludovico Lami. Non-classical correlations in quantum mechanics and beyond. *PhD thesis. arXiv preprint arXiv:1803.02902*, 2018. [3](#), [6](#)
- [Lep21] Leevi Leppäjärvi. Measurement simulability and incompatibility in quantum theory and other operational theories. *PhD thesis. arXiv preprint arXiv:2106.03588*, 2021. [3](#)
- [MG23] Markus P. Müller and Andrew J. P. Garner. Testing quantum theory by generalizing noncontextuality. *Physical Review X*, 13(4):041001, 2023. [2](#)
- [Mül21] Markus Müller. Probabilistic theories and reconstructions of quantum theory. *SciPost Physics Lecture Notes*, page 028, 2021. [3](#)
- [NB10] L. Narici and E. Beckenstein. *Topological Vector Spaces*. CRC Press, 2010. [66](#)
- [NDN⁺22] David P. Nadlinger, Peter Drmota, Bethan C. Nichol, Gabriel Araneda, Dougal Main, Raghavendra Srinivas, David M. Lucas, Christopher J. Ballance, Kirill Ivanov, Ernest Y.-Z. Tan, Pavel Sekatski, Rüdiger L. Urbanke, Renato Renner, Nicolas Sangouard, and Jean-Daniel Bancal. Experimental quantum key distribution certified by Bell’s theorem. *Nature*, 607(7920):682–686, 2022. [2](#)
- [OGWA17] Michał Oszmaniec, Leonardo Guerini, Peter Wittek, and Antonio Acín. Simulating positive-operator-valued measures with projective measurements. *Physical Review Letters*, 119:190501, 2017. [2](#), [13](#)
- [OMP19] Michał Oszmaniec, Filip B. Maciejewski, and Zbigniew Puchała. Simulating all quantum measurements using only projective measurements and postselection. *Physical Review A*, 100:012351, 2019. [2](#)

- [PG24] Martin Plávala and Otfried Gühne. Contextuality as a precondition for quantum entanglement. *Physical Review Letters*, 132(10):100201, 2024. [2](#)
- [PGQ24] Martin Plávala, Otfried Gühne, and Marco Túlio Quintino. All incompatible measurements on qubits lead to multipartite Bell nonlocality. *arXiv preprint arXiv:2403.10564*, 2024. [5](#)
- [Pir14] Stefano Pironio. All Clauser–Horne–Shimony–Holt polytopes. *Journal of Physics A: Mathematical and Theoretical*, 47(42):424020, 2014. [5](#), [30](#), [38](#)
- [Pis20] Gilles Pisier. *Tensor Products of C^* -Algebras and Operator Spaces: The Connes–Kirchberg Problem*. Cambridge University Press, 2020. [66](#)
- [Plá17] Martin Plávala. Conditions for the compatibility of channels in general probabilistic theory and their connection to steering and Bell nonlocality. *Physical Review A*, 96(5):052127, 2017. [31](#)
- [Plá22] Martin Plávala. Incompatibility in restricted operational theories: connecting contextuality and steering. *Journal of Physics A: Mathematical and Theoretical*, 55(17):174001, 2022. [2](#)
- [Plá23] Martin Plávala. General probabilistic theories: An introduction. *Physics Reports*, 1033:1–64, 2023. [3](#), [6](#), [7](#), [66](#)
- [PLG25] Martin Plávala, Laurens T. Ligthart, and David Gross. The polarization hierarchy for polynomial optimization over convex bodies, with applications to nonnegative matrix rank. *Linear Algebra and its Applications*, 732:15–32, 2025. [45](#)
- [PR94] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994. [31](#)
- [Rob29] Howard Percy Robertson. The uncertainty principle. *Physical Review*, 34(1):163, 1929. [2](#)
- [Roc70] R. Tyrrell Rockafellar. *Convex Analysis*. Number 28 in Princeton Mathematical Series. Princeton University Press, 1970. [22](#)
- [Rya02] Raymond A Ryan. *Introduction to tensor products of Banach spaces*. Springer, 2002. [66](#), [67](#)
- [SB14] Neil Stevens and Paul Busch. Steering, incompatibility, and Bell-inequality violations in a class of probabilistic theories. *Physical Review A*, 89(2):022123, 2014. [5](#), [22](#)
- [SBC⁺15] Ana Belén Sainz, Nicolas Brunner, Daniel Cavalcanti, Paul Skrzypczyk, and Tamás Vértesi. Postquantum steering. *Physical Review Letters*, 115:190403, 2015. [33](#)
- [Sch35] Erwin Schrödinger. Die gegenwärtige Situation in der Quantenmechanik. *Die Naturwissenschaften*, 48:23, 1935. [2](#)
- [Spe05] Robert W. Spekkens. Contextuality for preparations, transformations, and unsharp measurements. *Physical Review A—Atomic, Molecular, and Optical Physics*, 71(5):052108, 2005. [2](#)
- [SSW⁺21] David Schmid, John H. Selby, Elie Wolfe, Ravi Kunjwal, and Robert W. Spekkens. Characterization of noncontextuality in the framework of generalized probabilistic theories. *PRX Quantum*, 2(1):010331, 2021. [2](#)
- [TDS03] Barbara M. Terhal, Andrew C. Doherty, and David Schwab. Symmetric extensions of quantum states and local hidden variable theories. *Phys. Rev. Lett.*, 90:157903, Apr 2003. [54](#)
- [TPKLR22] Armin Tavakoli, Alejandro Pozas-Kerstjens, Ming-Xing Luo, and Marc-Olivier Renou. Bell nonlocality in networks. *Reports on Progress in Physics*, 85(5):056001, 2022. [2](#)
- [TU20] Armin Tavakoli and Roope Uola. Measurement incompatibility and steering are necessary and sufficient for operational contextuality. *Physical Review Research*, 2(1):013011, 2020. [2](#)
- [UBGP15] Roope Uola, Costantino Budroni, Otfried Gühne, and Juha-Pekka Pellonpää. One-to-one mapping between steering and joint measurability problems. *Physical Review Letters*, 115(23):230402, 2015. [5](#)
- [UCNG20] Roope Uola, Ana C. S. Costa, H. Chau Nguyen, and Otfried Gühne. Quantum steering. *Reviews of Modern Physics*, 92:015001, 2020. [2](#), [20](#)
- [UMG14] Roope Uola, Tobias Moroder, and Otfried Gühne. Joint measurability of generalized measurements implies classicality. *Physical Review Letters*, 113:160403, 2014. [2](#)
- [vDdB20] Josse van Dobben de Bruyn. Tensor products of convex cones, part II: Closed cones in finite-dimensional spaces. *arXiv preprint arXiv:2009.11843*, 2020. [66](#)
- [VPLU24] Isadora Veeren, Martin Plávala, Leevi Leppäjärvi, and Roope Uola. Semi-device-independent certification of the number of measurements. *Physical Review A*, 109(6):062203, 2024. [2](#)
- [WF23] Victoria J. Wright and Máté Farkas. Invertible map between Bell nonlocal and contextuality scenarios. *Physical Review Letters*, 131(22):220202, 2023. [2](#)
- [Wol21] Ramona Wolf. *Quantum key distribution*, volume 988 of *Lecture notes in physics*. Springer, 2021. [2](#)
- [WPGF09] Michael M. Wolf, David Pérez-García, and Carlos Fernández. Measurements incompatible in quantum theory cannot be measured jointly in any other no-signaling theory. *Physical Review Letters*, 103:230402, 2009. [5](#), [30](#), [37](#)
- [WZ82] William K. Wootters and Wojciech H. Zurek. A single quantum cannot be cloned. *Nature*, 299(5886):802–803, 1982. [2](#)

- [ZKSP26] Julius A Zeiss, Gereon Koßmann, René Schwonnek, and Martin Plávala. Finite de finetti for convex bodies and polynomial optimization. 2026. [44](#)
- [ZvLR⁺22] Wei Zhang, Tim van Leent, Kai Redeker, Robert Garthoff, René Schwonnek, Florian Fertig, Sebastian Eppelt, Wenjamin Rosenfeld, Valerio Scarani, Charles C.-W. Lim, and Harald Weinfurter. A device-independent quantum key distribution system for distant users. *Nature*, 607(7920):687–691, 2022. [2](#)

APPENDIX A. CONES, TENSOR PRODUCTS, AND POSITIVE MAPS

A.1. Cones and ordered vector spaces. In [Plá23] the notion and several basic results of cones were reviewed. More information on this topic can be found in [vDdB20, NB10] and more about tensor products is available in [Rya02, Pis20]. A central part of the theory of cones, included in the above references, is the equivalence of the existence of a cone in a vector space and the vector space having an order relation. Therefore, it is reasonable to associate positivity with cones. However, this is what makes the definition of tensor products of cones non trivial and in fact ambiguous as we will discuss shortly.

For the upcoming definitions let V , V_A and V_B be finite dimensional real vector spaces. We start with the definition of a cone.

Definition A.1 (Cones of vector spaces). *A subset $V^+ \subset V$ of a vector space is called a cone if for all $v \in V^+$ it holds that $\lambda v \in V^+$ for any $\lambda \in \mathbb{R}_+$.*

A cone V^+ is:

- *convex if V^+ is a convex set*
- *closed if V^+ is closed in the Euclidean topology on V*
- *pointed if $V^+ \cap -V^+ = \{0\}$*
- *generating if $\text{span}(V^+) = V$, i.e., if $V^+ - V^+ = \{w - v \mid w \in V^+, v \in V^+\} = V$*

If a cone V^+ is convex, closed, pointed, and generating, it is called a proper cone.

Throughout this whole work, to avoid pathologies all cones will be proper. Furthermore V^+ , V_A^+ and V_B^+ will denote the corresponding cones of V , V_A and V_B . An *ordered vector space* is a tuple (V, V^+) of a vector space and a proper cone.

The dual cone is a subset of the dual space V^* of V and defined by collecting all linear functionals that map elements of the cone to a positive number. In this way, the notion of positivity is preserved.

Definition A.2 (Dual cone). *The dual cone is a subset of all positive functionals on a vector space, i.e., $(V^*)^+ := (V^+)^* = \{\varepsilon \in V^* \mid \varepsilon(v) \geq 0 \ \forall v \in V^+\}$.*

An example of a cone is the set of $d \times d$ positive semi-definite complex matrices PSD_d (for $d < \infty$). That is, for $A \in \text{PSD}_d$ it holds that $\langle x, Ax \rangle \geq 0$ for all $x \in \mathbb{C}^d$. The underlying vector space for PSD_d is the real vector space of $d \times d$ self-adjoint (or Hermitian) complex matrices $\mathcal{M}(\mathbb{C})_d^{\text{sa}}$.

Proposition A.3. *The set of positive semi-definite matrices is self dual up to isomorphism, i.e., $\text{PSD}_d \cong \text{PSD}_d^*$.*

Up to some normalization, the set PSD_d can be thought of as the set of quantum states, that is, for every $A \in \text{PSD}_d$ we have $A = \lambda \rho$ for some density matrix $\rho \in D(\mathbb{C}^d)$. Density matrices are consequently the base of the cone PSD_d since any element of the cone is a nonnegative multiple of a density matrix.

A.2. Tensor products of cones. When considering the tensor product of ordered vector spaces, we need to preserve the notion of positivity. Put differently, the result of a tensor product between cones should again give a cone. In general, there are infinitely many ways of constructing the tensor product of two cones, but there is a minimal and a maximal way to do this:

Definition A.4 (Minimal tensor product). *Let V_A^+ and V_B^+ be two cones. Their minimal tensor product is then defined as $V_A^+ \otimes V_B^+ = \text{conv}\{x_A \otimes x_B \mid x_A \in V_A^+, x_B \in V_B^+\}$.*

It can be verified that the minimal tensor product indeed gives a cone. Inspired by quantum mechanics, we call elements of the minimal tensor product of cones *separable*, since the normalized elements of $\text{PSD}_{d_A} \dot{\otimes} \text{PSD}_{d_B}$ are the separable states.

Definition A.5 (Maximal tensor product). *The maximal tensor product of two cones is the set $V_A^+ \hat{\otimes} V_B^+ = ((V_A^*)^+ \dot{\otimes} (V_B^*)^+)^*$.*

It can again be verified that the above definition gives a cone. Normalized elements of the cone $\text{PSD}_{d_A} \hat{\otimes} \text{PSD}_{d_B}$ are entanglement witnesses. These definitions can immediately be generalized to any finite number of cones. We will recall next some basic but useful results about cones and their tensor products:

Lemma A.6. *Let (V, V^+) , (V_A, V_A^+) , (V_B, V_B^+) be ordered vector spaces. Then,*

- (1) $V^+ \cong (V^{**})^+$
- (2) *Let $v \in V$ and $\varepsilon(v) \geq 0 \forall \varepsilon \in (V^*)^+$. Then $v \in V^+$.*
- (3) *The sets $V_A^+ \dot{\otimes} V_B^+$ and $V_A^+ \hat{\otimes} V_B^+$ are proper cones.*
- (4) $V_A^+ \dot{\otimes} V_B^+ \subseteq V_A^+ \hat{\otimes} V_B^+$. *Equality holds if and only if at least one of the cones is simplicial, i.e., isomorphic to $\mathbb{R}_+^n$ for some $n \in \mathbb{N}$ [ALPP21].*
- (5) $(V_A^+ \hat{\otimes} V_B^+)^* = (V_A^*)^+ \dot{\otimes} (V_B^*)^+$

We will omit the isomorphism between V and V^{**} when we make use of the fact $V \cong V^{**}$ for real, finite-dimensional vector spaces. A tensor product of cones V_A and V_B is any cone V_{AB}^+ such that $V_A^+ \dot{\otimes} V_B^+ \subseteq V_{AB}^+ \subseteq V_A^+ \hat{\otimes} V_B^+$. Note that $\text{PSD}_{d_A} \dot{\otimes} \text{PSD}_{d_B} \subsetneq \text{PSD}_{d_A \cdot d_B} \subsetneq \text{PSD}_{d_A} \hat{\otimes} \text{PSD}_{d_B}$.

A.3. Positive maps. So far, we have focused on the notion of positivity and emphasized the connection to cones of vector spaces. We can now define what it means for a linear map to be positive.

Definition A.7 (Positive maps). *Let $L : V_A \rightarrow V_B$ be a linear map. Then, L is positive if $L(V_A^+) \subset V_B^+$.*

We can in fact connect tensor products between cones and positive maps. For this, we will use that one can identify a linear map $L : V_A \rightarrow V_B$ with a tensor $\xi_L \in V_A^* \otimes V_B$. As pointed out, e.g., in [Rya02], the one-to-one correspondence is given via

$$\psi_B(L(v_A)) = (v_A \otimes \psi_B)(\xi_L), \text{ where } v_A \in V_A, \psi_B \in V_B^*. \quad (\text{A.1})$$

Lemma A.8. *Let V_A^+ and V_B^+ be two cones. Then $\xi_L \in (V_A^*)^+ \dot{\otimes} V_B^+$ if and only if $L : V_A \rightarrow V_B$ is positive in the sense of Definition A.7.*

The next proposition shows again that the maximal tensor product is in general strictly bigger, which means that the conditions on its elements are less strong. Here in the case of positive maps this is demonstrated by not having an *if and only if* statement in contrast to the above lemma.

Proposition A.9. $L \in (V_A^*)^+ \dot{\otimes} V_B^+ \not\subseteq L(V_A^+) \subset V_B^+$

Quantum theory already provides us with a counterexample for the direction from right to left, that is, set $(V_A^*)^+ = \text{PSD}_{d_A}^*$ and $V_B^+ = \text{PSD}_{d_B}$. Separable maps can be thought of as being entanglement breaking and not every positive map is entanglement breaking, but every entanglement breaking map is positive.

It is crucial for this work to understand that inseparable operators can be distinguished from separable operators by functionals (“witnesses”), which can map some of the former to negative numbers and always map the latter to positive numbers. This is formalized by the next lemma.

Lemma A.10. *For $v \in V_A^+ \hat{\otimes} V_B^+$, such that $v \notin V_A^+ \dot{\otimes} V_B^+$, there is a witness type functional $W \in (V_A^*)^+ \dot{\otimes} (V_B^*)^+$ with $W(v) < 0$.*

The proof can be done with the help of the strict hyperplane separation theorem, using that we only consider proper cones which are in particular closed.

Finally, we extend a positive, linear map to a map between tensor products while not breaking the positivity.

Lemma A.11. *Let V_A^+ , V_B^+ as well as V_C^+ be cones and $\Phi : V_A^+ \rightarrow V_B^+$ be a linear map. Then Φ induces the maps:*

- (i) $\Phi \otimes \text{id} : V_A^+ \hat{\otimes} V_C^+ \rightarrow V_B^+ \hat{\otimes} V_C^+$
- (ii) $\Phi \otimes \text{id} : V_A^+ \hat{\otimes} V_C^+ \rightarrow V_B^+ \hat{\otimes} V_C^+$

Note that the induced maps are still positive because a cone gets mapped to a cone. Creating a tensor product between cones and hence adding an identity to the map Φ preserves the notion of positivity.

Remark A.12. *At this point one needs to be careful when talking about positive maps. A map from Lemma A.11 might be positive in the sense that it maps a cone to a cone, though this is not the same as complete positivity, which is a term also found in quantum theory. If we let V_B^+ and V_C^+ be sets of positive operators on a Hilbert space, then the operators in the maximal tensor product are not always positive in the sense that they map all bipartite states to positive numbers. There might be entangled states that get mapped to something negative. An example is the partial transposition which is a positive, however not completely positive map. This fact is the quintessence of the PPT-criterion for entanglement detection [HHH96].*

Email address, Tim Achenbach: `tim.n.achenbach@jyu.fi`

(Tim Achenbach and Leevi Leppäjärvi) FACULTY OF INFORMATION TECHNOLOGY, UNIVERSITY OF JYVÄSKYLÄ, 40100 JYVÄSKYLÄ, FINLAND

(Tim Achenbach) NATURWISSENSCHAFTLICH-TECHNISCHE FAKULTÄT, UNIVERSITÄT SIEGEN, WALTER-FLEX-STRASSE 3, 57068 SIEGEN, GERMANY

Email address, Andreas Bluhm: `andreas.bluhm@univ-grenoble-alpes.fr`

(Andreas Bluhm) UNIV. GRENoble ALPES, CNRS, GRENoble INP, LIG, 38000 GRENoble, FRANCE

Email address, Leevi Leppäjärvi: `leevi.i.leppajarvi@jyu.fi`

Email address, Ion Nechita: `nechita@irsamc.ups-tlse.fr`

(Ion Nechita) LABORATOIRE DE PHYSIQUE THÉORIQUE, UNIVERSITÉ DE TOULOUSE, CNRS, UPS, FRANCE

Email address, Martin Plávala: `martin.plavala@itp.uni-hannover.de`

(Martin Plávala) INSTITUT FÜR THEORETISCHE PHYSIK, LEIBNIZ UNIVERSITÄT HANNOVER, HANNOVER, GERMANY