Logical Methods in Computer Science
Volume 15, Issue 2, 2019, pp. 3:1–3:51
https://lmcs.episciences.org/

Submitted Apr. 25, 2017
Published Apr. 26, 2019

GENERALISED MERMIN-TYPE NON-LOCALITY ARGUMENTS

STEFANO GOGIOSO AND WILLIAM ZENG

Quantum Group, University of Oxford, UK
e-mail address: stefano.gogioso@cs.ox.ac.uk

Rigetti Computing, Berkeley, CA
e-mail address: zeng.will@gmail.com

Abstract. We broadly generalise Mermin-type arguments on GHZ states, and we provide exact group-theoretic conditions for non-locality to be achieved. Our results are of interest in quantum foundations, where they yield a new hierarchy of quantum-realisable All-vs-Nothing arguments. They are also of interest to quantum protocols, where they find immediate application to a non-trivial extension of the hybrid quantum-classical secret sharing scheme of Hillery, Bužek and Berthiaume (HBB). Our proofs are carried out in the graphical language of string diagrams for dagger compact categories, and their validity extends beyond quantum theory to any theory featuring the relevant algebraic structures.

Introduction

Non-locality is a defining feature of quantum mechanics, and its connection to the structure of phase groups is a key foundational question. A particularly crisp example of this connection is given by Mermin's argument for qubit GHZ states [Mer90], which finds practical application in the HBB quantum secret sharing protocol.

In Mermin's argument, N qubits are prepared in a GHZ state (with Pauli Z as computational basis), then a controlled phase gate is applied to each, followed by measurement in the Pauli X observable. Even though the N outcomes (each valued in $\mathbb{Z}_2$) are probabilistic, their parity turns out to satisfy certain deterministic equations. Mermin shows that the existence of a local hidden variable model would imply a joint solution for the equations, which however form an inconsistent system. Mermin concludes that the scenario is non-local.

Mermin's argument has sparked a number of lines of enquiry, and this work is concerned with two in particular: one leading to All-vs-Nothing arguments, and the other investigating the role played by the phase group. All-vs-Nothing arguments [ABK$^+$15] arise in the context of the sheaf-theoretic framework for non-locality and contextuality [AB11], and generalise the idea of a system of equations which is locally consistent but globally inconsistent. The second line of research is brought forward within the framework of categorical quantum

Key words and phrases: Mermin non-locality, categorical quantum mechanics, strongly complementary observables, All-vs-Nothing arguments, quantum secret sharing.

An extended abstract for a small part of this work has appeared in the Proceedings of QPL 2015, and can be found in EPTCS 195, 2015, pages 228-246, doi:10.4204/EPTCS.195.17.

DOI:10.23638/LMCS-15(2:3)2019

mechanics [AC09, CK15, CK17], and it focuses on the algebraic characterisation of phase gates and strongly complementary observables.

A detailed analysis of Mermin's argument shows that the special relationship between the Pauli X and Pauli Z observables, known as strong complementarity, is key to its success [CDKW12]. A pair of complementary observables corresponds to mutually unbiased orthonormal bases: for example, both Pauli X and Pauli Y are complementary to Pauli Z. Strong complementarity [CD11, DD16] amounts to a strictly stronger requirement: if one observable is taken as the computational basis, the other must correspond to the Fourier basis for some finite abelian group. Pauli X fits the bill, for the abelian group $\mathbb{Z}_2$, but Pauli Y doesn't (Pauli X is the only single-qubit observable strongly complementary to Pauli Z).

In [CDKW12], Mermin's argument is completely reformulated in terms of strongly complementary observables (using †-Frobenius algebras) and phase gates. It can therefore be tested on theories different from quantum mechanics, to better understand the connection between non-locality and the structure of phase groups. A particularly insightful comparison is given by qubit stabiliser quantum mechanics [CD11, Bac14] vs Spekkens' toy model [Spe07, CE12]: both theories sport very similar operational and algebraic features, but the difference in phase groups ($\mathbb{Z}_4$ for the former vs $\mathbb{Z}_2 \times \mathbb{Z}_2$ for the latter) results in the former being non-local and the latter being local (both models have $\mathbb{Z}_2$ as group of measurement outcomes, like Mermin's original argument). The picture arising from comparing qubit stabiliser quantum mechanics and Spekkens' toy model is iconic, and provides a first real glimpse into the connection between phase groups and non-locality [CES10].

While presenting an extremely compelling case for stabiliser qubits and Spekkens' toy qubits, the work of Refs. [CDKW12, CES10] does not treat the general case (i.e. beyond $\mathbb{Z}_2$ as group of measurement outcomes), nor does it provide a complete algebraic characterisation of the conditions guaranteeing non-locality. In this work, we fully generalise Mermin's argument from $\mathbb{Z}_2$ to arbitrary finite abelian groups, in arbitrary theories and for arbitrary phase groups (we will refer to these as **generalised Mermin-type arguments**). We also provide exact algebraic conditions for non-locality to be exhibited by our generalised Mermin-type arguments, thus bringing this line of investigation to a satisfactory conclusion.

We proceed to make contact with the All-vs-Nothing line of enquiry [ABK$^+$15], showing that the non-local generalised Mermin-type arguments yield a new hierarchy of quantum-realisable All-vs-Nothing empirical models (and hence they are strongly contextual). As a corollary, we manage to show that the hierarchy of quantum-realisable All-vs-Nothing models over finite fields does not collapse.

Mermin's argument for the qubit GHZ states also finds practical application in the quantum-classical secret sharing scheme of Hillery, Bužek and Berthiaume [HBB99]. We extend the scheme to our generalised Mermin-type arguments, and use strong contextuality to provide some device-independent security guarantees (which apply to the original HBB scheme as a special case).

Synopsis of the paper. In Section 1, we provide a brief recap of the technical background for this paper. We quickly review dagger compact categories, the CPM construction, Frobenius algebras, the CP* construction and the sheaf-theoretic framework for non-locality and contextuality. We introduce a new flavour of CP* categories, and prove some results connecting them to the sheaf-theoretic framework for non-locality and contextuality.

In Section 2, we present Mermin's original argument in detail, deconstructing it in the interest of our upcoming generalisation.

In Sections 3 and 4, we introduce complementarity, strong complementarity and phase groups, and elaborate on the relationship that ties them together. Some of the simpler results are common knowledge in the field, and have appeared in similar form in other works (such as [CDKW12, Kis12, CK17]): they are re-proposed here to achieve a uniform, coherent presentation of the material.

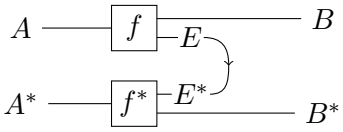
In Section 5, we use strong complementarity and phase groups to formulate our generalised Mermin-type arguments, and we prove the exact algebraic conditions for the models to be non-local. In Section 6, we prove that our arguments are all quantum realisable. In Section 7, we connect with the All-vs-Nothing framework, showing that our arguments always result in All-vs-Nothing models whenever they are non-local.

In Section 9, finally, we present an extension of the HBB quantum-classical secret sharing scheme to our newly generalised Mermin-type arguments, and we provide some novel device-independent guarantees of security.

1. Background

Categories for quantum theory. The framework of dagger compact categories captures some of the most fundamental structural features of pure-state quantum mechanics [AC09, CK15, CK17]: symmetric monoidal structure captures its characterisation as a theory of processes, composing sequentially and in parallel; the dagger captures the state-effect duality induced by the inner product; the compact closed structure captures operator-state duality. Dagger compact categories are commonplace in the practice of categorical quantum mechanics, and we will assume the reader is familiar with them. We recommend [Sel09] for a detailed treatment of many subtle technicalities in the various constructions. The Hilbert space model of (finite-dimensional) pure-state quantum mechanics corresponds to the dagger compact category fdHilb of (finite-dimensional) complex Hilbert spaces and linear maps between them, while the operator model of mixed-state quantum mechanics corresponds to the CPM category CPM[fdHilb].

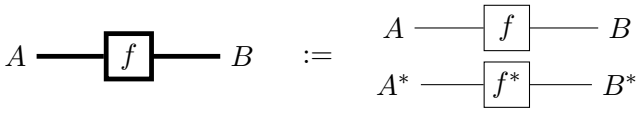
Categories of completely positive maps, also known as **CPM categories**, can be constructed for all dagger compact categories, in a process which mimics the way in which the operator model of mixed-state quantum mechanics is constructed from the Hilbert space model of pure-state quantum mechanics [Sel07]. Given a dagger compact $\mathcal{C}$, the corresponding CPM category CPM[$\mathcal{C}$] is defined as the subcategory of $\mathcal{C}$ having morphisms in the following form:



(1.1)

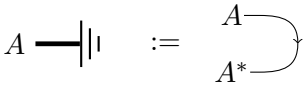
where $f : A \to B \otimes E$ is a morphism of $\mathcal{C}$, and $f^* : A^* \to E^* \otimes B^*$ is its conjugate, obtained via the dagger compact structure. Processes in the CPM category are called **completely positive (CP) maps**. The system E in Diagram 1.1 is often interpreted as an **environment** which is operationally inaccessible, and hence must be "discarded" after the process has taken place. In the case of CPM[fdHilb], i.e. in the operator model of mixed-state quantum mechanics, Diagram 1.1 can then be seen as an alternative formulation of Kraus decomposition.

Because diagrammatic reasoning about categories of completely positive maps often involves two distinct SMCs (the original category $\mathcal{C}$ and the CPM category CPM[$\mathcal{C}$]), a stylistic choice is adopted where systems and processes of the CPM category are denoted by thicker wires, boxes and decorations. For example, the "doubled" version $f \otimes f^*$ of a process $f : A \to B \otimes E$ will be denoted as f with thicker wires and box:



(1.2)

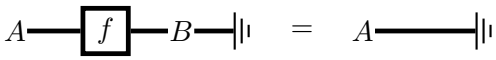
The caps from compact closed structure play a particularly important role in the definition of the CPM category, and are given their own decoration:



(1.3)

The CP map **double**$[f]$ defined by Equation 1.2 is called the **double** of process f, while the CP map $\dashv\!\Vert_A$ defined by Equation 1.3 is called the **discarding map** on system A. In mixed-state quantum mechanics, the double of a linear map f is the rank-1 CP map **double**$[f] : \rho \mapsto f \circ \rho \circ f^\dagger$, while the discarding map $\dashv\!\Vert_A$ sends a positive state $\rho \in \mathcal{L}[A]$ to its trace $\mathrm{Tr}[\rho] \in \mathcal{L}[\mathbb{C}] \cong \mathbb{R}^+$.

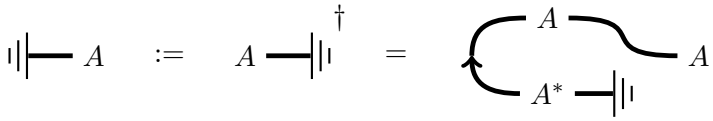
The discarding maps are also called an *environment structure* in the literature [CP10], and are tightly related to *causality*, an important feature arising at the interface between quantum theory and relativity [CL13, CK15, CDP11]. We say that a process is **normalised** (sometimes also called **causal**) if performing it and then discarding the output is the same as discarding the input:



(1.4)

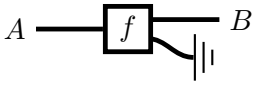
In particular, discarding normalised states results in the scalar 1.

CPM categories CPM[$\mathcal{C}$] are dagger compact, and the rules of diagrammatic reasoning for dagger compact categories apply to them. The compact structure for CPM[$\mathcal{C}$] is given by the doubles of the cups and caps of $\mathcal{C}$, while the adjoint of a process in the form of Diagram 1.6 is given by first taking the adjoint in $\mathcal{C}$, and then using the following equation for the adjoint of the discarding map:



(1.5)

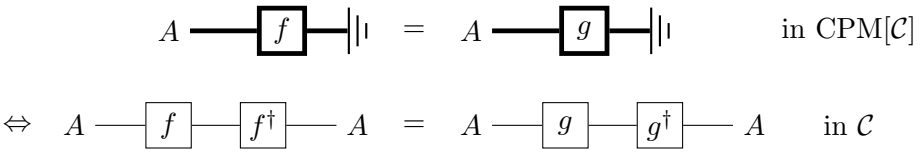
Because the doubled processes **double**$[f]$ and the discarding maps $\dashv\!\Vert_A$ are well-defined CP maps, it is legitimate to rephrase the very definition of the CPM category by saying that its processes are exactly those in the following form:



(1.6)

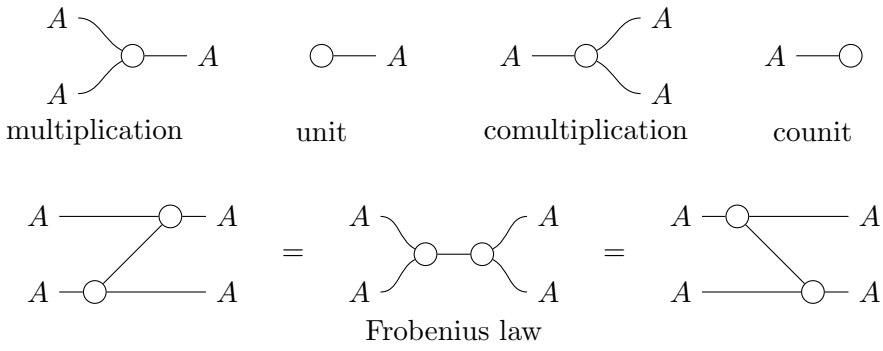
This means that doubled processes and discarding maps are enough to *express* all CP maps, but to *prove results about* CP maps we need a graphical axiom relating a generic CPM category CPM[$\mathcal{C}$] to the corresponding original category $\mathcal{C}$. The required relationship is

encoded by the following **CPM axiom**, which characterises the action of discarding maps in CPM[$\mathcal{C}$] in terms of the dagger structure of $\mathcal{C}$:



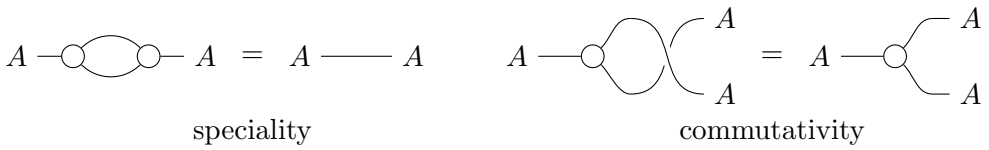
(1.7)

Frobenius algebras. Frobenius algebras are a fundamental ingredient of quantum theory, where they are intimately related to the notion of observable. A **†-Frobenius algebra** on an object A of a dagger symmetric monoidal category (henceforth †-SMC) is given by a monoid $(A, \triangleright, \circ\!-)$ on A (i.e. $\triangleright : A \otimes A \to A$ is associative and has $\circ\!- : I \to A$ as its bilateral unit), and a corresponding comonoid $(A, \prec, -\!\circ)$, which are related by the following Frobenius law:



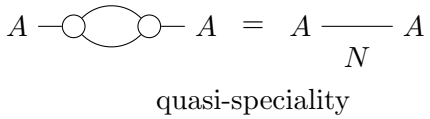
(1.8)

A †-Frobenius algebra is said to be **special** if the comultiplication $\prec$ is an isometry, and **commutative** if the monoid and comonoid are commutative:



(1.9)

More in general, a **quasi-special** †-Frobenius algebra is one with comultiplication $\prec$ which is an isometry up to a **normalisation factor** N, where N is in the form $n^\dagger n$ for some invertible scalar n:

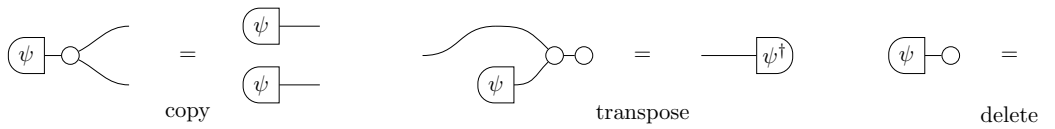


(1.10)

Because a several combinations of these properties will play a role in this work, we now introduce a number of short-hands for †-Frobenius algebras:

†-Frobenius algebras	commutative	arbitrary
special	†-SCFA	†-SFA
quasi-special	†-qSCFA	†-qSFA
arbitrary	†-CFA	†-FA

The importance of †-SCFAs in categorical quantum mechanics comes from the fact that they correspond to orthonormal bases, i.e. non-degenerate quantum observables. Key to this correspondence is the notion of **classical states** for a †-FA $\circ$, those states ψ which are copied/transposed/deleted by $\multimap$ in the following sense:

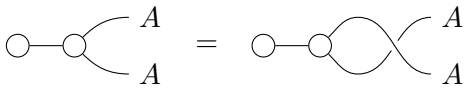


(1.11)

Theorem 1.1 [CPV13]**.** *We denote the set of classical states for a †-FA $\circ$ by $K(\circ)$. In* fdHilb*, the classical states for a †-SCFA $\circ$ always form an orthonormal basis*[1]*. Furthermore, any orthonormal basis arises this way for a unique †-SCFA. More in general, if $\circ$ is a †-qSCFA, with normalisation factor N, then the classical states for $\circ$ form an orthogonal basis, each state having norm $\sqrt{N}$. Furthermore, any orthogonal basis where all states have the same norm $\sqrt{N}$ arises this way for a unique †-qSCFA.*

The concept of classical states forming a basis is generalised to arbitrary †-SMCs by the notion of enough classical states. A †-FA $\circ$ on an object A is said to have **enough classical states** if its classical states separate morphisms from A, i.e. two morphisms $f, g : A \to B$ are equal whenever they satisfy $f \circ \psi = g \circ \psi$ for all classical states ψ of $\circ$. Because of the copy condition, a †-FA with enough classical states is always commutative.

This algebraic characterisation of quantum observables is not limited to the non-degenerate case of orthonormal bases, but can be extended to the more general case of complete families of orthogonal projectors. To do so, one considers **balanced-symmetric** †-Frobenius algebras[2], i.e. those satisfying the following equation:



(1.12)

Theorem 1.2 [Vic11]**.** *In* fdHilb*, balanced-symmetric †-SFAs are in bijective correspondence with C*-algebras, and hence with complete families of orthogonal projectors.*

The correspondence between balanced-symmetric †-SFAs and C*-algebras will not play a role in this work, but it helps to frame the broad role played by †-Frobenius algebras in categorical quantum mechanics.

Further to their correspondence with quantum observables, †-Frobenius algebras find direct use as fundamental building blocks of quantum algorithms and protocols [Vic12, BH12, CK17, GK17]. When designing quantum protocols, classical data is often encoded into quantum systems using orthonormal bases. In this context, the four processes in a †-SCFA can be seen as coherent versions of the basic data manipulation primitives:

(a) the comultiplication $\multimap = |\psi_x\rangle \mapsto |\psi_x\rangle \otimes |\psi_x\rangle$ is the coherent copy of classical data;
(b) the counit $\multimap = |\psi_x\rangle \mapsto 1$ is the coherent deletion of classical data;
(c) the multiplication $\multimap = |\psi_x\rangle \otimes |\psi_y\rangle \mapsto \delta_{xy}|\psi_x\rangle$ is the coherent matching of classical data;
(d) the unit $\multimap = \sum_x |\psi_x\rangle$ is the coherent superposition of classical data (up to normalisation).

In this sense, †-SCFAs in general †-SMC are often interpreted as modelling coherent copy/delete/matching operations on some kind of classical data (usually modelled by their

[1]The copy and delete conditions are sufficient to characterise classical states in the case of fdHilb.
[2]Commutative †-FAs are a special case of balanced-symmetric †-FAs.

classical states)[3]. More in general, balanced-symmetric †-SFAs can be though of as coherent manipulation of data which carries some residual entanglement after being copied[4].

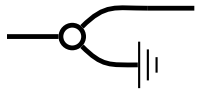
Because of their diagrammatic definition, †-FA in a dagger compact category $\mathcal{C}$ give rise to †-FA in the CPM category CPM[$\mathcal{C}$] by doubling: the †-FA in CPM[$\mathcal{C}$] that arise this way are said to be **canonical**. In this work, we will only consider canonical †-FA when working with CPM categories.

The CP* construction. In the framework of mixed-state quantum mechanics, classical systems can be though of as quantum systems which are constantly undergoing decoherence in some basis. In CPM[fdHilb], the **decoherence map** $\mathrm{dec}_\bigcirc$ in an orthonormal basis $(|x\rangle)_x$ is the one zeroing out all non-diagonal elements of a positive state:

$$\mathrm{dec}_\bigcirc := \rho \mapsto \sum_x |x\rangle(\langle x|\rho|x\rangle)\langle x| \tag{1.13}$$

The decoherence map can be written as follows in terms of the associated †-SCFA $\bigcirc$:

$\mathrm{dec}_\bigcirc$:=



(1.14)

This means that a decohered quantum system can be though of as having undergone a coherent copy operation, with one copy lost in the environment. We take Equation 1.14 to be the definition of decoherence maps for arbitrary canonical balanced-symmetric †-SFAs in arbitrary CPM categories.

Starting from a CPM category CPM[$\mathcal{C}$], we wish to construct a new category which includes some kind of classical systems, as defined by decoherence maps. This new category CP*[$\mathcal{C}$], known as the **CP* category**, can be defined as follows:

(i) the objects are pairs (A, a) of an object A of CPM[$\mathcal{C}$] and a normalised self-adjoint idempotent process $a : A \to A$ taking either the form $a := id_A$ or the form $a := \mathrm{dec}_\bigcirc$ for some balanced-symmetric †-SFA $\bigcirc$ on A;

(ii) the morphisms $(A, a) \to (B, b)$ in CP* are the morphisms $f : A \to B$ in CPM[$\mathcal{C}$] in CPM which are invariant under the specified idempotents on A and B, i.e. those which satisfy $b \circ f \circ a = f$.

Note that this definition is a hybrid of the perspective of Ref. [Sel08], based on decoherence maps and the Karoubi envelope, and the perspective of Refs. [CHK14, CH15], based on C* algebras and quantum logic. Because we have picked processes to be invariant under self-adjoint idempotents, CP*[$\mathcal{C}$] is always dagger compact.

The CP* category contains CPM[$\mathcal{C}$] as the full subcategory associated with objects in the form (A, id_A): we refer to the latter as **CPM objects**[5], and we simply denote them by A for simplicity. We refer to the CP* objects in the form $(A, \mathrm{dec}_\bigcirc)$ as **super-selected objects**, and often denote them by $(A, \bigcirc)$ for simplicity.

In the case of CP*[fdHilb], a balanced-symmetric †-SFA $\bigcirc$ on a finite-dimensional Hilbert space A corresponds, by Theorem 1.2, to a complete family $(P_j)_j$ of orthogonal projectors[6].

[3]This extends straightforwardly to †-qSCFA and their unnormalised classical states.

[4]In quantum mechanics, this is because non-demolition measurements in a degenerate observable only breaks entanglement between the subspaces associated with distinct projectors, but not within each subspace.

[5]We will keep using the doubled notation for them and morphisms between them.

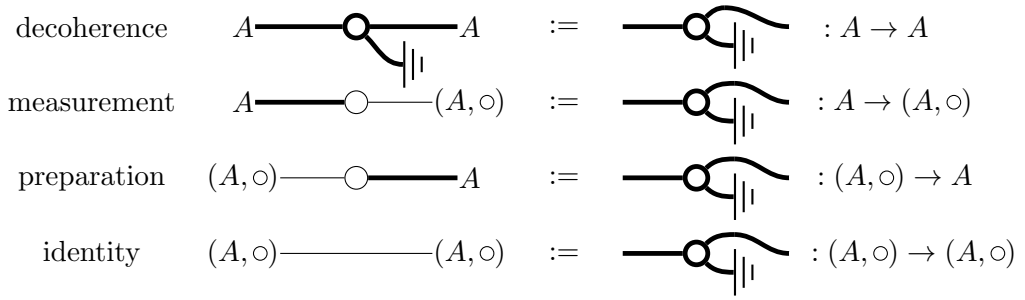
[6]I.e. $P_i \circ P_j = \delta_{ij} P_i$ and $\sum_i P_i = id_A$.

The associated decoherence map $\mathrm{dec}_\bigcirc$ takes the following concrete form:

$$\mathrm{dec}_\bigcirc = \rho \mapsto \sum_j P_j \, \rho \, P_j^\dagger \tag{1.15}$$

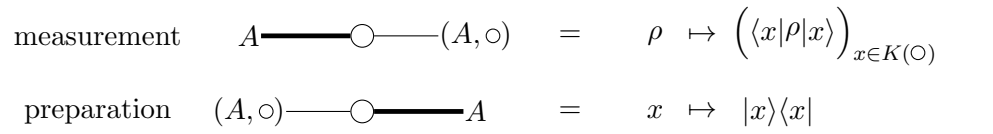
This means that objects in CP*[fdHilb] truly are super-selected quantum systems, with super-selection sectors given by the domains of the projectors $(P_j)_j$. In particular, super-selected objects associated to †-SCFAs (corresponding to non-degenerate observables, i.e. families of 1-dimensional projectors) behave as classical probabilistic systems.

If $\circ$ is a balanced-symmetric canonical †-SFA on an object A, the decoherence map $\mathrm{dec}_\circ$ is always a process $\mathrm{dec}_\circ : A \to A$ in CP*[$\mathcal{C}$]. Because of idempotence, however, it is also a process $A \to (A, \circ)$ and a process $(A, \circ) \to A$: we will refer to the former as the **measurement** in $\circ$, and the latter as the **preparation** in $\circ$. The single and doubled notation distinguish between the different cases:



(1.16)

In CP*[fdHilb], preparations and measurements for a †-SCFA $\circ$ associated to an orthonormal basis $(|x\rangle)_{x \in X}$ of a finite-dimensional Hilbert space A take the familiar form traditionally adopted by the literature:



(1.17)

Demolition measurements are traditionally thought to result in some kind of classical (probabilistic) data. Unfortunately, our framework does not yet allow us to conclude anything of the sort, as we lack an appropriate definition of classical systems to work with. Following the footsteps of the sheaf-theoretic framework for non-locality and contextuality [AB11], we generalise probabilities from $\mathbb{R}^+$ to some arbitrary commutative semiring R. For a fixed commutative semiring R, we define our category of **classical R-probabilistic systems** to be the category R-Mat of free, finite-dimensional R-semimodules and R-semilinear maps between them:

(a) the objects of R-Mat are in the form R^X for all finite sets X;
(b) the morphisms $R^X \to R^Y$ in R-Mat are $Y \otimes X$ matrices with values in R
(c) R-Mat is a SMC, with fSet (finite sets and functions) as a sub-SMC;
(d) R-Mat inherits the discarding maps $\dashv\!\shortmid_X := x \mapsto \star$ of fSet;
(e) R-Mat is enriched in itself, with morphisms $R^{Y \otimes X}$ forming the free finite-dimensional R-semimodule $R^{Y \otimes X}$.

If we generalise this definition to *involutive* commutative semirings R, i.e. those coming with an involution $\dagger : R \to R$, the category R-Mat is in fact a dagger compact category.

The traditional definition of **classical probabilistic systems** corresponds to working in $\mathbb{R}^+$-Mat: normalised states are probability distributions over finite sets, and normalised processes are stochastic maps (also, we always think of $\mathbb{R}^+$ as coming with the trivial involution $id_{\mathbb{R}^+}$). However, using arbitrary semirings opens the way to interesting generalisations: a prominent example is that of **classical possibilistic systems**, which are associated to the semiring $R = \mathbb{B}$ of the booleans and play a large role in the sheaf-theoretic framework for non-locality and contextuality.

Definition 1.3. We say that a SMC $\mathcal{D}$ is **distributively** CMon**-enriched** if the following conditions hold:

(1) the category is CMon-enriched, i.e. morphisms $A \to B$ form a commutative monoid $\left(\mathrm{Hom}_{\mathcal{D}}[A, B], +, 0\right)$ for any fixed objects A, B;
(2) the tensor product $\otimes$, associators and unitors are all linear.

The definition can be extended to a †-SMC (or dagger compact category) $\mathcal{D}$ by asking that the dagger also be linear.

The scalars of SMCs which are distributively CMon-enriched always form a commutative semiring R (which is furthermore involutive in the case of †-SMCs), and all homsets automatically inherit the structure of R-semimodules. We use this observation to define classical systems within the context of CP* categories.

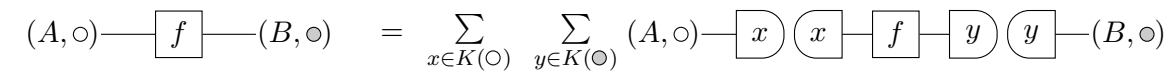
Definition 1.4. We say that a $\mathrm{CP}^*[\mathcal{C}]$ is an **R-probabilistic CP* category**, or an **R-probabilistic theory**, if it satisfies the following conditions.

(i) The dagger compact category $\mathrm{CP}^*[\mathcal{C}]$ is distributively CMon-enriched, with R as its involutive semiring of scalars[7].
(ii) For each $n \in \mathbb{N}$, there is some super-selected system $(A, \circ)$ in $\mathrm{CP}^*[\mathcal{C}]$ such that:
  (a) $\circ$ is a †-SCFA with enough classical states;
  (b) the classical states of $\circ$ are mutually orthogonal.
  (c) $\circ$ has exactly n classical states;

In this context, we refer to super-selected systems $(A, \circ)$ satisfying conditions (a) and (b) above as **classical systems**. Hence requirement (ii) above can be rephrased to say that for every $n \in \mathbb{N}$ there is a classical system with n classical states.

Theorem 1.5. *The full sub-SMC of an R-probabilistic CP* category spanned by the classical systems is equivalent to R-*Mat*.*

Proof. All we really need to show is that processes $(A, \circ) \to (B, \bullet)$ between two classical systems in the CP* category form an R-module which is isomorphic to the R-module of processes $R^{K(\bullet)} \to R^{K(\circ)}$ in the category R-Mat of classical R-probabilistic systems. Firstly, every process $f : (A, \circ) \to (B, \bullet)$ is determined by the R-valued matrix obtained by testing against classical states of the two †-SCFAs:



(1.18)

[7]Equivalently, we can ask for $\mathrm{CPM}[\mathcal{C}]$ to be enriched, as the two categories mutually inherit enrichment, scalars and discarding maps.

Secondly, every matrix $\left(F_x^y\right)_{x\in K(\bigcirc)}^{y\in K(\bigcirc\!\!\!\bullet)}$ corresponds to a unique process $(A,\bigcirc)\to(B,\bigcirc\!\!\!\bullet)$:

$$\sum_{x\in K(\bigcirc)}\ \sum_{y\in K(\bigcirc\!\!\!\bullet)} (A,\bigcirc)\text{---}\boxed{x}\ F_x^y\ \boxed{y}\text{---}(B,\bigcirc\!\!\!\bullet) \tag{1.19}$$

Because we have required that for each $n\in\mathbb{N}$ there be a classical system with n classical states, the bijective correspondence above establishes the desired equivalence of categories. □

As a special case, classical systems in CP* categories with $\mathbb{R}^+$ as their semiring of scalars behave exactly like classical probabilistic systems, and the entire toolbox of probability theory becomes available: we will refer to these as **probabilistic theories**, and they will come into play in the context of our very last result. Since the time this work was first written, a stand-alone framework for probabilistic theories capturing the constructions above has been developed by one of the authors, and can be found in Ref. [GS17]. Examples of notable quantum-like theories captured by this framework can be found in Ref. [Gog17].

Non-locality and contextuality. Consider the abstract setup of a Bell-type scenario:

(i) N parties are given devices $B_1,...,B_N$ which might share some global state ρ;
(ii) each device B_j takes an input, the **measurement choice**, freely chosen by party j from some finite set M_j;
(iii) upon receiving input $m_j\in M_j$, the device B_j produces some output o_j in some finite set O_j, the **measurement outcome**;
(iv) no signalling is possible between the devices from before the first input is given to after the last outputs has been produced.

The sheaf-theoretic framework for non-locality and contextuality [AB11] characterises the distribution of joint outputs conditional to joint inputs from the point of view of sheaf theory, showing that non-locality and contextuality are related to the (non-) existence of global sections for a particular presheaf. The framework does not rely on any concrete description of the state ρ or the devices $B_1,...,B_N$, focusing instead on the distributional properties of joint outputs/measurement outcomes $\underline{o_j}:=(o_1,...,o_N)$ conditional to the choice $\underline{M}:=(m_1,...,m_N)$ of joint inputs/measurement choices.

The framework begins by identifying a finite set $\mathcal{X}$ of inputs, which in the Bell-type scenario setup above (the one used in this work) would be $\mathcal{X}=\sqcup_{j=1}^N M_j$. The disjoint union preserves information about which party each measurement is associated to, so we will adopt the notation m_j for generic elements of $\mathcal{X}$, where m is the measurement and j is the party. For each subset $U\subseteq\mathcal{X}$, the family of all potential[8] **joint outcomes** takes the following form:

$$\mathcal{E}[U]:=\prod_{m_j\in U} O_j \tag{1.20}$$

The powerset $\mathcal{P}(\mathcal{X})$ is a poset (hence a poset category) under inclusion $V\subseteq U$ of subsets. We can define a functor $\mathcal{E}:\mathcal{P}(\mathcal{X})^{\text{op}}\to\text{Set}$, i.e. a **presheaf**, by setting:

(i) if $U\in\mathcal{P}(\mathcal{X})$, then we define $\mathcal{E}[U]:=\prod_{m_j\in U}O_j$ as above
(ii) if $V\subseteq U$, then we define $\mathcal{E}[V\subseteq U]:=\text{res}_V^U$ to be the **restriction map** $U\xrightarrow{\text{Set}}V$:

$$\text{res}_V^U = s\mapsto s|_V \tag{1.21}$$

[8]Not all subsets of measurements need be compatible in each concrete scenario: see below for the definition of measurement contexts.

A **section** s **over** U is a U-indexed family of outcomes in the following form:

$$s = \{(m_j, s(m_j)) \mid m_j \in U\} \in \prod_{m_j \in U} O_j \tag{1.22}$$

The restriction map then sends a section s over U to its restriction over V:

$$s|_V = \{(m_j, s(m_j)) \mid m_j \in V\} \in \prod_{m_j \in V} O_j \tag{1.23}$$

The definition of the set of possible joint inputs requires further consideration: it is a fundamental feature of quantum mechanics that not all measurements on a system are compatible, and we should not expect different measurement choices in each M_j to have a consistent assignment of outputs. Instead, the framework requires us to specify a set $\mathcal{M}$ of **measurement contexts**, subsets $C \subseteq \mathcal{X}$ of measurements which are mutually compatible (and therefore have a well-defined notion of joint outcome). Even though more general setups are allowed, we will assume that our measurement contexts all take the form $C = \{m_1, ..., m_N\}$ for $m_j \in M_j$, which we will denote by $\underline{m}$: each party chooses exactly one input for their device, but we allow the possibility that not all combinations of inputs might be allowed/interesting. The only requirement is that $\cup_{C \in \mathcal{M}} C = \mathcal{X}$, i.e. that $\mathcal{M}$ be a **global cover** of $\mathcal{X}$ (each measurement choice for each player appears in at least one measurement context), which we assume to be endowed with the discrete topology. One can also define the **local covers** for any $U \subseteq \mathcal{X}$ as the families $(U_i)_{i \in I}$ such that $\cup_{i \in I} U_i = U$.

The choice of the discrete topology on $\mathcal{X}$ makes $\mathcal{P}(\mathcal{X})$ its locale of open subsets, and one can define a notion of **sheaf** on it. Because it is defined in terms of sections[9], the presheaf $\mathcal{E}$ is in fact a sheaf on the locale $\mathcal{P}(\mathcal{X})$, and we shall refer to it as the **sheaf of events**. The measurement cover and the sheaf of events are the two ingredients required to define a **measurement scenario** $(\mathcal{E}, \mathcal{M})$: the former gives the compatible joint measurement choices, while the latter gives the joint measurement outcomes conditional on all possible measurement choices.

The next step in the framework sees the introduction of generalised notions of probabilities and distributions. In quantum mechanics, probabilities can be seen as taking values in the commutative semiring $R = (\mathbb{R}^+, +, 0, \cdot, 1)$ of the non-negative reals (in fact they fall within the interval $[0, 1]$, a consequence in the semiring R of the normalisation condition requiring that probabilities add up to 1). In other circumstances, one may be interested in the **possibilities** associated with events, living in the commutative semiring $\mathbb{B} = (\{0, 1\}, \vee, 0, \wedge, 1)$ of the booleans. In the sheaf-theoretic treatment of contextuality, one works with an arbitrary commutative semiring $R = (|R|, +, 0, \cdot, 1)$.

Given a set U, an R**-distribution** on U is a function $d : U \to R$ which has finite **support** $\operatorname{supp} d := \{s \in U \mid d(s) \neq 0\}$ and such that $\sum_{s \in \operatorname{supp} d} d(s) = 1$. One can then define a functor $\mathcal{D}_R : \mathrm{Set} \to \mathrm{Set}$ as follows:

(i) for any set U, define $\mathcal{D}_R[U]$ to be the set of R-distributions of U

(ii) for any function $f : U \to V$, define $\mathcal{D}_R[f] := d \mapsto \left[t \mapsto \sum_{f(s)=t} d(s)\right]$.

Composing this functor with the sheaf of events yields the **presheaf of distributions** $\mathcal{D}_R\mathcal{E} : \mathcal{P}(\mathcal{X})^{\mathrm{op}} \to \mathrm{Set}$, which captures the structure of R-distributions on joint measurement outcomes under marginalisation. The presheaf sends each set U of measurements (the

[9]Compatibility of local sections amounts to compatibility over the intersection of the domains, and hence compatible local sections can always be glued together by taking their union as relations.

objects of the presheaf category $\mathcal{P}(\mathcal{X})$) to the set $\mathcal{D}_R\mathcal{E}[U]$ of **R-distributions on U-sections**, and sends any inclusion $V \subseteq U$ (the morphisms of the presheaf category $\mathcal{P}(\mathcal{X})$) to the corresponding marginalisation of distributions:

$$\mathcal{D}_R\mathcal{E}[V \subseteq U] = d \mapsto d|_V := \left[t \mapsto \sum_{s|_V = t} d(s) \right] \tag{1.24}$$

We will refer to $d|_V$ as the **marginal** of d.

In quantum mechanics, if C is a set of compatible measurements on some state $|\psi\rangle$, then there is a probability distribution $d \in \mathcal{D}_{\mathbb{R}^+}\mathcal{E}[C]$ on the joint outcomes of the measurements, and the typical contextuality argument involves showing that the probability distributions on different contexts cannot be obtained, in a no-signalling scenario, as marginals of some non-contextual hidden variable. In the sheaf-theoretic framework, a **(no-signalling) empirical model** is defined to be a compatible family of distributions $(\zeta_C)_{C \in \mathcal{M}}$ for the global cover $\mathcal{M}$ of measurement contexts; the usual no-signalling property is shown in [AB11] to be a special case of the compatibility condition. In other literature (usually treating probabilistic models), empirical models for Bell-type scenarios are usually given explicitly as conditional (probability) distributions, in a format akin to the following:

$$\zeta_{\underline{m}}(\underline{o}) := \mathbb{P}\left[\underline{o} \,\middle|\, \underline{m} \right] \tag{1.25}$$

where $\underline{m} = (m_1, ..., m_N) \in \mathcal{M}$ are the measurement contexts used by the scenario and $\underline{o} \in \prod_j O_j$ are the joint outcomes. This is the format we will use in the last section of this work. In the probabilistic case, empirical models for a fixed scenario form a polytope. However, this need not be the same as the no-signalling polytope which usually studied in quantum information theory, because the set of measurement contexts need not include all possible combinations of all possible measurements for each party (i.e. it need not be the case that $\mathcal{M} = \prod_j M_j$, although it is the case that $\mathcal{M} \subseteq \prod_j M_j$).

A **global section** for an empirical model[10] $(\zeta_C)_{C \in \mathcal{M}}$ is a distribution $d \in \mathcal{D}_R\mathcal{E}[\mathcal{X}]$ over the joint outcomes of all measurements which marginalises to the distributions specified by the empirical model:

$$d|_C = \zeta_C \text{ for all } C \in \mathcal{M} \tag{1.26}$$

The fundamental observation behind the sheaf-theoretic framework is that the existence of a global section for an empirical model is equivalent to the existence of a **non-contextual hidden variable model** (also known as a **local hidden variable model**). Concretely, the existence of a global section d means that there is a finite set Λ, an R-distribution $q(\lambda) : \Lambda \to R$ and a family of functions $f_j^\lambda : M_j \to O_j$ such that:

$$\zeta_{\underline{m}}(\underline{o}) = \sum_{\lambda \in \Lambda} q(\lambda) \prod_j \delta_{f_j^\lambda(m_j) = o_j} \tag{1.27}$$

We will say that an empirical model $(\zeta_C)_{C \in \mathcal{M}}$ is **contextual** (or **non-local**) if it doesn't admit a global section.

[10] From now on, no-signalling is implicitly assumed.

Strong contextuality. Contextuality of probabilistic models is interesting in itself, but more refined notions can be obtained by relating $\mathbb{R}^+$ to two other semirings: the reals, modelling signed probabilities, and the booleans, modelling possibilities. Observe that the construction $\mathcal{D}_R$ is functorial in R, so that for any morphism of semirings $r : R \to R'$ we can define the following:

$$\mathcal{D}_r[U] = [d : U \to R] \mapsto \left[r \circ d : U \to R'\right] \tag{1.28}$$

In particular, there is an injective morphism of semirings $i^+ : \mathbb{R}^+ \hookrightarrow R$ sending $x \in \mathbb{R}^+$ to $+x \in \mathbb{R}$, as well as a surjective morphism of semirings $p : \mathbb{R}^+ \to \mathbb{B}$ sending $0 \mapsto 0$ and $x \neq 0 \mapsto 1$ (the latter is well defined for all positive semirings, not just for $\mathbb{R}^+$).

If $(\zeta_C)_{C\in\mathcal{M}}$ is a probabilistic empirical model, i.e. one in the semiring $\mathbb{R}^+$, then $(\zeta_C)_{C\in\mathcal{M}}$ can be seen as an empirical model $(i^+ \circ \zeta_C)_{C\in\mathcal{M}}$ in the semiring $\mathbb{R}$: regardless of whether $(\zeta_C)_{C\in\mathcal{M}}$ was contextual or not over $\mathbb{R}^+$, it can be shown [AB11] that over the reals it always admits a global section. On the other hand, any probabilistic empirical model $(\zeta_C)_{C\in\mathcal{M}}$ can be assigned a corresponding possibilistic empirical model $(p \circ \zeta_C)_{C\in\mathcal{M}}$ in the semiring $\mathbb{B}$ of the booleans (and each boolean function $p \circ \zeta_C$ can equivalently be seen as the characteristic function of the subset $\operatorname{supp} \zeta_C \subseteq \mathcal{E}[C]$).

Note that contextuality is a contravariant property with respect to change of semiring: if $(\zeta_C)_{C\in\mathcal{M}}$ is an empirical model in a semiring R and $r : R \to R'$ is a morphism of semiring, then contextuality of $(r \circ \zeta_C)_{C\in\mathcal{M}}$ implies contextuality of $(\zeta_C)_{C\in\mathcal{M}}$ (because a global section d of the latter is mapped to a global section $r \circ d$ of the former). We will say that a probabilistic empirical model $(\zeta_C)_{C\in\mathcal{M}}$ is **possibilistically contextual** if the corresponding possibilistic model $(p \circ \zeta_C)_{C\in\mathcal{M}}$ is contextual (as opposed to **probabilistically contextual**, which we use to say that $(\zeta_C)_{C\in\mathcal{M}}$ is contextual over $\mathbb{R}^+$). Because of contravariance, possibilistic contextuality implies probabilistic contextuality, but the opposite is not true: the Bell model given in [AB11] is probabilistically contextual but not possibilistically contextual.

Seeing distributions $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}[U]$ as indicator functions of the subsets $\operatorname{supp} d \subseteq \mathcal{E}[U]$ endows them with a partial order:

$$d' \preceq d \text{ if and only if } \operatorname{supp} d' \subseteq \operatorname{supp} d \tag{1.29}$$

The existence of a global section $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}[U]$ for a possibilistic empirical model $(\zeta_C)_{C\in\mathcal{M}}$ implies that:

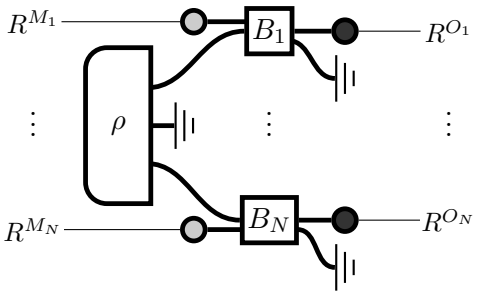
$$d|_C \preceq \zeta_C \text{ for all } C \in \mathcal{M} \tag{1.30}$$

We say that a possibilistic empirical model $(\zeta_C)_{C\in\mathcal{M}}$ is **strongly contextual** if there is no distribution $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}[\mathcal{X}]$ such that Equation 1.30 holds. In particular, the GHZ model given in [AB11], corresponding to Mermin's original non-locality argument, is strongly contextual. Because of Equation 1.29, strong contextuality implies contextuality, but the opposite is not true: the possibilistic Hardy model give in [AB11] is contextual, but not strongly contextual. We will say that a probabilistic empirical model is strongly contextual if the associated possibilistic empirical model is strongly contextual, yielding the following strict hierarchy of notions of contextuality for probabilistic empirical models:

$$\text{probabilistically contextual} \Leftarrow \text{possibilistically contextual} \Leftarrow \text{strongly contextual} \tag{1.31}$$

Empirical models within the CP* construction. The relevance of the sheaf-theoretic framework to this work stems from the following result: in any R-probabilistic CP* category, all Bell-type scenarios give rise to an empirical model.

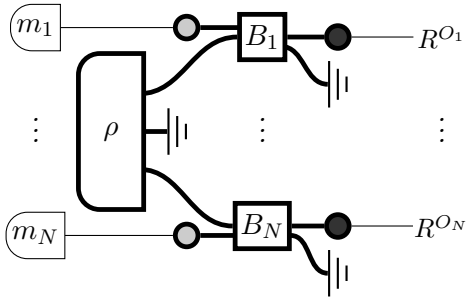
Definition 1.6. Consider an R-probabilistic CP* category. A **Bell-type scenario** is a process $\Phi : (A_1, \circ_1) \otimes ... \otimes (A_N, \circ_N) \to (B_1, \bullet_1) \otimes ... \otimes (B_N, \bullet_N)$, where all $(A_j, \circ_j)$ and all $(B_j, \bullet_j)$ are classical systems, which takes the following form, for some normalised state ρ and some normalised processes $B_1, ..., B_N$:



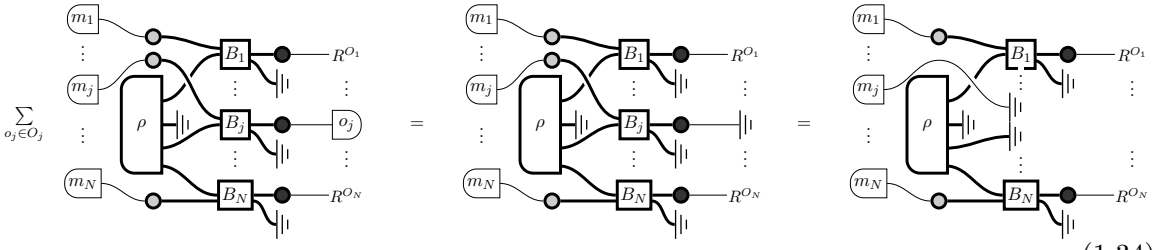
(1.32)

Theorem 1.7. *Consider a Bell-type scenario Φ in the form given by Definition 1.6. Let M_j be the finite set of classical states for $\circ_j$, and O_j be the finite set of classical states for $\bullet_j$. Then the process Φ gives rise to a no-signalling empirical model $(\zeta_{\underline{m}})_{\underline{m} \in \mathcal{M}}$ as follows, for any cover $\mathcal{M}$:*

$$\zeta_{\underline{m}} \quad := \quad \in \quad \mathcal{D}_R\mathcal{E}[\{m_1, ..., m_N\}] \tag{1.33}$$



Proof. We need to show that the states in Equation 1.33 (indexed by the measurement contexts $\underline{m} \in \mathcal{M}$) satisfy no-signalling and are normalised (i.e. are R-distributions). To do so, we (i) marginalise over party j, (ii) use the fact that the discarding map on the classical systems $(B_j, \bullet_j)$ can be written as $\dashv\!\Vdash_{(B_j, \bullet_j)} = \sum_{o_j} \langle o_j|$, and (iii) use the fact that the measurement on $\bullet_j$ and the process B_j are both normalised to show that the resulting state is independent of m_j:



(1.34)

Marginalising over all outputs leaves us with $\dashv\!\Vdash \circ \rho$, which equals 1 (independently of the measurement context $\underline{m}$) because ρ is normalised. Hence the state $(\zeta_{\underline{m}})_{\underline{m} \in \mathcal{M}}$ is also an R-distribution as desired, completing our proof. □

2. Mermin's Original Argument

The parity argument. In the original [Mer90], Mermin considers a 3-qubit GHZ state in the computational basis, the basis of eigenstates for the single-qubit Pauli Z observable, together with the following four joint measurements[11]:

(a) the GHZ state is measured in the observable $X_1 \otimes X_2 \otimes X_3$;
(b) the GHZ state is measured in the observable $Y_1 \otimes Y_2 \otimes X_3$;
(c) the GHZ state is measured in the observable $Y_1 \otimes X_2 \otimes Y_3$;
(d) the GHZ state is measured in the observable $X_1 \otimes Y_2 \otimes Y_3$.

We will denote the eigenstates of the Pauli Z observable by $|z_0\rangle, |z_1\rangle$, the eigenstates of the Pauli X observable by $|\pm\rangle := \frac{1}{\sqrt{2}}(|z_0\rangle \pm |z_1\rangle)$ and the eigenstates of the Pauli Y observable by $|\pm i\rangle := \frac{1}{\sqrt{2}}(|z_0\rangle \pm i|z_1\rangle)$. Mermin's argument is a parity argument, where measurement outcomes are valued in the abelian group $\mathbb{Z}_2 = \{0, 1\}$ according to the following bijections:

(i) for the X observable, $|+\rangle \mapsto 0$ and $|-\rangle \mapsto 1$
(ii) for the Y observable, $|+i\rangle \mapsto 0$ and $|-i\rangle \mapsto 1$

The argument then proceeds as follows. While the joint measurement outcomes are probabilistic, the $\mathbb{Z}_2$ sum of the three outcomes turns out to be deterministic, yielding the following system of equations ($\oplus$ here denotes the sum in $\mathbb{Z}_2$):

$$\begin{cases} X_1 \oplus X_2 \oplus X_3 &= 0 \\ Y_1 \oplus Y_2 \oplus X_3 &= 1 \\ Y_1 \oplus X_2 \oplus Y_3 &= 1 \\ X_1 \oplus Y_2 \oplus Y_3 &= 1 \end{cases} \tag{2.1}$$

If there was a non-contextual assignment of outcomes for all measurements (X_1, X_2, X_3, Y_1, Y_2 and Y_3), i.e. if there existed a non-contextual hidden variable model, then System 2.1 would have a solution in $\mathbb{Z}_2$, and in particular it would have to be consistent. However, the sum of the left hand sides yields 0 in $\mathbb{Z}_2$:

$$2X_1 \oplus 2X_2 \oplus ... \oplus 2Y_3 = 0X_1 \oplus ... \oplus 0Y_3 = 0 \tag{2.2}$$

while the sum of the right hand sides yields $0 \oplus 1 \oplus 1 \oplus 1 = 3 = 1$ in $\mathbb{Z}_2$. This shows the system to be inconsistent. Equivalently, one could observe that the sum of the LHS from Equation 2.2 can be written as $2(Y_1 \oplus Y_2 \oplus Y_3)$, and that inconsistency of the system is witnessed by the fact that the equation $2y = 1$ has no solution in $\mathbb{Z}_2$.

The first point of view, where contextuality is witnessed by an inconsistent system where each equation individually admits a solution, is behind the generalisation of Mermin's argument to All-vs-Nothing arguments, presented in [ABK+15]. The second point of view, where contextuality is witnessed by the single unsatisfiable equation $2y = 1$, will inspire the generalisation presented in this work.

[11] Where X_j and Y_j are the single-qubit Pauli X and Y observables on qubit j, for $j = 1, 2, 3$.

The role of phases. To understand the role played by the equation $2y = 1$ in the original Mermin argument, we need to take a step back. First of all, we observe that the Pauli Y measurement can be equivalently obtained as a Pauli X measurement preceded by an appropriate unitary. A single-qubit **phase gate**, in the computational basis (the Pauli Z observable), is a unitary transformation in the following form:

$$P_\alpha := \begin{pmatrix} 1 & 0 \\ 0 & e^{i\alpha} \end{pmatrix} \tag{2.3}$$

where we eliminated global phases by setting the first diagonal element to 1. Measuring in the single-qubit Y observable is equivalent to first applying the single-qubit phase gate $P_{\frac{\pi}{2}}$, and then measuring in the Pauli X observable.

Because they pairwise commute, phase gates come with a natural abelian group structure given by composition, resulting in an isomorphism $\alpha \mapsto P_\alpha$ between them and the abelian group[12] $\mathbb{R}/(2\pi\mathbb{Z})$. Of all the phase gates, P_0 (the identity element of the group) and P_π stand out because of their well-defined action on the (unnormalised) eigenstates of the Pauli X observable:

$$\begin{aligned} P_0 &= |\pm\rangle \mapsto |\pm\rangle \\ P_\pi &= |\pm\rangle \mapsto |\mp\rangle \end{aligned} \tag{2.4}$$

If we see $|\pm\rangle$ as the subgroup[13] $\{0, \pi\} < \mathbb{R}/(2\pi\mathbb{Z})$, then Equation 2.4 looks a lot like the regular action of $\{0, \pi\}$ on itself. This is not a coincidence. Each phase gate P_α can be (faithfully) associated the unique **phase state** $|\alpha\rangle := |z_0\rangle + e^{i\alpha}|z_1\rangle$ obtained from its diagonal, and these phase states can be abstractly characterised in terms of the Pauli Z observable, with no reference to the phase gates they came from (*cf.* section 4). The phase states inherit the abelian group structure of the phase gates, and their regular action coincides with the action of the group of phase gates on them. In particular, the phase gates P_0 and P_π have orthogonal eigenstates of the Pauli X observable as their associated phase states $|0\rangle$ and $|\pi\rangle$, which coincide with $\sqrt{2}|+\rangle$ and $\sqrt{2}|-\rangle$ respectively: this endows the outcomes of Pauli X measurements with the natural $\mathbb{Z}_2$ abelian group structure arising[14] from the inclusion $\{0, \pi\} < \mathbb{R}/(2\pi\mathbb{Z})$. We will henceforth refer to the group of phase states as the **group of Z-phase states**, and to the subgroup $\{0, \pi\}$ as the **subgroup of X-classical states**; the latter will also be used to label the corresponding measurement outcomes.

In order to pave the way to our generalisation, we now proceed to show how Mermin's original argument can be re-constructed from the following statement:

> *the equation $2y = \pi$ has no solution in the subgroup $\{0, \pi\}$ of X-classical states, but a solution[15] $y = \frac{\pi}{2}$ can be found in the larger group $\mathbb{R}/(2\pi\mathbb{Z})$ of Z-phase states.*

We begin by observing that tripartite qubit GHZ state used in Mermin's argument has a special property when it comes to the application of phase gates followed by measurements in the Pauli X observable.

[12]The abelian group $\mathbb{R}/(2\pi\mathbb{Z})$ is isomorphic to the circle group S^1. We prefer the former because of its additive notation, as opposed to the traditionally multiplicative notation of the latter (which is a subgroup of the non-zero multiplicative complex numbers $\mathbb{C}^\times$).

[13]Corresponding to $\{\pm 1\} < S^1$ in the circle group.

[14]Natural because there is a unique isomorphism $\mathbb{Z}_2 \cong \{0, \pi\}$.

[15]Corresponding to $y = e^{i\frac{\pi}{2}} = +i$ in the circle group S^1.

Lemma 2.1 [CDKW12]**.** *If $\alpha_j \in \mathbb{R}/(2\pi\mathbb{Z})$, denote by $X_j^{\alpha_j}$ the measurement outcome on qubit j obtained by first applying phase gate P_{α_j}, and then measuring in the Pauli X observable. If $\alpha_1 \oplus \alpha_2 \oplus \alpha_3 = 0$ or π (mod 2π), then $X_1^{\alpha_1} \oplus X_2^{\alpha_2} \oplus X_3^{\alpha_3} = 0$ or π (mod 2π) respectively.*

Now consider System 2.1 again, with values on the the RHS now obtained by applying Lemma 2.1 to $X_j := X_j^0$ and $Y_j := X_j^{\frac{\pi}{2}}$ (and valued in $\{0, \pi\}$ instead of the original $\mathbb{Z}_2$):

$$\begin{cases} X_1^0 \oplus X_2^0 \oplus X_3^0 & = 0, \text{ the control} \\ X_1^{\frac{\pi}{2}} \oplus X_2^{\frac{\pi}{2}} \oplus X_3^0 & = \pi, \text{ the first variation} \\ X_1^{\frac{\pi}{2}} \oplus X_2^0 \oplus X_3^{\frac{\pi}{2}} & = \pi, \text{ the second variation} \\ X_1^0 \oplus X_2^{\frac{\pi}{2}} \oplus X_3^{\frac{\pi}{2}} & = \pi, \text{ the third variation} \end{cases} \tag{2.5}$$

There are two complementary parts to the Mermin non-locality argument: (i) System 2.5 above must be inconsistent, to rule out the existence of a non-contextual hidden variable model, and (ii) joint measurements yielding the individual equations must be possible (in quantum theory). For the first part, inconsistency of the system is witnessed by the fact that the equation $2y = \pi$ has no solution in the subgroup of X-classical states. For the second part, notice that only measurements in the Y observable contribute to the sum for each equation, as measurements in the X observable are associated with the group unit 0 of the group of Z-phase states. As a consequence, the existence of measurements implementing each individual equation reduces to the existence of a Z-phase state $|y\rangle$ satisfying equation $2y = \pi$: the Y observable is chosen exactly because $y = \pi/2$ gives one such Z-phase state.

The following steps summarise the skeleton of the argument, and open the way to our generalisation:

1. consider a non-degenerate observable, call it Z, on an arbitrary quantum system;
2. consider another non-degenerate observable, call it X, such that the X-classical states are a subgroup (call it K) of the abelian group of Z-phase states (call it P);
3. consider an equation in the following form, generalising $2y = \pi$:
$$n_1 y_1 \oplus ... \oplus n_M y_M = a \tag{2.6}$$
(here $a \in K$, $n_1, ..., n_M$ are integers[16], and $\oplus$ is the group addition in P);
4. construct an appropriate system of equations, generalising System 2.5, with inconsistency witnessed by non-existence of solutions for Equation 2.6 in K, and consistency of the individual equations witnessed by the existence of solutions in P;
5. a measurement scenario can be implemented if and only if a solution exists in P;
6. the measurement scenario is contextual if and only if no solutions exist in K.

To give a first example of how such an appropriate system of equations might be constructed, we consider the simple generalisation of the argument from a 3-partite to an N-partite GHZ state, for appropriate values of $N \geq 2$. Our requirements are as follows:

(i) we want the phases in the control to sum to 0, and hence we will take them all to be 0 (i.e. measurements in the X observable), just as in the original argument;
(ii) we also want the phases in each variation to sum to π, and hence we will take two measurements in each variation to be with phase $\pi/2$ (i.e. measurements in the Y observable), and all the other ones to be with phase 0;
(iii) we want an odd number V of variations, so that the RHSs will sum to $0 \oplus V\pi = \pi$;

[16] This is a general equation in abelian groups, seen equivalently as $\mathbb{Z}$-modules.

(iv) we want the LHSs to sum to an even multiple of $X_1^{\frac{\pi}{2}} \oplus ... \oplus X_N^{\frac{\pi}{2}}$;

An appropriate choice is given by the following system of equations, where $V := N$ and all variations are cyclic permutations of the first one:

$$\begin{cases} X_1^0 \oplus X_2^0 \oplus X_3^0 \oplus \quad ... \quad \oplus X_{N-1}^0 \oplus X_N^0 & = 0, \text{ the control} \\ X_1^{\frac{\pi}{2}} \oplus X_2^{\frac{\pi}{2}} \oplus X_3^0 \oplus \quad ... \quad \oplus X_{N-1}^0 \oplus X_N^0 & = \pi, \text{ the } 1^{st} \text{ variation} \\ X_1^{\frac{\pi}{2}} \oplus X_2^0 \oplus \ ... \ \oplus X_{N-2}^0 \oplus X_{N-1}^0 \oplus X_N^{\frac{\pi}{2}} & = \pi, \text{ the } 2^{nd} \text{ variation} \\ \qquad\qquad\qquad \vdots & \\ X_1^0 \oplus X_2^{\frac{\pi}{2}} \oplus X_3^{\frac{\pi}{2}} \oplus \quad X_4^0 \quad \oplus \quad ... \quad \oplus X_N^0 & = \pi, \text{ the } N^{th} \text{ variation} \end{cases} \tag{2.7}$$

As long as $N = 1 \pmod k$, where $k = 2$ is the exponent[17] of K, the RHSs will sum to π in K. Having chosen our variations by cyclic permutation also makes for the desired sum of the LHSs, since each $X_j^{\pi/2}$ will be counted exactly twice:

$$\underset{\text{control}}{\left(X_1^0 \oplus ... \oplus X_N^0\right)} \quad \oplus \quad \underset{X_j^{\frac{\pi}{2}}\text{s from the variations}}{2 \cdot \left(X_1^{\frac{\pi}{2}} \oplus ... \oplus X_N^{\frac{\pi}{2}}\right)} \quad \oplus \quad \underset{X_j^0\text{s from the variations}}{(N-2) \cdot \left(X_1^0 \oplus ... \oplus X_N^0\right)}$$

Writing x for $X_1^0 \oplus ... \oplus X_N^0$ and y for $X_1^{\frac{\pi}{2}} \oplus ... \oplus X_N^{\frac{\pi}{2}}$, the sum above can be rearranged to take the form $(N-1)x \oplus 2y$, which is equal to $2y$ in K (since $(N-1) = 0 \pmod k$)[18]. Hence summing all the LHSs and RHSs leaves us with the equation $2y = \pi$, which we know to be unsatisfiable in K.

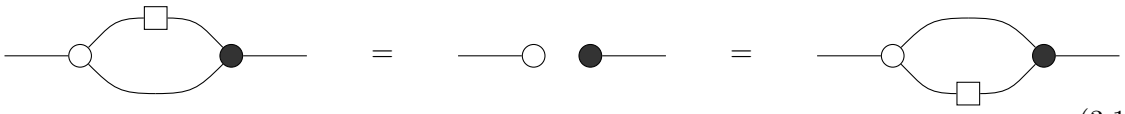
3. Strong Complementarity

Mermin's parity argument is fundamentally group-theoretic, and it depends almost entirely on the special relationship between the Pauli Z and Pauli X observables. Fixing the eigenstates of the Pauli Z observable as the computational basis, the requirement that the X-classical states are Z-phase states is satisfied by the Pauli X observable, but also by the Pauli Y: in fact, the Z-phase states are exactly the **unbiased states** for the Pauli Z observables, the states lying on the equator of the Bloch sphere, and hence any observable **complementary**, or **mutually unbiased**, to Pauli Z would do the trick; because their eigenstates lie on the equator of the Bloch sphere, we will refer to observables complementary to Pauli Z as **equatorial observables**. Definition 3.1 gives an algebraic/diagrammatic presentation of complementarity using Hopf's Law, and Lemma 3.3 shows that observables which are complementary observables under this definition are always mutually unbiased. A more general result relating complementarity and mutual unbias in †-SMCs will be given by Theorem 4.10 in the next section.

[17] The smallest positive integer such that $kx = 0$ for all $x \in K$.

[18] In this specific case, it is also true that $2 = 0 \pmod k$, but this is not key to the argument.

Definition 3.1. Two †-qSFAs ○ and ● on the same object $\mathcal{H}$ of a †-SMC are said to be **complementary** if they satisfy the following **Hopf's Law**:



(3.1)

where the **antipode** $-\square- : \mathcal{H} \to \mathcal{H}$ is the unitary defined as follows, which we require to be self-adjoint (or equivalently self-inverse) as part of the definition of complementarity:

(3.2)

Definition 3.2. Let ◎ be a †-qSFA in a dagger compact category. Then a state u is a **◎-unbiased** state if the following holds:

(3.3)

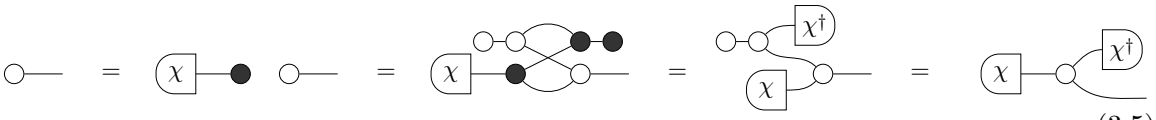
Equation 3.3 can be unfolded into the following more general definition, which holds in an arbitrary †-SMCs:

(3.4)

In fdHilb, the ◎-unbiased states are those which, once normalised, yield the uniform distribution upon measurement in the ◎ observable.

Lemma 3.3. *Consider a complementary pair of †-qSFAs ○ and ●. The ●-classical states are ○-unbiased, and the ○-classical states are ●-unbiased.*

Proof. We prove that a ●-classical state χ is ○-unbiased:



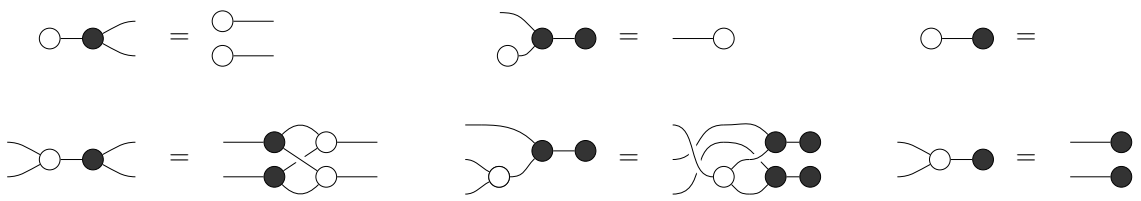
(3.5)

The first equality is by ●-classicality (delete condition), the second equality is Hopf's law (together with the self-adjoint requirement for the antipode), the third equality is again by ●-classicality (copy condition for the bottom state, transpose condition for the top state), the last equality is by Frobenius law and unit law for ○. The proof for ○-classical states is the same, with colours swapped. □

Complementarity is not sufficient for Mermin's argument: Lemma 2.1 only holds if we measure the GHZ state in the Pauli X observable, not in any other equatorial observable. The algebraic relationship between the Pauli X, Y and Z observables is vividly captured by the ZX calculus [CD11]: there, the special property relating the Pauli Z and X observables is axiomatised under the name of **strong complementarity**, to distinguish it from the complementarity of Pauli Z and any other equatorial observable (such as Pauli Y). Strong complementarity is behind the proof of Lemma 2.1, which lies at core of the fully diagrammatic treatment of Mermin's original argument appearing in [CDKW12].

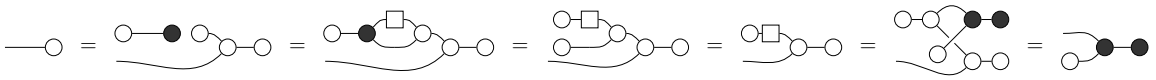
Definition 3.4 gives an algebraic/diagrammatic presentation of strong complementarity, while Theorem 3.6 provides an exact correspondence in finite-dimensional Hilbert spaces between complementarity and the representation theory of finite abelian groups. A more general characterisation of strong complementarity in †-SMCs will be given by Theorem 4.10 in the next section.

Definition 3.4. Two †-qSFAs ○ and ● on the same object $\mathcal{H}$ of a †-SMC are said to be **strongly complementary** if they are complementary and furthermore satisfy the following equations[19]:



(3.6)

Remark 3.5. *Technically speaking, the central equations of both rows are not necessary: indeed, they do not usually feature in the literature. The central equation on the top row of 3.6 is in fact a consequence of Hopf's law and the other two equations of the top row:*



(3.7)

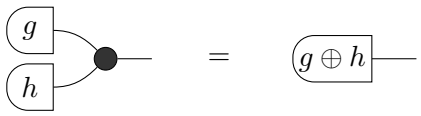
Similarly, the central equation of the top row with colours swapped is a consequence of Hopf's law, the rightmost equation of the top row, and the rightmost equation of the bottom row. The central equation of the bottom row can also be obtained using Hopf's law, self-adjointness of the antipode, and the other equations.

Strong complementarity means that the eigenstates of the Pauli X observable are very specific equatorial states, given by the two multiplicative characters $\chi : \mathbb{Z}_2 \to S^1$ of the abelian group $\mathbb{Z}_2$:

$$|\chi\rangle := \chi(0)|z_0\rangle + \chi(1)|z_1\rangle = \begin{cases} \sqrt{2}|+\rangle & \text{if } \chi \text{ is the trivial character } \chi(j) := +1 \\ \sqrt{2}|-\rangle & \text{if } \chi \text{ is the alternating character } \chi(j) := (-1)^j \end{cases} \tag{3.8}$$

This group-theoretic characterisation of strong complementarity fully generalises to arbitrary finite-dimensional Hilbert spaces and finite abelian groups.

Theorem 3.6 [CDKW12, Kis12]**.**
Let ○ and ● be a †-SCFA and a †-qSCFA on the same finite-dimensional Hilbert space $\mathcal{H}$. Then ○ and ● are strongly complementary iff there exists an abelian group $(G, \oplus, 0)$ such that (,) endows the set of ○-classical states with the abelian group structure of G, i.e. iff we can label the ○-classical states as $(|g\rangle)_{g\in G}$ in a way such that:



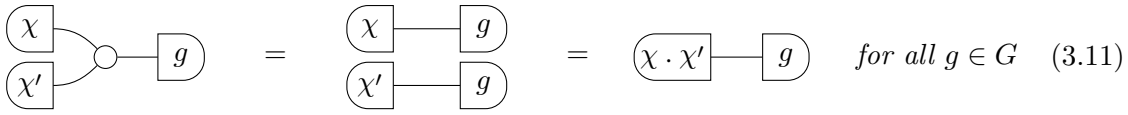
(3.9)

[19] The empty diagram on the RHS of the top right equation is the scalar 1.

If ○ and ● are strongly complementary, then the ●-classical are labelled by the multiplicative characters $\chi : G \to S^1$ of the abelian group G, and take the following form:

$$|\chi\rangle := \sum_{g \in G} \chi(g)|g\rangle \tag{3.10}$$

Furthermore, $(\succ, \circ\!-)$ turns the set of X-classical states into the finite abelian group $(G^\wedge, \cdot, \mathbb{1})$ of multiplicative characters[20] of G:



for all $g \in G$ (3.11)

The importance of strong complementarity for quantum algorithms [Vic12, GK17] mostly lies in the following observation: if the ○-classical states are taken to form the computational basis, then the ●-classical states form the (unnormalised) Fourier basis, and measuring in the ● observable amounts to performing the quantum Fourier transform. The situation with Mermin-type arguments, however, is different: the relevant facet of strong complementarity will be the special relationship between ●-classical points and ○-phase states, explored in detail in the coming section.

Remark 3.7. *Theorem 3.6 extend to the case where ● is a †-qSFA, i.e. not commutative, as long as the adjective abelian is dropped for the group G (which becomes a generic finite group); it should however be noted that the ●-classical states form a basis if and only if G is abelian (if and only if ● is commutative, i.e. a non-degenerate observable). The group $G^\wedge$ of multiplicative characters of a finite group G is always abelian, and Pontryagin duality[21] necessarily fails when G is not abelian. Theorem 3.6 also extends to the case where ○ is a †-qSCFA, in which case the ○-classical states form an orthogonal basis, rather than an orthonormal one. Most of Theorem 3.6 can be further extended to the case where ○ is a generic †-qSFA: it is still true that $(\succ, \bullet\!-)$ endows the ○-classical states with the structure of a finite group, and that $(\succ, \circ\!-)$ endows the ●-classical states with the structure of a finite group, but Equation 3.10 need not hold, and the group structure given by $(\succ, \circ\!-)$ need not be that of the group of multiplicative characters.*

4. The phase group

If strong complementarity is the fundamental algebraic property at work in Mermin's argument, phase gates and GHZ states are the operational components key to its implementation. Phase gates arise in the context of quantum-to-classical transitions, where they provide a characterisation, in the spirit of groups and symmetries, of how much information is lost by performing a (demolition) measurement in a non-degenerate observable.

[20] Multiplicative characters form an abelian group under pointwise multiplication $(\chi \cdot \chi')(g) := \chi(g) \cdot \chi'(g)$ and with the trivial character $\mathbb{1} := g \mapsto 1$ as group unit. This is known as the **Pontryagin dual** $G^\wedge$ of G.

[21] The existence of a (canonical) isomorphism $(G^\wedge)^\wedge \cong G$.

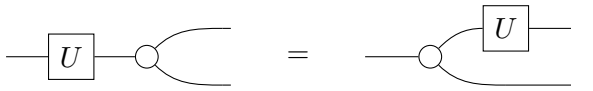
Definition 4.1. Let $\circ$ be a †-qSFA on an object $\mathcal{H}$ of a dagger compact category. Then the **$\circ$-phase gates** are the unitaries $U : \mathcal{H} \to \mathcal{H}$ which are annihilated by the measurement:

(4.1)

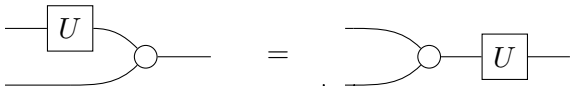
Equation 4.1 can be unfolded into the following equivalent definition, which extends to an arbitrary †-SMC:

(4.2)

Remark 4.2. *A simpler algebraic characterisation of phase gates is given by the following two equations, which are equivalent to Equation 4.2 (because U is assumed to be unitary):*



(4.3)



(4.4)

Both equations will play a pivotal role in this section: Equation 4.3 will features shortly in Lemma 4.5, the result relating phase gates and GHZ states, while Equation 4.4 will feature later on in Theorem 4.6, the result relating phase gates and unbiased states.

From Equation 4.1, it is not hard to see that $\circ$-phase gates form a group: we will refer to this as the **$\circ$-phase group**, and we will denote it by $P(\circ)$. If $\circ$ is a †-SFA on a finite-dimensional Hilbert space $\mathcal{H}$, associated with a direct sum decomposition $\mathcal{H} = \oplus_j \mathcal{H}_j$, then the phase group $P(\circ)$ is given by the corresponding direct sum of unitary groups, modulo a global phase:

$$P(\circ) = \Big(\oplus_j U(\mathcal{H}_j) \Big) / S^1 \tag{4.5}$$

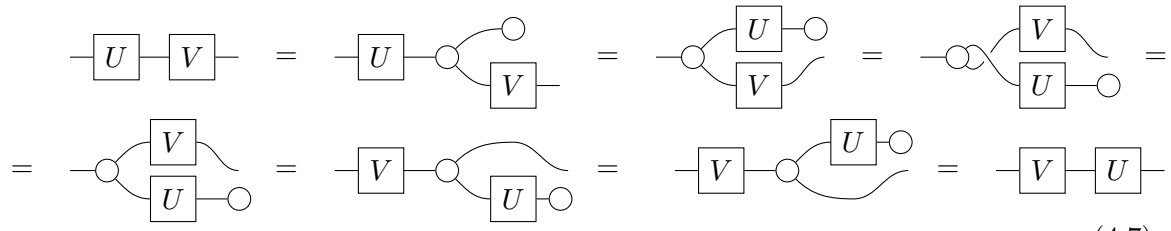
In the special case where $\circ$ is a †-SCFA on $\mathcal{H}$, i.e. when all $\mathcal{H}_j$ subspaces are 1-dimensional, the phase group is abelian, the translation group of a torus:

$$P(\circ) = \Big(\oplus_{j=1}^{\dim \mathcal{H}} U(1) \Big) / S^1 \cong T^{\dim \mathcal{H} - 1} \tag{4.6}$$

The connection between abelian phase groups and commutative Frobenius algebras generalises from fdHilb to arbitrary dagger compact categories. The following result shows that the phase group of a commutative Frobenius algebra is always abelian, while the converse will be proven later on in Corollary 4.9 (conditional to the existence of enough unbiased states)

Lemma 4.3. *Let $\circ$ be a †-qSFA on an object $\mathcal{H}$ of a dagger compact category. If $\circ$ is commutative, then the $\circ$-phase group $P(\circ)$ is abelian.*

Proof.

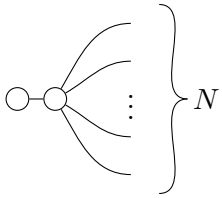


(4.7)

The first equality is by unit law for $\circ$; the second equality is by Equation 4.3; the third equality is some topological manipulation; the fourth equality (top right to bottom left) is by commutativity of $\circ$; the fifth equality is by Equation 4.3; the sixth equality is commutativity of $\circ$; the seventh and last equality is by Equation 4.3, followed by unit law for $\circ$. □

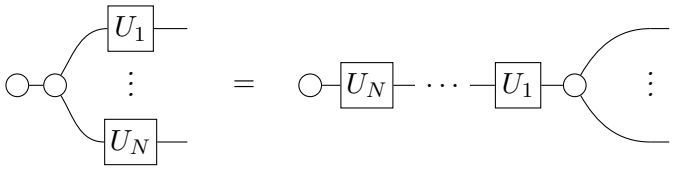
Having defined the phase group and proven Lemma 4.3, we are now in a position to state the first important result of this section. Lemma 4.5 characterises the sates that can be obtained by application of phases gates to a GHZ state: in the context of our generalised Mermin-type arguments, it will play the same role that Lemma 2.1 played in Mermin's original argument.

Definition 4.4. If $\circ$ is a †-qSFA on an object $\mathcal{H}$ of a dagger compact category, the **N-partite $\circ$-GHZ state** is the following state of $\mathcal{H}^{\otimes N}$:



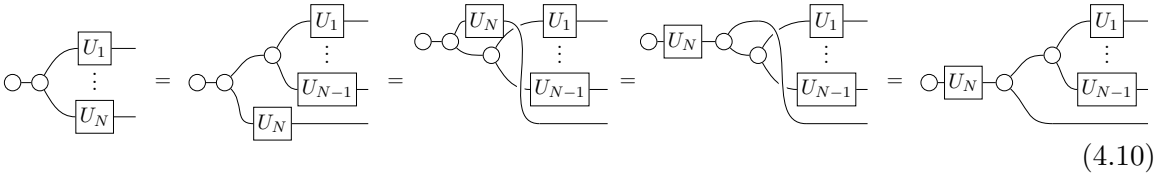
(4.8)

Lemma 4.5. *Let $\circ$ be a †-qSCFA on an object $\mathcal{H}$ of a dagger compact category. Then the state obtained by applying $\circ$-phase gates $U_1, ..., U_N$ to the N-partite $\circ$-GHZ state only depends on the composition $U_1 \cdot ... \cdot U_N$ of the phase gates:*



(4.9)

Proof. Each $\circ$-phase gate is pushed down by using Equation 4.3 and commutativity of $\circ$. Formally, the proof is by induction, with inductive step given by the following equality:



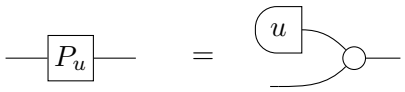
(4.10)

□

We have remarked before that the phase gates in Mermin's original argument are associated to certain phase states, extracted from their diagonalisation, which are also unbiased states for the relevant observable. As the following Theorem 4.6 shows, the connection between $\circ$-phase gates and $\circ$-unbiased states holds true in full generality, and

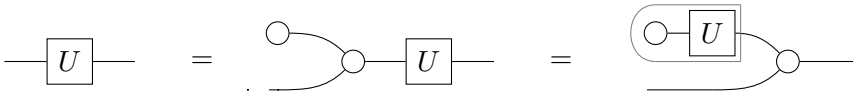
as a consequence we will also refer to ○-unbiased states as **○-phase states**. In the case of fdHilb, the decomposition of a ○-phase gate U given by Equation 4.11 for a †-SCFA ○ is equivalent to saying that U is diagonal in the orthonormal basis $(|x\rangle)_x$ associated with ○, and has diagonal encoded by state $|u\rangle$ as $U_{xx} = \langle x|u\rangle$.

Theorem 4.6. *Let ○ be a †-qSFA on an object $\mathcal{H}$ of a dagger compact category. Then the ○-phase gates are exactly the maps P_u taking the following form for a ○-unbiased state u:*



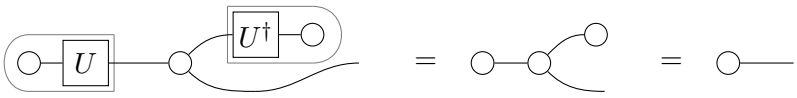
(4.11)

Proof. First we prove that any phase gate U takes the form above, for some ○-unbiased state u. An appropriate state u can then be obtained by unit law for ○:



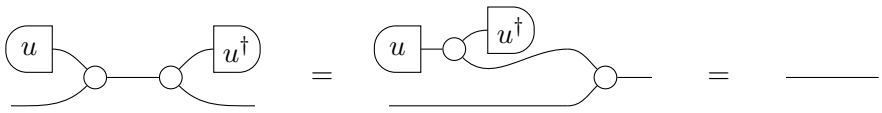
(4.12)

By using Equation 4.2, we can prove that the state we obtained is ○-unbiased:



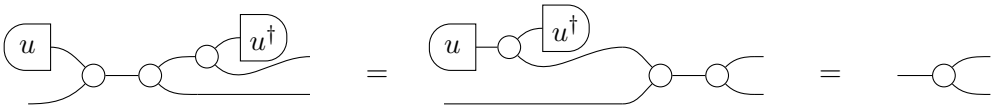
(4.13)

Then we prove that any U in the form above with u a ○-unbiased state is a unitary:



(4.14)

Finally, we prove that any unitary U in the form above with u a ○-unbiased state is a ○-phase gate:



(4.15)

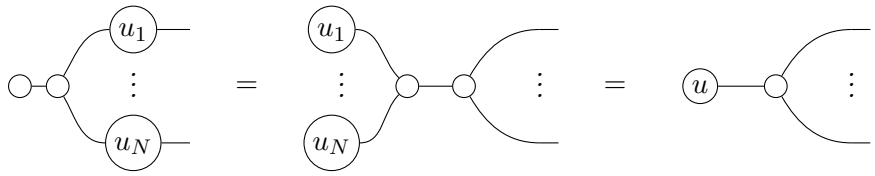
□Because of the correspondence above, we will adopt a uniform notation for phase gates and phase states, known in the literature as **decorated spider** notation [CD11, CK17]:

(4.16)

phase gate P_u phase state u

Corollary 4.7. *Let ○ be a †-qSCFA on an object $\mathcal{H}$ of a dagger compact category. Then the state obtained by applying ○-phase gates $P_{u_1}, ..., P_{u_N}$ to the N-partite ○-GHZ state takes*

the following form in terms of the corresponding ○-phase states $u_1, ..., u_N$:



(4.17)

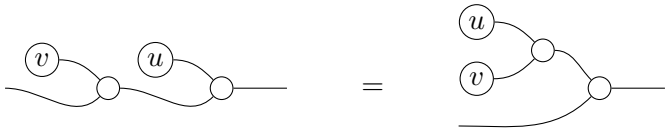
That is, the states that can be obtained by applying ○-phase gates to the N-partite ○-GHZ state are exactly those obtained by comultiplying N-times some ○-unbiased state u (specifically, above we have $u = u_1 \cdot ... \cdot u_N$, and all ○-unbiased states can be obtained this way).

Proof. From Lemma 4.5, by re-writing each ○-phase gate in terms of the corresponding ○-phase state using Theorem 4.6, and then using associativity to group the ○-phase states. □

The group structure of phase gates transfers to unbiased states via the correspondence given by Theorem 4.6. Albeit not surprising, this result plays an important role in our generalisation of Mermin-type arguments, where it connects the operational side of phase gates and GHZ states to the algebraic side of strong complementarity (in the group-theoretic characterisation given by Theorem 3.6 and Theorem 4.10 below).

Lemma 4.8. *Let ○ be a †-qSFA on an object $\mathcal{H}$ of a dagger compact category. Then $(\succ, \circ-)$ endows the set of ○-unbiased states with the structure of $P(○)$.*

Proof. The ○-phase gate corresponding to the ○-unbiased state $\circ-$ is the identity, the unit of $P(○)$, so all we need to show is that composition of phase gates is the same as multiplication under $\succ$ of the corresponding ○-unbiased states:



(4.18)

□

As a bonus, the correspondence between the ○-phase group and the group structure on ○-unbiased states can be used to prove a converse to Lemma 4.3.

Corollary 4.9. *Let ○ be a †-qSFA on an object $\mathcal{H}$ of a dagger compact category, and assume that ○ has* ***enough unbiased states***[22]*. Then ○ is commutative iff $P(○)$ is abelian.*

Proof. We already know from Lemma 4.3 that if ○ is commutative then the ○-phase group $P(○)$ must be abelian. Conversely, if $P(○)$ is abelian then so is the group structure induced by $(\succ, \circ-)$ on the ○-unbiased states: in particular, this means that $\succ$ is commutative whenever it is applied to ○-unbiased states, and the existence of enough unbiased states allows us to conclude that ○ is always commutative. □

With Theorem 4.6 we have proven a general correspondence between phase gates and unbiased states, while with Lemma 4.5 and Corollary 4.7 we have characterised the states that can be obtained by applying phase gates to GHZ states. Phase gates and the GHZ state for the Pauli Z observable are the key operational ingredients for Mermin's original argument. However, just as important is the special algebraic standing of those phase gates derived from the eigenstates of the Pauli X observable (an observable strongly complementary to Pauli Z), as opposed to the phase gates derived from other equatorial states (the eigenstates of observables complementary to Pauli Z).

[22] I.e. that two morphisms $F, G : \mathcal{H} \to \mathcal{K}$ are equal whenever $F \circ u = G \circ u$ for all ○-unbiased states u.

The last result of this section, Theorem 4.10, provides a general characterisation of complementarity and strong complementarity in terms of the relation between classical states of one observable and unbiased states of the other. Together with Theorem 4.6 and 4.7, it will form the basis for the formulation of our generalised Mermin-type arguments in the next section.

Theorem 4.10.
Let ○ and ● be †-qSFAs on an object $\mathcal{H}$ of a †-SMC. The following implications always hold:

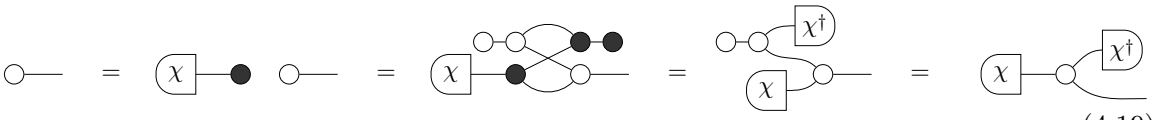
(i) *if ○ and ● are complementary, then the ●-classical states form a subset of the ○-unbiased states;*
(ii) *if ○ and ● are strongly complementary, then the ●-classical states form a subgroup of the ○-unbiased states.*

The converse implications hold if ● has enough classical states:

(i) *if the ●-classical states form a subset of the ○-unbiased states, then ○ and ● are complementary;*
(ii) *if the ●-classical states form a subgroup of the ○-unbiased states, then ○ and ● are strongly complementary.*

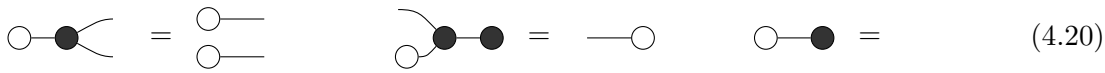
Note that the existence of enough ●-classical states implies the existence of enough ○-unbiased states when the former are a subset/subgroup of the latter.

Proof. Implication (i) is the statement of Lemma 3.3: in its proof, Equation 3.5 is shown that Hopf's law is equivalent to the defining equation of a ○-unbiased state when it applied to ●-classical state.



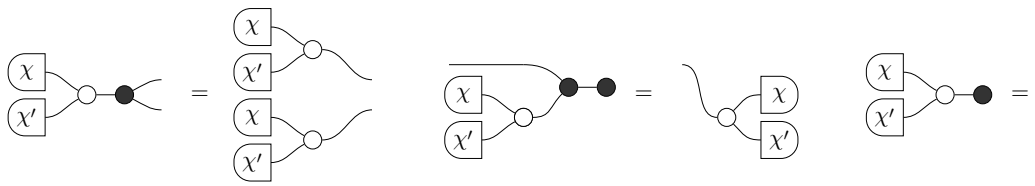
(4.19)

Implication (ii) follows by applying the four defining equations of strong complementarity, together with Hopf's law, to ●-classical states. The top row in 3.6 holds if and only if the unit ○ is a ●-classical state: it is coherently copied, transposed and deleted by ●.



(4.20)

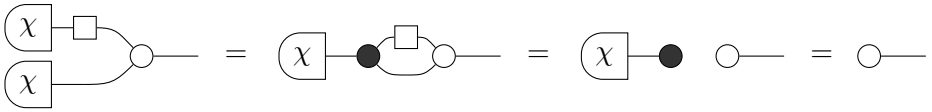
The bottom row in 3.6 holds applied to two ●-classical states if and only if the multiplication under ⊳ of two ●-classical states is a ●-classical state.



(4.21)

Conditional to $(○, ○)$ endowing the ●-classical states with the structure of a monoid, Hopf's law applied to a ●-classical state is equivalent to the antipode acting as group inverse on

●-classical states.



(4.22)

Implications (iii) and (iv) follow the same lines as implications (i) and (ii). Under the assumptions of (iii) we can conclude that Equation 4.19 holds, and under the assumption of (iv) we can conclude that Equations 4.20, 4.21 and 4.22 hold: from the existence of enough ●-classical states, we can conclude that the laws of complementarity and strong complementarity hold as desired. □

5. Generalised Mermin-type Arguments

Armed with the necessary results relating the classical and unbiased states of strongly complementary observables, we are now in a position to formulate our generalised Mermin-type arguments. To do so, we first review the ingredients of Mermin's original parity argument for qubit GHZ states:

(a) a 3-partite qubit GHZ state for the Pauli Z observable;
(b) the abelian group $P(Z) \cong \mathbb{R}/(2\pi\mathbb{Z})$ of phase states for the Pauli Z observable;
(c) the finite subgroup $\{0, \pi\} \cong \mathbb{Z}_2$ given by the eigenstates of the Pauli X observable;
(d) an equation $2x = 1$ with no solution in the subgroup $\{0, \pi\}$ given by the Pauli X eigenstates, but with a solution $\pi/2$ in the group $\mathbb{R}/(2\pi\mathbb{Z})$ of Pauli Z phase states;
(e) measurements in the Pauli X observable.

Similarly, our generalised Mermin-type arguments will involve the following ingredients:

(a) an N-partite GHZ state for a †-qSCFA ○;
(b) the abelian group $(P(○), \oplus, 0)$ of ○-phase states[23];
(c) the subgroup $(K(●), \oplus, 0)$, assumed to be finite, of ●-classical states for a †-qSFA ● which is strongly complementary to ○;
(d) a finite system of $\mathbb{Z}$-module equations, together with a solution in the group $P(○)$;
(e) measurements in the ● observable.

The non-existence of a solution in the subgroup $K(●)$ of ●-classical states is not part of our generalised setup: it will be explicitly characterised as the necessary and sufficient condition for contextuality. Also, N will not be a free parameter, being instead determined by the exponent of the finite abelian group $K(●)$.

Definition 5.1. Consider an R-probabilistic CP* Category $\mathrm{CP}^*[\mathcal{C}]$. A **generalised Mermin-type argument** in $\mathrm{CP}^*[\mathcal{C}]$ is specified by the following data:

(i) a strongly complementary pair $(○, ●)$ of a canonical †-qSCFA ○ and a canonical †-SCFA ● on some object $\mathcal{H}$ of $\mathcal{C}$, such that ● has enough classical states; we furthermore assume that the set $K(●)$ of ●-classical states is finite[24], and that $|K(●)|$ is invertible as an element of the semiring R of scalars of $\mathcal{C}$;

[23] Isomorphic, by Theorem 4.6, to the ○-phase group, which we will denote by $(P(○), \cdot, id)$.
[24] This, together with commutativity of ○, means that $(K(●), \oplus, 0)$ is a finite abelian group.

(ii) a finite system of $\mathbb{Z}$-module equations[25] in the following form, with $a^1, ..., a^S \in K(\bullet)$:

$$\mathcal{S} = \begin{cases} \bigoplus_{r=1}^{M} n_r^1 \, y_r = a^1 \\ \quad \vdots \\ \bigoplus_{r=1}^{M} n_r^S \, y_r = a^S \end{cases} \tag{5.1}$$

(iii) a given solution $(y_r := \beta_r)_{r=1}^M$ in the abelian group $P(\circ)$ of $\circ$-phase states;

(iv) a positive integer N such that $N \geq \sum_{r=1}^{M} n_r^s$ for all $s = 1, ..., S$, and satisfying $\gcd(N, \exp[K(\bullet)]) = 1$, where $\exp[K(\bullet)]$ is the exponent[26] of $K(\bullet)$.

Therefore a generalised Mermin-type argument is specified by a quintuple $(\circ, \bullet, \mathcal{S}, \beta, N)$.

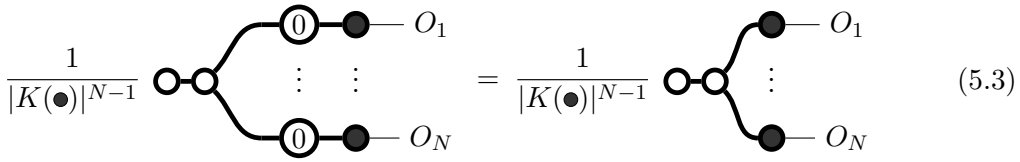
The quintuple $(\circ, \bullet, \mathcal{S}, \beta, N)$ contains all the algebraic and operational ingredients we need to formulate a measurement scenario, which sees N no-signalling parties sharing an N-partied $\circ$-GHZ state. Each party makes a measurement choice $m_j \in \{0, 1, ..., M\}$, applies the phase gate $P_{\beta_{m_j}}$ to her system, and then measures it in the $\bullet$ observable (i.e. measurement outcomes are valued in the set $K(\bullet)$ of $\bullet$-classical states).

Not all combinations of measurement choices are needed for the argument, and the measurement contexts will be determined by System 5.1. We begin by zero-padding the system as follows, so that exactly N phase states are involved in each equation:

$$\begin{cases} n_0^0 \, y_0 \oplus \quad 0 \, y_1 ... \oplus \quad 0 \, y_M = 0 \\ n_0^1 \, y_0 \oplus n_1^1 \, y_1 ... \oplus n_M^1 \, y_M = a^1 \\ \qquad \vdots \\ n_0^S \, y_0 \oplus n_1^S \, y_1 ... \oplus n_M^S \, y_M = a^S \end{cases} \tag{5.2}$$

where we have defined $a^0 := 0$, $n_0^s := N - \sum_{r=1}^{M} n_r^s$ for all $s = 1, ..., S$, $n_0^0 := N$ and $n_r^0 := 0$ for all $r = 1, ..., M$; we will also extend the given solution by setting $\beta_0 := 0$. The first equation in System 5.2 (which we will refer to by the special value $s = 0$ of the parameter s) will contribute to a single measurement context, the **control**; each further equation (i.e. for each value $s = 1, ..., S$ of the parameter s) will give rise to N measurement contexts, the **variations**, for a total of $1 + S \cdot N$ measurement contexts involved in the scenario.

In the control, all parties choose $m_j^0 = 0$, i.e. perform no phase gate before measuring. They obtain the following global state (where $1/|K(\bullet)|^{N-1}$ is the normalisation factor required to obtain a R-distribution):



(5.3)

The first variation for each value $s = 1, ..., S$ is specified by the corresponding equation in System 5.2: the first n_0^s parties choose $m_j^s = 0$, the next n_1^s parties choose $m_j^s = 1$, the next

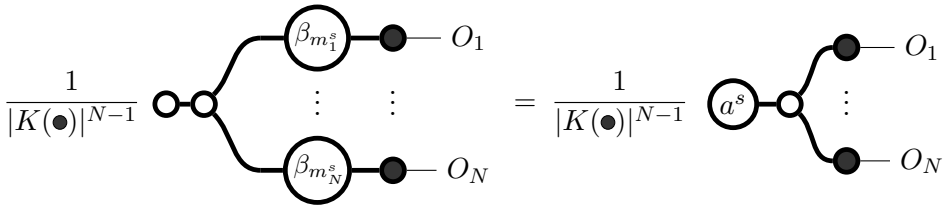
[25] I.e. equations with integer coefficients $n_r^s \in \mathbb{Z}$ and valued in abelian groups (aka $\mathbb{Z}$-modules).

[26] The smallest positive integer e such that $e \cdot g = 0$ for all $g \in K(\bullet)$.

n_2^s parties choose $m_j^s = 2$ and so on, until the last n_M^s parties choose $m_j^s = M$:

$$m_j^s := \text{the largest } m \in \{0, ..., M\} \text{ such that } j \geq \sum_{r=0}^{m-1} n_r^s \tag{5.4}$$

They obtain the following global state, where the equality results from an application of Corollary 4.7, using the relevant equation from System 5.2:



(5.5)

For each fixed value of s, the next $N-1$ variations are cyclic permutations of the first: the measurement choice for the j^{th} party at the k^{th} variation of a given s is $m^s_{j+(k-1)}$, where the sum $j + (k-1)$ is taken modulo N:

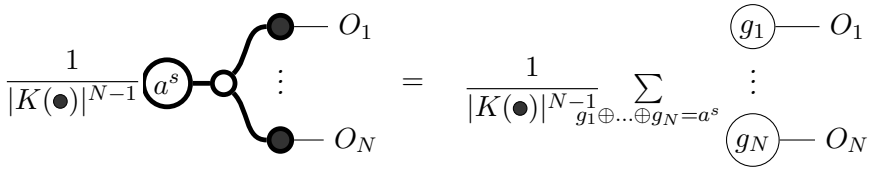
Parties:	1	2	...	$N-1$	N
1^{st} variation for s	m_1^s	m_2^s	...	m_{N-1}^s	m_N^s
2^{nd} variation for s	m_2^s	m_3^s	...	m_N^s	m_1^s
3^{rd} variation for s	m_3^s	m_4^s	...	m_1^s	m_2^s
⋮	⋮	⋮		⋮	⋮
N^{th} variation for s	m_N^s	m_1^s	...	m_{N-2}^s	m_{N-1}^s

(5.6)

Because ○ is commutative, the global state obtained is the same as that for the first variation for that value of s (shown on the RHS of Equation 5.5).

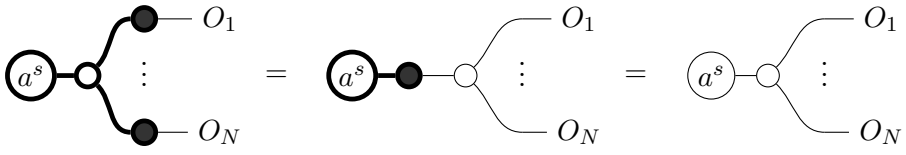
By using strong complementarity and Theorem 4.10, we rewrite the global state obtained by the N parties in the control and variations, obtaining an explicit R-distribution over the set $K(\bullet)^N$ of joint measurement outcomes (from now on, the parameter s can take any value in $\{0, 1, ..., S\}$, unless otherwise specified).

Lemma 5.2.



(5.7)

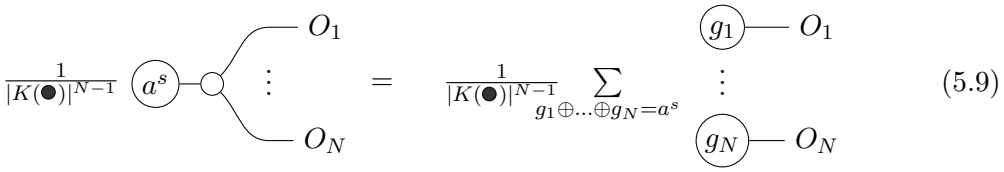
Proof. Strong complementarity can be used to swap ○ and ●, as shown in Corollary 4.1 of [CDKW12], and then a^s can be pushed through because it is a ●-classical state (we have left normalisation aside, and we use $a^0 := 0$ to treat control and variations uniformly):



(5.8)

Using fact that ● has enough classical states, and recalling from Theorem 4.10 that $(\multimap, \circ\!\!-)$ acts as the group multiplication of $K(\bullet)$ when restricted to the ●-classical states, we can

further decompose the state on the RHS of Equation 5.8 into an R-distribution over the set $K(\bullet)^N$:



(5.9)

□

The joint outcome of measurements for the control is uniformly distributed over the subgroup $H_0 \trianglelefteq K(\bullet)^N$ specified by $H_0 := \{(g_1, ..., g_N) \mid g_1 \oplus ... \oplus g_N = 0\}$, while the joint outcome of any of the N variations for each specific value of s is uniformly distributed over the coset $H_{a_s} := (a^s, 0, ..., 0) \oplus H_0$. For each $s, s' \in \{0, 1, ..., S\}$, the cosets H_{a_s} and $H_{a_{s'}}$ are disjoint whenever $a^s \neq a^{s'}$. All in all, we get the following empirical model for the generalised Mermin-type argument:

$$\mathbb{P}[(g_1, ..., g_N)|\text{control}] = \begin{cases} \dfrac{1}{|K(\bullet)|^{N-1}} & \text{if } g_1 \oplus ... \oplus g_N = 0 \\ 0 & \text{otherwise} \end{cases} \tag{5.10}$$

$$\mathbb{P}[(g_1, ..., g_N)|k^{th} \text{ variation for } s] = \begin{cases} \dfrac{1}{|K(\bullet)|^{N-1}} & \text{if } g_1 \oplus ... \oplus g_N = a^s \\ 0 & \text{otherwise} \end{cases} \tag{5.11}$$

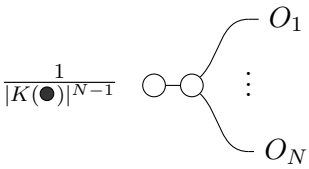
One of the catchy features of Mermin's original argument is that it is entirely deterministic: instead of relying on the violation of some probabilistic inequality, the proof of contextuality shows that the existence of a local hidden variable (LHV) model leads would lead to existence of solutions to an unsatisfiable parity equation (i.e. one which doesn't admit solutions in the finite abelian group $\mathbb{Z}_2$). The proof of contextuality for our generalised Mermin-type arguments goes by similar lines, showing that the existence of a LHV model is equivalent to System 5.1 admitting solutions in the finite abelian group $K(\bullet)$.

Theorem 5.3. *Consider an R-probabilistic CP* category $\mathrm{CP}^*[\mathcal{C}]$, and let $(\circ, \bullet, \mathcal{S}, \beta, N)$ be a generalised Mermin-type argument in it. If the associated empirical model is contextual, then the system $\mathcal{S}$ admits no solution in the finite abelian group $K(\bullet)$. Conversely, if the system $\mathcal{S}$ admits no solution in $K(\bullet)$ and R is a positive semiring, then the empirical model is contextual.*

Proof. The proof comes in two parts: ($\Rightarrow$) we show that any solution in $K(\bullet)$ can be turned into a LHV model; ($\Leftarrow$) we show that, as long as R is a positive semiring, any LHV model can be turned into a solution in $K(\bullet)$.

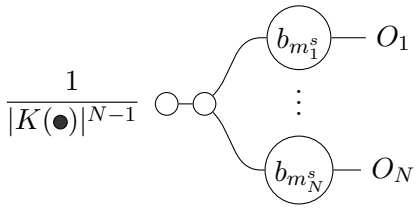
Proof of ($\Rightarrow$). Assume that the system $\mathcal{S}$ (in the form of System 5.1) admits a solution $(y_r := b_r)_{r=1}^M$, and define $b_0 := 0$. A LHV model can be obtained as follows:

(i) the uniform R-distribution on $H_0 \trianglelefteq K(\bullet)^N$ is taken as a shared classical state amongst the N parties:



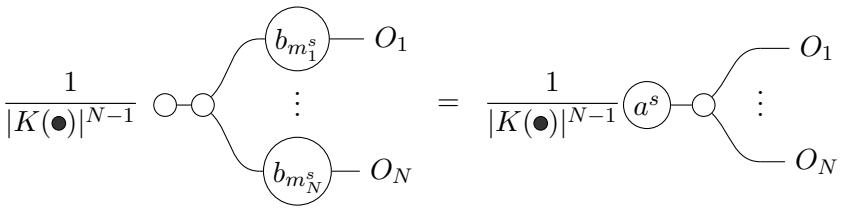
(5.12)

(ii) upon measurement choice $m_j \in \{0, 1, ..., M\}$ for the j^{th} party, a translation by b_{m_j} in the group $K(\bullet)$ is applied to the respective classical subsystem, independently of the measurement choices of the other parties:



(5.13)

All we need to show is that the procedure above produces the same R-distributions on $K(\bullet)^N$ as those given by the empirical model of Equations 5.10 and 5.11. To do so, we simply observe that the global state obtained with the procedure above is the same as the global states obtained in the control 5.3 and in the variations 5.5 (which we treat uniformly by considering $s = 0, 1, ..., S$), because $b_0, b_1, ..., b_N$ satisfy the same equations satisfied by the phases $\beta_0, \beta_1, ..., \beta_N$:



(5.14)

Proof of (⇐). Now assume that R is a positive semiring, and that the scenario admits a LHV model:

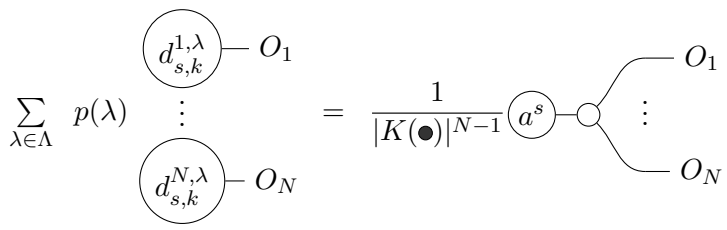
(i) there is a some finite set Λ, the set of values for the hidden variable, coming with an R-distribution $p : \Lambda \to R$;

(ii) for each possible measurement choice $r = 0, 1, ..., M$ that each party $i = 1, ..., N$ can make, there is a family $(c_r^{i,\lambda})_{\lambda\in\Lambda}$ of $\bullet$-classical states, the deterministic local outcomes for each value of the hidden variable;

(iii) for each measurement context (either $s = 0$, $k = 1$ for the control, or $(s, k) \in \{1, ..., S\} \times \{1, ..., N\}$ for the $N \cdot S$ variations), a definite $\bullet$-classical outcome $d_{s,k}^{i,\lambda}$ is obtained by each party $i = 1, ..., N$ at each definite value $\lambda \in \Lambda$ of the hidden variable:

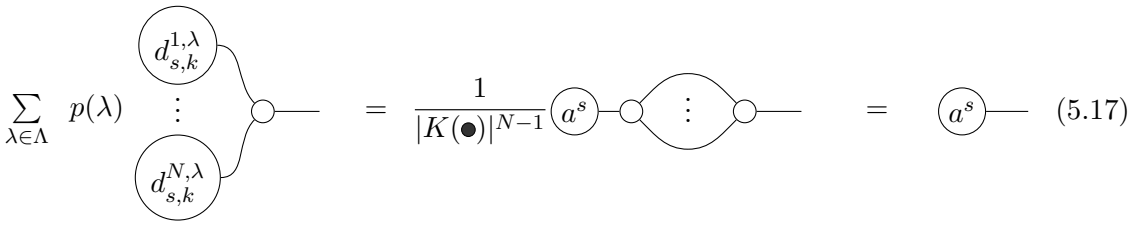
$$d_{s,k}^{i,\lambda} := c_{m_{i+(k-1)}^s}^{i,\lambda} \tag{5.15}$$

(iv) if these definite $\bullet$-classical global states are weighted based on the R-distribution p on Λ, one obtains the same R-distribution on joint measurement outcomes that would be expected from the measurement context:



(5.16)

Given a LHV model, we can sum up all N outcomes of each side of Equation 5.16 in $(K(\bullet), \oplus, 0)$ to obtain an equation between R-distribution over $K(\bullet)$:



(5.17)

The last equation used the fact that $\bullet$ was chosen to be special[27], and hence the normalisation factor for the †-qSCFA $\circ$ is $|K(\bullet)|$ (because $\bullet$ has enough classical states)[28]. Equation 5.17 can be turned into the following conditions on the LHV:

$$\sum_{\lambda \text{ s.t. } \bigoplus_{i=1}^N d^{i,\lambda}_{s,k}=a^s} p(\lambda) = 1 \qquad \sum_{\lambda \text{ s.t. } \bigoplus_{i=1}^N d^{i,\lambda}_{s,k}\neq a^s} p(\lambda) = 0 \tag{5.18}$$

Because R is a positive semiring, $p(\lambda) = 0$ for any λ such that $\oplus_{i=1}^N d^{i,\lambda}_{s,k} \neq a^s$ for some s. Conversely, picking any λ_+ such that $p(\lambda_+) > 0$ (and at least one such λ_+ exists, because p is an R-distribution) yields a family $(d^{i,\lambda_+}_{s,k})_{s,k,i}$ such that $\oplus_{i=1}^N d^{i,\lambda_+}_{s,k} = a^s$ for all s and k. For the control ($s = 0$ and $k = 1$), we obtain the following equation:

$$\oplus_{i=1}^N c_0^{i,\lambda_+} = 0 \tag{5.19}$$

For each variation $(s, k) \in \{1, ..., S\} \times \{1, ..., N\}$, we obtain the following equation:

$$\oplus_{i=1}^N c^{i,\lambda_+}_{m^s_{i+(k-1)}} = a^s \tag{5.20}$$

If c_r^{i,λ_+} was independent of the party i for all $r = 1, ..., M$, this equation would yield a solution to system $\mathcal{S}$ in the form of $b_r := c_r^{i,\lambda_+}$ for any i; unfortunately, this need not be the case. This is where our cyclic definition of the N variations for each value of s comes into play. For each fixed value of s, we add up the N equations for $k = 1, ..., N$:

$$\oplus_{k=1}^N \oplus_{i=1}^N c^{i,\lambda_+}_{m^s_{i+(k-1)}} = Na^s \tag{5.21}$$

Because $\gcd(N, \exp[K(\bullet)]) = 1$, we can take the inverse of N modulo $\exp[K(\bullet)]$, and the equation above has solutions if and only if the equation below does:

$$\oplus_{k=1}^N \oplus_{i=1}^N N^{-1} c^{i,\lambda_+}_{m^s_{i+(k-1)}} = a^s \tag{5.22}$$

Now refer to the Table 5.6 defining the N variations for s, as well as to Equation 5.4 which defines the measurement choices,. The LHS of Equation 5.21 is a sum by rows of the N^2 measurement choices in Table 5.6: each $r = 0, 1, ..., M$ appears n_r^s times in each row, but the changing value of i along each row stops us from turning it into a solution to system $\mathcal{S}$. However, we can switch the summations in Equation 5.21 to obtain a sum by columns of

[27]The special $\bullet$ could have been replaced by a more general †-qSCFA, but at the price of an additional normalisation factor in all global states.

[28]The normalisation factor $|K(\bullet)|$ refers to two wires: each additional wire is an additional copy of $|K(\bullet)|$, for a total of $|K(\bullet)|^{N-1}$ in the N-wire case here.

the table, where each $r = 0, 1, ..., M$ still appears n_r^s times in each column (by the cyclic definition), but now i is constant along each column:

$$\oplus_{i=1}^{N} \oplus_{k=1}^{N} c^{i,\lambda_+}_{m^s_{i+(k-1)}} = \oplus_{i=1}^{N} \oplus_{r=0}^{M} n_r^s c_r^{i,\lambda_+} \tag{5.23}$$

We can then sum up all $(c_r^{i,\lambda_+})_{i=1}^N$ for each $r = 0, 1, ..., M$, and use Equation 5.21 (together with Equation 5.19 to cancel out the contribution from $r = 0$) to finally obtain the desired solution $(b_r)_{r=1}^M$ to system $\mathcal{S}$:

$$\oplus_{r=1}^{M} n_r^s \underbrace{\left(\oplus_{i=1}^{N} N^{-1} c_r^{i,\lambda_+}\right)}_{b_r} = a^s \tag{5.24}$$

□

6. Quantum realisability

In quantum theory, i.e. in the probabilistic CP* category CP*[fdHilb], many of the requirements of generalised Mermin-type arguments are automatically satisfied: canonical †-SCFAs in CPM[fdHilb] (i.e. †-SCFAs in fdHilb) always have enough classical states (and finitely many so), the semiring $\mathbb{R}^+$ of scalars is positive, and any non-zero integer is invertible in it. Hence, only strong complementarity is required in point (i) of the definition of generalised Mermin-type arguments, and Theorem 5.3 establishes an unconditional equivalence between contextuality of a generalised Mermin-type argument $(\circ, \bullet, \mathcal{S}, \beta, N)$ and the existence of solutions to system $\mathcal{S}$ in the finite abelian group $K(\bullet)$ of $\bullet$-classical states.

The remarks above show that the correspondence between systems of equations in finite abelian groups and generalised Mermin-type arguments is particularly tight in the case of quantum theory, but an important question remains unanswered: which systems of $\mathbb{Z}$-module equations lead to arguments which can be realised in quantum theory? As it turns out, all of them (but an obvious caveat applies).

Theorem 6.1. *Let $(K, \oplus, 0)$ be a finite abelian group, and $\mathcal{S}$ be a finite system of $\mathbb{Z}$-module equations in the following form, with $a^1, ..., a^S \in K$:*

$$\mathcal{S} = \begin{cases} \bigoplus_{r=1}^{M} n_r^1 \, y_r = a^1 \\ \quad \vdots \\ \bigoplus_{r=1}^{M} n_r^S \, y_r = a^S \end{cases} \tag{6.1}$$

*Assume that the system is **consistent** in the following sense, where by $\underline{n}^s \in \mathbb{Z}^M$ we denoted the row vectors of System 6.1:*

$$\bigoplus_{s=1}^{S} c_s \cdot \underline{n}^s =_{\mathbb{Z}^M} \underline{0} \implies \bigoplus_{s=1}^{S} c_s \cdot a^s =_K 0, \tag{6.2}$$

Then for every $|K|$-dimensional quantum system $\mathcal{H}$ and every †-qSCFA $\circ$ on $\mathcal{H}$ with normalisation factor $|K|$, there exists a generalised Mermin-type argument $(\circ, \bullet, \mathcal{S}, \beta, N)$ corresponding to System 6.1, i.e. we can always find:

(i) *a †-SCFA $\bullet$, strongly complementary to $\circ$, such that $(K(\bullet), \succ, \prec) \cong (K, \oplus, 0)$;*
(ii) *a solution $(y_r := \beta_r)_{r=1}^M$ to $\mathcal{S}$ in $P(\circ) \cong T^{|K|-1}$;*

(iii) *a positive integer N (infinitely many, in fact) such that $N \geq \sum_{r=1}^{M} n_r^s$ for all $s = 1, ..., S$, and such that* $\gcd(N, \exp[K(\bullet)]) = 1$.

Proof. Point (iii) is trivial: there are infinitely many N such that $\gcd(N, \exp[K(\bullet)]) = 1$, and hence we can always find one such that $N \geq \sum_{r=1}^{M} n_r^s$ for all $s = 1, ..., S$. Point (i) is more interesting, and relies on Theorem 3.6 and Pontryagin duality for finite abelian groups. Point (ii) is perhaps the most interesting, and relies on the possibility of solving consistent systems of $\mathbb{Z}$-module equations in the torus $T^{|K|-1}$.

Proof of point (i). Because $\circ$ is a †-qSCFA with normalisation factor $|K|$ on a $|K|$-dimensional Hilbert space $\mathcal{H}$, it is associated with a basis of $|K|$ vectors, each having norm $\sqrt{|K|}$. Label the basis vectors by the $|K|$ multiplicative characters $\chi \in K^\wedge$ of the finite abelian group K, and construct an orthonormal basis by using the multiplicative characters $\tau \in (K^\wedge)^\wedge$ of the finite abelian group $K^\wedge$:

$$|\tau\rangle := \frac{1}{|K|} \sum_{\chi \in K^\wedge} \tau(\chi) |\chi\rangle \tag{6.3}$$

By Pontryagin duality, there is a canonical isomorphism $(K^\wedge)^\wedge \cong K$, so that the new orthonormal basis given by Equation 6.3 is canonically labelled by elements of K. Consider the †-SCFA $\bullet$ associated to the orthonormal basis thus defined to obtain the desired $(K(\bullet), \succ, \prec) \cong (K, \oplus, 0)$.

Proof of point (ii). The phase group $P(\circ)$ for a canonical †-qSCFA on a $|K|$-dimensional Hilbert space in CPM[fdHilb] is isomorphic to the $(|K|-1)$-dimensional torus, an abelian Lie group. To find a solution $(y_r := \beta_r)_{r=1}^M$ to System 6.1, we will show that one can always find solutions to arbitrary consistent systems of $\mathbb{Z}$-module equations in a torus.

While all K-valued systems with solutions in some super-group of K must necessarily be consistent, the converse is not true in general: given a super-group P of K there may be consistent systems with no solutions in P. Certainly if P is finite then at least one such system exists (because of the finite exponent), and certainly if $P = \mathbb{Q}^d$ then no such system exists; in fact, every divisible torsion-free abelian group P is canonically a $\mathbb{Q}$-vector space, and thus every consistent system of $\mathbb{Z}$-modules equations (and, in fact, of $\mathbb{Q}$-vector space equations) valued in a divisible torsion-free abelian group P has solutions in P (e.g. by Gaussian elimination over the field $\mathbb{Q}$). Unfortunately, while tori are divisible, they are not torsion-free, and in particular not $\mathbb{Q}$-vector spaces: as a consequence, the reasoning above does not apply.

However, a more general argument can be used to show that any consistent system of equations can be solved in any divisible abelian group, regardless of whether the group is torsion-free or not [Fuc15] (although uniqueness of solution need not hold for systems with linearly independent row vectors). As tori are divisible abelian groups, all consistent systems of $\mathbb{Z}$-module equations can be solved in them, and in particular we can find our solution $(y_r := \beta_r)_{r=1}^M$ to System 6.1. □

7. All-vs-Nothing Arguments

Strong contextuality can be reformulated directly in terms of the supports of the distributions. The supports of the global sections, i.e. the $d \in \mathcal{D}_{\mathbb{B}}\mathcal{E}[\mathcal{X}]$ satisfying Equation 1.30, form a (possibly empty) lattice, and thus a probabilistic empirical model is strongly contextual iff

the following set is empty:

$$\mathbb{S}[\mathcal{X}] := \{s \in \mathcal{E}[\mathcal{X}] \mid s|_C \in \operatorname{supp} \zeta_C \text{ for all } C \in \mathcal{M}\} \tag{7.1}$$

For a possibilistic (no-signalling) empirical model $(\zeta_C)_{C\in\mathcal{M}}$, we can define [ABK$^+$15] a **support subpresheaf** $\mathbb{S} \subseteq \mathcal{E}$ by setting:

$$\mathbb{S}[U] := \{s \in \mathcal{E}[U] \mid s|_{C\cap U} \in \operatorname{supp} \zeta_C|_{U\cap C} \text{ for all } C \in \mathcal{M}\} \tag{7.2}$$

Then a possibilistic empirical model is strongly contextual if and only if $\mathbb{S}[\mathcal{X}] = \emptyset$.

The fundamental observation behind the **All-vs-Nothing arguments** of [ABK$^+$15] is that contextuality of Mermin's original argument follows from the existence of the system of $\mathbb{Z}_2$ equations which has no global solution (corresponding to $\mathbb{S}[\mathcal{X}] = \emptyset$ in the sheaf-theoretic framework for contextuality [AB11]), but where each equation admits a solution (i.e. we have $\mathbb{S}[C] \neq \emptyset$ for the measurement context C associated to each equation). In this section we summarise the basic framework of All-vs-Nothing arguments from [ABK$^+$15], taking the liberty of slightly generalising the definitions therein from rings to modules over rings.

Let $\mathcal{R}$ be a commutative ring with unit: we will denote by $+$ the addition in the ring $\mathcal{R}$, and by $\oplus$ the addition in $\mathcal{R}$-modules. The ring $\mathcal{R}$ should not be confused with the semiring R over which the distributions are taken (i.e. the semiring of scalars of the enriched CPM category the arguments take place in). If G is some $\mathcal{R}$-module, we will define an **$\mathcal{R}$-linear equation valued in** G to be a triple $\phi = (C, n, b)$ where:

(i) C is some finite set, and we define $\operatorname{index}(\phi) := C$;
(ii) $n : C \to \mathcal{R}$ is any function;
(iii) $b \in G$ is a given element of G.

If $\phi = (C, n, b)$ is an $\mathcal{R}$-linear equation valued in G, we will say that a function $s : C \to G$ (henceforth an **assignment**) **satisfies** ϕ, written $s \models \phi$, if and only if the following equation holds in G:

$$\bigoplus_{m\in C} n_m s_m = b \tag{7.3}$$

where we denoted $n_m := n(m)$ and $s_m := s(m)$. Any set W of assignments $C \to G$ can be associated a corresponding set $\mathbb{T}_{\mathcal{R}}(W)$ of satisfied equations, which is itself an $\mathcal{R}$-module[29]:

$$\mathbb{T}_{\mathcal{R}}(W) := \{\phi \mid s \models \phi \text{ for all } s \in W\} \tag{7.4}$$

Let $(\zeta_C)_{C\in\mathcal{M}}$ be a possibilistic empirical model for a measurement scenario $(\mathcal{E}, \mathcal{M})$, such that all measurements have the same $\mathcal{R}$-module G as their set of outcomes (for example we had $G = \mathbb{Z}_2$, a $\mathbb{Z}$-module, for Mermin's original argument). Let $\mathbb{S} \subseteq \mathcal{E}$ be the support subpresheaf for the empirical model and define its **R-linear theory** to be:

$$\mathbb{T}_{\mathcal{R}}(\mathbb{S}) := \bigcup_{C\in\mathcal{M}} \mathbb{T}_{\mathcal{R}}(\mathbb{S}[C]) \tag{7.5}$$

We say that a possibilistic empirical model is **All-vs-Nothing** with respect to ring $\mathcal{R}$ and $\mathcal{R}$-module G, written $\mathrm{AvN}_{\mathcal{R},G}$, iff the $\mathcal{R}$-linear theory admits no solution in G, i.e. iff there exists no global assignment $s : \mathcal{X} \to G$ such that:

$$s|_C \models \phi \text{ for all } C \in \mathcal{M} \text{ and all } \phi \in \mathbb{T}_{\mathcal{R}}(\mathbb{S}[C]) \tag{7.6}$$

To connect back with the notation in [ABK$^+$15], we will simply write $\mathrm{AvN}_{\mathcal{R}}$ for $\mathrm{AvN}_{\mathcal{R},\mathcal{R}}$.

A straightforward generalisation (from rings to modules) of a result by [ABK$^+$15] proves that any possibilistic empirical model which is $\mathrm{AvN}_{\mathcal{R},G}$ for some ring $\mathcal{R}$ and some $\mathcal{R}$-module

[29] This gives rise to some interesting results on affine closures, see [ABK$^+$15].

G is strongly contextual: if the model weren't strongly contextual, then there would be some global section $s \in \mathbb{S}[\mathcal{X}]$, and this would imply $s|_C \in \mathbb{S}[C]$ for all $C \in \mathcal{M}$, which in turn would prove that global assignment s satisfies Equation 7.6 (by appealing to Equation 7.4).

A result by [AB11] shows that a probabilistic empirical model is strongly contextual if and only if it is maximally contextual, i.e. if and only if it lies on a face of the no-signalling polytope with no local vertices. As a consequence, showing that our generalised Mermin-type arguments are $\mathrm{AvN}_{\mathcal{R},G}$ is a particularly neat way of proving that they are maximally contextual, a highly desirable property for the device-independent security of the quantum-classical secret sharing protocol which we will present in the next section.

Theorem 7.1. *Consider a R-probabilistic CP* category* $\mathrm{CP}^*[\mathcal{C}]$*, where R is a positive semiring, and let $(\circ, \bullet, \mathcal{S}, \beta, N)$ be a generalised Mermin-type argument in it. If the associated empirical model is contextual, then it is* $\mathrm{AvN}_{\mathbb{Z},K}$*.*

Proof. The associated probabilistic empirical model is given by Equations 5.10 and 5.11: the only scalars appearing are 0 and the invertible $\frac{1}{|K(\bullet)|}$, which are (necessarily) sent to 0 and 1 respectively in the passage to the possibilistic empirical model. The possibilistic empirical model is as follows:

$$\mathbb{P}[(g_1, ..., g_N)|\text{control}] = \begin{cases} 1 & \text{if } g_1 \oplus ... \oplus g_N = 0 \\ 0 & \text{otherwise} \end{cases} \tag{7.7}$$

$$\mathbb{P}[(g_1, ..., g_N)|k^{th} \text{ variation for } s] = \begin{cases} 1 & \text{if } g_1 \oplus ... \oplus g_N = a^s \\ 0 & \text{otherwise} \end{cases} \tag{7.8}$$

The possibilistic empirical model has the following support subpresheaf $\mathbb{S} \subseteq \mathcal{E}$:

$$\mathbb{S}[\text{control}] = \left\{ (c^i_{m^0_i})_{i=1}^N \in K^N \;\middle|\; \oplus_{i=1}^N c^i_{m^0_i} =_K 0 \right\} \tag{7.9}$$

$$\mathbb{S}[k^{th} \text{ variation for } s] = \left\{ (c^i_{m^s_{i+(k-1)}})_{i=1}^N \in K^N \;\middle|\; \oplus_{i=1}^N c^i_{m^s_{i+(k-1)}} =_K a^s \right\} \tag{7.10}$$

Amongst the (many) equations in $\mathbb{T}_{\mathbb{Z}}(\mathbb{S})$ we can find the following $1 + N \cdot S$ equations:

$$\bigoplus_m s_m = 0, \text{ satisfied by all } s \in \mathbb{S}[\text{control}] \tag{7.11}$$

$$\bigoplus_m s_m = a^s, \text{ satisfied by all } s \in \mathbb{S}[k^{th} \text{ variation for } s] \tag{7.12}$$

Any global assignment satisfying all equations in $\mathbb{T}_{\mathbb{Z}}(\mathbb{S})$ would in particular satisfy the $1 + N \cdot S$ equations above, and hence provide a solution in K to the system $\mathcal{S}$. By Theorem 5.3, if the empirical model is contextual then no such solution exists: hence no global assignment satisfying all equations in $\mathbb{T}_{\mathbb{Z}}(\mathbb{S})$ can exist, proving that the model is in particular $\mathrm{AvN}_{\mathbb{Z},K}$. □

Corollary 7.2. *The generalised Mermin-type arguments provide an infinite family of quantum realisable* $\mathrm{AvN}_{\mathbb{Z},K}$ *empirical models, indexed by all finite abelian groups K and all finite consistent systems $\mathbb{S}$ of $\mathbb{Z}$-module equations valued in K which admit no solution in K. Furthermore, all* $\mathrm{AvN}_{\mathbb{Z},K}$ *arguments for some fixed K are equivalently* $\mathrm{AvN}_{\mathbb{Z}_n,K}$ *for any positive integer n divisible by the exponent of K: as a consequence, there are generalised Mermin-type arguments providing quantum realisable* $\mathrm{AvN}_{\mathbb{Z}_n}$ *models for positive integers $n \geq 2$.*

Proof. The first part is a straightforward consequence of Theorems 5.3, 6.1 and 7.1. The second part is a consequence of the fact that any $\mathbb{Z}$-module equation valued in a finite abelian group K is equivalent to a $\mathbb{Z}_{\exp[K]}$-module equation (by taking remainders modulo $\exp[K]$ of all coefficients), and to a $\mathbb{Z}_n$-module equation for any n divisible by the exponent $\exp[K]$ (taking reminders modulo n of coefficients). The last part is the special case where we consider the finite abelian group $K = \mathbb{Z}_n$ as a module over the ring $\mathcal{R} = \mathbb{Z}_n$. □

One open question about All-vs-Nothing arguments asks whether all quantum realisable $\mathrm{AvN}_{\mathbb{Z}}$ models are in fact $\mathrm{AvN}_{\mathbb{Z}_2}$. The following result answers the question negatively, showing that the infinite family of $\mathrm{AvN}_{\mathbb{Z}}$ models provided by the previous corollary form a non-collapsing hierarchy of $\mathrm{AvN}_{\mathbb{Z}_p}$ models for all $n \geq 2$.

Theorem 7.3. *For each $n \geq 2$, there is a quantum realisable $\mathrm{AvN}_{\mathbb{Z}_n}$ (and hence also $\mathrm{AvN}_{\mathbb{Z},\mathbb{Z}_n}$) empirical model which is not $\mathrm{AvN}_{\mathbb{Z}_m,K'}$ for any $m \geq 2$ coprime with n and any non-trivial abelian group K' with exponent dividing m; in particular, it is not $\mathrm{AvN}_{\mathbb{Z}_m}$.*

Proof. The next Section fully works out the example of $K := \mathbb{Z}_n$ with the system $\mathcal{S}$ consisting of a single $\mathbb{Z}$-module equation $ty = 1$. If we pick a $t \in \{2, ..., n-1\}$ which divides n, the equation cannot be satisfied for $K = \mathbb{Z}_n$, giving rise to a model which is both $\mathrm{AvN}_{\mathbb{Z},\mathbb{Z}_n}$ and $\mathrm{AvN}_{\mathbb{Z}_n}$ (because the equation can be replaced by an equivalent $\mathbb{Z}_n$-module equation). Now consider some m coprime with n, and some abelian group K' with exponent dividing m. Then the equation has solutions in K', giving rise to a model which is not $\mathrm{AvN}_{\mathbb{Z},K'}$ nor $\mathrm{AvN}_{\mathbb{Z}_m,K'}$ (nor $\mathrm{AvN}_{\mathbb{Z}_m}$, in the case $K' := \mathbb{Z}_m$). Indeed, we must have $K' \cong \prod_{l=1}^{L} \mathbb{Z}_{p_l^{e_l}}$ for some primes p_l not dividing n and some exponents $e_l \geq 1$, and the equation has solutions in $\mathbb{Z}_{p_l^{e_l}}$ for all l (because t has the same prime factors of n, and hence no p_l can divide t). □

8. A fully worked-out example

In this Section, we fully work out a generalised Mermin-type argument, for the group $K := \mathbb{Z}_d$ and the system $\mathcal{S}$ consisting of a single $\mathbb{Z}$-module equation $ty = 1$ (i.e. we have $S = M = 1$); we take $d \geq 2$ and $t \in \{1, ..., d-1\}$. This can equivalently be seen as a $\mathbb{Z}_d$-module equation $ty = 1 \pmod{d}$. Our presentation goes through four distinct Subsections, covering all aspects from abstract definition of the argument to concrete realisation in quantum theory. In the first Subsection, we present the measurement scenario and empirical model, by writing down the table of measurement choices (for the measurement scenario) and the table of outcome probabilities (for the empirical model). In the second Subsection, we characterise those cases in which the empirical model admits a local hidden variable model, which we describe in full detail. In the third Subsection, we write down the equations turning the empirical model into an All-vs-Nothing argument, in those cases in which a local hidden variable model is ruled out. In the fourth and final Subsection, we give an explicit quantum realisation in terms of GHZ states and phase gates on qudits (i.e. d-dimensional quantum systems). When restricted to the special case $d = 2$ and $t = 1$, the setup we describe coincides with the one originally proposed by Mermin [Mer90], as long as we take the observable ○ in the quantum realisation to correspond to Pauli Z, and the observable ● to correspond to Pauli X (recall that performing the Pauli Z phase gate $P_{e^{i\frac{\pi}{2}}}$ and then measuring in Pauli X is the same as measuring in Pauli Y).

8.1. **Measurement scenario and empirical model.** Firstly, the exponent of $\mathbb{Z}_d$ is d, and we fix a number of parties N such that $\gcd(N,d)=1$ (e.g. $N := d+1$). Each party $i = 1, ..., N$ can make a measurement choice m_i in the set $\{0,1\}$, and the measurement contexts take the following form: in the control, all parties make measurement choice 0, while the variations take the form of N cyclic permutations, each one featuring $N-t$ contiguous parties making measurement choice 0 and t parties making measurement choice 1. This is summarised by the table below:

Party:	1	2	...	$N-t-1$	$N-t$	$N-t+1$	...	$N-1$	N
control	0	0	...	0	0	0	...	0	0
1^{st} variation	0	0	...	0	0	1	...	1	1
2^{nd} variation	0	0	...	0	1	1	...	1	0
3^{rd} variation	0	0	...	1	1	1	...	0	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$
N^{th} variation	1	0	...	0	0	0	...	1	1

(8.1)

The joint measurement outcomes $(g_1, ..., g_N)$ for the N parties are valued in $\mathbb{Z}_d^N$, and the generalised Mermin-type argument is associated with the following empirical model:

	$g_1 \oplus ... \oplus g_N = 0$	$g_1 \oplus ... \oplus g_N = 1$	$g_1 \oplus ... \oplus g_N \neq 0, 1$
control	$\frac{1}{d^{N-1}}$	0	0
1^{st} variation	0	$\frac{1}{d^{N-1}}$	0
2^{nd} variation	0	$\frac{1}{d^{N-1}}$	0
3^{rd} variation	0	$\frac{1}{d^{N-1}}$	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$
N^{th} variation	0	$\frac{1}{d^{N-1}}$	0

(8.2)

8.2. **Local hidden variable models.** When t and d are coprime, the equation $ty = 1 \pmod d$ has a (unique) solution $y := t^{-1} \pmod d$, and a local hidden variable model for the empirical model 8.2 can be obtained as follows. Consider the set Λ of all the $(g_1, ..., g_N) \in \mathbb{Z}_d^N$ such that $g_1 \oplus ... \oplus g_N = 0$, together with the uniform probability distribution $p : \Lambda \to \mathbb{R}^+$ on Λ (i.e. $p(g_1, ..., g_N) = \frac{1}{d^{N-1}}$). Also, consider deterministic local outcomes for each fixed value $\underline{g} \in \Lambda$ of the hidden variable such that, upon measurement choice m_i for party i, the measurement outcome is g_i whenever $m_i = 0$ and $g_i \oplus t^{-1}$ whenever $m_i = 1$. In the control, all parties $i = 1, ..., N$ will choose $m_i = 0$, and the joint measurement outcome will be uniformly distributed over the subgroup $\Lambda \subset \mathbb{Z}_d^N$. In any variation, t parties will choose $m_i = 1$ and $N-t$ parties will choose $m_i = 0$, and the joint measurement outcome will be uniformly distributed over the coset $(1, 0, ..., 0) \oplus \Lambda \subset \mathbb{Z}_d^N$ (using the fact that $t \cdot t^{-1} = 1$ in $\mathbb{Z}_d$). Hence this really defines a local hidden variable model for the empirical model 8.2 associated with the generalised Mermin-type argument.

8.3. **All-vs-Nothing arguments.** When t and d are not coprime, the equation $ty = 1 \pmod d$ cannot have solutions in $K = \mathbb{Z}_d$ (by a standard argument from number theory). The possibilistic empirical model associated with the argument has the following support

subpresheaf $\mathbb{S} \subseteq \mathcal{E}$ (only the control and the first three variations are explicitly shown here, to exemplify the pattern):

$$\mathbb{S}[\text{control}] = \text{ the set of all } (g_0^1, g_0^2, ..., g_0^{N-t-1}, g_0^{N-t}, g_0^{N-t+1}, ..., g_0^{N-1}, g_0^N) \in \mathbb{Z}_d^N$$
$$\text{such that } \bigoplus_{i=1}^{N} g_0^i = 0 \tag{8.3}$$

$$\mathbb{S}[1^{st} \text{ var'n}] = \text{ the set of all } (g_0^1, g_0^2, ..., g_0^{N-t-1}, g_0^{N-t}, g_1^{N-t+1}, ..., g_1^{N-1}, g_1^N) \in \mathbb{Z}_d^N$$
$$\text{such that } \Big(\bigoplus_{i=1}^{N-t} g_0^i\Big) \oplus \Big(\bigoplus_{i=N-t+1}^{N} g_1^i\Big) = 1 \tag{8.4}$$

$$\mathbb{S}[2^{nd} \text{ var'n}] = \text{ the set of all } (g_0^1, g_0^2, ..., g_0^{N-t-1}, g_1^{N-t}, g_1^{N-t+1}, ..., g_1^{N-1}, g_0^N) \in \mathbb{Z}_d^N$$
$$\text{such that } \Big(g_0^N \oplus \bigoplus_{i=1}^{N-t-1} g_0^i\Big) \oplus \Big(\bigoplus_{i=N-t}^{N-1} g_1^i\Big) = 1 \tag{8.5}$$

$$\mathbb{S}[3^{rd} \text{ var'n}] = \text{ the set of all } (g_0^1, g_0^2, ..., g_1^{N-t-1}, g_1^{N-t}, g_1^{N-t+1}, ..., g_0^{N-1}, g_0^N) \in \mathbb{Z}_d^N$$
$$\text{such that } \Big(g_0^{N-1} \oplus g_0^N \oplus \bigoplus_{i=1}^{N-t-2} g_0^i\Big) \oplus \Big(\bigoplus_{i=N-t-1}^{N-2} g_1^i\Big) = 1 \tag{8.6}$$

Amongst the (many) equations in $\mathbb{T}_{\mathbb{Z}}(\mathbb{S})$ we can find the $N+1$ equations equations above, one for the control (Equation 8.3) and N for the variations (Equations 8.4, 8.5 and 8.6, corresponding to the first three variations, exemplify the pattern). Any global assignment $(g_r^i)_{r=0,1}^{i=1,...,N}$ which satisfies all equations in $\mathbb{T}_{\mathbb{Z}}(\mathbb{S})$ would in particular satisfy those $N+1$ equations. However, adding up the N equations corresponding to the variations yields, after a bit of rearranging, the following equation:

$$(N-t)\Big(\bigoplus_{i=1}^{N} g_0^i\Big) \oplus t\Big(\bigoplus_{i=1}^{N} g_1^i\Big) = N \tag{8.7}$$

Taking this together with the equation $\bigoplus_{i=1}^{N} g_0^i = 0$ associated with the control then results in the following equation (using $\gcd(N,d) = 1$ to obtain the inverse N^{-1} modulo d):

$$t\Big(\bigoplus_{i=1}^{N} N^{-1} g_1^i\Big) = 1 \tag{8.8}$$

But this means that setting $y := \bigoplus_{i=1}^{N} g_1^i$ would yield a solution to the equation $ty = 1$ in $\mathbb{Z}_d$, which we assumed not to exist. Hence we cannot have any global assignment satisfying all equations in $\mathbb{T}_{\mathbb{Z}}(\mathbb{S})$, and the model is both $\text{AvN}_{\mathbb{Z},\mathbb{Z}_d}$ and $\text{AvN}_{\mathbb{Z}_d}$.

8.4. **Quantum realisation.** Now consider the d-dimensional quantum system $\mathbb{C}[\mathbb{Z}_d]$. Pick $\bullet$ to be the special commutative †-Frobenius algebra associated with the orthonormal basis $|g\rangle_{g\in\mathbb{Z}_d}$, and $\circ$ to be the quasi-special commutative †-Frobenius algebra associated with the orthogonal basis $|\chi\rangle_{\chi\in\mathbb{Z}_d^\wedge}$ defined as follows in terms of the multiplicative characters $\chi : \mathbb{Z}_d \to S^1$ of the finite abelian group $\mathbb{Z}_d$:

$$|\chi\rangle := \sum_{g\in\mathbb{Z}_d} \chi(g)|g\rangle \tag{8.9}$$

Let H_0^N be the subgroup of $\mathbb{Z}_d^N$ defined by $H_0 := \left\{\underline{g} \in \mathbb{Z}_d^N \;\middle|\; g_1 \oplus ... \oplus g_N = 0\right\}$. Then the (normalised) N-party GHZ state $|GHZ_\circ\rangle$ in the observable $\circ$ takes the following form:

$$|GHZ_\circ\rangle := \frac{1}{d^{N-1}} \sum_{\underline{g}\in H_0} |g_1\rangle \otimes ... \otimes |g_N\rangle \tag{8.10}$$

Let $P_{\alpha_1}, ..., P_{\alpha_N}$ be phase gates for the $\circ$ observable, where each α_i is a function $\mathbb{Z}_d^\wedge \to S^1$ (not necessarily a group homomorphism): writing the phase gates explicitly we get $P_{\alpha_i} := \frac{1}{d}\sum_{\chi\in\mathbb{Z}_d^\wedge} \alpha_i(\chi)\,|\chi\rangle\langle\chi|$. Applying the N phase gates $P_{\alpha_1}, ..., P_{\alpha_N}$ to the N subsystems of the GHZ state is the same as applying their composition $P_\alpha := P_{\alpha_1} \circ ... \circ P_{\alpha_N}$ (where $\alpha := \alpha_1 \cdot ... \cdot \alpha_N$) to a single subsystem.

Amongst the many functions $\alpha : \mathbb{Z}_d^\wedge \to S^1$ are the group homomorphisms, which form the group $\left(\mathbb{Z}_d^\wedge\right)^\wedge$ under pointwise product. By Pontryagin duality, there is a canonical isomorphism $\mathbb{Z}_d \cong \left(\mathbb{Z}_d^\wedge\right)^\wedge$, given explicitly by $h \mapsto (\chi \mapsto \chi(h))$, and we can consider phase gates P_h corresponding to all $h \in \mathbb{Z}_d$. The phase gate P_h acts as $P_h|g\rangle = |g \oplus h\rangle$ on the basis $|g\rangle_{g\in\mathbb{Z}_d}$, and hence applying P_h to a single subsystem of the GHZ state results in the following state, involving the coset $H_h = \left\{\underline{g} \in \mathbb{Z}_d^N \;\middle|\; g_1 \oplus ... \oplus g_N = h\right\}$:

$$\Big(P_h \otimes id \otimes ... \otimes id\Big)|GHZ_\circ\rangle = \frac{1}{d^{N-1}} \sum_{\underline{g}\in H_h^N} |g_1\rangle \otimes ... \otimes |g_N\rangle \tag{8.11}$$

We wish to find a solution to $ty = 1$ in the group of phase gates for $\circ$, i.e. we want some $\beta : \mathbb{Z}_d^\wedge \to S^1$ such that $(P_\beta)^t = P_1$. To do so, first note that the multiplicative characters $\chi : \mathbb{Z}_d \to S^1$ can equivalently be formulated in terms of elements $k \in \mathbb{Z}_d$, e.g. by considering $\chi_k := g \mapsto e^{i2\pi\frac{k\cdot g}{d}}$, and then rewriting the equation $(P_\beta)^t = P_1$ as follows:

$$\frac{1}{d}\sum_{k\in\mathbb{Z}_d} \beta(\chi_k)^t |\chi_k\rangle\langle\chi_k| = \frac{1}{d}\sum_{k\in\mathbb{Z}_d} e^{i2\pi\frac{k}{d}} |\chi_k\rangle\langle\chi_k| \tag{8.12}$$

It is now easy to see that a solution to $(P_\beta)^t = P_1$ is given by $\beta := \chi_k \mapsto e^{i2\pi\frac{1}{t}\frac{k}{d}}$.

Remark 8.1. *The explicit construction presented here for the case of cyclic groups $\mathbb{Z}_d$ in ordinary quantum theory is related to the recent work of Ref.* [RLZL13] *(as well as previous work by Refs.* [LLK06, CMP02, KZ02, ZK99]*), and the relationship between the dimension d of the quantum systems and the allowed numbers N of parties (i.e.* $\gcd(N, d) = 1$*) is the same here and in Ref.* [RLZL13]*. Even when restricted to the special case of quantum theory, however, the work presented here is a significant generalisation of the work of Ref.* [RLZL13]*: in the latter, the authors focus on a specific family of $M = 2, S = 1$ systems with values in a finite cyclic group $\mathbb{Z}_d$; in this work, we provide necessary and sufficient algebraic conditions for arbitrary systems and arbitrary finite abelian groups.*

9. Quantum-classical Secret Sharing

In contrast to other information security protocols, classical secret sharing comes with the intrinsic assumption that some participants cannot, to some extent, be trusted. A *dealer* is interested in sharing some *secret* with a number of *players*, with the caveat that the secret be revealed to the players only when all players agree to cooperate[30]. Integrity and availability of communications is guaranteed by the existence of authenticated classical channels between dealer and players, and the protocol is only concerned with confidentiality, defined as the impossibility of recovering the secret unless all players cooperate.

The quantum-classical scheme of Hillery, Bužek and Berthiaume [HBB99] introduces a new layer of security to secret sharing, employing entangled states and non-commuting observables to detect eavesdropping. The HBB scheme is based on the same measurement contexts of Mermin's original parity argument: a dealer and $N-1$ players share N qubits in a GHZ state (with respect to the computational basis associated with the Pauli Z observable), and randomly choose to measure their qubit in either of the mutually unbiased Pauli X or Pauli Y observables. It can be shown [Zam12] that confidentiality is an immediate consequence of strong complementarity of the Pauli Z and X observables, while eavesdropping detection follows from mutual unbias of the Pauli X and Y observables.

We extend the HBB scheme from Mermin's original parity argument to our generalised Mermin-type arguments, and we use our result on contextuality to provide a number of device-independent security guarantees. For the remainder of this section, we will consider a generalised Mermin-type argument $(\circ, \bullet, \mathcal{S}, \beta, N)$, on an object $\mathcal{H}$ of a R-probabilistic CP* category $\mathrm{CP}^*[\mathcal{C}]$, where R is a positive semiring.

Consider a **dealer**, call her Alice, who wishes to share a **secret** with N' *players*, where $2 \leq N' < N$. As the owner of the secret, Alice is always a trusted party, the *only* trusted party in the protocol. The secret is assumed to take the form of a string of elements of $K(\bullet)$, the **plaintext** (at most one element of $K(\bullet)$, the **round plaintext**, transmitted for each round of the protocol). We wish to ensure that the plaintext can be decoded from the information Alice sends, the **cyphertext**, if and only if all players agree to cooperate (by which we mean that they all reveal their secret keys to some party in possession of the cyphertext). Alice and the players are given N devices (one per player, and $N-N'$ for Alice): at each round w, each device B_j is fed an **input** $m_j^w \in \{0,1,...,M\}$ and returns an **output** $g_j^w \in K(\bullet)$ (we also refer to the outputs $g_1^w, ..., g_{N'}^w$ as the **secret keys** of the players for round w). We furthermore assume the following **security conditions** to hold.

(i) Alice and the players share an authenticated classical channel, ensuring integrity and availability of all classical communications involved in the protocol.

(iia) Alice and the players are in possession of N secure independent classical sources of randomness, to generate independent inputs at each round which are uniformly distributed in $\{0,1,...,M\}$.

(iib) Alice is in possession of a secure classical source of randomness, independent from all other, to decide which rounds will be **secret rounds** (with probability $(1-\tau) > 0$) and which rounds will be **test rounds** (with probability $\tau > 0$).

(iii) During step 2 of the protocol below, no signalling is possible between distinct parties/devices[31].

[30] More in general, a minimum number of cooperating players can be specified.

[31] This can be achieved, for example, by ensuring the devices are operated in conditions controlled by Alice (trusted laboratories, synchronized time-stamp servers, etc).

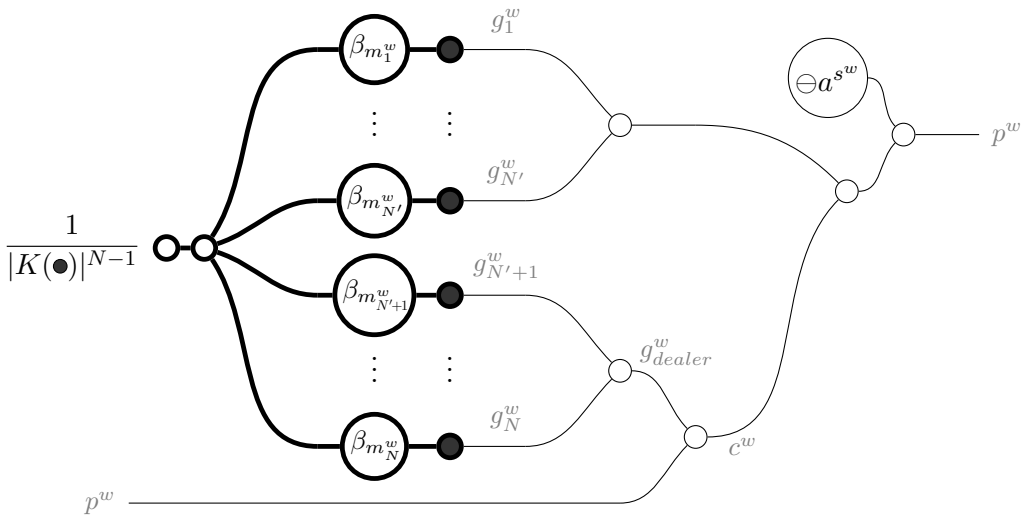


Figure 1: Graphical presentation of a noiseless, trusted implementation.

(iv) We will assume that in step 3 Alice is communicated the measurement choices faithfully[32]

Because tampering can only be determined after the protocol has ended and the entirety (or an otherwise significant portion) of the plaintext has been transmitted, we distinguish between the **plaintext**, the data that can be decoded using the secret keys, and the actual **secret** that Alice wants the players to share. Before the protocol begins, Alice will obtain the plaintext by encrypting the secret with a secure symmetric encryption protocol[33], using a freshly generated ephemeral key which she will broadcast only if the protocol is successful. If the protocol fails, the random key will not be broadcast and the secret will be unrecoverable even if the plaintext is decoded.

The quantum-classical secret sharing protocol then proceeds as follows for each round $w = 1, ..., W$, until the entire secret has been transmitted. An individual round for a noiseless, trusted implementation is presented in Figure 1. Throughout the protocol, Alice keeps a count of occurrences of joint outputs $g_1, ..., g_N$ conditional to each joint input $m_1, ..., m_N$ that she observes in test rounds.

1. Alice and the players share N subsystems of a state ρ: each player has an individual subsystem and Alice keeps the remaining $N - N'$ subsystems. In a noiseless, trusted implementation, ρ is the N-partite $\circ$-GHZ state. For the purposes of a device-independent security analysis, ρ can be potentially any state.
2. Alice and the players each sample their classical source of randomness and obtain inputs $m_1^w, ..., m_N^w$ which are passed to the devices $B_1, ..., B_N$ and result in outputs $g_1^w, ..., g_{N'}^w \in K(\bullet)$ for the players (the secret keys for the round) and $g_{N'+1}^w, ..., g_N^w \in K(\bullet)$

[32]This can be achieved by entrusting the laboratory setup with the communication of the random measurement choices to Alice, the player and the device.

[33]If the secret is in the form of a string of elements of $K(\bullet)$, the natural choice for this protocol, then the plaintext can be obtained by generating a string of uniformly random k^w elements of $K(\bullet)$, obtaining the round plaintext p^w from the corresponding "round secret" q^w as $p^w = q^w \oplus k^w$. Once the string of random elements is broadcast, upon successful completion of the protocol, the secret can be recovered from the decoded plaintext as $q^w = p^w \ominus k^w$.

for Alice. In a noiseless, trusted implementation, B_j with input m_j^w applies the phase gate $P_{\beta_{m_j^w}}$ to the subsystem j and then measures it in the $\bullet$ observable.

3. The inputs for the players are communicated to Alice. She checks that $m_1^w, ..., m_N^w$ define a valid **measurement context** (either the control ($s = 0$) or a variation for some $s = 1, ..., S$).
4. Alice samples her source of randomness to decide whether the round will be a test round or a secret round.

4a. If the round is a test round, Alice requests all players to communicate their secret keys, and she increases the occurrence count for joint output $(g_1^w, ..., g_N^w)$ conditional to joint input $m_1^w, ..., m_N^w$.

4b. If the round is a secret round, Alice computes $g_{dealer}^w := \bigoplus_{j=N'+1}^{N} g_j^w$ and broadcasts the **round ciphertext** $c^w := p^w \oplus g_{dealer}^w$ to the players, where the **round plaintext** p^w is the next element of the plaintext to be sent. She also broadcasts the relevant value $s^w \in \{0, 1, ..., S\}$ she obtained from the joint inputs $m_1^w, ..., m_N^w$.

5. Anyone in possession of the round ciphertext c^w and all secret keys $g_1^w, ..., g_{N'}^w$ can obtain the round plaintext p^w by computing $p^w = (c^w \oplus g_1^w \oplus ... \oplus g_{N'}^w) \ominus a^{s^w}$, where s^w is the value broadcast in Step 4.

The chosen generalised Mermin-type argument determines the following **promised conditional distribution** $\mathbb{P}_{promised}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$, the one which Alice and the players expect to observe (asymptotically) in a trusted noiseless implementation (we use the more compact notation $\underline{g} := (g_1, ..., g_N)$ for the joint output and $\underline{m} := (m_1, ..., m_N)$ for the joint input):

$$\mathbb{P}_{promised}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] = \begin{cases} \dfrac{1}{|K(\bullet)|^{N-1}} & \text{if } g_1 \oplus ... \oplus g_N = \beta m_1 \oplus ... \oplus \beta m_N \\ 0 & \text{otherwise} \end{cases} \tag{9.1}$$

At the end of the protocol, Alice normalises her joint output counts for each joint input to obtain the **observed conditional distribution** $\mathbb{P}_{observed}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ (which need not be no-signalling). She then computes the **noise parameter** ϵ as follows:

$$\epsilon := 1 - |K(\bullet)|^{N-1} \min \left\{ \mathbb{P}_{observed}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] \,\middle|\, g_1 \oplus ... \oplus g_N = \beta_{m_1} \oplus ... \oplus \beta_{m_N} \right\} \tag{9.2}$$

The error parameter as defined above is the smallest $\epsilon \in [0, 1]$ such that the observed conditional distribution can be decomposed as the following convex combination of promised conditional distribution and some **noise conditional distribution** $\mathbb{P}_{noise}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$:

$$\mathbb{P}_{observed}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] = (1 - \epsilon)\, \mathbb{P}_{promised}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] + \epsilon\, \mathbb{P}_{noise}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] \tag{9.3}$$

Before a run of the protocol begins, Alice sets a maximum ϵ_{max} that she is going to accept for the noise parameter. Alice chooses as low an ϵ_{max} as possible compatibly with the specifications of the device provider (and any other beliefs she might have) on the amount of noise she should expect from the devices and states in the absence of any tampering from Eve. At the end of the protocol run, Alice compares the noise parameter ϵ she computed with the maximum ϵ_{max} she decided to accept: if $\epsilon \leq \epsilon_{max}$, she declares the protocol run a success and broadcasts the ephemeral key she used to encode the secret into the plaintext; if $\epsilon > \epsilon_{max}$, she declares the protocol run a failure and she destroys the ephemeral key, rendering the secret unrecoverable even if the plaintext is at some point obtained by the players or by Eve.

The HBB quantum-classical secret sharing protocol comes with two security guarantees: (i) ignorance about any one secret key for a round denies knowledge about the plaintext for that round; (ii) successful, undetected eavesdropping has low probability. It can be shown [Zam12] that in a noiseless and trusted implementation the first guarantee follows abstractly from strong complementarity of the Pauli Z and X observables, and the proof straightforwardly transfers to the strongly complementary pairs $(\circ, \bullet)$ appearing in our generalised protocol. Instead of treating eavesdropping directly, we will present a more general, device-independent proof of security, based solely on contextuality of the generalised Mermin-type argument used by the protocol.

Works on device-independent security (such as [BHK05, VV14] on quantum key distribution) usually posit Eve to be an adversary who can arbitrarily tamper with the shared state and measurement devices, and is only bound in her attempts by the physical theory under consideration[34] and by the security conditions explicitly enforced by the protocol (including no-signalling). Examples of things that the Eve is allowed to do include:

(i) the measurement outcomes broadcast at a test round can reveal to Eve information about measurement outcomes in previous secret rounds;
(ii) Eve can keep a subsystem of the shared state to herself, which she can optimally measure, once all inputs and test round outputs have been broadcast, to obtain information about the secret keys.

Our choice of a device-independent setting comes from the more modest desire to show that the security guarantees follow from contextuality of the generalised Mermin-type argument, regardless of the specific implementation; as a consequence, we will be content with a more restricted model of attack. We assume that Alice and the players might be provided with noisy or imperfect states and devices, which might give Eve a variety of security loopholes to exploit. However, we assume that the device provider shows no malice:

(i) the devices are memoryless and operate independently at each round;
(ii) the states used at different rounds are independent and identical;
(iii) the states are not entangled with any additional system.

However, Eve might possess classical information about the states which is unavailable to the players (such as information leaked through noise or side channels, information acquired via eavesdropping, etc).

Although not fully general, this setup subsumes a variety of more specialised security scenarios that are of interest in classical and quantum cryptography:

(i) Real-world implementations are unavoidably noisy, and one should consider any noise as a potential source of cryptophthora. Our setup allows for the possibility that both the shared state and the measurement devices be noisy, with no dependence on a specific model of noise; it also allows for the possibility that what looks like random noise to Alice and the players might actually carry side-channel information to Eve.
(ii) Eavesdropping detection is a typical desideratum in quantum cryptography, where Eve intercepts the local state of a player[35], measures it in some basis to obtain classical information, and forwards the resulting collapsed state to the player. Our setup allows

[34] Eve is often assumed to be bound by the laws of quantum theory, but sometimes super-quantum attackers are also considered, bound only by causality and no-signalling.

[35] In our secret sharing protocol, a single player's secret key is all that Eve needs to break confidentiality, as we may freely assume that the remaining players are colluding and asked Eve to help them.

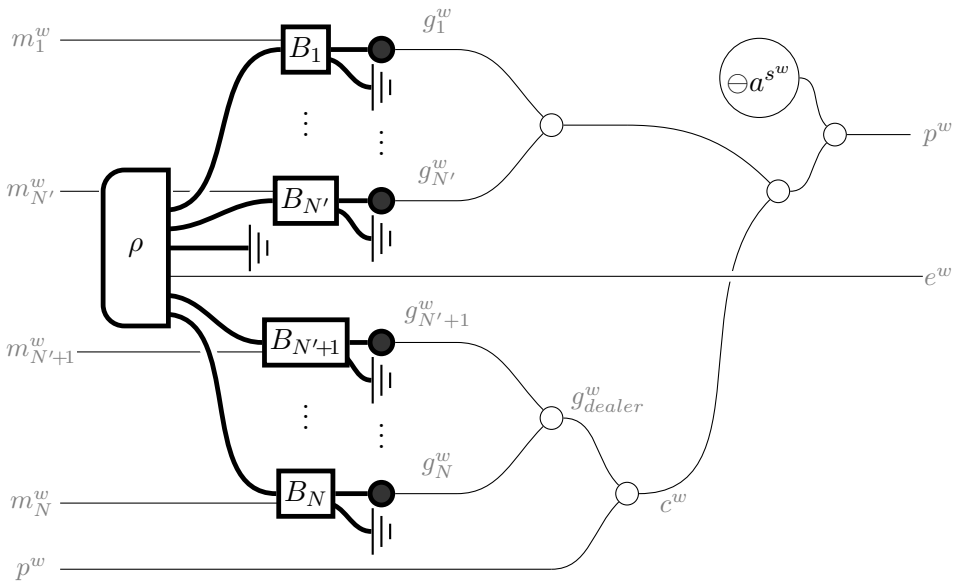


Figure 2: Graphical presentation of a generic, untrusted implementation at a single round of the protocol. Eve might have some classical information e^w about the states which is unknown to Alice and the players. The classical side of the protocol is entirely in the hands of Alice and the players, and proceeds as in the trusted noiseless case.

for the possibility of eavesdropping[36]: the classical information that Eve possesses about the state can be used to model the information she acquired by eavesdropping. Our security proof then has eavesdropping detection as a special case of protocol failure.

Figure 2 displays a single round w of the protocol in a generic, untrusted implementation. An N-partite state ρ is shared between Alice and the players at a given round of the protocol, with no additional subsystem accessible to Eve (who might however be in possession of classical information e^w about it). The measurement devices $B_1, ..., B_N$ operate independently at each round, with no memory or shared resource other than the state ρ. At each round w, device B_j takes measurement choice m_j^w as a classical input and returns measurement outcome g_j^w as a classical output. The rest of the protocol is entirely in the hands of Alice and the players, and proceeds as in the trusted noiseless case.

Our first result shows that lack of contextuality implies the existence of a scenario in which a perfect undetectable attack may take place. In fact, the scenario is not particularly remote: it might well happen happen that the device provider inadvertently chose phases states $\beta_1, ..., \beta_M$ which happen to be $\bullet$-classical states (maybe she did not notice, maybe she was tricked by Eve into choosing them), and that the GHZ state decoheres (spontaneously or with a malicious helping hand) in the $\bullet$ observable. In that case, Alice and the player will notice nothing wrong with their protocol, and Eve will obtain the entirety of the secret all by herself.

[36]However, it does not cover a more advanced attack in which Eve sends through a subsystem of an entangled state, keeping the rest of the state to herself and measuring it in the future to obtain more information about the player's outcome.

Theorem 9.1. *Consider a quantum-classical secret sharing protocol based on a generalised Mermin-type argument $(\circ, \bullet, \mathcal{S}, \beta, N)$, in an R-probabilistic CP* category $\mathrm{CP}^*[\mathcal{C}]$, where R is a positive semiring. If the associated empirical model is non-contextual, then there are shared states ρ and measurement devices $B_1, ..., B_N$ such that test rounds will succeed with certainty, and Eve will always know all the secret keys.*

Proof. By Theorem 5.3, if the empirical model is non-contextual then there exists a solution $(y_r := b_r)_{r=1}^M$ in $K(\bullet)$ to the system $\mathcal{S}$ (which we take to be in the form of System 5.1). For each round w, Eve samples a random variable uniformly distributed over the following set:

$$\left\{(h_1^w, ..., h_N^w) \in K(\bullet)^N \;\middle|\; h_1^w \oplus ... \oplus h_N^w = 0\right\} \tag{9.4}$$

Now assume that the separable pure state $\rho^w = |h_1\rangle \otimes ... \otimes |h_N\rangle$ is given in input to the measurement devices $B_1, ..., B_N$, which are designed so that B_j returns $g_j^w := h_j^w \oplus b_{m_j^w}$ upon measurement choice m_j^w (i.e. applies a phase $b_{m_j^w}$ which happens to be $\bullet$-classical). Once the measurement $(m_j^w)_{j=1}^{N'}$ choices for the players are broadcast, Eve can compute all the secret keys $(g_j^w)_{j=1}^{N'}$. Furthermore, since $(b_r)_{r=1}^M$ is a solution to $\mathcal{S}$, the measurement outcomes obtained from this setup will have the same distribution as the ones from a noiseless trusted implementation, and all test rounds will succeed with certainty. □

Our second result is restricted to probabilistic theories, i.e. distributively CMon-enriched CPM categories having $\mathbb{R}^+$ as their semiring of scalars. Consider the no-signalling polytope associated with the measurement scenario of a contextual generalised Mermin-type argument $(\circ, \bullet, \mathcal{S}, \beta, N)$, and let F be the face of the polytope specified by the support of the empirical model (the one defined by Equation 9.1). For each vertex $v \in F$ of that face, corresponding to empirical model $\mathbb{P}_v\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$, let H_v be the average entropy across all measurement contexts:

$$H_v := \frac{1}{1 + N \cdot S} \sum_{\underline{m} \in \mathcal{M}} H\Big[\, \mathbb{P}_v\big[\, \_ \,\big|\, \underline{m} \,\big] \,\Big] \tag{9.5}$$

Let $H_{promised}^{(min)} := \min_{v \in F} H_v$ be the minimum average entropy across all vertices of the face: because the generalised Mermin-type argument is strongly contextual, the face cannot contain any local vertices, and hence the minimum average entropy $H_{promised}^{(min)}$ is always strictly positive; a tighter estimation of this quantity is left to future work. Call $\eta := \left(1 - \frac{H_{promised}^{(min)}}{|K(\bullet)|^{N-1}}\right) \in [0, 1)$ the **information leakage fraction** for the face: it is the maximum fraction of plaintexts that Eve can expect to decipher when the empirical model she sees lies on face F.

We will now show that protocols based on contextual generalised Mermin-type arguments always provide a certain amount of security: for observed noise parameter ϵ small enough, the maximum expected fraction of plaintexts that Eve can expect to decipher is sharply peaked somewhere between η and $c \cdot \epsilon$, where c is some constant depending on the geometry of the no-signalling polytope. In one extreme, we may have $\eta = 0$, i.e. all empirical model on the face carry the same maximal amount of entropy. In this case, Eve's chances of learning some parts of the secret rely entirely on the noise parameter ϵ: in her best case scenario, she observes a deterministic empirical model for some fraction ϵ of rounds, in which case she can gain complete knowledge about the round plaintext. In the other extreme, we have $\eta \gg \epsilon$, i.e. there are empirical models on the face F which might lead to more leakage of plaintext

information than any number of deterministic model which might be lurking in the noise ϵ. In this case, Eve's best bet might just be to exploit the empirical models on the face F itself.

Theorem 9.2. *Consider a quantum-classical secret sharing protocol based on a generalised Mermin-type argument $(\circ, \bullet, \mathcal{S}, \beta, N)$, in a probabilistic CP* category $\mathrm{CP}^*[\mathcal{C}]$ (with $\mathbb{R}^+$ as its positive semiring of scalars). Consider a run of the protocol with a large number W of rounds, of which P secret rounds and T test rounds (with $P \to (1-\tau)W$ and $T \to \tau W$ almost certainly as $W \to \infty$). Let ϵ be the noise parameter observed by Alice at the end (a random variable), and let P_{Eve} be maximum number of round plaintexts that Eve expects to successfully decipher (another random variable). Then the maximum fraction of plaintexts P_{Eve}/P that Eve expects to successfully decipher is sharply peaked around some value between η and $O(\epsilon)$, with variance bounded above by $O(\frac{\tau(1-\tau)}{W})$ almost certainly for $W \to \infty$ (where the big-O notation hides a constant depending on the geometry of the polytope alone).*

Proof. As part of this proof, a number of different conditional distributions will be considered:

(i) the no-signalling conditional distribution $\mathbb{P}_{true}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ determined by ρ and the devices $B_1, ..., B_N$ conditional to Eve obtaining information e (this is the conditional distribution as seen from Eve's vantage point);

(ii) the no-signalling conditional distribution $\mathbb{P}_{true}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] := \sum_e \mathbb{P}[e] \cdot \mathbb{P}_{true}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ determined by ρ and the devices $B_1, ..., B_N$, averaged over Eve's information (this is the **true conditional distribution** as seen from Alice's vantage point, which her tests will estimate);

(iii) the no-signalling conditional distribution $\mathbb{P}_{promised}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ derived from the generalised Mermin-type argument (this is what Alice would expect to estimate in the absence of any noise or tampering);

(iv) the conditional distribution $\mathbb{P}_{observed}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ estimated by Alice.

Alice's estimate of the true conditional distribution $\mathbb{P}_{true}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ can be modelled by considering the vector-valued random variables $\underline{X}^w := \big(X^w_{(\underline{g},\underline{m})}\big)$ for all test rounds w, where $X^w_{(\underline{g},\underline{m})}$ is the real-valued random variable defined as follows (note that $\underline{g}^w$ is a random element of $K(\bullet)^N$, and $\underline{m}^w$ is a uniformly random element of the set of $1 + NS$ measurement contexts):

$$X^w_{(\underline{g},\underline{m})} = \begin{cases} 1 & \text{if } \underline{g} = \underline{g}^w \text{ and } \underline{m} = \underline{m}^w \\ 0 & \text{otherwise} \end{cases} \tag{9.6}$$

The vector $\underline{X}^w$ takes the value 1 over the joint input/joint output pair recorded by Alice for round w, and 0 everywhere else: Alice's estimate of the true conditional distribution is then obtained from the average random variable $\frac{1}{T} \sum_{w \text{ test}} \underline{X}^w$. By the central limit theorem, Alice's estimate $\mathbb{P}_{observed}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ will be normally distributed around the true conditional distribution, with variance $O(\frac{1}{T})$; because the noise parameter ϵ observed by Alice is obtained from this estimate, it will similarly be distributed around the true noise parameter ϵ_{true} defined below, with variance bounded above by $O(\frac{1}{T})$ (almost certainly for $T \to \infty$).

We define the **true noise parameter** ϵ_{true} to be obtained from the conditional distribution $\mathbb{P}_{true}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ in the same way that ϵ is obtained from the conditional distribution $\mathbb{P}_{observed}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$. This means ϵ_{true} is the largest such that $\mathbb{P}_{true}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ decomposes as follows, for some conditional distribution $\mathbb{P}_{true,noise}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$:

$$(1 - \epsilon_{true})\, \mathbb{P}_{promised}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] + (\epsilon_{true}) \mathbb{P}_{true,noise}\big[\, \underline{g} \,\big|\, \underline{m} \,\big] \tag{9.7}$$

For each value $e \in E$ that Eve's information can take, we define the parameter $\xi(e) \in [0,1]$ to be the smallest possible such that the conditional distribution $\mathbb{P}_{true}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ decomposes as follows:

$$(1-\xi(e))\mathbb{P}_F(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big] + \xi(e)\mathbb{P}_{F,noise}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big] \tag{9.8}$$

for some distribution $\mathbb{P}_F(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ lying on the face F and some distribution $\mathbb{P}_{F,noise}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ lying outside of face F. To Eve, in possession of information e, the conditional distribution $\mathbb{P}_{true}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ looks like a biased coin deciding between the two following scenarios:

(a) with probability $(1-\xi(e))$, she observes a distribution $\mathbb{P}_F(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ lying on face F, which means that the fraction of the round plaintext that she expects to learn is bounded above by η;
(b) with probability $\xi(e)$, she observes some other distribution $\mathbb{P}_{F,noise}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$, which in the best case scenario could give her full knowledge of the round plaintext.

Because marginalising over Eve's knowledge[37] must result in the distribution $\mathbb{P}_{true}\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$, the geometry of the polytope implies that the convex combination $\sum_e \mathbb{P}[e]\xi(e)$ must go to zero as $O(\epsilon_{true})$ (i.e. there must be some constant $c > 0$ such that $\sum_e \mathbb{P}[e]\xi(e) \leq c \cdot \epsilon_{true}$).

It should be noted that the information e obtained by Eve is random to Eve herself: sometimes she will obtain information giving her better guessing probability, sometimes she will obtain information giving her worse guessing probability. When the distribution of e is taken into account, the fraction of round plaintexts that Eve can expect to decipher is bounded above by the following value, falling somewhere between η and $O(\epsilon_{true})$:

$$\sum_e \mathbb{P}[e]\Big((1-\xi(e))\eta + \xi(e)\Big) \tag{9.9}$$

Again by central limit theorem, the maximum fraction P_{Eve}/P of round plaintexts that Eve expects to successfully decipher is normally distributed around the value above, with variance $O(\frac{1}{P})$ (almost certainly for $P \to \infty$).

Finally, because P_{Eve}/P is sharply peaked around some value between η and $O(\epsilon_{true})$, with variance $O(\frac{1}{P})$, and because ϵ is sharply peaked around ϵ_{true}, with variance bounded above by $O(\frac{1}{T})$, we can conclude that P_{Eve}/P is sharply peaked around some value between η and $O(\epsilon)$, with variance bounded above by $O(\frac{1}{T}+\frac{1}{P})$ (which tends to $O(\frac{1}{\tau(1-\tau)W})$ almost certainly as $W \to \infty$). □

Conclusions and future work

Conclusions. Using phase groups and strongly complementary observables, we have fully generalised Mermin-type non-locality arguments, and we have provided the exact group-theoretic conditions required for non-locality to arise. In this sense, our results complete the line of enquiry on the connection between phase groups and non-locality started in Refs. [CES10, CDKW12]. We have shown that all generalised Mermin-type argument can be realised in quantum mechanics, using GHZ states and appropriate sets of phase gates. Furthermore, the abstract, diagrammatic nature of our proofs makes our results immediately applicable to a much larger class of quantum-like theories, such as real quantum theory,

[37] I.e. taking the convex combination of the conditional distributions $\mathbb{P}_{true}(e)\big[\, \underline{g} \,\big|\, \underline{m} \,\big]$ with respect to the probability distribution $\mathbb{P}[e]$ of Eve's side-channel information.

relational quantum theory, hyperbolic quantum theory, p-adic quantum theory, and modal quantum theory [Gog17].

We have proceeded to investigate the empirical models arising from our generalised arguments, using the sheaf-theoretic framework for non-locality and contextuality [AB11]. We have shown the models to provide a new infinite family of inequivalent quantum-realisable All-vs-Nothing arguments [ABK$^+$15], and we have concluded that the hierarchy of quantum-realisable All-vs-Nothing arguments over rings of modular integers does not collapse. As a corollary, we have established that generalised Mermin-type arguments give rise to strong contextuality whenever they are contextual in the first place.

Our generalisations have found practical application in the formulation of an extension of the quantum-classical secret sharing scheme of Hillery, Bužek and Berthiaume [HBB99], which was originally based on Mermin's non-locality argument for qubit GHZ states [Mer90]. We have provided a diagrammatic description of the scheme in its most general form, and we have used our results on strong contextuality to provide some device-independent security guarantees (which apply to the original HBB scheme as a special case).

Future work. In the process of pursuing the connection between phase groups and non-locality, this work has left several questions unanswered.

Firstly, our generalised arguments are restricted to finite abelian group algebras, while it is known that strongly complementary pairs extend to all finite group algebras (where one Frobenius algebra is commutative, and one is potentially non-commutative), and even further to certain finite-dimensional compact quantum groups (where both Frobenius algebras are potentially non-commutative). From a quantum perspective, this corresponds to using possibly degenerate observables in place of the non-degenerate ones employed in this work. In the future, we will be interested in investigating the impact that the introduction of degenerate observables—in the form of non-abelian group algebras and compact quantum groups—might have on the connection between non-locality and phase groups that was established here in the abelian group case.

Secondly, we have shown that our generalised Mermin-type arguments are All-vs-Nothing, but the converse need not hold in general: there are many examples of All-vs-Nothing arguments which do not come in the format of Mermin-type arguments [ABK$^+$15]. In the future, we will be interested in fleshing out an exact description of which All-vs-Nothing are equivalent to / arise as deformation of Mermin-type arguments, and in exploring whether the techniques used in this work can be applied in a more general context.

Finally, the model of attack we used in the security proof for our quantum-classical secret sharing scheme is somewhat more restrictive than the gold standard employed in device-independent quantum cryptography. In the future, we will be interested in obtaining a more complete proof of security, covering broader modes of attack (such as memory and side-channel attacks). Also, our current proof relies on some rather lax parameters, which we expect to be tightened as part of upcoming work.

Acknowledgements

The authors would like to thank Samson Abramsky, Bob Coecke, Aleks Kissinger and Rui Soares Barbosa for comments, suggestions and useful discussions, as well as Sukrita Chatterji and Nicolò Chiappori for their support. The authors would especially like to thank Samson Abramsky for pointing out a mistake in a previous version of this work. Funding from

EPSRC (OUCL/2013/SG) and Trinity College (Williams Scholarship) for the first author is gratefully acknowledged.

References

[AB11] Samson Abramsky and Adam Brandenburger. The sheaf-theoretic structure of non-locality and contextuality. *New Journal of Physics*, 13, 2011.

[ABK+15] Samson Abramsky, Rui Soares Barbosa, Kohei Kishida, Raymond Lal, and Shane Mansfield. Contextuality, Cohomology and Paradox. *24th EACSL Annual Conference on Computer Science Logic (CSL)*, pages 211–228, 2015.

[AC09] Samson Abramsky and Bob Coecke. Categorical Quantum Mechanics. *Handbook of Quantum Logic and Quantum Structures*, pages 261–323, 2009.

[Bac14] Miriam Backens. The ZX-calculus is complete for stabilizer quantum mechanics. *New Journal of Physics*, 16(9), 2014.

[BH12] Sergio Boixo and Chris Heunen. Entangled and sequential quantum protocols with dephasing. *Physical Review Letters*, 108(12):1–5, 2012.

[BHK05] Jonathan Barrett, Lucien Hardy, and Adrian Kent. No signaling and quantum key distribution. *Physical Review Letters*, 95(1):1–4, 2005.

[CMP02] Nicolas J. Cerf, Serge Massar, and Stefano Pironio. Greenberger-Horne-Zeilinger Paradoxes for Many Qudits. *Physical Review Letters*, 89:080402, 2002.

[CD11] Bob Coecke and Ross Duncan. Interacting quantum observables: Categorical algebra and diagrammatics. *New Journal of Physics*, 13, 2011.

[CDKW12] Bob Coecke, Ross Duncan, Aleks Kissinger, and Quanlong Wang. Strong complementarity and non-locality in categorical quantum mechanics. *Proceedings of the 2012 27th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2012*, pages 245–254, 2012.

[CDP11] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Informational derivation of quantum theory. *Physical Review A - Atomic, Molecular, and Optical Physics*, 84(1):1–39, 2011.

[CE12] Bob Coecke and Bill Edwards. Spekkens's toy theory as a category of processes. *Proceedings of Symposia in Applied Mathematics*, 71, 2012.

[CES10] Bob Coecke, Bill Edwards, and Robert W. Spekkens. Phase groups and the origin of non-locality for qubits. *Electronic Notes in Theoretical Computer Science*, 270(2):15–36, 2010.

[CH15] Oscar Cunningham and Chris Heunen. Axiomatizing complete positivity. *Electronic Proceedings in Theoretical Computer Science*, (Qpl 2015):148–157, 2015.

[CHK14] Bob Coecke, Chris Heunen, and Aleks Kissinger. Categories of quantum and classical channels. *Quantum Information Processing*, pages 1–31, 2014.

[CK15] Bob Coecke and Aleks Kissinger. Categorical Quantum Mechanics I: Causal Quantum Processes, 2015.

[CK17] Bob Coecke and Aleks Kissinger. *Picturing Quantum Processes*. Cambridge University Press, 2017.

[CL13] Bob Coecke and Raymond Lal. Causal Categories: Relativistically Interacting Processes. *Foundations of Physics*, 43(4):458–501, 2013.

[Coe12] Bob Coecke. The logic of quantum mechanics - Take II. *Logic and Algebraic Structures in Quantum Computing*, 2012.

[CP10] Bob Coecke and Simon Perdrix. Environment and classical channels in categorical quantum mechanics. *Logical Methods in Computer Science*, 8(4:14):1–24, 2010.

[CPV13] Bob Coecke, Dusko Pavlovic, and Jamie Vicary. A new description of orthogonal bases. *Mathematical Structures in Computer Science*, 23(03), 2013.

[DD16] Ross Duncan and Kevin Dunne. Interacting Frobenius Algebras are Hopf. *Proceedings of the 31st Annual ACM/IEEE Symposium on Logic in Computer Science - LICS '16*, 2016.

[Fuc15] László Fuchs. *Abelian groups*. Springer, 2015.

[Gog17] Stefano Gogioso. Fantastic Quantum Theories and Where to Find Them, 2017.

[GK17] Stefano Gogioso and Aleks Kissinger. Fully graphical treatment of the quantum algorithm for the Hidden Subgroup Problem, 2017

[GS17] Stefano Gogioso and Carlo Maria Scandolo. Categorical Probabilistic Theories, 2017

[HBB99] Mark Hillery, Vladimír Bužek, and André Berthiaume. Quantum secret sharing. *Physical Review A*, 59(3):1829–1834, 1999.

[KZ02] Dagomir Kaszlikowski and Marek Zukowski. Greenberger-Horne-Zeilinger paradoxes for N N-dimensional systems. *Physical Review A*, 66:042107, 2002.

[Kis12] Aleks Kissinger. *Pictures of Processes: Automated Graph Rewriting for Monoidal Categories and Applications to Quantum Computing*. PhD thesis, 2012.

[LLK06] Jinhyoung Lee, Seung-Woo Lee, and M. S. Kim. Greenberger-Horne-Zeilinger nonlocality in arbitrary even dimensions. *Physical Review A*, 73:032316, 2006.

[Mer90] David Mermin. Quantum mysteries revisited. *American Journal of Physics*, 58(8):731–734, 1990.

[RTHH16] Ravishankar Ramanathan, Jan Tuziemski, Michał Horodecki, and Paweł Horodecki. No Quantum Realization of Extremal No-Signaling Boxes. *Physical Review Letters*, 117(5):050401, 2016.

[RLZL13] Junghee Ryu, Changhyoup Lee, Marek Zukowski, and Jinhyoung Lee. Greenberger-Horne-Zeilinger theorem for N qudits. *Physical Review A*, 88:042101, 2013.

[Sel07] Peter Selinger. Dagger Compact Closed Categories and Completely Positive Maps. *Electronic Notes in Theoretical Computer Science*, 170:139–163, 2007.

[Sel08] Peter Selinger. Idempotents in dagger categories (extended abstract). *Electronic Notes in Theoretical Computer Science*, 210:107–122, 2008.

[Sel09] Peter Selinger. A survey of graphical languages for monoidal categories. *Lecture Notes in Physics*, 813:289–355, 2009.

[Spe07] Robert W. Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110, 2007.

[Vic11] Jamie Vicary. Categorical Formulation of Finite-Dimensional Quantum Algebras. *Communications in Mathematical Physics*, 304(3):765–796, 2011.

[Vic12] Jamie Vicary. Topological structure of quantum algorithms. In *Proceedings of the 2013 28th Annual ACM/IEEE Symposium on Logic in Computer Science*, 2012.

[VV14] Umesh Vazirani and Thomas Vidick. Fully Device-Independent Quantum Key Distribution. *Physical Review Letters*, 113(14):140501, sep 2014.

[Zam12] Vladimir Nikolaev Zamdzhiev. An Abstract Approach towards Quantum Secret Sharing, 2012.

[ZK99] Marek Zukowski and Dagomir Kaszlikowski. Greenberger-Horne-Zeilinger paradoxes with symmetric multiport beam splitters. *Physical Review A*, 59:3200, 1999.