

Preface

This is informal proceedings of the 11th International Workshop on Quantum Physics and Logic (QPL 2014), which is held June 4–6, 2014 at Kyoto University.

The goal of the QPL workshop series is to bring together researchers working on mathematical foundations of quantum physics, quantum computing and spatio-temporal causal structures, and in particular those that use logical tools, ordered algebraic and category-theoretic structures, formal languages, semantic methods and other computer science methods for the study of physical behavior in general. Over the past few years, there has been growing activity in these foundational approaches, together with a renewed interest in the foundations of quantum theory, which complement the more mainstream research in quantum computation. Earlier workshops in this series, with the same acronym under the name Quantum Programming Languages, were held in Ottawa (2003), Turku (2004), Chicago (2005), and Oxford (2006). The first QPL under the new name Quantum Physics and Logic was held in Reykjavik (2008), followed by Oxford (2009 and 2010), Nijmegen (2011), Brussels (2012) and Barcelona (2013).

This edition of the workshop attracted 53 submissions. We wish to thank all their authors for their interest in QPL 2014. After careful discussions, the Program Committee selected 32 papers for presentation at the workshop. Each submission was refereed by at least two reviewers, who delivered detailed and insightful comments and suggestions. The Program Chairs thank all the Program Committee Members and all the additional reviewers for their excellent service.

The workshop program is enriched by three invited lectures:

- Giulio Chiribella (Tsinghua), *Pure, reversible and sharp: a tale of systems in interaction with their environment*
- Masahito Hasegawa (Kyoto), *Denotational semantics and quantum topology*
- Masanao Ozawa (Nagoya), *Quantum set theory extending standard probabilistic interpretation of quantum theory*

The workshop enjoys partial support from Research Institute for Mathematical Sciences (RIMS), Kyoto University; from the EPSRC Network on Structures at the Interface of Physics and Computer Science (EP/I03596X/1); and the Support Center for Advanced Telecommunications Technology Research (SCAT).

June 2014

Bob Coecke
Ichiro Hasuo
Prakash Panangaden

Workshop Organization

Program Chairs

Bob Coecke (Oxford), Ichiro Hasuo (Tokyo), Prakash Panangaden (McGill)

Program Committee

Dan Browne (UCL), Giulio Chiribella (Tsinghua), Ross Duncan (Strathclyde), Simon Gay (Glasgow), Chris Heunen (Oxford), Matty Hoban (ICFO), Bart Jacobs (Nijmegen), Viv Kendon (Leeds), Simon Perdrix (CNRS Grenoble), Mehrnoosh Sadrzadeh (QMUL), Peter Selinger (Dalhousie), Rob Spekkens (Perimeter), Bas Spitters (Nijmegen), Jamie Vicary (Oxford & CQT Singapore), Mingsheng Ying (UTS Sydney & Tsinghua)

Steering Committee

Bob Coecke (Oxford), Prakash Panangaden (McGill), Peter Selinger (Dalhousie)

Local Organizers

Ichiro Hasuo (Tokyo), Naohiko Hoshino (Kyoto), Yoshihiko Kakutani (Tokyo), Susumu Nishimura (Kyoto)

Additional Reviewers

Daisuke Bekki, Paul Busch, Kentaro Honda, Matthew Leifer, Timothy Proctor, Francisco Rios, and Michael Westmoreland

Table of Contents

Pure, reversible and sharp: a tale of systems in interaction with their environment	5
<i>Giulio Chiribella</i>	
Denotational semantics and quantum topology	7
<i>Masahito Hasegawa</i>	
Quantum set theory extending standard probabilistic interpretation of quantum theory	9
<i>Masanao Ozawa</i>	
The dagger lambda calculus	10
<i>Philip Atzemoglou</i>	
Depicting qudit quantum mechanics and mutually unbiased qudit theories	27
<i>Andre Ranchin</i>	
Entropic formulation of Heisenberg's measurement-disturbance relation	38
<i>Patrick Coles and Fabian Furrer</i>	
The ZX -calculus is approximately complete for single qubits	41
<i>Miriam Backens</i>	
Tensors, !-graphs, and non-commutative quantum structures	52
<i>Aleks Kissinger and David Quick</i>	
On modifications of Reichenbach's principle of common cause in light of Bell's theorem	64
<i>Eric Cavalcanti and Raymond Lal</i>	
Terminality implies non-signalling	67
<i>Bob Coecke</i>	
Contextuality and Noncommutative Geometry	75
<i>Nadish de Silva</i>	
Complexity of Grammar Induction for Quantum Types	89
<i>Antonin Delpeuch</i>	
Observational Equivalence Using Schedulers for Quantum Processes	101
<i>Kazuya Yasuda, Takahiro Kubota and Yoshihiko Kakutani</i>	
Equivalence of wave-particle duality to entropic uncertainty	113
<i>Patrick Coles, Jędrzej Kaniewski and Stephanie Wehner</i>	
An equational characterization of quantum computation	116
<i>Sam Staton</i>	
Mixed quantum states in higher categories	133
<i>Chris Heunen, Jamie Vicary and Linde Wester</i>	
A 2-Categorical Analysis of Complementary Families, Quantum Key Distribution and the Mean King Problem	145
<i>Krzysztof Bar and Jamie Vicary</i>	
Reflections on the PBR Theorem: Reality Criteria & Preparation Independence	162
<i>Shane Mansfield</i>	

Abstract structure of unitary oracles for quantum algorithms	175
<i>William Zeng and Jamie Vicary</i>	
Dichromatic and Trichromatic Calculus for Qutrit Systems	189
<i>Quanlong Wang and Xiaoning Bian</i>	
General probabilistic theories on arbitrary causal structures	199
<i>Joe Henson, Raymond Lal and Matthew Pusey</i>	
A Study of Entanglement in a Categorical Framework of Natural Language	211
<i>Dimitri Kartsaklis and Mehrnoosh Sadrzadeh</i>	
The ZX-calculus is incomplete for quantum mechanics	223
<i>Christian Schröder de Witt and Vladimir Zamdzhiev</i>	
Belief propagation in monoidal categories	231
<i>Jason Morton</i>	
Translating measurement-based quantum computations with gflow into quantum circuits	241
<i>Jisho Miyazaki, Michal Hajdusek and Mio Murao</i>	
On monogamy of non-locality and macroscopic averages: examples and preliminary results	244
<i>Rui Soares Barbosa</i>	
Stochastic Relational Presheaves and Dynamic Logic for Contextuality	262
<i>Kohei Kishida</i>	
Circuit model implementation of controllization functional on unitary with and without fractional query	276
<i>Akihito Soeda, Shojun Nakayama and Mio Murao</i>	
On Gacs' quantum algorithmic entropy	283
<i>Toru Takisaka</i>	
Parallelized adiabatic gate teleportation	295
<i>Mio Murao, Kousuke Nakago, Michal Hajdusek and Shojun Nakayama</i>	
Globalness of separable maps in terms of time and space resources	299
<i>Seiseki Akibue, Masaki Owari, Go Kato and Mio Murao</i>	
Completeness of Hardy Non-locality: Consequences & Applications	312
<i>Shane Mansfield</i>	
QPEL: Quantum Program and Effect Language	329
<i>Robin Adams</i>	
Semantics for a Quantum Programming Language by Operator Algebras	341
<i>Kenta Cho</i>	
A Kochen-Specker system has at least 21 vertices	355
<i>Bas Westerbaan and Sander Uijlen</i>	

Pure, reversible and sharp: a tale of systems in interaction with their environment

Giulio Chiribella

Center for Quantum Information, Institute for Interdisciplinary Information Sciences,
Tsinghua University, Beijing, 100084, China
gchiribella@mail.tsinghua.edu.cn

Few theories in physics have brought as many successes and surprises as Quantum Theory. Both successes and surprises come from a simple mathematical framework of virtually universal applicability, which blends physics and information theory in peculiar and often puzzling way. At the heart of this framework there is a set of mathematical theorems, known as *dilation theorems* [10], which allow one to reduce all possible states, evolutions, and measurements allowed by quantum mechanics to some privileged subsets. Specifically,

1. mixed states are reduced to pure states (by the GNS construction [7, 11], familiar to the quantum information community as *purification*)
2. general evolutions are reduced to reversible evolutions (by Stinespring's theorem [13])
3. general measurements are reduced to sharp measurements (by Naimark's [8] and Ozawa's [9] theorems).

For finite dimensional quantum systems and for the simplest examples of infinite dimensional systems, the reductions 1-3 are achieved by introducing an auxiliary system (the *environment*), which is eventually discarded. This fact lends itself to an operational interpretation: The ignorance about the preparation of a system, the irreversibility of an evolution, and the unsharpness of a measurement can always be explained as resulting from the lack of control over some degree of freedom in the surrounding environment.

Dilation theorems are usually regarded just as a consequence of the mathematical framework of Quantum Theory. They are heavily used as technical tool by researchers in quantum information theory, to the extent that one can hardly find results that do not invoke any of them in a more or less implicit way. However, the operational content of these theorems is independent of the quantum framework: Even forgetting about Hilbert spaces and operator algebras, one can still express the notions of pure/mixed state, reversible/irreversible evolution, and sharp/unsharp measurement in a general framework of operational-probabilistic theories. In this broader framework the dilation of states, evolutions and measurements can be promoted to the rank of *axioms*, from which (a number of features of) the theory is derived [1, 2]. There are at least three good reasons to follow this route. First, given the amount of results that invoke dilation theorems in quantum information processing, turning these theorems into axioms seems to be a convenient way to restructure the landscape of quantum information and to facilitate the discovery of new protocols. Second, the dilation approach sheds light on the old question “Why the quantum?”, the question of finding a set of well-motivated axioms that single out quantum theory among all possible theories. This is the path followed by Ref. [2] where the finite-dimensional Hilbert space framework has been reconstructed from the purification of mixed states—the so-called *Purification Principle*. In the light of this result, Quantum Theory appears as the golden standard of theory where information-theoretic notions admit a description within the framework of fundamental physics, which aims at providing a picture

of the world in terms of pure states and fundamentally reversible interactions [3]. Third, the approach of abstracting dilation theorems from the Hilbert space framework yielded deep insights in category theory, leading to the formulation of Selinger’s CPM construction [12] and to its axiomatization in terms of interaction with the environment [5, 6].

In this talk I will review the state of the art in the programme of reconstructing Quantum Theory from the Purification Principle and other dilation-type axioms. I will start with a brief introduction to the framework of operational-probabilistic theories, which allows one to talk about pure states, reversible transformations and sharp measurements without assuming Quantum Theory from the outset. Then, I will show how some of the key features of Quantum Theory, such as steering, no-cloning, no-information without disturbance, teleportation and the no-programming, can be derived directly from the Purification Principle, without detouring into Hilbert spaces. Similarly, many upper bounds on quantum nonlocality and contextuality can be derived from the dilation of measurements to sharp measurements and from the compositional structure of sharp measurements [4]. If time permits, I will also discuss how the Purification Principle can be extended to new scenarios, like perspective quantum gravity scenarios, where time and causal structure may not be a priori defined.

References

- [1] G. Chiribella, G.M. D’Ariano & P. Perinotti (2010): *Probabilistic theories with purification*. *Phys. Rev. A* 81, p. 062348.
- [2] G. Chiribella, G.M. D’Ariano & P. Perinotti (2011): *Informational derivation of quantum theory*. *Phys. Rev. A* 84, p. 012311.
- [3] G. Chiribella, G.M. D’Ariano & P. Perinotti (2012): *Quantum Theory, namely the pure and reversible theory of information*. *Entropy* 14(10), pp. 1877–1893.
- [4] G. Chiribella & X. Yuan (2014): *Measurement sharpness cuts nonlocality and contextuality in every physical theory*. *arXiv preprint arXiv:1404.3348*.
- [5] B. Coecke (2008): *Axiomatic description of mixed states from Selinger’s CPM-construction*. *Electronic Notes in Theoretical Computer Science* 210, pp. 3–13.
- [6] B. Coecke & S. Perdrix (2010): *Environment and classical channels in categorical quantum mechanics*. In: *Computer Science Logic*, Springer, pp. 230–244.
- [7] I.M. Gelfand & M.A. Naimark (1943): *On the imbedding of normed rings into the ring of operators in Hilbert space*. *Matematicheskij sbornik* 54(2), pp. 197–217.
- [8] M. Naimark (1940): *Spectral functions of a symmetric operator*. *Izvestiya Rossiiskoi Akademii Nauk. Seriya Matematicheskaya* 4(3), pp. 277–318.
- [9] M. Ozawa (1984): *Quantum measuring processes of continuous observables*. *Journal of Mathematical Physics* 25(1), pp. 79–87.
- [10] V. Paulsen (2002): *Completely bounded maps and operator algebras*. 78, Cambridge University Press.
- [11] I.E. Segal (1947): *Irreducible representations of operator algebras*. *Bulletin of the American Mathematical Society* 53(2), pp. 73–88.
- [12] P. Selinger (2007): *Dagger compact closed categories and completely positive maps*. *Electronic Notes in Theoretical Computer Science* 170, pp. 139–163.
- [13] W.F. Stinespring (1955): *Positive functions on C^* -algebras*. *Proceedings of the American Mathematical Society* 6(2), pp. 211–216.

Denotational Semantics and Quantum Topology

Masahito Hasegawa

Research Institute for Mathematica Sciences
Kyoto University
Kyoto, Japan
hassei@kurims.kyoto-u.ac.jp

In last two decades, *traced monoidal categories* [9] have found many important applications in theoretical computer science, especially in the area of *semantics of logic and computation*. The notion of trace nicely captures various forms of feedback or iteration [2] and circular or recursive structure [5, 14], which have been extensively studied in denotational semantics for more than 40 years. Moreover, there is a canonical way of constructing a *ribbon category* [15] (*tortile monoidal categories* [13]) from any traced monoidal category called *Int-construction* [9] which captures the key aspect of *Geometry of Interaction* [1, 4], that is, an abstract implementation of bi-directional information flow using feedback. A wide range of semantic models of programming languages as well as proof systems have been obtained using Int-construction [4, 7, 12].

On the other hand, since a free ribbon category is equivalent to the category of (oriented framed) tangles [13, 16], each ribbon category gives rise to an invariant of tangles in a functorial way [16]. (This situation can be compared with the role of cartesian closed categories in denotational semantics: a free cartesian closed category is equivalent to the term model of the simply typed lambda calculus.) In particular, many important ribbon categories arise as the categories of linear representations of *quantum groups* [3] or *ribbon Hopf algebras*, and they give so-called *quantum invariants* of (oriented framed) tangles. In this way, ribbon categories play key roles in the development of *quantum topology* [15, 16].

However, despite the importance of traced monoidal categories and ribbon categories in denotational semantics and quantum topology, there were not much interaction between these two areas. Specifically, we had no non-trivial example of traced monoidal categories or ribbon categories which are interesting for *both* of denotational semantics and quantum topology.

In this talk, we give an overview of our recent attempts to fill this gap between denotational semantics and quantum topology. Namely, we start with some familiar monoidal categories used in denotational semantics, and try to find a well-behaved, non-trivial Hopf algebras in these categories which play the role of quantum groups in quantum topology. In some categories such as **Rel** of sets and binary relations, this approach works very much like the case of linear representations of quantum groups and gives rise to braided monoidal categories which can provide semantics of programs and invariants of tangles at the same time [6]. On the other hand, recently Kenji Maillard [11] has shown that there is *no* Hopf algebra in the compact closed category of Conway games [8]. This result suggests that it is quite hard (if not impossible) to obtain braided monoidal categories of sequential games by this approach. We also have some negative result for the categories of domains. Thus it is not always possible to literally copy the ideas of quantum topology in denotational semantics.

As of writing this, all our results are of purely mathematical nature, and we are yet to find concrete computational applications. A promising direction would be that of *topological quantum computation* [10], for which *modular tensor categories* [15] (certain class of well-behaved ribbon categories) play the central role. It would be interesting to see if our approach would provide a way of relating program semantics with topological quantum computation, like a compilation scheme for (quantum or classical) programs into a topological quantum computing architecture.

References

- [1] S. Abramsky, E. Haghverdi & P.J. Scott (2002): *Geometry of Interaction and linear combinatory algebras*. *Mathematical Structures in Computer Science* 12, pp. 625–665, doi:10.1017/S0960129502003730.
- [2] S. Bloom & Z. Ésik (1993): *Iteration Theories*. Springer-Verlag.
- [3] V.G. Drinfel'd (1987): *Quantum groups*. In: *Proceedings of 1986 International Congress of Mathematicians*, pp. 798–820. Available at <http://www.mathunion.org/ICM/ICM1986.1/Main/icm1986.1.0798.0820.ocr.pdf>.
- [4] J.-Y. Girard (1989): *Geometry of Interaction I: interpretation of system F*. In: *Logic Colloquium 88*, Elsevier, pp. 221–260.
- [5] M. Hasegawa (1999): *Models of Sharing Graphs: A Categorical Semantics of let and letrec*. Springer-Verlag. Originally published as Ph.D. thesis, ECS-LFCS-97-360, University of Edinburgh, 1997.
- [6] M. Hasegawa (2012): *A quantum double construction in Rel*. *Mathematical Structures in Computer Science* 22(4), pp. 618–650, doi:10.1017/S0960129511000703. Available at <http://hdl.handle.net/2433/159905>.
- [7] I. Hasuo & N. Hoshino (2011): *Semantics of higher-order quantum computation via Geometry of Interaction*. In: *LICS 2011*, IEEE Press, pp. 237–246, doi:10.1109/LICS.2011.26.
- [8] A. Joyal (1977): *Remarques sur la théorie des jeux à deux personnes*. *Gazette des Sciences Mathématiques du Québec* 1, pp. 46–52.
- [9] A. Joyal, R. Street & D. Verity (1996): *Traced monoidal categories*. *Mathematical Proceedings of the Cambridge Philosophical Society* 119(3), pp. 447–468, doi:10.1017/S0305004100074338.
- [10] A. Kitaev (2003): *Fault-tolerant quantum computation by anyons*. *Annals of Physics* 303(1), pp. 3–20, doi:10.1016/S0003-4916(02)00018-0. Available at <http://arxiv.org/abs/quant-ph/9707021>.
- [11] K. Maillard (2013): *Looking for the lost Hopf monoids*. Report of Internship at RIMS in Kyoto University, ENS Paris.
- [12] Ulrich Schöpp (2013): *On interaction, continuations and defunctionalization*. In: *TLCA 2013, Lecture Notes in Computer Science* 7941, Springer-Verlag, pp. 205–220, doi:10.1007/978-3-642-38946-7.
- [13] M.-C. Shum (1994): *Tortile tensor categories*. *Journal of Pure and Applied Algebra* 93, pp. 57–110, doi:10.1016/0022-4049(92)00039-T.
- [14] G. Ştefănescu (2000): *Network Algebra*. Springer-Verlag.
- [15] V.G. Turaev (1994): *Quantum Invariants of Knots and 3-Manifolds*. de Gruyter.
- [16] D.N. Yetter (2001): *Functorial Knot Theory*. World Scientific.

Quantum set theory extending standard probabilistic interpretation of quantum theory

Masanao Ozawa
Nagoya University
ozawa@is.nagoya-u.ac.jp

Set theory provides foundations of mathematics in the sense that all the mathematical notions like numbers, functions, relations, structures are defined in the axiomatic set theory, ZFC. Quantum set theory naturally extends the logical basis of ZFC from classical logic to quantum logic. Hence, we can expect that quantum set theory provides much more systematic foundations of quantum mechanics than Hilbert spaces and operator algebras. In this talk, I will show a useful application of quantum set theory to quantum mechanics based on the fact that the real numbers constructed in quantum set theory exactly corresponds to the quantum observables. The standard formulation of quantum mechanics answers the question as to in what state an observable A has the value in an interval I . However, the question is not answered as to in what state two observables A and B have the same value. The notion of equality between the values of observables will play many important roles in foundations of quantum mechanics such as the notion of measurement and disturbance. It is shown that all the observational propositions on a quantum system corresponds to some propositions in quantum set theory and the equality relation naturally provides the proposition that two observables have the same value. It has been broadly accepted that we cannot speak of the values of quantum observables without assuming a hidden variable theory, which severely constrained by Kochen-Specker type no-go theorems. However, quantum set theory enables us to do so without assuming hidden variables but alternatively under the consistent use of quantum logic.

References

- [1] M. Ozawa: *Orthomodular-valued models for quantum set theory*. Available at <http://arxiv.org/abs/0908.0367>. ArXiv:0908.0367.
- [2] M. Ozawa (2006): *Quantum perfect correlations*. *Ann. Phys.* 321, pp. 744–769. Available at <http://arxiv.org/abs/quant-ph/0501081>.
- [3] M. Ozawa (2007): *Transfer principle in quantum set theory*. *Journ. Symb. Logic* 72, pp. 625–648. Available at <http://arxiv.org/abs/math.LO/0604349>.
- [4] M. Ozawa (2011): *Quantum reality and measurement: A quantum logical approach*. *Found. Phys.* 41, pp. 592–607. Available at <http://arxiv.org/abs/0911.1147>.

The dagger lambda calculus

Philip Atzemoglou

Department of Computer Science,
University of Oxford,
Oxford, UK
abb.research@the-judges.com

We present a novel lambda calculus that casts the categorical approach to the study of quantum protocols [4] into the rich and well established tradition of type theory. Our construction extends the linear typed lambda calculus [6] with a linear negation [1] of "trivialised" De Morgan duality [5]. Reduction is realised through explicit substitution, based on a symmetric notion of binding of global scope, with rules acting on the entire typing judgement instead of on a specific subterm. Proofs of subject reduction, confluence, strong normalisation and consistency are provided, and the language is shown to be an internal language for dagger compact categories.

1 Introduction

1.1 Motivation

Since the turn of the century, the study of quantum protocols and quantum computation has gained new momentum through the introduction of a category theoretic approach in the works of [4] and [21]. This approach has primarily been using dagger compact categories. In addition to introducing categories to the study of quantum computation, however, the line of work that sprang from this approach has been instrumental in driving a new breed of diagrammatic calculi [11, 14, 12, 13, 9, 10].

In parallel to this approach, another very prominent line of research was seen in the works of [19, 20, 27, 26, 23, 24, 25] and was geared towards the development of a quantum programming language. This approach was seminal in establishing a semantic approach to quantum programming language design and focused primarily in designing a higher order lambda calculus for quantum computation. More specifically, in [25], a quantum lambda calculus with a complicated set of rules is presented, whose structural equations nevertheless allow for higher-order structures. The rest of the work towards constructing a concrete model for the language's semantics remains an open problem.

The purpose of this paper is to bridge these two approaches, bringing the programming languages approach closer to the categorical approach, by casting the diagrammatic formalism into the rich and well established tradition of type theory.

1.2 Summary of results

Since Symmetric Monoidal Closed categories are the precursor to Compact Closed and Dagger Compact categories, we begin our construction by extending the linear typed lambda calculus of [6]. Similarly to the approach used by [1], we introduce a linear negation operator. Contrary to [1], however, because *quantum logics* equate \otimes with \wp [5], our linear negation operator only allows for a "trivialised" form of De Morgan duality. We also redefine the notion of binding, as a symmetric relation whose scope spans the entire sequent. Reduction works by means of an explicit substitution, in the spirit of the operational

semantics of the linear chemical abstract machine [1]. The rules for explicit substitution act globally on the entire typing judgement, instead of limiting their scope to a specific subterm.

By designing our calculus in this way we manage to deconstruct lambda abstraction, one of the traditional primitives of computation, into finer notions of tensor-based binding. This allows us to easily reason with binding operations, such as teleportation, even when they are performed on compound terms. The representation of those operations remains the same, regardless of whether they are teleporting a state or an entire function. A detailed example of this is presented in the end of the Appendix.

An elimination procedure allows us to reconstruct Application using Cut, hence removing it from our primitive rule set. The new rules allow for a fully symmetric language, where inputs and outputs are treated as elements of a symmetric relation, and give rise to a new structural rule called the *dagger-flip*. The resulting set of rules is minimal and simple to use, which allows us to easily prove the properties of subject reduction, confluence, strong normalisation and consistency. Our analysis of the language's semantics is completed by a proof that the dagger lambda calculus is an internal language for dagger compact categories.

2 The dagger lambda calculus

Dagger compact categories were first introduced in [3], albeit under a different name, using some of the terminology of [15]. They were later proposed by [4] and [21] as an axiomatic framework for the study of quantum protocols. Though a lot of work has been done on categorically driven quantum programming languages [23], [24] and [25], these lambda calculi did not provide a way of modelling the dagger functor of dagger compact categories. The work of [8] highlighted the importance of dagger compact categories for the semantics of quantum computation; it presented a rough correspondence between quantum computation, logic and the lambda calculus, yet its type theory fell short of providing a correspondence to the entire structure of dagger compact categories. This section fills this gap by presenting the *dagger lambda calculus*: a computational interpretation for dagger compact categories.

2.1 Language construction

We will now construct a language for *dagger compact categories* by defining well formed formulas for terms, types and sequents. The rules for deriving these formulas will be given in the form of Gentzen-style inference rules. In order to give computational meaning to our language, we will reformalise the typing dynamics of the linear typed lambda calculus [6] with the explicit substitution of the linear chemical abstract machine [1]. The linear negation we will use causes a significant collapse between conjunction and disjunction, extends tensor to a (potentially) binding operator, and provides us with a semantics similar to that of the proof nets in [5]. The set of rules is kept at a minimum, allowing for clean proofs of the various desired properties. Many familiar computational notions do not appear as primitives, but they do arise as constructed notions in good time.

Definition 2.1 (Variables, constants and terms in the dagger lambda calculus). The fundamental building blocks of our language are *variables*; they are denoted by single letters and are traditionally represented using the later letters of the alphabet (i.e. x, y, z). We also allow for the use of *constant terms* (i.e. c_1, c_2, c_3); these are terms with an inherent value and cannot serve as placeholders for substitution. These primitives can then be combined with each other to form composite *terms*, denoted by different combinations of the following forms:

$$\langle term \rangle ::= variable \mid \langle term \rangle_* \mid \langle term \rangle \otimes \langle term \rangle \mid constant$$

Definition 2.2 (Types in the dagger lambda calculus). Every term in our language, regardless of whether it is a variable, a constant or composite, has a *type*. We will first start by defining a set of *atomic types*; these are traditionally represented using capital letters (i.e. A, B, C). Atomic types can then be combined to give us types of the following forms:

$$\langle \text{type} \rangle ::= \text{atomic} \mid \langle \text{type} \rangle^* \mid \langle \text{type} \rangle \otimes \langle \text{type} \rangle$$

The star operator that we use is not a repetition operator; instead, it corresponds to a particular form of *linear negation*. As one would expect from a negation operation, the star operator is involutive $(a_*)_* \equiv a$ and $(A^*)^* \equiv A$. Abramsky [1] proposed using linear negation as the passageway between Intuitionistic Linear Logic and Classical Linear Logic. The linear negation used in [5] "trivialized" the notion of De Morgan duality of [1] by setting $(A \otimes B)^* := A^* \otimes B^*$. The linear negation that we use is similar to the one used in [13]; it distributes differently over tensor by performing a swap of the terms/types at hand and allows for a more "planar" representation. An exchange rule, presented later in this section, will maintain the symmetry of the language's tensors.

Definition 2.3 (Linear negation). The star operator is a form of linear negation whose De Morgan duality is defined by: $(a \otimes b)_* := b_* \otimes a_*$ on terms and $(A \otimes B)^* := B^* \otimes A^*$ on types.

Definition 2.4 (Scalars). One of the language's atomic types, denoted by I , acts as the tensor unit. One of the very important properties of the type I is *negation invariance*, whereby $I \equiv I^*$. We say that a term i is a *scalar* iff it is of type I .

Definition 2.5 (Dimensions). For every type A , we will define a scalar constant $D_A : I$, referring to it as the *dimension* of type A . The dimension of I is defined to be $D_I = 1 : I$, where $1 = 1_* : I \equiv I^*$.

Definition 2.6 (Soup connection). A *soup connection* is an ordered pair of equityped terms. A soup connection between two terms of type A is written as $t_1 :_A t_2$ and is an element of the cartesian product of the terms of type A with themselves. To simplify our notation, we write the connection as $t_1 : t_2$, omitting the type, whenever there is no ambiguity about the type of the connected terms. Soup connections do not form a symmetric relation; we use the property $a_1 : a_2 \equiv a_{2*} : a_{1*}$ to equate some soup terms by collapsing them into the same congruence class. Moreover, soup connections are not self-dual; we define a *negation* on soup connections as $(t : u)_* := t_* : u_* \equiv u : t$.

Definition 2.7 (Soup). A *soup* is a set of soup connections, where not all of the connections have to be of the same type. The resulting soup is of the form $S = \{v_1 : v_2, \dots, v_{m-1} : v_m\}$. All of the computation in our language is performed inside the relational soup, by treating its constituent soup connections as a form of explicit substitution. Our negation extends naturally into a *soup negation* whereby $(S \cup S')_* := S_* \cup S'_*$.

Definition 2.8 (Typing judgements in the dagger lambda calculus). The *typing judgements*, or *sequents*, of our language are composed of terms, their respective types, and a relational soup. A typing judgement is thus represented by:

$$t_1 : A_1, t_2 : A_2, \dots, t_n : A_n \vdash_S t : B$$

Now that we know which formulas are well formed in our language, we can proceed by defining a notion of binding. Contrary to what we are used to from the lambda calculus, where the notion of binding is restricted in scope to the confines of a single term, the dagger lambda calculus supports a binding that is global and whose scope spans the entire typing judgement. The computational interpretation of classical linear logic, which was provided by [1] in his linear chemical abstract machine, views two occurrences of the same variable as two ends of a communication channel. Adhering to the spirit of that definition, we define binding as follows:

Definition 2.9 (Bound variables and terms in the dagger lambda calculus). For any variable x , we say that it is a *bound variable* when it appears twice within a given sequent, regardless of where in the sequent those instances appear. We can also say that an instance of that variable is *captured* by the other instance of the variable in the sequent. As such, variable capture is not limited to the scope of a single term but spans the entire sequent. For any term t that does not contain any occurrences of constants, we say that that term is captured when it consists entirely of variables that are captured within the scope of the current sequent. We use the phrases *bound term* and *bundle of bound variables* interchangeably when referring to captured terms. Trivially, a bound variable is also a bound term.

Remark. As will become obvious from our language's sequent rules, which will impose linearity constraints on the introduction of variables, the nature of linearity in our language mandates that all of the variables within a given sequent occur exactly twice. This means that all of the free variables in a given term will occur once more in the sequent within which they reside, hence becoming captured in the scope of that sequent. Within that scope, all terms will essentially consist of captured variables and constants.

Definition 2.10 (α -renaming on variables in the dagger lambda calculus). A bound variable x can be α -renamed by replacing all of its instances, in a given sequent, with a bundle of bound variables t . The term t has to be of the same type as x , must not contain any constants (since it will be a bundle of bound variables), and it must consist of variables that do not already appear in the sequent.

We can now extend the operation of α -renaming to operate on captured terms:

Definition 2.11 (α -renaming on terms in the dagger lambda calculus). A bound term t can be α -renamed by either α -renaming its constituent variables or, in cases where t appears twice in a given sequent, by replacing all of its instances with a variable x . The variable x has to be of the same type as t and it must not already appear in the sequent.

Definition 2.12 (α -equivalence in the dagger lambda calculus). We define a notion of α -equivalence as the reflexive, symmetric and transitive closure of α -renaming. In other words, we say that two sequents are α -equivalent, or *equivalent up to α -renaming*, when one can be transformed to the other by α -renaming zero or more terms.

Definition 2.13 (Typing contexts in the dagger lambda calculus). The left-hand-side of a typing judgement is actually a list of typed terms. We use the letters Γ and Δ as shorthand for arbitrary (possibly empty) lists of such terms. Let Δ be the list $t_1 : T_1, t_2 : T_2, \dots, t_n : T_n$. We define $\otimes \Delta$ to be the term $((t_1 \otimes t_2) \otimes \dots) \otimes t_n : (((T_1 \otimes T_2) \otimes \dots) \otimes T_n)$, referring to it as Δ in *tensor form*.

Our language exposition features a Gentzen-style Sequent Calculus, which provides us with the inference rules used to produce judgements. Rules with a double line are bidirectional; sequents matching the top of the rule can be used to derive sequents matching the bottom and vice versa. The rules are formed in a way that allows composite terms to appear to the left of the turnstile. The sequent rules are:

$$\begin{array}{c}
 \frac{}{x : A \vdash x : A} \text{Id}, \quad \frac{a : A \vdash_S b : B}{a_* : A^* \vdash_{S_*} b_* : B^*} \text{Negation}, \\
 \frac{\Gamma \vdash_{S_1} a : A \quad a' : A, \Delta \vdash_{S_2} b : B}{\Gamma, \Delta \vdash_{S_1 \cup S_2 \cup \{a : a'\}} b : B} \text{Cut}^*, \quad \frac{a : A, \Gamma \vdash_S b : B}{\Gamma \vdash_S a_* \otimes b : A^* \otimes B} \text{Curry}, \\
 \frac{\Gamma \vdash_{S_1} a : A \quad \Delta \vdash_{S_2} b : B}{\Gamma, \otimes \Delta \vdash_{S_1 \cup S_2} a \otimes b : A \otimes B} \otimes R^*, \quad \frac{\Gamma, a : A, b : B \vdash_S c : C}{\Gamma, a \otimes b : A \otimes B \vdash_S c : C} \otimes L.
 \end{array}$$

*: The sequents merged by the Cut and $\otimes R$ rules must not share any common variables. Whenever we use these two rules on sequents whose variables overlap, we have to α -rename them first to prevent capturing the variables.

Remark. The identity axiom (Id) is the only inference rule we have for introducing variables into our expressions. Consequently, variables are always introduced as bound pairs. The Cut rule establishes a connection between the output of one sequent and the input of another. The $\otimes R$ rule tensors two sequents together, preserving tensor associativity by turning Δ into $\otimes \Delta$. Given the capturing restriction for Cut and $\otimes R$, no other bindings can be introduced in our expressions. As such, variables will appear exactly twice in a sequent. We call this property *linearity*, the sequents *linear*, and the restrictions on Cut and $\otimes R$ *linearity constraints*.

We sometimes use sequents with an empty right-hand-side, for instance $a : A, \Gamma \vdash$ as shorthand for $a : A, \Gamma \vdash 1 : I$. Such sequents are easy to produce by using *Uncurrying*, the inverse of the *Curry* rule, together with the constant $1 : I$:

$$\frac{\frac{\Gamma \vdash a_* : A^* \quad \vdash 1 : I}{\Gamma \vdash a_* \otimes 1 : A^* \otimes I} \otimes R}{a : A, \Gamma \vdash 1 : I} \text{Uncurry}$$

The language has a structural exchange rule that can be used to swap terms on the left hand side of a sequent. When navigating through a proof tree, instances of the exchange rule can be used to keep track of which terms were swapped and at which points during a derivation:

$$\frac{\Gamma, a : A, b : B, \Delta \vdash c : C}{\Gamma, b : B, a : A, \Delta \vdash c : C} \text{Exchange.}$$

Our language also has two unit rules, λ_Γ and ρ_Γ , that are used to more accurately represent scalars:

$$\frac{\Gamma \vdash_{S \cup \{i_* : 1\}} b : B}{i : I, \Gamma \vdash_S b : B} \lambda_\Gamma, \quad \frac{\Gamma \vdash_{S \cup \{i_* : 1\}} b : B}{\Gamma, i : I \vdash_S b : B} \rho_\Gamma.$$

Our language dynamics are defined through soup rules. These rules explain how the relational connections propagate within the soup, giving rise to an operational semantics for a form of "global substitution" that resembles pattern matching on terms. The soup propagation rules, called *bifunctoriality*, *trace* and *cancellation* respectively, are:

$$\begin{aligned} S \cup \{a \otimes b : c \otimes d\} &\longrightarrow S \cup \{a : c, b : d\} \\ S \cup \{x :_A x\} &\longrightarrow S \cup \{D_A : 1\} \\ S \cup \{1 : 1\} &\longrightarrow S \end{aligned}$$

where ψ is a constant and x is a variable. Our soup rules also contain a *consumption rule*. This rule uses up a relational connection between $\{t : u\}$ to perform a substitution in the typing judgement. Note, however, that the term we are substituting for has to be one that was captured in the scope of the sequent:

$$\Gamma \vdash_{S \cup \{t : u\}} b : B \longrightarrow \left(\Gamma \vdash_S b : B \right) \begin{cases} [t/u], & \text{if } u \text{ does not contain constants} \\ [u/t], & \text{if } t \text{ does not contain constants} \end{cases}$$

If t and u are both without constants, linearity implies that their constituent variables were all captured in the scope of the original sequent. In such a case, we can choose the way in which we want to substitute. This gives us a symmetric notion of substitution, where our choice of substitution does not affect the typing judgement, as the sequents will be equivalent up to alpha renaming.

Definition 2.14 (Soup reduction). We use the term *soup reduction* to refer to the binary relation that extends α -equivalence with the sequent transformations that are caused by applying one of the soup rules. Thus, for two sequents $\Gamma \vdash_{S_1} t : T$ and $\Gamma \vdash_{S_2} t : T$, if the soup S_1 is transformed into S_2 through the application of one of the soup propagation rules, $S_1 \rightarrow S_2$, then we say that one sequent reduces to the other via *soup reduction*. Similarly, if a sequent J_1 is transformed into J_2 by using the consumption rule to perform a substitution, we say that J_1 reduces to J_2 via *soup reduction*.

Definition 2.15 (Soup equivalence). We define a notion of *soup equivalence* as the reflexive, symmetric, and transitive closure of soup reduction. In other words, we say that two sequents J_1 and J_2 are *soup-equivalent*, or *equivalent up to soup-reduction*, when we can convert one to the other by using zero or more instances of α -renaming and soup reduction.

We can now use the rules that we have defined so far in order to express the computational notion of application:

Definition 2.16 (Application in the dagger lambda calculus). Let t and f be terms such that $t : A$ and $f : A^* \otimes B$ for some types A and B . We define the *application* ft as a notational shorthand for representing a variable $x : B$, along with a connection in our soup. The origins of the application affect the structure of its corresponding soup connection:

$$ft : B, \Gamma \vdash c : C := x : B, \Gamma \vdash_{\{f:t_* \otimes x\}_*} c : C \quad \text{and} \quad \Gamma \vdash ft : B := \Gamma \vdash_{\{f:t_* \otimes x\}} x : B$$

For an application originating inside our soup, we have:

$$\{ft : c\} := \{x : c\} \cup \{f : t_* \otimes x\} \quad \text{and} \quad \{c : ft\} := \{c : x\} \cup \{f : t_* \otimes x\}_*$$

Corollary 2.1 (Beta reduction). *This immediately allows us to represent a form of beta reduction. Instead of relying on an implicit meta-concept of substitution, our beta reduction is going to express the binding and reduction of terms by connecting them in the soup by setting $(a_* \otimes b)t \xrightarrow{B} b$, while causing $\{t : a\}$ or $\{t : a\}_*$ to be added to the relational soup.*

Proof. This is derived from our definition of application because $(a_* \otimes b)t$ represents a variable x along with one of two possible connections in our soup. The soup connection can be manipulated into:

$$\{a_* \otimes b : t_* \otimes x\} \rightarrow \{a_* : t_*, b : x\} \rightarrow \{t : a\} \cup \{b : x\}$$

$$\{a_* \otimes b : t_* \otimes x\}_* \rightarrow \{a_* : t_*, b : x\}_* \rightarrow \{t : a\}_* \cup \{b : x\}$$

The connection between b and x can then be consumed to change the variable x into a b . All that remains is $\{t : a\}$ or $\{t : a\}_*$. \square

Now that all of the language's rules are in place, we can demonstrate how the familiar notion of lambda abstraction can be reconstructed from the finer notions of linear negation and tensor, by defining it to be a notational shorthand:

Definition 2.17 (Lambda abstraction in the dagger lambda calculus). Let $\lambda a.b := a_* \otimes b$ and $A \multimap B := A^* \otimes B$

The following combinators are used in the rest of this paper:

$$\begin{aligned} id_A &:= \lambda a.a \text{ (where } a : A) & \bar{b} &:= \lambda g.\lambda f.\lambda a.g(fa) \\ \bar{s} &:= \lambda (a \otimes b).(b \otimes a) & \bar{t} &:= \lambda f.\lambda g.\lambda (x_1 \otimes x_2).(fx_1 \otimes gx_2) \end{aligned}$$

Theorem 2.1 (Admissibility of $\multimap E$). *We can also use the definition of application to demonstrate that an implication elimination rule ($\multimap E$) is admissible within our set of rules.*

Proof. A proof for this can be found in the Appendix. \square

We define some additional notational conventions, so that we can more easily describe the reversal in the causal order of computation:

Definition 2.18 (Complex conjugation). Let $f : A^* \otimes B$ be an arbitrary function. As a notational convention, we set $f^* := \bar{s}f : B \otimes A^*$.

Theorem 2.2 (Admissibility of \dagger -flip). *We can use the language's rules and definitions in order to admit a new structural rule called the \dagger -flip. This rule contains all the computational symmetry that we will later need in order to model the dagger functor.*

Proof. A proof for this can be found in the Appendix. □

Lemma 2.1 (Interchangeability of \dagger -flip and Negation). *Alternatively, we could have defined the language by including \dagger -flip in our initial set of sequent rules. That would have allowed us to admit the Negation rule as a derived rule.*

Proof. A proof for this can be found in the Appendix. □

2.2 Language properties

Our lambda calculus was designed with a minimal set of rules. This has led to a tractable language, where most of the properties are easy to prove by structural induction. Throughout the rest of this section, we establish that our lambda calculus satisfies the following important properties of a calculus: subject reduction, confluence, strong normalisation, and consistency. Sketches of the proofs are presented in the Appendix and more detailed versions can be found in [7].

2.2.1 Subject reduction

The first thing we have to prove, in order to demonstrate that our typing system is well defined, is the consistency of our typing dynamics. In other words, we have to verify that the way in which relational connections propagate through our soup preserves type assignments. This is easy to observe since our soup only connects *equitytyped* terms. Pair consumption substitutes a term for another of the same type, thus preserving types.

Theorem 2.3 (Subject reduction). *Let J_1 and J_2 be two typing judgements such that $J_1 = \Gamma \vdash_S t_1 : A_1$ and $J_2 = \Delta \vdash_{S'} t_2 : A_2$. Suppose that these two judgements are such that we can use a soup reduction rule $S \longrightarrow S'$ to reduce one to the other: $J_1 \longrightarrow J_2$. Then, the reduction will not alter type assignments in any way: $\text{types}(\Gamma) = \text{types}(\Delta)$ and $A_1 \equiv A_2$.*

2.2.2 Normalisation

Strong normalisation is a highly sought after property for lambda calculi, primarily because of the implications it has on the practical implementation of the language. A reduction that is strongly normalising implies that every sequent has a normal form. Furthermore, it requires that the normal form is attained after a finite number of steps, without any chance of running into an infinite reduction loop.

Theorem 2.4 (Strong normalisation). *Every sequence of soup reduction steps is finite and ends with a typing judgement that is in normal form.*

2.2.3 Confluence

Another very important property for our language is the Church-Rosser property. It ensures that we can end up with the same sequent regardless of the reduction path we choose to follow. A careful observation of our rewrite rules will reveal that the rules are all left-linear.

Lemma 2.2 (Left-linearity). *All of our soup rewrite rules are left-linear.*

One should note, at this point, that our soup rules do exhibit a form of "harmless" overlap. More specifically, the consumption rule $(S \cup \{t : u\} \longrightarrow S)$ forms a critical pair with itself in cases where t and u are both bound. Fortunately, as we will see in the next lemma, these pairs prove to be *trivial* as they correspond to sequents that are equivalent up to α -renaming.

Lemma 2.3 (Symmetry of substitution). *Let J be a typing judgement of the form $J := \Gamma \vdash_{S \cup \{t:u\}} a : A$, where t and u are both bound. The connection $\{t : u\}$ can be consumed in either of two ways; one substitutes t for u and the other substitutes u for t in the typing judgement. Let's call these J_1 and J_2 respectively. J_1 will then be α -equivalent to J_2 .*

Corollary 2.2 (No overlap). *The rewrite rules have no overlap up to α -equivalence of typing judgements.*

Theorem 2.5 (Confluence). *Our reduction rules have the Church-Rosser property.*

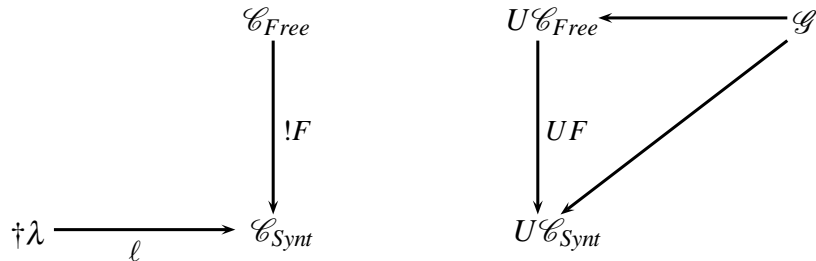
2.2.4 Consistency

In order to show that our type theory is consistent, we have to show that our soup dynamics do not collapse all equityped terms to the same element.

Theorem 2.6 (Consistency). *There exist two terms of the same type, henceforth referred to as t_1 and t_2 , such that $\Gamma \vdash_{S_1} t_1 : A$ and $\Gamma \vdash_{S_2} t_2 : A$ could never reduce to the same typing judgement.*

2.3 Correspondence to dagger compact categories

The purpose of this section is to provide a full Curry-Howard-Lambek correspondence between the dagger lambda calculus and dagger compact categories. We start by defining a directed graph \mathcal{G} , representing a signature for dagger compact categories. We then show how that graph can be interpreted to define the free dagger compact category \mathcal{C}_{Free} and the dagger lambda calculus $\dagger\lambda$. An appropriate Cut-elimination procedure is defined to partition the sequents of the dagger lambda calculus into equivalence classes up to soup equivalence. The resulting equivalence classes are modular proof invariants represented by denotations. We show that the types and denotations can be used to form a syntactic category, \mathcal{C}_{Synt} , and prove that the category is dagger compact. The diagram below, fashioned to resemble the diagram at the bottom of page 49 in [17], is provided to help visualise the Curry-Howard-Lambek correspondence. In this diagram, $U\mathcal{C}_{Free}$ and $U\mathcal{C}_{Synt}$ are the underlying graphs of their respective categories, where identities, composition, natural isomorphisms and other structural elements of the parent categories have been "forgotten" by applying the forgetful functor U . F is the unique functor between the free and the syntactic category, that satisfies the rest of the conditions in the diagram.



We will prove an equivalence between the free category and the syntactic category. We should note at this point that our typing conventions of an involutive negation ($A \equiv (A^*)^*$) and negation invariance of the tensor unit ($I \equiv I^*$) implicitly introduce equivalence classes on types. Our proof of equivalence will be achieved by fully exhibiting the correspondence in objects and arrows between the two categories, showing that their notions of equality overlap, up to the equivalence classes that are induced by our typing conventions.

2.3.1 A signature for dagger compact categories

The notion of signature that we will use combines the algebraic signature of [22] with the directed graph used by [17]. Consider a set of object variables Σ_0 . Using the tensor operation, an associated tensor identity, and the duality operator star, we can construct the free (\otimes, I, \square^*) -algebra over Σ_0 . This corresponds to the set of all object terms or vertices in a compact closed category and will be denoted by $Dagger(\Sigma_0)$. Now consider a set Σ_1 of morphism variables or edges between those vertices. Let dom, cod be a pair of functions such that $dom, cod : \Sigma_1 \longrightarrow Dagger(\Sigma_0)$. Throughout the rest of this section, we will be referring to the graph \mathcal{G} as the directed graph whose vertices and edges are defined by $Dagger(\Sigma_0)$ and Σ_1 . This graph forms the signature upon which we will base both the dagger lambda calculus and our description of the free dagger compact category; it includes all of the symbols but none of the logic of the languages that we want to describe.

2.3.2 The free dagger compact category

We will now show how to define the free dagger compact category \mathcal{C}_{Free} as an interpretation of the graph \mathcal{G} . A highly intuitive introduction to free categories and how they can be generated from directed graphs can be found in [17]. Furthermore, a more extensive presentation of the process of constructing various kinds of free categories can be found in [22]. A more detailed presentation of the incremental buildup to the construction of free dagger compact categories can also be found in [2].

The set of objects for the free category in this section will be the same as the set of vertices $Dagger(\Sigma_0)$ in the graph \mathcal{G} . The set of edges Σ_1 in the graph is used to generate morphisms for the free category. Thus, an edge of the form $f : A \rightarrow B$ generates an arrow in \mathcal{C}_{Free} which we will denote as $\langle A, f, B \rangle$. The identities are represented by: $\langle A \rangle, \langle B \rangle, \langle C \rangle, \dots$

The free category over a directed graph, also referred to as a path category, includes morphisms that correspond to the paths generated by combining adjoining edges in \mathcal{G} . These morphisms are formed using the free category's composition operation. Given two morphisms $\langle A, f, B \rangle$ and $\langle B, g, C \rangle$, we write their composition in \mathcal{C}_{Free} as $\langle A, f, B, g, C \rangle$. The *tensor* operation, *monoidal isomorphisms*, *unit* and *counit*, and the *dagger* and *star* operations are described in the Appendix.

2.3.3 The dagger lambda calculus

This section demonstrates how the graph signature \mathcal{G} can be interpreted to derive the dagger lambda calculus. The set of types used by $\dagger\lambda$ is precisely the set of vertices $Dagger(\Sigma_0)$ used in graph \mathcal{G} . Every edge $f : A \rightarrow B$ in Σ_1 is interpreted as a sequent $a : A \vdash_{\{f : a_* \otimes b\}} b : B$ up to alpha-equivalence. These interpretations essentially introduce constants, in our case $f : A^* \otimes B$, written as sequents that are reminiscent of η -expanded forms. The rest of the rules of the dagger lambda calculus can be used to process and combine sequents, yielding a richer logical structure.

2.3.4 The syntactic category

Following a method that is similar to [18], we will define a process of Cut-elimination by using the soup reduction relation to partition the sequents of the dagger lambda calculus into equivalence classes. The resulting equivalence classes are modular proof invariants called *denotations*. This section demonstrates how these denotations give rise to the *syntactic category* $\mathcal{C}_{\text{Synt}}$, a dagger compact category. Sketches of the proofs are presented in the Appendix and more detailed versions can be found in [7].

Definition 2.19 (Denotations). We will use the term *denotations* to refer to the equivalence classes that are formed by partitioning the sequents of the lambda calculus according to soup equivalence. Hence, two sequents will correspond to the same denotation if and only if they are equivalent up to soup reduction.

Theorem 2.7 (The syntactic category). *The types of the lambda calculus and the denotations generated by soup equivalence form a category whose objects are types and whose arrows are denotations.*

Theorem 2.8 (Dagger compact closure). *The syntactic category is a dagger compact category.*

2.3.5 Proof of equivalence

We will now prove that the free dagger compact category $\mathcal{C}_{\text{Free}}$ is equivalent to the syntactic category $\mathcal{C}_{\text{Synt}}$.

Lemma 2.4 (Essentially surjective on objects). *The set of objects in the free category and the set of objects in the syntactic category are surjective, up to isomorphism.*

Proof. Recall $\text{Dagger}(\Sigma_0)$; the free (\otimes, I, \square^*) -algebra over the set of object variables Σ_0 . The sets of objects in $\mathcal{C}_{\text{Free}}$ and $\mathcal{C}_{\text{Synt}}$ both correspond to $\text{Dagger}(\Sigma_0)$, up to the equivalence classes induced by $(A^*)^* \equiv A$ and $I^* \equiv I$. \square

Lemma 2.5 (Equal arrows correspond to equal denotations). *If two arrows, $\langle A, f, B \rangle$ and $\langle A, f', B \rangle$ are equal in the free category, then they will also be equal in the syntactic category: $[f] = [f'] : A \rightarrow B$.*

Proof. The structure of the free category $\mathcal{C}_{\text{Free}}$ imposes the minimum number of equalities for a category to be dagger compact. Moreover, both the free category and the syntactic category derive their symbols from the same signature graph \mathcal{G} . Since we have already shown that $\mathcal{C}_{\text{Synt}}$ is dagger compact, the same steps can be used to show that any arrows $\langle A, f, B \rangle$ and $\langle A, f', B \rangle$ that are equal in the free category correspond to equal denotations $[f] = [g]$ in the syntactic category. \square

Lemma 2.6 (Equal denotations correspond to equal arrows). *Any denotations that are equal in the syntactic category correspond to equal arrows in the free category.*

Proof. Let $[f] : \Gamma \rightarrow B$ and $[g] : \Gamma \rightarrow B$ be denotations in the syntactic category such that $[f] = [g]$. Since the two denotations are equal, the sequents they represent in the dagger lambda calculus must be equivalent up to soup reduction. Without loss of generality, let's assume that $[f]$ represents a sequent J_1 and that $[g]$ represents a sequent J_2 , where $J_1 \rightarrow J_2$. The soup reduction relation consists of four soup rules: *bifunctionality*, *trace*, *cancellation*, and *consumption*. We prove this lemma by induction on the structure of the soup reduction linking J_1 and J_2 . The details of the induction have been omitted in this extended abstract; they are, available in [7]. This shows that $\langle \Gamma, f, B \rangle = \langle \Gamma, g, B \rangle$. \square

Theorem 2.9 (Equivalence between the free category and the syntactic category). *The free dagger compact category $\mathcal{C}_{\text{Free}}$ and the syntactic category $\mathcal{C}_{\text{Synt}}$ are equivalent.*

Proof. The two categories derive their symbols from a common signature graph \mathcal{G} . As we have already shown, bearing in mind the equivalence classes that we have induced on types, the categories are essentially surjective on objects. Moreover, arrows that are equal in the free category are equal in the syntactic category and vice versa. This means that the functor F is *full* and *faithful*, causing the notions of equality between arrows to overlap in these two categories. Consequently, the categories are equivalent. \square

Corollary 2.3 (Internal language). *The dagger lambda calculus is an internal language for dagger compact categories.*

3 Conclusion

This paper has presented a lambda calculus for dagger compact categories. As we have seen from [4], this language can be used to represent a subset of quantum computation, namely, quantum protocols. The dagger lambda calculus was shown to satisfy subject reduction, confluence, strong normalisation, and consistency, while the language was shown to be an internal language for dagger compact categories.

In order to be able to cover all of quantum computation, commonly referred to as universal quantum computation, we need a language with classical control. One way of adding this feature in a denotationally sound way is by extending our language's axiomatisation to include classical basis states. This can be achieved by introducing complementary classical structures, like the ones built on top of the dagger compact structure in [12], [10] and [13]. This work is partly covered by [7] and will be included in a forthcoming paper.

References

- [1] Samson Abramsky (1993): *Computational interpretations of linear logic*. *Theoretical Computer Science* 111, pp. 3–57. (DOI:10.1.1.16.2984).
- [2] Samson Abramsky (2005): *Abstract Scalars, Loops, and Free Traced and Strongly Compact Closed Categories*. In: *In Proceedings of the First Conference on Algebra and Coalgebra in Computer Science (CALCO 2005)*, 3629, Springer Lecture Notes in Computer Science, pp. 1–31. (arXiv:0910.2931v1 [quant-ph]).
- [3] Samson Abramsky, Rick Blute & Prakash Panangaden (1999): *Nuclear and trace ideals in tensored *-categories*. *Journal of Pure and Applied Algebra* 143, pp. 3–47.
- [4] Samson Abramsky & Bob Coecke (2004): *A categorical semantics of quantum protocols*. In: *Proceedings of the 19th IEEE conference on Logic in Computer Science (LiCS'04)*, IEEE Computer Science Press. (arXiv:quant-ph/0402130v5).
- [5] Samson Abramsky & Ross Duncan (2006): *A Categorical Quantum Logic*. *Mathematical Structures in Computer Science* 16, pp. 469–489. (arXiv:quant-ph/0512114v1).
- [6] Samson Abramsky & Nikos Tzevelekos (2010): *Introduction to categories and categorical logic*. In Bob Coecke, editor: *New Structures for Physics*, Springer Lecture Notes in Physics. (arXiv:1102.1313v1 [math.CT]).
- [7] Philip Atzemoglou (2013): *Higher-order semantics for quantum programming languages with classical control*. Ph.D. thesis, Oxford University Computing Laboratory. (arXiv:1311.6563v1 [cs.LO]).
- [8] John Baez & Michael Stay (2010): *Physics, topology, logic and computation: A Rosetta Stone*. In Bob Coecke, editor: *New Structures for Physics*, Springer Lecture Notes in Physics. (arXiv:0903.0340v3 [quant-ph]).
- [9] Bob Coecke & Ross Duncan (2008): *Interacting quantum observables*. In: *Proceedings of the 35th International Colloquium on Automata, Languages and Programming*, Lecture Notes in Computer Science 5126, Springer-Verlag, pp. 298–310. (arXiv:0906.4725v1 [quant-ph]).

- [10] Bob Coecke & Ross Duncan (2011): *Interacting quantum observables: Categorical algebra and diagrammatics*. *New Journal of Physics* 13, p. 043016. (arXiv:0906.4725v3 [quant-ph]).
- [11] Bob Coecke & Éric Oliver Paquette (2006): *POVMs and Naimark's theorem without sums*. *Electronic Notes in Theoretical Computer Science*. (arXiv:quant-ph/0608072).
- [12] Bob Coecke, Éric Oliver Paquette & Duško Pavlović (2010): *Classical and quantum structuralism*. In S. Gay & I. Mackie, editors: *Semantic Techniques in Quantum Computation*, Cambridge University Press. (arXiv:0904.1997v2 [quant-ph]).
- [13] Bob Coecke, Éric Oliver Paquette & Simon Perdrix (2008): *Bases in diagrammatic quantum protocols*. *Electronic Notes in Theoretical Computer Science* 218, pp. 131–152. (arXiv:0808.1029v1 [quant-ph]).
- [14] Bob Coecke & Duško Pavlović (2007): *Quantum measurements without sums*. In G. Chen, L. Kauffman & S. Lamonaco, editors: *Mathematics of Quantum Computing and Technology*, Taylor and Francis, pp. 567–604. (arXiv:quant-ph/0608035).
- [15] Sergio Doplicher & John E. Roberts (1989): *A new duality theory for compact groups*. *Inventiones mathematicae* 98(1), pp. 157–218. Available at <http://eudml.org/doc/143725>.
- [16] Jan Willem Klop (1992): *Term rewriting systems*. In S. Abramsky, D.M. Gabbay & T.S.E. Maibaum, editors: *Handbook of Logic in Computer Science*, 2, Oxford University Press, pp. 1–116. (DOI:10.1.1.35.425).
- [17] Saunders Mac Lane (1998): *Categories for the Working Mathematician*, second edition. Springer.
- [18] Paul-André Melliès (2009): *Categorical Semantics of Linear Logic*. *Panoramas et synthèses - Société mathématique de France* (27), pp. 1–196. (DOI:10.1.1.62.5117).
- [19] Peter Selinger (2004): *A brief survey of quantum programming languages*. In: *Proceedings of the 7th International Symposium on Functional and Logic Programming*, 2998, Springer Lecture Notes in Computer Science, Nara, Japan, pp. 1–6. (DOI:10.1.1.94.463).
- [20] Peter Selinger (2004): *Towards a quantum programming language*. *Mathematical Structures in Computer Science* 14(4), pp. 527–586. (DOI:10.1.1.144.6380).
- [21] Peter Selinger (2007): *Dagger compact closed categories and completely positive maps*. In: *Proceedings of the 3rd International Workshop on Quantum Programming Languages (QPL 2005)*, 170, *Electronic Notes in Theoretical Computer Science*, Chicago, pp. 139–163. (DOI:10.1.1.134.2476).
- [22] Peter Selinger (2010): *A survey of graphical languages for monoidal categories*. In Bob Coecke, editor: *New Structures for Physics*, Springer Lecture Notes in Physics. (arXiv:0908.3347v1 [math.CT]).
- [23] Peter Selinger & Benoît Valiron (2006): *A lambda calculus for quantum computation with classical control*. *Mathematical Structures in Computer Science* 16(3), pp. 527–552. (arXiv:cs/0404056v2 [cs.LO]).
- [24] Peter Selinger & Benoît Valiron (2008): *A linear-non-linear model for a computational call-by-value lambda calculus (extended abstract)*. In: *Proceedings of the Eleventh International Conference on Foundations of Software Science and Computation Structures (FOSSACS 2008)*, 4962, Springer Lecture Notes in Computer Science, Budapest, pp. 81–96. (arXiv:0801.0813v1 [cs.LO]).
- [25] Peter Selinger & Benoît Valiron (2010): *Quantum lambda calculus*. In S. Gay & I. Mackie, editors: *Semantic Techniques in Quantum Computation*, Cambridge University Press. [Http://www.mscs.dal.ca/selinger/papers.html#qlambdabook](http://www.mscs.dal.ca/selinger/papers.html#qlambdabook).
- [26] André van Tonder (2004): *A Lambda Calculus for Quantum Computation*. *SIAM Journal on Computing* 33(5), pp. 1109–1135. (arXiv:quant-ph/0307150v5).
- [27] André van Tonder & Miquel Dorca (2003): *Quantum Computation, Categorical Semantics and Linear Logic*. Archive. (arXiv:quant-ph/0312174v4).

A Appendix

A.1 Language construction

Theorem 2.1 (Admissibility of $\multimap E$). *We can also use the definition of application to demonstrate that an implication elimination rule ($\multimap E$) is admissible within our set of rules.*

$$\begin{array}{c}
 \text{Proof.} \quad \frac{\Gamma \vdash_{S_1} t : A}{\Gamma, \Delta \vdash_{S_1 \cup S_2 \cup \{t : a, f : a_* \otimes b\}} b : B} \text{Cut} \\
 \frac{\Delta \vdash_{S_2} f : A^* \otimes B \quad \frac{\frac{\frac{a : A \vdash a : A}{a_* : A^* \vdash a_* : A^*} \quad \frac{b : B \vdash b : B}{a_* : A^*, b : B \vdash a_* \otimes b : A^* \otimes B}}{a_* \otimes b : A^* \otimes B \vdash a_* \otimes b : A^* \otimes B} \text{Cut}}{\Delta \vdash_{S_2 \cup \{f : a_* \otimes b\}} a_* \otimes b : A^* \otimes B} \text{Uncurry} \\
 \frac{\Gamma, \Delta \vdash_{S_1 \cup S_2 \cup \{t : a, f : a_* \otimes b\}} b : B}{\Gamma, \Delta \vdash_{S_1 \cup S_2 \cup \{f : t_* \otimes b\}} b : B} \text{Cut} \\
 \frac{\Gamma, \Delta \vdash_{S_1 \cup S_2} ft : B}{\Gamma, \Delta \vdash_{S_1 \cup S_2} ft : B}
 \end{array}$$

□

Theorem 2.2 (Admissibility of \dagger -flip). *We can use the language's rules and definitions in order to admit a new structural rule called the \dagger -flip. This rule contains all the computational symmetry that we will later need in order to model the dagger functor.*

$$\begin{array}{c}
 \text{Proof.} \quad \frac{\frac{\frac{a : A \vdash_S b : B}{a_* : A^* \vdash_{S_*} b_* : B^*} \text{Negation}}{b : B, a_* : A^* \vdash_{S_*} b_* : B^*} \text{Uncurry}}{\frac{a_* : A^*, b : B \vdash_{S_*} b_* : B^*}{b : B \vdash_{S_*} a : A} \text{Curry}} \text{Exchange}
 \end{array}$$

□

Lemma 2.1 (Interchangeability of \dagger -flip and Negation). *Alternatively, we could have defined the language by including \dagger -flip in our initial set of sequent rules. That would have allowed us to admit the Negation rule as a derived rule.*

$$\begin{array}{c}
 \text{Proof.} \quad \frac{\frac{\frac{a : A \vdash_S b : B}{b : B \vdash_{S_*} a : A} \dagger\text{-flip}}{a_* : A^*, b : B \vdash_{S_*} b_* : B^*} \text{Uncurry}}{\frac{b : B, a_* : A^* \vdash_{S_*} b_* : B^*}{a_* : A^* \vdash_{S_*} b_* : B^*} \text{Curry}} \text{Exchange}
 \end{array}$$

□

A.2 Language properties

Theorem 2.3 (Subject reduction). *Let J_1 and J_2 be two typing judgements such that $J_1 = \Gamma \vdash_S t_1 : A_1$ and $J_2 = \Delta \vdash_{S'} t_2 : A_2$. Suppose that these two judgements are such that we can use a soup reduction rule $S \longrightarrow S'$ to reduce one to the other: $J_1 \longrightarrow J_2$. Then, the reduction will not alter type assignments in any way: $\text{types}(\Gamma) = \text{types}(\Delta)$ and $A_1 \equiv A_2$.*

Proof. A longer version of this proof can be found in [7]. The only soup rule that could affect the premises and conclusion of a typing judgement is the consumption rule. The resulting substitution may be global in scope, but it does not affect the sequent's typing, since it is substituting one term for another one of the same type. \square

Theorem 2.4 (Strong normalisation). *Every sequence of soup reduction steps is finite and ends with a typing judgement that is in normal form.*

Proof. A longer version of this proof can be found in [7], using an induction on the size and structure of the soup reduction. A sequent not in normal form will have a soup with at least one usable connection, for which there are four possible reduction steps. A step using the *trace*, *cancellation* or *consumption* rule will use up that soup connection, the soup being a finite set, leaving us with a smaller usable soup. A step using the *bifunctionality* rule, bounded in its application by the number of atomic types, will split the soup connection into simpler subtypes. \square

Lemma 2.2 (Left-linearity). *All of our soup rewrite rules are left-linear.*

Proof. In accordance with the linearity constraints of our language, no variable appears more than twice on the left hand side of any of our soup reduction rules. \square

Lemma 2.3 (Symmetry of substitution). *Let J be a typing judgement of the form $J := \Gamma \vdash_{S \cup \{t:u\}} a : A$, where t and u are both bound. The connection $\{t : u\}$ can be consumed in either of two ways; one substitutes t for u and the other substitutes u for t in the typing judgement. Let's call these J_1 and J_2 respectively. J_1 will then be α -equivalent to J_2 .*

Proof. Since t and u are both bound, by linearity, we know that they appear exactly once in $\Gamma \vdash_S a : A$. After substitution is performed, J_1 will have two occurrences of t where t and u used to be, so t will be a bound term in that judgement. Similarly, J_2 will have two occurrences of u where t and u used to be, so u will be a bound term in that judgement. These bound terms occur in the exact same spots, so we can *alpha*-rename J_1 to J_2 and vice versa. \square

Theorem 2.5 (Confluence). *Our reduction rules have the Church-Rosser property.*

Proof. Our set of rewrite rules is *left-linear* and has no significant overlap, since it only gives rise to critical pairs that are *trivial* up to α -equivalence. Therefore, our rewrite rules constitute a *weakly orthogonal* rewrite system, which is *weakly confluent* according to [16] (Consider the variation of Theorem 2.1.5 for *weakly orthogonal* TRS's on page 72). Since the rewrite system is both strongly normalising and weakly confluent, we can use Newman's lemma to conclude that it also possesses the Church-Rosser property. See [16] for a more detailed explanation of the properties of orthogonal rewriting systems. \square

Theorem 2.6 (Consistency). *There exist two terms of the same type, henceforth referred to as t_1 and t_2 , such that $\Gamma \vdash_{S_1} t_1 : A$ and $\Gamma \vdash_{S_2} t_2 : A$ could never reduce to the same typing judgement.*

Proof. Consider two combinators of the same type, $t_1 = id_{A \otimes A}$ and $t_2 = \bar{s}_{A \otimes A}$. Both terms are closed, containing no free variables or constants. The sequents $\vdash id_{A \otimes A} : (A \otimes A) \multimap (A \otimes A)$ and $\vdash \bar{s}_{A \otimes A} : (A \otimes A) \multimap (A \otimes A)$ are distinct normal forms: They are clearly distinct from one another and cannot be further reduced using any of our rules, thereby proving that they could never reduce to the same typing judgement. \square

A.3 Correspondence to dagger compact categories

A.3.1 The free dagger compact category

Since the free category is a monoidal category, it allows us to consider two of the graph's edges concurrently by bringing together their corresponding categorical morphisms using a monoidal tensor product. Given two morphisms $\langle A, f, B \rangle$ and $\langle C, h, D \rangle$, we write their tensor product as $\langle A \otimes C, f \otimes h, B \otimes D \rangle$.

The free category generated by the graph \mathcal{G} also includes a number of morphisms that are part of the dagger compact logical structure. The monoidal natural isomorphisms are written as:

$$\langle A \otimes (B \otimes C), \alpha_{A,B,C}, (A \otimes B) \otimes C \rangle \quad \langle I \otimes A, \lambda_A, A \rangle \quad \langle A \otimes I, \rho_A, A \rangle$$

The symmetry isomorphism, and the units and counits are written as:

$$\langle A \otimes B, \sigma_{A,B}, B \otimes A \rangle \quad \langle I, \eta_A, A^* \otimes A \rangle \quad \langle A \otimes A^*, \epsilon_A, I \rangle$$

For every map $\langle A, f, B \rangle$ in the free category, the dagger compact logical structure contains maps f_* and f^\dagger , represented by $\langle A^*, f_*, B^* \rangle$ and $\langle B, f^\dagger, A \rangle$ respectively. When acting on compositions of paths, such as $\langle A, f, B, g, C, \dots, X, h, Y, t, Z \rangle$, the dagger operator reverses the order of operations, yielding:

$$\langle Z, t^\dagger, Y, h^\dagger, X, \dots, C, g^\dagger, B, f^\dagger, A \rangle$$

A.3.2 The syntactic category

Theorem 2.7 (The syntactic category). *The types of the lambda calculus and the denotations generated by soup equivalence form a category whose objects are types and whose arrows are denotations.*

Proof. As we noticed during the proof of the subject reduction property, soup reduction rules do not affect our language's type assignments. Consequently, the type of the premises used by a sequent will be the same across all sequents in a given denotation. Similarly, the type of the conclusion produced by a sequent will be the same across all sequents in a given denotation. For any sequent $\Gamma \vdash_S b : B$, corresponding to a denotation $[\pi_1]$, we will say that its *domain* is Γ and its *codomain* is B , writing this as $[\pi_1] : \Gamma \rightarrow B$.

Let $[f] : A \rightarrow B$ and $[g] : B \rightarrow C$ be denotations representing the soup equivalent forms of some sequents $a : A \vdash_{S_1} b : B$ and $b' : B \vdash_{S_2} c : C$ respectively. For any two such denotations, where the codomain of the first matches the domain of the second, we will define a *composition* operator \circ that can combine them into $[g] \circ [f] : A \rightarrow C$. The new denotation will represent all the soup equivalent forms of the sequent that is generated by combining the two sequents using the Cut rule:

$$\frac{a : A \vdash_{S_1} b : B \quad b' : B \vdash_{S_2} c : C}{a : A \vdash_{S_1 \cup S_2 \cup \{b:b'\}} c : C} \text{Cut}$$

The composition operation we just defined inherits associativity from the Cut rule; the order in which Cuts are performed does not matter since the connected terms are allowed to "float" freely within the soup. Therefore, $[h] \circ ([g] \circ [f]) = ([h] \circ [g]) \circ [f]$. Moreover, for every type A , there is a denotation $[id_A]$ that represents the sequent generated by the Identity axiom (Id): $x : A \vdash x : A$.

Composing a denotation $[f] : A \rightarrow B$ with an identity yields $[f] \circ [id_A]$ or $[id_B] \circ [f]$ depending on whether we compose with an identity on the right or on the left. The two resulting denotations represent

$$\frac{x : A \vdash x : A \quad a : A \vdash_S b : B}{x : A \vdash_{S \cup \{x:a\}} b : B} \quad \text{and} \quad \frac{a : A \vdash_S b : B \quad x : B \vdash x : B}{a : A \vdash_{S \cup \{b:x\}} x : B}$$

both of which are soup equivalent to $a : A \vdash_S b : B$ and the rest of the sequents represented by $[f]$. Hence $[id_B] \circ [f] = [f] = [f] \circ [id_A]$ \square

Definition A.1 (Syntactic category notational conventions). For notational convenience, we define the following combinators:

$$\begin{aligned} \alpha_{A,B,C} &:= \lambda (a \otimes (b \otimes c)). ((a \otimes b) \otimes c) : (A \otimes (B \otimes C)) \multimap ((A \otimes B) \otimes C) \\ \eta_A &:= \lambda 1. (x_* \otimes x) : I \multimap (A^* \otimes A) & \lambda_A &:= \lambda (1 \otimes a). a : (I \otimes A) \multimap A & \rho_A &:= \lambda (a \otimes 1). a : (A \otimes I) \multimap A \\ \varepsilon_A &:= \lambda (x \otimes x_*). 1 : (A \otimes A^*) \multimap I & \sigma_{A,B} &:= \lambda (a \otimes b). (b \otimes a) : (A \otimes B) \multimap (B \otimes A) \end{aligned}$$

Theorem A.1 (Monoidal category). *The syntactic category is a monoidal category.*

Proof. Let $[f] : A \rightarrow B$ and $[g] : C \rightarrow D$ be denotations representing the soup equivalent forms of $a : A \vdash_{S_1} b : B$ and $c : C \vdash_{S_2} d : D$. For any such $[f]$ and $[g]$, we define a monoidal product $[f] \otimes [g] : A \otimes B \rightarrow C \otimes D$. The product represents all the soup equivalent sequents generated by using the right tensor rule to combine the sequents for $[f]$ and $[g]$. We can now use soup reduction to show that $([g] \circ [f]) \otimes ([t] \circ [h]) = ([g] \otimes [t]) \circ ([f] \otimes [h])$, $[id_A] \otimes [id_B] = [id_{A \otimes B}]$, $[\alpha_{A \otimes B, C, D}] \circ [\alpha_{A, B, C \otimes D}] = ([\alpha_{A, B, C}] \otimes [id_D]) \circ [\alpha_{A, B \otimes C, D}] \circ ([id_A] \otimes [\alpha_{B, C, D}])$, and $([\rho_A] \otimes [id_B]) \circ [\alpha_{A, I, B}] = [id_A] \otimes [\lambda_B]$. The syntactic category, therefore, satisfies all of the requirements and coherence conditions of a monoidal category. \square

Theorem A.2 (Symmetric monoidal category). *The syntactic category is a symmetric monoidal category.*

Proof. We can use soup reduction to show that $[\sigma_{B,A}] \circ [\sigma_{A,B}] = [id_{A \otimes B}]$, $[\rho_A] = [\lambda_A] \circ [\sigma_{A,I}]$, and $[\alpha_{C,A,B}] \circ [\sigma_{A \otimes B, C}] \circ [\alpha_{A, B, C}] = ([\sigma_{A,C}] \otimes [id_B]) \circ [\alpha_{A, C, B}] \circ ([id_A] \otimes [\sigma_{B,C}])$. The syntactic category thus satisfies all of the requirements and coherence conditions of a symmetric monoidal category. \square

Theorem A.3 (Compact closure). *The syntactic category is a compact closed category.*

Proof. Using our soup reduction rules, we can show that $[\lambda_A] \circ ([\varepsilon_A] \otimes [id_A]) \circ [\alpha_{A, A^*, A}] \circ ([id_A] \otimes [\eta_A]) \circ [\rho_A]^{-1} = [id_A]$ and $[\rho_{A^*}] \circ ([id_{A^*}] \otimes [\varepsilon_A]) \circ [\alpha_{A^*, A, A^*}]^{-1} \circ ([\eta_A] \otimes [id_{A^*}]) \circ [\lambda_{A^*}]^{-1} = [id_{A^*}]$, by reducing the sequents represented by the denotations on the left hand sides to identities. The syntactic category thus satisfies both of the yanking conditions that are required of a compact closed category. \square

Theorem 2.8 (Dagger compact closure). *The syntactic category is a dagger compact category.*

Proof. For every denotation $[f] : A \rightarrow B$, we define its dagger $[f]^\dagger : B \rightarrow A$, as the denotation representing the soup equivalent sequents of the \dagger -flipped sequents for $[f]$. It is now easy to show that $([f]^\dagger)^\dagger = [f]$ and $[\sigma_{A, A^*}] \circ [\varepsilon_A]^\dagger = [\eta_A]$, by showing that the sequents they represent are soup equivalent. The syntactic category, therefore, satisfies all of the requirements of a dagger compact category. \square

A.4 Example

We will examine the differences in representation between teleportation¹ of a single state and teleportation of an entire function. The "yanking" action of teleportation can be witnessed by considering the reduction:

$$\begin{aligned}
x_1 : T &\vdash_{\{x_1 \otimes x_{2*} \otimes 1 : \varepsilon, \eta : 1 \otimes x_{2*} \otimes x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{x_1 \otimes x_{2*} \otimes 1 : x_4 \otimes x_{4*} \otimes 1, \eta : 1 \otimes x_{2*} \otimes x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{x_1 : x_4, x_{2*} : x_{4*}, 1 : 1, \eta : 1 \otimes x_{2*} \otimes x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{x_{2*} : x_{1*}, \eta : 1 \otimes x_{2*} \otimes x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{\eta : 1 \otimes x_{1*} \otimes x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{1 \otimes x_{5*} \otimes x_5 : 1 \otimes x_{1*} \otimes x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{1 : 1, x_{5*} : x_{1*}, x_5 : x_3\}} x_3 : T \\
x_1 : T &\vdash_{\{x_1 : x_3\}} x_3 : T \\
x_1 : T &\vdash x_1 : T
\end{aligned}$$

For a state of type A , we could replace the type T with A and leave the rest of the sequents in the derivations as they are. Similarly, for a function of type $A \multimap B$, we could replace T with $A \multimap B$ and keep the rest of the derivation intact. This reveals the power of the dagger lambda calculus; we are essentially using the same syntax to represent all types of teleportation.

¹Our analysis will not include the unitary corrections that are typically applied at the end of the teleportation protocol, as the classical control they require is beyond the scope of this paper.

Depicting qudit quantum mechanics and mutually unbiased qudit theories

André Ranchin

University of Oxford, Department of Computer Science, Quantum Group
Imperial College London, Department of Physics, Controlled Quantum Dynamics

We generalize the ZX calculus to quantum systems of dimension higher than two. The resulting calculus is sound and universal for quantum mechanics. We define the notion of a mutually unbiased qudit theory and study two particular instances of these theories in detail: *qudit stabilizer quantum mechanics* and *Spekkens-Schreiber toy theory for dits*. The calculus allows us to analyze the structure of qudit stabilizer quantum mechanics and provides a geometrical picture of qudit stabilizer theory using D-toruses, which generalizes the Bloch sphere picture for qubit stabilizer quantum mechanics. We also use our framework to describe generalizations of Spekkens toy theory to higher dimensional systems. This gives a novel proof that qudit stabilizer quantum mechanics and Spekkens-Schreiber toy theory for dits are operationally equivalent in three dimensions. The qudit pictorial calculus is a useful tool to study quantum foundations, understand the relationship between qubit and qudit quantum mechanics, and provide a novel, high level description of quantum information protocols.

1 Introduction

An interesting approach to understanding the foundations of quantum mechanics is to study sets of alternative theories which exhibit similar structural or physical features as quantum theory. Several mathematical formalisms for operational physical theories have been proposed [1, 3] which encompass quantum mechanics as one possible theory within a space of different potential theories. These provide a setting in which we can determine which features are truly particular to quantum theory and which ones are more generic. This approach can pave the way towards novel axiomatizations of quantum mechanics and could yield precious clues about future physical theories which may supersede quantum theory, such as a theory of quantum gravity. As Lewis Carroll aptly put it: “If you don’t know where you are going, any road will get you there”.

Symmetric monoidal categories (SMCs) provide a general framework for physical theories since they contain two interacting modes, \otimes and \circ , of composing systems and processes. Previous work has investigated which additional structure must be imposed on a SMC in order to recover the structure of quantum theory [1]. This approach has yielded an intuitive graphical language, called the **ZX calculus**, which allows us to explicitly formulate quantum mechanics within a symmetric monoidal category [6].

More precisely, the ZX calculus is a two coloured pictorial calculus for qubits whose diagrams are generated by composing basic process diagrams and which has a set of rule equations specifying how one diagram can be transformed into another. The calculus is sound and universal for quantum mechanics and it has been shown that it is complete for stabilizer quantum mechanics, given a certain choice of phases [2]. The ZX calculus has proven useful in the study of quantum foundations [7], quantum computation [10] and quantum error-correction [17].

In the present article, we generalize the ZX calculus to qudit systems and show that the resulting calculus is universal for quantum mechanics. We anticipate that the new calculus will provide a practical

tool to study quantum information and computation from a high-level point of view. For example, the qudit calculus for dimensions higher than two should be well suited to understanding structural properties of quantum algorithms, quantum key distribution and quantum error-correction. Moreover, as the complexity of the quantum systems we study will grow, computer software such as Quantomatic [18], which allows automated reasoning within the calculus, may play an important role in the design of future quantum networks.

For the time being, we focus on using key ideas from the ZX qudit calculus to study quantum foundations. In particular, we define the notion of a **mutually unbiased qudit theory** (MUQT), which corresponds to a symmetric monoidal category whose observable structures are all mutually unbiased. These MUQTs can be classified in terms of a particular Abelian group, called the **phase group**.

Previous work has shown that in the case of qubits [8], there are essentially two MUQTs: stabilizer quantum mechanics [14], which has phase group \mathbb{Z}_4 , and Spekken's toy theory for bits [23], which has phase group $\mathbb{Z}_2 \times \mathbb{Z}_2$. Furthermore, the phase groups of these theories determine whether or not they admit a local hidden variable model. We aim to generalize this work to higher dimensional systems.

This article focuses on the study of two interesting families of MUQTs, corresponding to *stabilizer quantum theory for qudits* [15] and *Spekkens-Schreiber's toy theory for dits* [21]. This is a first step towards a full classification of MUQTs and a thorough study of the relationship between physical features of these theories and the properties of their phase groups.

2 The ZX calculus for qudit quantum mechanics:

We now present the ZX calculus for qudit quantum mechanics. This is a generalization of the standard qubit ZX calculus [6]. The mathematical background upon which the calculus is built is presented in [20]. We briefly repeat a few essential definitions. An observable structure, which is a generalization of the Hilbert space concept of an orthonormal basis, consists of a copying map $\delta : \bullet \rightarrow \bullet$ and a deleting map

$\varepsilon : \bullet \rightarrow \bullet$ satisfying certain algebraic conditions. A state (or point) ψ is *classical* (or an eigenstate) for an observable structure if it is copied by the copying map and deleted by the deleting map. ψ is *unbiased* with respect to an observable structure if: $s(\delta^\dagger \circ (\psi \otimes \psi^*)) = \varepsilon^\dagger$ for some scalar s .

Given an observable structure, each state ψ has a corresponding *phase map*: $\Lambda(\psi) := \delta^\dagger \circ (\psi \otimes \mathbb{I})$. The set of all phase maps corresponding to unbiased states for an observable structure, together with map composition, form a group called the **phase group**. We will now present the rules of the calculus and its relationship to quantum theory.

General network diagrams are built out of parallel (tensor product) and downward compositions of generating diagrams from Figure 1.

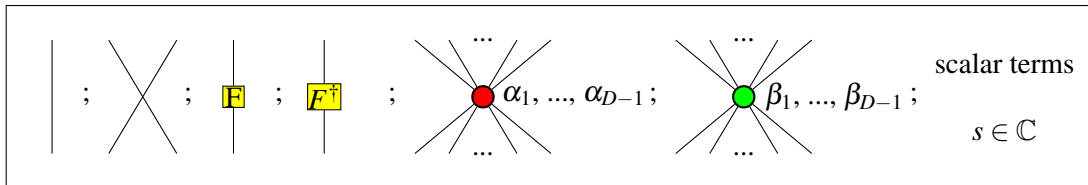


Figure 1: Generating diagrams for the ZX network.

The rules of the *qudit ZX calculus* are the (S), (D), (B), (K), and (F) rules below (and their reversed

colour counterparts), together with a (T) rule which states that after identifying the inputs and outputs of any part of a ZX network, any topological deformation of the internal structure does not matter.

$$\begin{array}{c} \alpha_1, \alpha_2, \dots, \alpha_{D-1} \\ \dots \\ \beta_1, \beta_2, \dots, \beta_{D-1} \end{array} = \begin{array}{c} \alpha_1 + \beta_1, \alpha_2 + \beta_2, \dots, \alpha_{D-1} + \beta_{D-1} \end{array} \quad (S1)$$

$$\text{red dot on a line} := \text{red dot on a line} \quad 0, \dots, 0 = \quad (S2)$$

Diagrammatic equation (D) shows a bubble diagram with four external legs (two red, two green) equal to the product of two propagators (one red-green, one red-green) and another propagator (one red-green, one red-green), which is then simplified to a single propagator (one red-green, one red-green) multiplied by \sqrt{D} .

Diagrammatic equation (B1): A red node connected to a green node, and a green node connected to two red nodes, is equal to two red nodes.


$$\text{Diagram 1} + \text{Diagram 2} = \text{Diagram 3} \quad (\text{B2})$$

(K1)

(K2)

$$\begin{array}{c}
\cdots \\
\text{F}^{\dagger} \text{F}^{\dagger} \text{F}^{\dagger} \text{F}^{\dagger} \\
\diagup \quad \diagdown \\
\bullet \quad \alpha_1, \alpha_2, \dots, \alpha_{D-1} \\
\diagdown \quad \diagup \\
\text{F} \text{F} \text{F} \text{F} \\
\cdots
\end{array} = \begin{array}{c}
\cdots \\
\diagup \quad \diagdown \\
\bullet \quad \alpha_1, \alpha_2, \dots, \alpha_{D-1} \\
\diagdown \quad \diagup \\
\cdots
\end{array} \quad (\text{F1})$$

$$\begin{array}{c} \boxed{F^\dagger} \\ | \\ \boxed{F} \end{array} = \begin{array}{c} \boxed{F} \\ | \\ \boxed{F^\dagger} \end{array} = \quad (F2)$$

where $Neg(\alpha_1, \dots, \alpha_{D-1}) := \alpha_{k+1} - \alpha_k, \alpha_{k+2} - \alpha_k, \dots, \alpha_{D-1} - \alpha_k, -\alpha_k, \alpha_1 - \alpha_k, \dots, \alpha_{k-1} - \alpha_k$, and where there are D-1 different red k vertices which have phases $\alpha_1, \dots, \alpha_{D-1}$ such that  k are the phase maps corresponding to the D-1 classical points for Z whose phases are not all zero. In higher dimensions, the (K) rules give rise to more intricate interference phenomena, since the D classical points of an observable structure each permute the phase group elements.

Diagrammatic reasoning in the qudit calculus is identical to reasoning in the qubit calculus. As before, two network diagrams can be shown to be equal by locally replacing some part of a diagram with a diagram equal to it.

Note that the restricted case of the ZX calculus for qutrits has been studied independently (and synchronously) in [4].

As with the qubit case, we can model the calculus in Hilbert space. We interpret all diagram edges by \mathbb{C}^D and elements of the qudit calculus correspond to the following Hilbert space elements:

$$\begin{aligned}
 \left[\begin{array}{c} | \\ | \end{array} \right] &= \mathbb{I}_{D \times D} := \sum_{k=0}^{D-1} |k\rangle \langle k| \quad ; \quad \left[\begin{array}{c} \diagup \\ \diagdown \end{array} \right] = SWAP_{a,b} := \sum_{j,k=0}^{D-1} |k\rangle \langle j|_a \otimes |j\rangle \langle k|_b \\
 \left[\begin{array}{c} \boxed{F} \\ | \end{array} \right] &= Fourier := \frac{1}{\sqrt{D}} \sum_{j,k=0}^{D-1} \eta^{jk} |j\rangle \langle k| \quad ; \quad \left[\begin{array}{c} \boxed{F^\dagger} \\ | \end{array} \right] = Fourier^\dagger. \\
 \left[\begin{array}{c} \bullet \\ | \end{array} \right] &= |0\rangle := \sqrt{D} \begin{pmatrix} 1 \\ 0 \\ \dots \\ 0 \end{pmatrix} \quad ; \quad \left[\begin{array}{c} \bullet \\ | \end{array} \right] = |+\rangle := \begin{pmatrix} 1 \\ 1 \\ \dots \\ 1 \end{pmatrix} \quad ; \quad \left[\begin{array}{c} \bullet \\ | \end{array} \right] = \epsilon_X := \langle 0| \quad ; \quad \left[\begin{array}{c} \bullet \\ | \end{array} \right] = \epsilon_Z := \langle +| \\
 \left[\begin{array}{c} \bullet \\ \diagup \end{array} \right] &= \delta_X := \begin{pmatrix} \mathbb{I}_{D \times D} \\ P_1(\mathbb{I}_{D \times D}) \\ \dots \\ P_{D-1}(\mathbb{I}_{D \times D}) \end{pmatrix} \quad ; \quad \left[\begin{array}{c} \bullet \\ \diagdown \end{array} \right] = \delta_Z := (e_1| \ e_2| \ \dots \ |e_D) \\
 \left[\begin{array}{c} \bullet \\ | \end{array} \right] \alpha_1, \dots, \alpha_{D-1} &= \Lambda_X(\alpha_1, \alpha_2, \dots, \alpha_{D-1}) := \frac{1}{D} \begin{pmatrix} c_0 & c_{D-1} & c_{D-2} & \dots & c_2 & c_1 \\ c_1 & c_0 & c_{D-1} & \dots & c_3 & c_2 \\ c_2 & c_1 & c_0 & \dots & c_4 & c_3 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ c_{D-1} & c_{D-2} & c_{D-3} & \dots & c_1 & c_0 \end{pmatrix} \\
 \left[\begin{array}{c} \bullet \\ | \end{array} \right] \alpha_1, \dots, \alpha_{D-1} &= \Lambda_Z(\alpha_1, \alpha_2, \dots, \alpha_{D-1}) := \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & e^{i\alpha_1} & 0 & \dots & 0 \\ 0 & 0 & e^{i\alpha_1} & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & 0 & e^{i\alpha_{D-1}} \end{pmatrix} \\
 \left[\begin{array}{c} \bullet \\ | \end{array} \right] \left[\begin{array}{c} \bullet \\ | \end{array} \right] &= CNOT_{a,b} := \sum_{j,k=0}^{D-1} |k\rangle \langle j|_a \otimes |k\rangle \langle k+j|_b = \begin{pmatrix} \mathbb{I}_{D \times D} & 0 & 0 & \dots & 0 \\ 0 & P_1(\mathbb{I}_{D \times D}) & 0 & \dots & 0 \\ 0 & 0 & P_2(\mathbb{I}_{D \times D}) & \dots & \dots \\ \dots & \dots & \dots & \dots & 0 \\ 0 & 0 & \dots & 0 & P_{D-1}(\mathbb{I}_{D \times D}) \end{pmatrix}
 \end{aligned}$$

Figure 2: Hilbert space interpretation of the qudit ZX calculus elements.

where $P_j(\mathbb{I}_{D \times D})$ ($j=1,2, \dots, D-1$) are $D \times D$ matrices corresponding to the identity matrix $\mathbb{I}_{D \times D}$, with all its rows permuted to the right by j and where e_i are the vectors which have one 1 in row $D \times (i-1) + i$ and $D-1$ zeros in all the other rows.

The c_j elements of the Λ_X matrix are defined by: $|+\rangle + \sum_{k=1}^{D-1} e^{i\alpha_k} |+_k\rangle = \frac{1}{\sqrt{D}}(c_0 |0\rangle + \sum_{j=1}^{D-1} c_j |j\rangle)$ where $|+_k\rangle$ are the D eigenvectors of the $X = \sum_{j=0}^{D-1} |j\rangle \langle j+1|$ matrix.

Theorem 1: The qudit ZX calculus is universal for quantum mechanics.

Outline of proof[20]: We can show that the universal gate set [19, 5] consisting of the two sets of D single

qudit gates:

$$Z_j(b_0, b_1, \dots, b_{D-1}) : b_0 |0\rangle + b_1 |1\rangle + \dots + b_{D-1} |D-1\rangle \mapsto |j\rangle \quad (1)$$

$$X_j(\phi) : b_0 |0\rangle + b_1 |1\rangle + \dots + b_{D-1} |D-1\rangle \mapsto b_0 |0\rangle + \dots + e^{i\phi} b_j |j\rangle + b_{D-1} |D-1\rangle \quad (2)$$

for $j \in \{0, 1, \dots, D-1\}$, together with the qudit CNOT gate, is contained in the calculus.

By construction, any equation which can be shown to be true using the qudit ZX calculus is true in quantum mechanics so the qudit ZX calculus is **sound** for quantum mechanics. Moreover, extending the qudit ZX calculus to account for mixed states and general quantum evolution described by completely positive maps can be achieved by using the same standard constructions [22, 9, 10] as in the qubit case.

We know that [2] the qubit ZX calculus is **complete** for qubit stabilizer quantum theory, in the sense that any two equivalent qubit stabilizer processes can be shown to be equal by using the qubit ZX calculus. Backens' proof of this result [2], however, relies on results for qubit graph states and it is unclear whether it can be generalized to show completeness of the qudit ZX calculus for qudit stabilizer theory. Therefore, we leave this as an open question:

Is the qudit ZX calculus, with additional rules analogous to the Euler decomposition of the Hadamard vertex, complete for qudit stabilizer quantum mechanics? If it is not, then which other rules need to be added for completeness?

Another important question is how the qudit and qubit ZX calculi are related. More generally, it would be interesting to understand exactly how the ZX calculus for qudits of dimension m is related to the ZX calculus of dimension $n > m$. Perhaps, we could introduce maps which “create” and “annihilate” dimensions. This could lead to an interesting structure and provide insight into the relationship between qubit and qudit quantum mechanics.

3 Mutually unbiased qudit theories

One of the main goals of the present article is to use the abstract structures we introduced to study the foundation of quantum theory. In this respect, we aim to define a class of theories which exhibit many key features of quantum mechanics, within a single mathematical framework.

Therefore, we will generalize the previous approach of studying mutually unbiased qubit theories using dagger compact symmetric monoidal categories [8, 12] to the case of qudits.

Definition: A **mutually unbiased qudit theory** is a dagger symmetric monoidal category [1] with observable structures such that:

- (i) The objects of the category are the unit and finite tensor products of qudit-like systems Q .
- (ii) The observables on a given object are all mutually unbiased, have the same number of eigenstates and have the same phase groups.
- (iii) All states of Q are eigenstates of some observable.

We will study mutually unbiased qudit theories for dimensions higher than two. In the following two sections, we analyze in detail two key examples of mutually unbiased qudit theories: *qudit stabilizer quantum mechanics* and *Spekkens-Schreiber theory for dits*.

4 Picturing stabilizer quantum mechanics

We define the process category **DStab** as the \dagger -compact symmetric monoidal subcategory of **FHilb** corresponding to qudit stabilizer quantum mechanics (see [20]) which is generated by the unit, n -fold tensor products of \mathbb{C}^D , single qudit Clifford operations and the quantum copying and deleting maps. **DStab** can be depicted using the qudit ZX diagrams, where the allowed phases are restricted according to the phase group.

In the case of the standard qubit stabilizer quantum mechanics, the phase group is the cyclic group \mathbb{Z}_4 , which is a finite subgroup of the quantum qubit phase group S^1 (the circle group). Since the unbiased circles for the Z and X observables coincide on the points corresponding to $|+i\rangle$ and $|-i\rangle$, we can completely picture single qubit stabilizer quantum theory using the Bloch sphere.

Can one find an analogous picture for qutrit quantum mechanics?

Let $\{|0\rangle, |1\rangle, |2\rangle\}$ and $\{|+\rangle, |\top\rangle, |\perp\rangle\}$ be the eigenbases for the qutrit Z and X observables respectively. Then the unbiased states for the Z and X observable:

$$|\{\alpha_1, \alpha_2\}_Z\rangle = |0\rangle + e^{i\alpha_1}|1\rangle + e^{i\alpha_2}|2\rangle ; |\{\alpha_1, \alpha_2\}_X\rangle = |+\rangle + e^{i\alpha_1}|\top\rangle + e^{i\alpha_2}|\perp\rangle \quad (3)$$

under pairwise addition of phases form a torus group $S^1 \times S^1$.

All the single qutrit stabilizer states, corresponding to the eigenstates of the qutrit X, Z, XZ and XZ² operators, can be written as unbiased states for either the Z basis or the X basis since:

$$\begin{aligned} |0\rangle &= |\{0, 0\}_X\rangle, |1\rangle = \left|\left\{\frac{4\pi}{3}, \frac{2\pi}{3}\right\}_X\right\rangle, |2\rangle = \left|\left\{\frac{2\pi}{3}, \frac{4\pi}{3}\right\}_X\right\rangle; \\ |+\rangle &= |\{0, 0\}_Z\rangle, |\top\rangle = \left|\left\{\frac{2\pi}{3}, \frac{4\pi}{3}\right\}_Z\right\rangle, |\perp\rangle = \left|\left\{\frac{4\pi}{3}, \frac{2\pi}{3}\right\}_Z\right\rangle; \\ |-\rangle &= \left|\left\{\frac{4\pi}{3}, \frac{4\pi}{3}\right\}_Z\right\rangle = \left|\left\{\frac{2\pi}{3}, \frac{2\pi}{3}\right\}_X\right\rangle, |\top\rangle = \left|\left\{0, \frac{2\pi}{3}\right\}_Z\right\rangle = \left|\left\{\frac{4\pi}{3}, 0\right\}_X\right\rangle, |\top\rangle = \left|\left\{\frac{2\pi}{3}, 0\right\}_Z\right\rangle = \left|\left\{0, \frac{4\pi}{3}\right\}_X\right\rangle; \\ |\times\rangle &= \left|\left\{\frac{2\pi}{3}, \frac{2\pi}{3}\right\}_Z\right\rangle = \left|\left\{\frac{4\pi}{3}, \frac{4\pi}{3}\right\}_X\right\rangle, |\top\rangle = \left|\left\{\frac{4\pi}{3}, 0\right\}_Z\right\rangle = \left|\left\{\frac{2\pi}{3}, 0\right\}_X\right\rangle, |\top\rangle = \left|\left\{0, \frac{4\pi}{3}\right\}_Z\right\rangle = \left|\left\{0, \frac{2\pi}{3}\right\}_X\right\rangle \end{aligned} \quad (4)$$

Single qutrit stabilizer operations take subsets of these 12 states to other subsets of these 12 states. This shows that the phase group for qutrit stabilizer quantum mechanics is $\mathbb{Z}_3 \times \mathbb{Z}_3$.

Therefore, single qutrit stabilizer quantum theory can be pictured using 12 points on two toruses, which is a direct generalization of the Bloch sphere case, where the 4 elements on each of the two unbiased circles (coinciding on two elements) visualized in three dimensions are replaced by **9 elements on each of two unbiased toruses coinciding on six points** (the blue and yellow points in Figures (3a, 3b)).

In fact, this picture can easily be generalized to higher dimensional qudit stabilizer theories for prime dimensions. In that case, the single qudit states of qudit stabilizer quantum theory correspond to the vectors in the $D+1$ mutually unbiased eigenbases of the single qudit operators: $X, Z, XZ, XZ^2, \dots, XZ^{D-1}$. The mutually unbiased points with respect to each of these bases forms a D -torus. If we chose an observable structure, whose eigenstates are a basis, then all the other stabilizer states are on the unbiased D -torus of the chosen basis. In this way, qudit stabilizer theory for prime dimension D can be pictured using **D^2 points on each of two D -toruses** (unbiased toruses for the Z and X operators for example), which coincide on $D^2 - D$ points and can be visualized in $D+1$ dimensions.

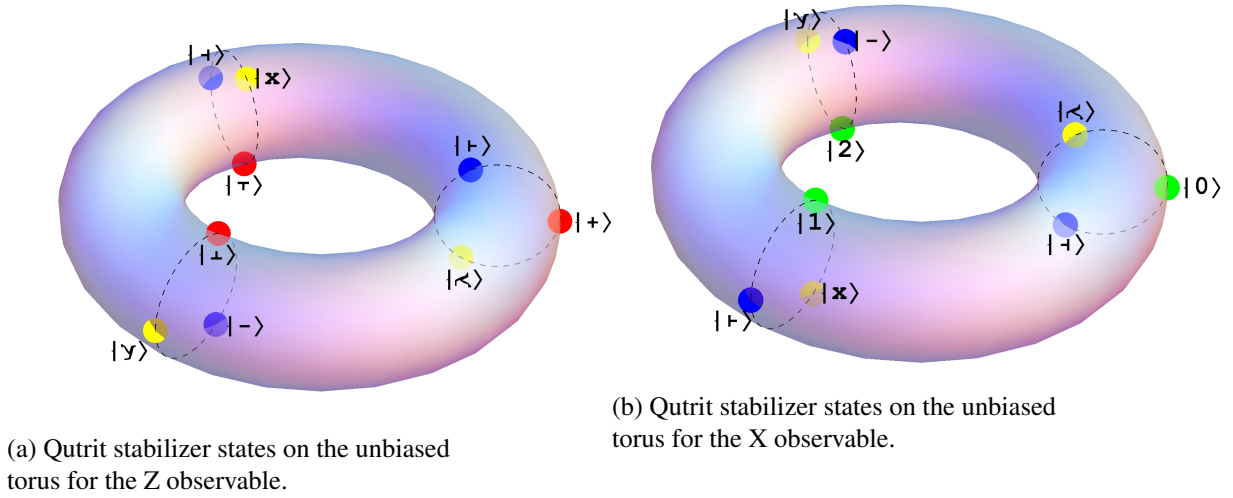


Figure 3: Depicting qutrit stabilizer theory on two tori.

In general, the phase group for qudit stabilizer quantum theory of dimension $D > 3$ is an Abelian subgroup of the group $\mathbb{Z}_D \times \mathbb{Z}_D \times \dots \times \mathbb{Z}_D$ ($D-1$ times). In fact, every finite dimensional closed subgroup of the torus group is isomorphic to a product of finite cyclic groups. Therefore, the phase group for mutually unbiased qudit theories which are also subtheories of quantum mechanics must be of the form $Z_{n_1} \times Z_{n_2} \times \dots \times Z_{n_k}$ for positive integers n_1, \dots, n_k . In further work, we will study how these integers n_1, \dots, n_k for stabilizer phase groups depend on the dimension D . In general, we would like a physical classification of all the mutually unbiased qudit subtheories of quantum mechanics in terms of n_1, \dots, n_k . Once we have determined their **phase group**, the qudit ZX calculus allows us to fully describe these physical theories.

5 Depicting Spekkens-Schreiber toy theory for dits

We define the category **FRel** whose objects are finite sets and whose morphisms are relations. By taking the Cartesian product of sets as the tensor product, the single element set $\{\star\}$ as the identity object and the relational converse as the dagger, **FRel** can be viewed as a SMC with dagger structure.

We can then define the category **DSpek** as a subcategory of **FRel** whose objects are the single element set $I = \{\star\}$ and n -fold Cartesian products of the D^2 -element set: $\mathcal{D} := \{1, 2, \dots, D^2\}$.

The morphisms of **DSpek** are those generated by composition, Cartesian product and relational converse from the following relations:

- (a) All $(D^2)!$ permutations $\sigma_i : \mathcal{D} \rightarrow \mathcal{D}$ of the D^2 -element set.
- (b) The copying relation: $\delta_Z : \mathcal{D} \rightarrow \mathcal{D} \times \mathcal{D}$ defined as:

1	2	...	D														
D	1	...	D-1														
...														
2	3	...	1														
				D+1	D+2	...	2D										
				2D	D+1	...	2D-1										
													
				D+2	D+3	...	D+1										
													
													
													
												D(D-1)+1	D(D-1)+2	...	D ²		
												D ²	D(D-1)+1	...	D ² -1		
													
												D(D-1)+2	D(D-1)+3	...	D(D-1)+1		

where there is x in the (y,z) location of the grid iff $\delta_Z : x \sim (y, z)$.

(c) The deleting relation: $\varepsilon_Z : \mathcal{D} \rightarrow I$ defined as: $\{1, D+1, 2D+1, \dots, D(D-1)+1\} \sim \star$.

(d) The relevant unit, associativity and symmetry natural isomorphisms.

If we interpret relations from I to n-fold tensor products of \mathcal{D} as epistemic states on phase space then this category corresponds to Spekkens-Schreiber theory for dits with only states of maximal knowledge. Adding the maximally mixed state $\perp_D :: \{\star\} \sim \{1, 2, \dots, D^2\}$ to **DSpek** yields the category **MDSpek**, corresponding to Spekkens-Schreiber theory for dits of dimension D (See [20]).

DSpek and **MDSpek** inherit symmetric monoidal and \dagger -compact structure from **FRel** since we can define Bell states (corresponding to compact structures) as:

$$\mu_D := \delta_Z \circ \varepsilon_Z^\dagger : I \rightarrow \mathcal{D} \times \mathcal{D} :: \star \sim \{(1, 1), (2, 2), \dots, (D^2, D^2)\} \quad (5)$$

We can define the other copying map as:

$$\delta_X = (\Pi_{k=1}^{D-1} \sigma_{(k+1, (kD)+1)}) \otimes \Pi_{k=1}^{D-1} \sigma_{(k+1, (kD)+1)}) \circ \delta_Z \circ (\Pi_{k=1}^{D-1} \sigma_{(k+1, (kD)+1)}) \quad (6)$$

where $\sigma_{(j,k)}$ permutes entries j and k of the input D^2 -element set (epistemic state). This map is explicitly: $\delta_X : \mathcal{D} \rightarrow \mathcal{D} \times \mathcal{D}$ such that: $\delta_X : x \sim (y, z)$ iff there is x in the (y,z) location of the following grid:

1				D+1			D(D-1)+1						
	2				D+2			D(D-1)+2					
					
			D				2D					D ²	
D(D-1)+1				1				D(D-2)+1					
	D(D-1)+2				2				D(D-2)+2				
			
			D ²				D					D(D-1)	
...
...
...
...
D+1				2D+1				1					
	D+2				2D+2				2				
			
			2D				3D					D	

and the other erasing map as: $\varepsilon_X : \mathcal{D} \rightarrow I$ such that: $\{1, 2, 3, \dots, D\} \sim \star$. It is easy to check that this then gives us two strongly complementary observable structures, analogous to the Z and X observable structures in quantum theory.

In fact, we can use the fact that **3Spek** can be depicted in the qutrit ZX calculus to provide a novel proof of the following known result:

Theorem 2[16, 21]: Spekkens-Schreiber theory for trits is operationally equivalent to stabilizer theory for qutrits.

Outline of proof[20]: We can show that **3Spek** and **3Stab** can both be expressed in the qutrit ZX calculus as mutually unbiased qutrit theories with twelve states and phase group $\mathbb{Z}_3 \times \mathbb{Z}_3$. This uniquely determines all the allowable preparations, measurements and transformations (compositions of spiders with phases adding according to $\mathbb{Z}_3 \times \mathbb{Z}_3$) and provides a one-to-one mapping between the process categories **3Spek** and **3Stab**.

In future, we can find the phase group of Spekkens-Schreiber theory for dits in any dimension D . This should allow us to depict these theories using (a version of) the qudit ZX calculus. We would then be able to study the relationship between Spekkens-Schreiber theory for dits and qudit stabilizer theory in the general case. We will return to address this question and the issue of analyzing mutually unbiased qudit theories with arbitrary Abelian phase groups in another article.

6 Further work

We will conclude this paper with a brief outline of possible avenues of research which follow from the work presented here. As we mentioned earlier, understanding how the qubit calculus fits into the general qudit calculus and proving the completeness of a slightly modified version of the generalized qudit ZX calculus for stabilizer quantum mechanics would certainly provide new insights into qudit stabilizer quantum mechanics. This might lead to modifications of the qudit ZX calculus before it reaches its final form.

For example, the relation between phase group structure and graph structure is still unclear. Can the qudit ZX calculus be expressed without angles by adding axioms relating to graph structure [11]? It might be interesting to analyze alternative calculi with multiple edges between vertices. This approach could simplify proofs of completeness or provide a graphical depiction of non-locality.

Another possible mathematical framework for studying the qudit ZX calculus would be to use product and permutation categories (PROPs). This approach may allow an elegant synthetic axiomatization of numerous physical process theories and could provide new completeness theorems for corresponding graphical calculi.

On a more practical note, the calculus for qudit stabilizer quantum theory can help describe quantum information protocols in a new light. One may generalize qubit protocols to qudits and try to understand new features of familiar quantum processes. For example, the formalism could be used to give a general description of error correction and fault tolerance for qudits. This could then be related to qubit error correction and links might be made between error correction in various dimensions. Furthermore, getting new insights into the abstract structure of qudit quantum mechanics could play a pivotal role in the development of new quantum algorithms.

There are also a number of quantum foundations questions which could be addressed next. For instance, we know that the single qudit stabilizer theory is operationally equivalent to Spekkens-Schreiber theory for dits for *finite odd dimensions* and therefore admits a non-contextual, local hidden variable model in those cases. But what is the relationship between qudit stabilizer theory and Spekkens-Schreiber toy theory in general? We could also study van Enk's toy model[24] as a MUQT and find its phase group.

More generally, it would be useful to classify all the mutually unbiased qudit theories and determine which quantum-like features each one exhibits. For example, what is the relationship between a theory's phase group and whether it admits a local hidden variable interpretation? The study of the qudit ZX calculus with different Abelian phase groups should produce a large class of interesting toy models.

In the future, we could also consider theories where distinct observable structures have different phase groups.

Moreover, the qubit ZX calculus, by providing a clear description of complementarity and phase, has helped clarify the relationship between complementarity and non-locality[7]. The qudit ZX calculus provides the ideal framework to study other similar foundational questions related to complementarity. We could, for example, use the categorical framework to study how various notions of complementarity arise in different dimensions. Can one find a pictorial calculus, like a ZXY calculus for qudits, which captures complementarity of more than two observables?

Finally, it would be interesting to understand the interpretation of the D-torus phase groups for qudit quantum mechanics observables from a physical point of view. What properties of the quantum phase group lead to specific quantum features? Perhaps studying the operational interpretation of phase [13] in physical theories could help us find the physical reason for each phase group taking the form it does. The study of phase and complementarity from an operational point of view may also provide insight into the relationship between categorical quantum mechanics and generalized probabilistic theories.

Acknowledgments

I would like to thank both of my supervisors Bob Coecke and Terry Rudolph for insightful discussions. I am also very grateful to Mihai Vidrighin for his help with the torus illustrations and for his useful comments. This research is funded by the EPSRC.

References

- [1] S. Abramsky and B. Coecke. A categorical semantics of quantum protocols. *University Computing*, 415(RR-04-02):21, 2004.
- [2] M. Backens. The ZX-calculus is complete for stabilizer quantum mechanics. *ArXiv e-prints*, July 2013.
- [3] J. Barrett. Information processing in generalized probabilistic theories. *Phys. Rev. A*, 75:032304, Mar 2007.
- [4] X. Bian and Q. Wang. Graphical calculus for qutrit systems. 2013.
- [5] J.-L. Brylinski and R. Brylinski. Universal quantum gates. *eprint arXiv:quant-ph/0108062*, Aug. 2001.
- [6] B. Coecke and R. Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13(4):043016, 2011.
- [7] B. Coecke, R. Duncan, A. Kissinger, and Q. Wang. Strong Complementarity and Non-locality in Categorical Quantum Mechanics. *ArXiv e-prints*, Mar. 2012.
- [8] B. Coecke, B. Edwards, and R. W. Spekkens. Phase groups and the origin of non-locality for qubits. *Electronic Notes in Theoretical Computer Science*, 270(2):15–36, 2011. arXiv:1003.5005.
- [9] B. Coecke and S. Perdrix. Environment and classical channels in categorical quantum mechanics. *ArXiv e-prints*, Apr. 2010.
- [10] R. Duncan and S. Perdrix. Rewriting measurement-based quantum computations with generalised flow. In *Proceedings of the 37th international colloquium conference on Automata, languages and programming: Part II*, ICALP’10, pages 285–296, Berlin, Heidelberg, 2010. Springer-Verlag.
- [11] R. Duncan and S. Perdrix. Pivoting makes the ZX-calculus complete for real stabilizers. *ArXiv e-prints*, July 2013.
- [12] B. Edwards. Phase groups and local hidden variables. Technical Report RR-10-15, September 2010.
- [13] A. J. P. Garner, O. C. O. Dahlsten, Y. Nakata, M. Murao, and V. Vedral. A general framework for phase and interference. *ArXiv e-prints*, Apr. 2013.

- [14] D. Gottesman. Stabilizer codes and quantum error correction. *Energy*, 2008:114, 1997.
- [15] D. Gottesman. Fault tolerant quantum computation with higher dimensional systems. *Chaos Solitons Fractals*, 10:1749–1758, 1999.
- [16] D. Gross. Hudson’s theorem for finite-dimensional quantum systems. *Journal of Mathematical Physics*, 47(12):122107, Dec. 2006.
- [17] C. Horsman. Quantum pictorialism for topological cluster-state computing. *New Journal of Physics*, 13(9):18, 2011.
- [18] A. Kissinger. Exploring a quantum theory with graph rewriting and computer algebra. 2009.
- [19] A. Muthukrishnan and C. R. Stroud, Jr. Multivalued logic gates for quantum computation. *Phys. Rev. A*, 62(5):052309, Nov. 2000.
- [20] A. Ranchin. Depicting qudit quantum mechanics and mutually unbiased qudit theories. *ArXiv e-prints*, Apr. 2014.
- [21] O. Schreiber and R. W. Spekkens. Reconstruction of the stabilizer formalism for qutrits from a statistical theory of trits with an epistemic restriction. *to be published*, 2012.
- [22] P. Selinger. Dagger compact closed categories and completely positive maps (extended abstract). *Electronic Notes in Theoretical Computer Science*, 170:139163, 2007.
- [23] R. W. Spekkens. Evidence for the epistemic view of quantum states: A toy theory. *Physical Review A*, 75(3):032110, 2007.
- [24] S. J. van Enk. A Toy Model for Quantum Mechanics. *Foundations of Physics*, 37:1447–1460, Oct. 2007.

Entropic formulation of Heisenberg's measurement-disturbance relation

Patrick Coles

National University of Singapore,
Singapore
Centre for Quantum Technologies
pat@nus.edu.sg

Fabian Furrer

University of Tokyo
Tokyo, Japan
Department of Physics, Graduate School of Science
furrer@eve.phys.s.u-tokyo.ac.jp

Heisenberg's original intuition was that there should be a tradeoff between measuring a particle's position with greater precision and disturbing its momentum. Recent formulations of this idea have primarily focused on the question of how well two complementary observables can be jointly measured. Here, we provide an alternative approach based on how enhancing the predictability of one observable necessarily disturbs a complementary one. Our measurement-disturbance relation refers to a clear operational scenario and is expressed by entropic quantities with clear statistical meaning, evading recent criticism directed at some previous formulations. We show that our relation is perfectly tight for all measurement strengths in an existing experimental setup involving qubit measurements. (See arXiv:1311.7637.)

Introduction. Heisenberg's uncertainty principle is one of the most central concepts in quantum physics and with increasing experimental abilities to control quantum degrees of freedom it is no longer only interesting from a theoretical point of view; it is now practically relevant. For instance, it provides limits on quantum metrology and can be used to prove security in quantum cryptography. Moreover, experimental setups are now capable of sensitively testing such formulations [6, 9]. These advances demand tight, operationally-meaningful formulations of the uncertainty principle.

The most common formulation of the uncertainty principle gives a limit on one's ability to prepare a system with low uncertainty for two complementary observables X and Z . Textbooks often illustrate this with Robertson's bound on the standard deviations $\Delta X \Delta Z \geq \frac{1}{2} |\langle \psi | [X, Z] | \psi \rangle|$, which generalized Kennard's earlier relation for position and momentum observables $\Delta Q \Delta P \geq \hbar/2$.

However, a conceptually different aspect of the uncertainty principle concerns not preparation limitations but rather measurement limitations [3]. There are at least two aspects of measurement uncertainty: (1) joint measurability - the idea that one cannot build a device that jointly measures X and Z , and (2) measurement-disturbance - the idea that measuring X necessarily disturbs Z . One approach to joint measurability considers state-dependent errors, e.g., [8], while another approach considers calibrating the apparatus on idealised input states associated with the X and Z observables [4, 5, 2].

Because sequential measurement can be thought of as an attempted joint measurement, joint measurement relations also have interpretations as measurement-disturbance relations. The measurement-disturbance relations associated with the calibration approach, e.g., using root-mean-square error in [4, 5] and entropic error in [2], have a clear interpretation, in that any device that extracts information about the X eigenstates must necessarily disturb the information about the Z eigenstates.

Besides these state-independent formulations, it seems an interesting question to ask how measuring X disturbs Z for an arbitrary input state. At first sight, the relations similar to Ozawa's [8] seem to treat this; however, their disturbance measure have some disadvantages, e.g., they are state-independent for recent qubit experiments [7, 5] and criticised as being in general non-operational [7, 5]. In this sense, we believe that state-dependent measurement-disturbance has not fully been treated.

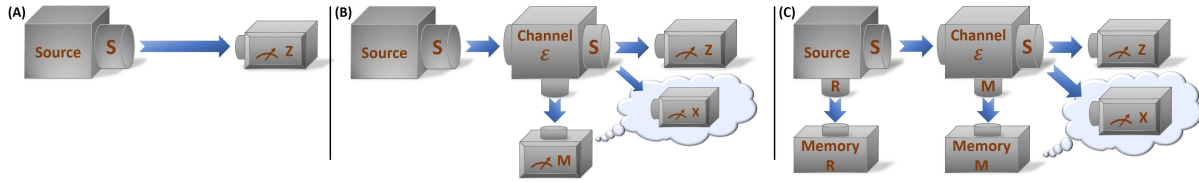


Figure 1: (A): We consider a source S sending a quantum state to a receiver measuring observable \mathbb{Z} . (B): During the transmission a channel combined with a measurement is used to extract information M about a second observable \mathbb{X} . Our relation captures the tradeoff between the disturbance of the Z distribution in (B) compared with the undisturbed situation (A) and the predictability of a hypothetical future measurement of \mathbb{X} given the information M . In the memory-assisted situation (C), M can be a quantum memory and the disturbance of the correlations between S and an isolated reference system R is included.

Recent observations [7] suggest that a state-dependent measurement-disturbance principle for an error quantifying the measurement accuracy and an operationally defined disturbance does not exist: for any input state and observables X and Z , one can perform a perfect X measurement and then simply reset the system's state such that it has the same probability distribution for Z as the original input state. Thus, both the X error and the Z disturbance can be made to vanish on any input state. Noting this, we propose that the Z disturbance is instead in tradeoff with a revised concept of X error: the residual ignorance about X after the measurement (i.e., predictive error). We then show that the residual error is indeed in trade-off with the disturbance by proving an inequality for error and disturbance measures with clear operational and statistical meaning. A similar relation was previously also shown by Appleby [1] for position and momentum, but only for a restrictive measurement model.

Results. In order to precisely state our trade-off relation, let us consider the situation as in Figure 1 (A) and (B). We have a system S prepared in state ρ_S and sent to a receiver who performs a measurement of the observable \mathbb{Z} . During the transmission of S to the receiver an interaction \mathcal{E} is applied that intends to extract classical information M about a complementary observable \mathbb{X} . For simplicity, we assume here that both observables \mathbb{X} and \mathbb{Z} are sharp with orthonormal eigenstates $\{|\mathbb{X}_x\rangle\}_{x \in X}$ and $\{|\mathbb{Z}_z\rangle\}_{z \in Z}$. We denote the Z distribution of ρ_S as P_Z and the one after the interaction \mathcal{E} as $P_Z^\mathcal{E}$. The joint probability distribution of M and X after the interaction is denoted by $Q_{MX}^\mathcal{E}$.

From an operational point of view, we can only detect a disturbance of the Z measurement if $P_Z \neq P_Z^\mathcal{E}$. This motivates to define the disturbance as the distance between P_Z and $P_Z^\mathcal{E}$ which we quantify with the relative entropy $D(\rho_S, \mathbb{Z}, \mathcal{E}) := D(P_Z || P_Z^\mathcal{E})$. The relative entropy is a common quantity in information theory and has direct application to hypothesis testing. The residual error should express how much information M contains about a (hypothetical) X measurement after the interaction \mathcal{E} (see Figure 1 (B)). For that we use the conditional max-entropy of $Q_{MX}^\mathcal{E}$: $E(\rho_S, \mathbb{X}, \mathcal{E}) := H_{\max}(X|M)_{Q_{MX}^\mathcal{E}}$. The conditional max-entropy is part of a family used to quantify resources beyond their behavior in the limit of infinitely many copies and is related to the amount of additional data that must be supplied to the observer, given that they have access to M , to learn the outcome of a future \mathbb{X} measurement.

Our trade-off relation is then given by

$$D(\rho_S, \mathbb{Z}, \mathcal{E}) + E(\rho_S, \mathbb{X}, \mathcal{E}) \geq \log 1/c - H(Z)_P, \quad (1)$$

where the lower bound composes from a state-independent term $c = \max_{x,z} |\langle \mathbb{X}_x | \mathbb{Z}_z \rangle|^2$ which measures

the complementary of X and Z minus the Shannon entropy of P_Z . Note that the dependence of the lower bound on the initial uncertainty is unavoidable since, roughly speaking, extraction of information about X induces noise in the Z basis which is not detectable if the uncertainty in Z was already high.

We further extend relation (1) into two directions, see Figure 1 (C). We show that extracting quantum information M instead of only classical does not change the lower bound. And secondly, that the bound can be strengthened if one includes the disturbance of the Z correlations with a memory (reference) system R . The disturbance is then given as the relative entropy of the joint system of outcome Z and reference R before (ρ_{ZR}) and after ($\rho_{ZR}^{\mathcal{E}}$) the interaction: $D(\rho_{SR}, \mathbb{Z}, \mathcal{E}) = D(\rho_{ZR} || \rho_{ZR}^{\mathcal{E}})$. Using the new error and disturbance measures, the trade-off then reads as in (1) except that $H(Z)_P$ is exchanged with the generally smaller quantum conditional von Neumann entropy $H(Z|R)_\rho$ further tightening the bound. We then show that for the setting in [9] the extended relation is tight over the entire parameter range of the interaction.

We also extend the relation (1) to position and momentum operators in which entropies are simply replaced by their differential versions, and proof tightness if the interaction is given by a covariant approximate position measurement. We consider further Heisenberg's example of a coarse grained position measurement and show a trade-off of the form $(Q \text{ precision}) \cdot (P \text{ disturbance}) \geq \hbar / (P \text{ initial uncertainty})$.

Conclusion. We have shown a state-dependent error-disturbance relation with clear operational meaning. We further introduced the novel concept of memory-assisted disturbance, where a quantum memory helps to reveal the disturbing effects of a measurement; this may lead to application to reverse reconciliation quantum key distribution.

Acknowledgments. PJC is funded by the Ministry of Education (MOE) and National Research Foundation Singapore, as well as MOE Tier 3 Grant "Random numbers from quantum processes" (MOE2012-T3-1-009). FF acknowledges support from Japan Society for the Promotion of Science (JSPS) by KAKENHI grant No. 24-02793.

References

- [1] D.M. Appleby (1998): *Concept of Experimental Accuracy and Simultaneous Measurements of Position and Momentum*. *International Journal of Theoretical Physics* 37(5), pp. 1491–1509, doi:10.1023/A:1026659601439.
- [2] Francesco Buscemi, Michael J. W. Hall, Masanao Ozawa & Mark M. Wilde (2014): *Noise and Disturbance in Quantum Measurements: An Information-Theoretic Approach*. *Phys. Rev. Lett.* 112, p. 050401, doi:10.1103/PhysRevLett.112.050401.
- [3] Paul Busch, Teiko Heinonen & Pekka Lahti (2007): *Heisenberg's uncertainty principle*. *Physics Reports* 452(6), pp. 155 – 176, doi:http://dx.doi.org/10.1016/j.physrep.2007.05.006.
- [4] Paul Busch, Pekka Lahti & Reinhard F. Werner (2013): *Proof of Heisenberg's Error-Disturbance Relation*. *Phys. Rev. Lett.* 111, p. 160405, doi:10.1103/PhysRevLett.111.160405.
- [5] Paul Busch, Pekka Lahti & Reinhard F. Werner (2014): *Heisenberg uncertainty for qubit measurements*. *Phys. Rev. A* 89, p. 012129, doi:10.1103/PhysRevA.89.012129.
- [6] Jacqueline Erhart, Stephan Sponar, Georg Sulyok, Gerald Badurek, Masanao Ozawa & Yuji Hasegawa (2012): *Experimental demonstration of a universally valid error-disturbance uncertainty relation in spin measurements*. *Nature Phys.* 8(3), pp. 185–189.
- [7] K. Korzekwa, D. Jennings & T. Rudolph: ArXiv:1311.5506.
- [8] Masanao Ozawa (2004): *Uncertainty relations for joint measurements of noncommuting observables*. *Phys. Lett. A* 320(5–6), pp. 367 – 374, doi:http://dx.doi.org/10.1016/j.physleta.2003.12.001.
- [9] Lee A. Rozema, Ardavan Darabi, Dylan H. Mahler, Alex Hayat, Yasaman Soudagar & Aephraim M. Steinberg (2012): *Violation of Heisenberg's Measurement-Disturbance Relationship by Weak Measurements*. *Phys. Rev. Lett.* 109, p. 100404, doi:10.1103/PhysRevLett.109.100404.

The ZX-calculus is approximately complete for single qubits

Miriam Backens

Department of Computer Science, University of Oxford
Wolfson Building, Parks Road, Oxford, OX1 3QD, UK
`miriam.backens@cs.ox.ac.uk`

The ZX-calculus is a graphical calculus for reasoning about pure state qubit quantum mechanics. It is complete for pure qubit stabilizer quantum mechanics, meaning any equality involving only stabilizer operations that can be derived using matrices can also be derived pictorially. Yet for general pure state qubit quantum mechanics, the ZX-calculus is incomplete: there exist equalities involving non-stabilizer operations on single qubits which can not be derived from the current rule set for the ZX-calculus. Here, we show that the ZX-calculus for single qubits remains complete upon adding the operator $T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix}$ to the single-qubit stabilizer operations. As any single-qubit operator can be approximated to arbitrary accuracy using only single-qubit stabilizer operations and T , we conclude that the ZX-calculus for single qubits is approximately complete.

1 Introduction

The ZX-calculus introduced by Coecke and Duncan [4] is a powerful graphical calculus for pure state qubit quantum mechanics. It combines the advantages of using both dimensions of a sheet of paper, as in quantum circuit notation, with a built-in system of rewrite rules. These allow computations to be done graphically without the need to re-state or re-derive circuit identities each time. Like quantum circuit notation, the ZX-calculus can be used to express any operation in pure state qubit quantum mechanics, i.e. it is *universal*. Furthermore the rewrite rules can easily be shown to hold true when translated into matrix mechanics; therefore any equality derived in the ZX-calculus can also be derived in matrix mechanics. This property is called *soundness*.

A more intricate question is that of *completeness*: Can any equality that is true in matrix mechanics also be derived graphically using the given rule set? As recently shown, the answer is no: there are equalities in pure state qubit quantum mechanics, even when restricted to single-qubit operators, that cannot be derived graphically using the current set of rewrite rules [7]. Yet when the set of allowed operations is restricted to stabilizer quantum mechanics, the answer is yes, even for multi-qubit states and operations [2].

The overall incompleteness result, together with the completeness result for a small fragment of pure state qubit quantum mechanics—there are only finitely many distinct stabilizer operations on any fixed finite number of qubits—pose the question of what is the largest fragment of quantum mechanics for which the ZX-calculus is complete. For example, there exist many *approximately universal* sets of unitary operators: finite sets of operators, which can nevertheless be used to approximate any unitary to arbitrary accuracy. We will call the ZX-calculus *approximately complete* if it is complete for such an approximately universal set.

Here, we make a first step towards an approximate completeness result for the ZX-calculus by showing that it is approximately complete for single-qubit operators. The approximately universal set used is

the Clifford+T group [3], which is generated by the operators

$$T = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\pi/4} \end{pmatrix} \quad \text{and} \quad H = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}. \quad (1)$$

The core of the completeness proof is a normal form theorem for single-qubit Clifford+T operators, based on a result for quantum circuits by Matsumoto and Amano [6]. We show how an arbitrary single-qubit Clifford+T operator in the ZX-calculus can be brought into normal form and also prove that the normal form is unique. This implies that any equality between single-qubit Clifford+T operators that can be derived in matrix mechanics can also be derived in the ZX-calculus: as all rewrite rules are invertible, bringing two diagrams into the same normal form directly yields a series of rewrite steps transforming one diagram into the other.

In section 2, we present the elements and rules of the ZX-calculus for single-qubit Clifford+T operators, as well as a number of definitions and lemmas used throughout the rest of the paper. Section 3 contains the completeness proof. Some conclusions and ideas for further work are given in section 4.

2 Preliminaries

2.1 Elements and rules of the ZX-calculus

In this paper, we are only considering a small fragment of the ZX-calculus and only introduce the components and rules needed to show the result. For a treatment of the full ZX-calculus see [5].

Diagrams of the ZX-calculus consist of nodes and wires between them. Wires may also end at “empty nodes”, representing inputs and outputs of the diagram. Here, we are interested only in line graphs, i.e. diagrams in which there are exactly two wires connected to any non-empty node. The three types of nodes are interpreted as follows:

$$\bullet \phi :: \begin{cases} |0\rangle \mapsto |0\rangle \\ |1\rangle \mapsto e^{i\phi} |1\rangle \end{cases} \quad \bullet \theta :: \begin{cases} |+\rangle \mapsto |+\rangle \\ |-\rangle \mapsto e^{i\theta} |-\rangle \end{cases} \quad \square :: \begin{cases} |0\rangle \mapsto |+\rangle \\ |1\rangle \mapsto |-\rangle \end{cases}$$

where the phases ϕ and θ are multiples of $\pi/4$, and $|\pm\rangle = \frac{1}{\sqrt{2}}(|0\rangle \pm |1\rangle)$. A single wire corresponds to an identity operation. Diagrams are read from bottom to top, i.e. the dangling wire at the bottom is considered the input, the one at the top the output. Connecting the output of one node to the input of another corresponds to serial composition of operators.

For example, $\bullet \pi$ is the Pauli-Z operator and $\bullet \pi$ is the Pauli-X operator. Thus



corresponds to applying first Z and then X. The node $\bullet \pi/4$ represents the T -operator defined in (1).

It is easy to see that the three types of nodes with the given phase values are all in the Clifford+T group. Furthermore, as $\{H, T\}$ is a generating set for this group, any Clifford+T operator can be expressed as a diagram in terms of \square and $\bullet \pi/4$.

The ZX-calculus is not just an alternative notation, it comes with rewriting rules that allow the derivation of equalities between diagrams. The rules relevant for this paper are the following, where n is an integer:

$$\begin{array}{c}
\text{green circle } 2n\pi = | \quad (\text{Id}) \qquad \begin{array}{c} \text{green circle } \phi \\ \text{green circle } \theta \end{array} = \text{green circle } \phi + \theta \quad (\text{S}) \qquad \begin{array}{c} \text{green circle } \pi \\ \text{red circle } \phi \end{array} = \begin{array}{c} \text{red circle } -\phi \\ \text{green circle } \pi \end{array} \quad (\text{P}) \\
\\
\begin{array}{c} \text{yellow box } H \\ \text{green circle } \phi \\ \text{yellow box } H \end{array} = \begin{array}{c} \text{red circle } \phi \end{array} \quad (\text{C}) \qquad \begin{array}{c} \text{yellow box } H \\ \text{green circle } \pi/2 \\ \text{red circle } \pi/2 \\ \text{green circle } \pi/2 \end{array} = \begin{array}{c} \text{green circle } \pi/2 \\ \text{red circle } \pi/2 \\ \text{green circle } \pi/2 \end{array} \quad (\text{Eu})
\end{array}$$

All of these rules also hold with the colours reversed. Furthermore, the rules are all sound—i.e. they hold true when translated back into Dirac or matrix notation—if equality is taken to be up to global scalar factors.

E.g. by rule (P) with $\phi = \pi$, we have

$$\begin{array}{c} \text{green circle } \pi \\ \text{red circle } \pi \end{array} = \begin{array}{c} \text{red circle } \pi \\ \text{green circle } \pi \end{array}$$

In matrix notation, the two diagrams correspond to

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

The two matrices are not equal, but as

$$-\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix},$$

the equality holds up to a scalar factor.

2.2 Some further definitions and lemmas

We shall denote the single-qubit Clifford group by \mathcal{C}_1 ; this group contains all diagrams built up from the components given in section 2.1 in which the phases of red and green nodes are restricted to integer multiples of $\pi/2$. It will also be useful to define two further sets of ZX-calculus operators:

$$\mathcal{W} = \left\{ \begin{array}{c} | \\ \text{red circle } \pi/2 \\ \text{green circle } \pi/2 \end{array} \right\} \quad \text{and} \quad \mathcal{V} = \left\{ \begin{array}{c} \text{green circle } \pi/4 \\ \text{red circle } \pi/2 \\ \text{green circle } 3\pi/4 \\ \text{red circle } \pi/2 \end{array} \right\}.$$

In the remainder of this section we will prove various lemmas about how operators in \mathcal{C}_1 , \mathcal{W} and \mathcal{V} compose.

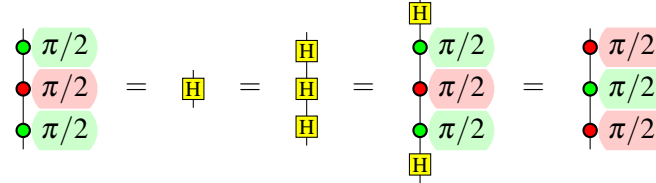
Lemma 1. *The following two sets each contain a unique representation for each operator $C \in \mathcal{C}_1$:*

$$\left\{ \begin{array}{c} \text{red circle } \alpha \\ \text{green circle } \beta \\ \text{green circle } \pi/2 \\ \text{red circle } \pm\pi/2 \\ \text{green circle } \gamma \end{array} \right\} \quad \text{and} \quad \left\{ \begin{array}{c} \text{green circle } \beta \\ \text{red circle } \alpha \\ \text{green circle } \gamma \\ \text{red circle } \pm\pi/2 \\ \text{green circle } \pi/2 \end{array} \right\}, \quad (2)$$

where in both cases $\alpha, \beta, \gamma \in \{0, \pi/2, \pi, -\pi/2\}$.

Proof. First note that any single-qubit Clifford operator can be written in terms of red and green nodes only, by substituting for the Hadamard nodes using the rule (Eu). Then, each such operator must have a representation with no more than three nodes: given any diagram with at least four nodes, either

- there are two adjacent nodes of the same colour, in which case they can be merged by rule (S), or
- there is a node with a phase that is a multiple of 2π , in which case it can be removed by rule (Id), or
- there is a node with a phase of π , in which case it can be moved past a node of the other colour using rule (P) and then merged with another node of the same colour, or
- there are three adjacent nodes with phases in the set $\{\pm\pi/2\}$. In this last case, note that



Similar results can be derived for any combination of plus and minus signs in the phases. Hence if there is a sequence of four nodes of alternating colours, all of which have phases in the set $\{\pm\pi/2\}$, we can change the colours of three of them, and thus get two adjacent nodes of the same colour, which can be merged.

In each of the cases listed above the number of nodes in the diagram can be reduced by applying suitable rewrite rules. The strategy works until there are no more than three nodes left. Having reduced all diagrams to at most three nodes, it is straightforward—albeit somewhat tedious—to check that the given sets indeed contain a unique representation of each Clifford operator. \square

Note that lemma 1 shows directly that the ZX-calculus is complete for single qubit Clifford operators.

Lemma 2. *The following set contains a unique representation of each operator of the form TC , where $C \in \mathcal{C}_1$:*

$$\mathcal{U} = \left\{ \begin{array}{c} \text{green } \pi/4 + \beta \\ \text{red } \alpha \\ \text{green } \pi/2 \end{array} \quad \begin{array}{c} \text{green } \pi/4 + \gamma \\ \text{red } \pm\pi/2 \\ \text{green } \pi/2 \end{array} \right\}$$

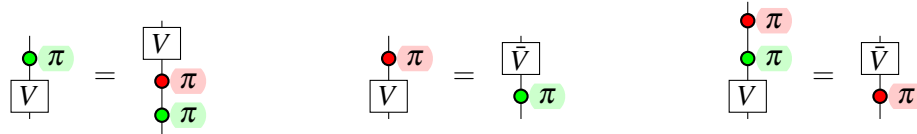
if $\alpha, \beta, \gamma \in \{0, \pi/2, \pi, -\pi/2\}$.

Proof. This follows immediately from lemma 1. \square

Lemma 3. *Let $C \in \mathcal{C}_1$, $U \in \mathcal{U}$ and $V \in \mathcal{V}$. Then*

$$\begin{array}{c} \boxed{C} \\ \boxed{V} \end{array} = \begin{array}{c} \boxed{W} \\ \boxed{V'} \end{array} \quad \text{and} \quad \begin{array}{c} \boxed{C} \\ \boxed{U} \end{array} = \begin{array}{c} \boxed{W} \\ \boxed{U'} \end{array}$$

for some $W \in \mathcal{W}$, $U' \in \mathcal{U}$, $V' \in \mathcal{V}$ and $a, b \in \{0, 1\}$. For the particular case of the first equality where C consists solely of π phase shifts, W is the identity and we have



with $\bar{V} \in \mathcal{V} \setminus \{V\}$.

Proof. Substitute for C using the first set of normal forms given in lemma 1 and for V and U using the definitions of \mathcal{V} and \mathcal{U} ; the results then follow from straightforward application of the rules of the ZX-calculus. \square

Lemma 4. Suppose $V_1, \dots, V_n \in \mathcal{V}$ for some positive integer n . Then if $a, b \in \{0, 1\}$,

for some $a', b' \in \{0, 1\}$ and $V'_1, \dots, V'_n \in \mathcal{V}$.

Proof. By induction on n , using the second part of lemma 3. \square

3 Completeness

Theorem 5. Any single-qubit operator consisting of phase shifts that are multiples of $\pi/4$ and Hadamard operators is either a Clifford operator or it can be written in the normal form

for some integer $n \geq 0$, where $W \in \mathcal{W}$, $V_1, \dots, V_n \in \mathcal{V}$ and $U \in \mathcal{U}$.

Proof. It is easy to see that any single-qubit Clifford+T operator can be written solely in terms of $\bullet \pi/2$ and $\bullet \pi/4$. To prove the theorem, it thus suffices to show that adding $\bullet \pi/2$ or $\bullet \pi/4$ to any Clifford operator or any diagram in normal form yields a diagram that can be rewritten to a Clifford operator or normal form diagram.

Consider first $\bullet \pi/2$. This is a Clifford operator, so adding it to a Clifford diagram yields another Clifford diagram. Furthermore,

for some $C \in \mathcal{C}_1$, so if $n > 0$,

by lemmas 3 and 4, where $a, b \in \{0, 1\}$, $W' \in \mathcal{W}$, $U' \in \mathcal{U}$ and $V'_1, \dots, V'_n \in \mathcal{V}$. From lemma 3, we also have that, if $n = 0$, the diagram resulting from the application of $\bullet \pi/2$ to a normal form diagram can be rewritten into normal form. This covers all the cases.

Now consider $\bullet \pi/4$ instead. Note that

$$\begin{array}{c} \bullet \pi/4 \\ \boxed{C} \end{array} = \boxed{U'} \quad \text{and} \quad \begin{array}{c} \bullet \pi/4 \\ \boxed{U} \end{array} = \boxed{C'}$$

for some $U' \in \mathcal{U}$ and $C' \in \mathcal{C}_1$. Furthermore, unless W is the identity,

$$\begin{array}{c} \bullet \pi/4 \\ \boxed{W} \end{array} = \boxed{V}$$

for some $V \in \mathcal{V}$. Thus adding $\bullet \pi/4$ to a Clifford operator or a normal form diagram with non-trivial W results in diagrams that can be rewritten to normal form. If W is the identity and $n = 0$, then the result of adding $\bullet \pi/4$ will be a Clifford diagram.

It remains to check what happens when W is the identity and $n > 0$. For any $V_n \in \mathcal{V}$, we can find $W \in \mathcal{W}$ and $a \in \{0, 1\}$ such that

$$\begin{array}{c} \bullet \pi/4 \\ \boxed{V_n} \end{array} = \begin{array}{c} \boxed{W} \\ \bullet a\pi \\ \bullet a\pi \end{array}$$

Then by lemmas 3 and 4, the entire diagram can be brought into normal form.

Thus, whenever $\bullet \pi/2$ or $\bullet \pi/4$ is added to a Clifford circuit or normal form diagram, the resulting diagram can be rewritten into a Clifford circuit or normal form diagram, completing the proof. \square

Theorem 6. *No normal form diagram as given in (3) is equal to the identity.*

Proof. We will show that in matrix mechanics, no normal form circuit is equal to a scalar multiple of the identity matrix. As the ZX-calculus is sound, this implies that no normal form circuit is equal to the identity within the ZX-calculus.

Following [6], we use an adaptation of the stabilizer formalism. Let $M_{(x,y,z)} := xX + yY + zZ$, where X, Y, Z are the Pauli matrices. We say that a single qubit state $|\psi\rangle$ is *stabilized* by (x, y, z) if $M_{(x,y,z)} |\psi\rangle = |\psi\rangle$. It is easy to see that if (x, y, z) stabilizes $|0\rangle$, then $(x, y, z) = (0, 0, 1)$.

Let S be the phase gate, and denote the $\bullet \pi/2$ operator by R , so $\mathcal{V} = \{TR, TSR\}$. Now suppose (x, y, z) stabilizes some state $|\psi\rangle$. Then for any $C \in \mathcal{C}_1$, $C|\psi\rangle$ is stabilized by some expression of the form $(a\sigma(x), b\sigma(y), c\sigma(z))$, where σ is some permutation on the set $\{x, y, z\}$ and $a, b, c \in \{\pm 1\}$. This is because $C|\psi\rangle = (CM_{(x,y,z)}C^{-1})C|\psi\rangle$ and conjugation by a Clifford operator maps the set of Pauli matrices to itself, up to factors of ± 1 . Furthermore,

- $T|\psi\rangle$ is stabilized by $\frac{1}{\sqrt{2}}(x - y, x + y, z)$,
- $TR|\psi\rangle$ is stabilized by $\frac{1}{\sqrt{2}}(x + z, x - z, y)$, and
- $TSR|\psi\rangle$ is stabilized by $\frac{1}{\sqrt{2}}(z - x, x + z, y)$.

We shall consider the effect of applying a normal form diagram to $|0\rangle$. First, consider the case where W is the identity and $n = 0$, i.e. the diagram is simply of the form TC for some Clifford operator C . Now $TC|0\rangle$ is stabilized by one of the expressions

$$\frac{1}{\sqrt{2}}(\pm 1, \pm 1, 0), \quad \frac{1}{\sqrt{2}}(\mp 1, \pm 1, 0), \quad \text{and} \quad (0, 0, \pm 1). \quad (4)$$

Even though one of the potential stabilizers is $(0, 0, 1)$, it is easy to see that TC is not a scalar multiple of the identity for any C .

Next consider the possible stabilizers for $V_1 TC|0\rangle$, where $V_1 \in \mathcal{V}$. These are

$$\begin{aligned} & \frac{1}{2}(\pm 1, \pm 1, \pm \sqrt{2}), \quad \frac{1}{2}(\mp 1, \pm 1, \pm \sqrt{2}), \quad \frac{1}{2}(\mp 1, \mp 1, \pm \sqrt{2}), \quad \frac{1}{2}(\pm 1, \mp 1, \pm \sqrt{2}), \\ & \frac{1}{\sqrt{2}}(\pm 1, \pm 1, 0), \quad \text{and} \quad \frac{1}{\sqrt{2}}(\mp 1, \pm 1, 0). \end{aligned}$$

Any stabilizer in the set above can be expressed as

$$\frac{1}{\sqrt{2^m}}(x_1 + x_2\sqrt{2}, y_1 + y_2\sqrt{2}, z_1 + z_2\sqrt{2}), \quad (5)$$

where $m, x_1, x_2, y_1, y_2, z_1, z_2 \in \mathbb{Z}$ with $m \geq 0$. Applying a transformation from \mathcal{V} maps that stabilizer to

$$\frac{1}{\sqrt{2^{m+1}}} \left((x_1 + z_1) + (x_2 + z_2)\sqrt{2}, (x_1 - z_1) + (x_2 - z_2)\sqrt{2}, 2y_2 + y_1\sqrt{2} \right)$$

or

$$\frac{1}{\sqrt{2^{m+1}}} \left((z_1 - x_1) + (z_2 - x_2)\sqrt{2}, (x_1 + z_1) + (x_2 + z_2)\sqrt{2}, 2y_2 + y_1\sqrt{2} \right).$$

Note that $\mathcal{W} \subset \mathcal{C}_1$, so the effect of $W \in \mathcal{W}$ is at most a permutation of the numbers x, y, z and the introduction of minus signs. Thus the stabilizer of $U|\psi\rangle$ for any normal form operator U can be written in the form (5).

Following [6], we consider the parity of x_1, x_2, y_1, y_2, z_1 and z_2 under the transformations given by repeated application of elements of \mathcal{V} . For the stabilizers given in (4), we have either x_1 and y_1 odd and the others even, or z_1 odd and the others even. For a given a, b , the parity of $|a - b|$ is the same as that of $a + b$, so the two transformations in \mathcal{V} have the same effects on the parity of x_1, x_2, y_1, y_2, z_1 and z_2 .

If x_1 and y_1 are odd and the others even, then after application of some $V \in \mathcal{V}$, x_1, y_1 , and z_2 are odd. A second application of V leads to a stabilizer where all factors are odd except for z_1 . A third application of V gives a stabilizer where once again x_1, y_1 , and z_2 are odd. Thus the parity of these factors changes cyclically.

If z_1 is odd in the beginning and the other factors are even, then after one application of V , x_1, y_1 and z_2 are odd, after which the same cyclical behaviour appears as above.

Note that if $WV_n \dots V_1 TC$ is to be a scalar multiple of the identity, then $V_n \dots V_1 TC|0\rangle$ must have a stabilizer in the set $\{(0, 0, c), (0, c, 0)\}$ for some non-zero c , i.e. either $x_1 = x_2 = y_1 = y_2 = 0$ or $x_1 = x_2 = z_1 = z_2 = 0$. In particular, $WV_n \dots V_1 TC$ can only be the identity if $V_n \dots V_1 TC|0\rangle$ has a stabilizer in which either x_1, x_2, y_1 , and y_2 are all even, or x_1, x_2, z_1 , and z_2 are all even. Yet, as shown above, for any $V_n \dots V_1 TC|0\rangle$, the factor x_1 in the stabilizer is always odd. Thus $WV_n \dots V_1 TC$ is never the identity, completing the proof. \square

Lemma 7. Consider a normal form diagram $D = WV_n \dots V_1 U$. Then D^\dagger is equal to some normal form diagram with the same number of copies of elements of \mathcal{V} , i.e. $D^\dagger = W'V'_n \dots V'_1 U'$ for some $W' \in \mathcal{W}$, $V'_1, \dots, V'_n \in \mathcal{V}$ and $U' \in \mathcal{U}$.

Proof. By the properties of the dagger functor, $D^\dagger = U^\dagger V_1^\dagger \dots V_n^\dagger W^\dagger$. Now for any $U \in \mathcal{U}$, we can find $C \in \mathcal{C}_1$ such that

$$\boxed{U^\dagger} = \boxed{C} \text{ --- } \text{green circle } \pi/4$$

and for any $V \in \mathcal{V}$, we have

$$\boxed{V^\dagger} = \boxed{V} \text{ --- } \text{red circle } \pi/2$$

Thus by lemmas 3 and 4,

$$\begin{array}{c} \boxed{U^\dagger} \\ \boxed{V_1^\dagger} \\ \vdots \\ \boxed{V_n^\dagger} \\ \boxed{W^\dagger} \end{array} = \begin{array}{c} \boxed{C} \\ \text{green circle } \pi/4 \\ \text{red circle } \pi/2 \\ \boxed{V_1} \\ \text{red circle } \pi/2 \\ \vdots \\ \text{red circle } \pi/2 \\ \boxed{V_n} \\ \text{red circle } \pi/2 \\ \boxed{W^\dagger} \end{array} = \begin{array}{c} \boxed{C} \\ \boxed{V_0} \\ \boxed{V_1} \\ \text{red circle } \pi \\ \boxed{V_2} \\ \vdots \\ \boxed{V_n} \\ \text{red circle } \pi \\ \boxed{V_n} \\ \text{red circle } \pi/2 \\ \boxed{W^\dagger} \end{array} = \begin{array}{c} \boxed{W'} \\ \boxed{V'_0} \\ \vdots \\ \boxed{V'_n} \\ \text{green circle } a\pi \\ \text{red circle } \pm\pi/2 \\ \boxed{W^\dagger} \end{array} = \begin{array}{c} \boxed{W'} \\ \boxed{V'_0} \\ \vdots \\ \boxed{V'_{n-1}} \\ \boxed{U'} \end{array} = \begin{array}{c} \boxed{W'} \\ \boxed{V''_1} \\ \vdots \\ \boxed{V''_n} \\ \boxed{U'} \end{array} \quad (6)$$

for some $W' \in \mathcal{W}$, $V'_0, \dots, V'_n, V''_1, \dots, V''_n \in \mathcal{V}$, and $U' \in \mathcal{U}$. Note that V''_1, \dots, V''_n is just a relabelling of V'_{n-1}, \dots, V'_0 . \square

Theorem 8. The normal form for Clifford+T diagrams given in (3) is unique.

Proof. Suppose there are two normal form diagrams which are equal but not identical. Pick a shortest pair of such diagrams, i.e. suppose the topmost nodes in the two digrams have different colours or different phases (or both). If the topmost nodes are the same, remove them both and keep going like this until a stage is reached where the remaining topmost nodes are different. As the two diagrams are not identical, this must be possible.

Call these two diagrams D_1 and D_2 . As $D_1 = D_2$ by assumption, and because any normal form diagram is unitary, it must be the case that $D_1^\dagger \circ D_2$ is equal to the identity. We will show that under the given assumptions, $D_1^\dagger \circ D_2$ must be equal to some non-trivial normal form diagram. By theorem 6, this normal form diagram cannot be equal to the identity, thus leading to a contradiction. From that we conclude that two normal form diagrams are equal if and only if they are identical.

Suppose D_1 can be written in normal form as $WV_n \dots V_1 U$ and D_2 as $W'V'_m \dots V'_1 U'$. The requirement that the topmost nodes of D_1 and D_2 be different can be satisfied in different ways. Where the conditions are not symmetric under interchange of D_1 and D_2 , by lemma 7 it nevertheless suffices to consider just one of the two options. We will hence distinguish the following cases:

- $W = W' = 1$, $n = m = 0$, and the topmost nodes of U and U' differ
- $W = W' = 1$, $n = 0 \neq m$, and the topmost nodes of U and V'_m differ
- $W = W' = 1$, $n, m \neq 0$, and $V_n \neq V'_m$
- $W \neq W'$, $n = m = 0$
- $W \neq W'$, $n = 0 \neq m$
- $W \neq W'$, $n, m \neq 0$

Firstly, if $W = W' = 1$ and $n = m = 0$, then $D_1 = U$ and $D_2 = U'$ with $U, U' \in \mathcal{U}$. Now any element of \mathcal{U} can be expressed as TC , for some $C \in \mathcal{C}$. Thus $D_1 = TC$ and $D_2 = TC'$, and as $U \neq U'$ we must have $C \neq C'$. Therefore,

$$\begin{array}{c} \boxed{U^\dagger} \\ \boxed{U'} \end{array} = \begin{array}{c} \boxed{C^\dagger} \\ \bullet -\pi/4 \\ \bullet \pi/4 \\ \boxed{C'} \end{array} \neq \left| \right.$$

Secondly, if $W = W' = 1$ and $n = 0 \neq m$, consider U and V'_m . Note that $U = TC$ for some Clifford operator C , and $V'_m = TC'$ for some Clifford operator C' . Again, the requirement that the topmost nodes of U and V'_m be different means that $C \neq C'$. As in the first case, we thus find $U^\dagger V'_m = C''$ for some C'' . Then by lemmas 3 and 4, $D_1^\dagger \circ D_2$ has a normal form $W''V''_{m-1} \dots V''_1 U''$. As $m > 0$, this is non-trivial.

The third case, $W = W' = 1$, $n, m \neq 0$, and $V_n \neq V'_m$, can be reduced to a case where $W \neq W'$ by applying $\bullet -\pi/4$ to both diagrams and using rule (S).

For $W \neq W'$, we have (after some rewriting),

$$\begin{array}{c} \boxed{W^\dagger} \\ \boxed{W'} \end{array} \in \left\{ \begin{array}{c} \bullet \pm\pi/2 \\ \bullet \pi/2 \\ \bullet \pi/2 \\ \bullet \pi/2 \end{array} \right\}. \quad (7)$$

Then if $n = m = 0$,

$$\begin{array}{c} \boxed{U^\dagger} \\ \boxed{W^\dagger} \\ \boxed{W'} \\ \boxed{U'} \end{array} = \begin{array}{c} \boxed{C} \\ \bullet \pi/4 \\ \boxed{W^\dagger} \\ \boxed{W'} \\ \boxed{U'} \end{array} = \begin{array}{c} \boxed{W''} \\ \bullet \pi/4 + \alpha \\ \bullet a\pi \\ \boxed{W^\dagger} \\ \boxed{W'} \\ \boxed{U'} \end{array} = \begin{array}{c} \boxed{W''} \\ \boxed{V} \\ \bullet \gamma \\ \bullet c\pi \\ \boxed{U'} \end{array} = \begin{array}{c} \boxed{W''} \\ \boxed{V} \\ \boxed{U''} \end{array}$$

since

$$\begin{array}{c} \bullet \pi/4 + \alpha \\ \bullet a\pi \\ \boxed{W^\dagger} \\ \boxed{W'} \end{array} \in \left\{ \begin{array}{c} \bullet \pi/4 + \beta \\ \bullet \pm\pi/2 \\ \bullet \pi/2 \end{array} \right\} = \left\{ \begin{array}{c} \boxed{V} \\ \bullet \gamma \\ \bullet c\pi \end{array} \right\}$$

for some $\alpha, \beta, \gamma \in \{0, \pi/2, \pi, -\pi/2\}$, $a, c \in \{0, 1\}$ and $V \in \mathcal{V}$.

The argument for the case $W \neq W'$ and $n = 0 \neq m$ is very similar, noting that for any $c \in \{0, 1\}$, $\gamma \in \{0, \pi/2, \pi, -\pi/2\}$, and $V \in \mathcal{V}$

$$\begin{array}{c} \text{green dot } \gamma \\ \text{red dot } c\pi \\ \boxed{V} \end{array} = \begin{array}{c} \boxed{V'} \\ \text{green dot } a\pi \\ \text{red dot } b\pi \end{array}$$

for some $V' \in \mathcal{V}$ and $a, b \in \{0, 1\}$. Hence by lemmas 3 and 4, the diagram can be rewritten into normal form.

Lastly, consider the case where $W \neq W'$ and $n, m \neq 0$. By lemma 7, we can rewrite D_1^\dagger to

$$\begin{array}{c} \boxed{W'} \\ \boxed{V'_0} \\ \vdots \\ \boxed{V'_n} \\ \text{green dot } a\pi \\ \text{red dot } \pm\pi/2 \\ \boxed{W^\dagger} \end{array}$$

Now

$$\begin{array}{c} \boxed{V'_n} \\ \text{green dot } a\pi \\ \text{red dot } \pm\pi/2 \end{array} = \begin{array}{c} \text{green dot } \pi/4 + \beta \\ \text{red dot } b\pi \end{array}$$

for some $\beta \in \{0, \pi/2, \pi, -\pi/2\}$ and $b \in \{0, 1\}$. Thus the argument concludes in the same way as in the previous case.

We have shown that for any pair of normal form diagrams D_1 and D_2 , $D_1^\dagger \circ D_2$ has a non-trivial normal form unless the two diagrams are identical. Therefore, by theorem 6 and by unitarity of Clifford+T operators, two normal form diagrams are equal if and only if they are identical, i.e. the normal form is unique. \square

4 Conclusions

We have shown that the ZX-calculus is approximately complete for unitary single-qubit operations by showing that it is complete for the approximately universal single-qubit Clifford+T group. The proof yields a unique normal form for single-qubit Clifford+T diagrams. An obvious next step is to attempt to extend the approximate completeness proof to multiple qubits, possibly by combining the results from this paper with [2]. It would also be useful to implement the normalisation algorithm in the automated graph rewriting system *Quantomatic* [1].

While we have used the fact that the Clifford+T group is approximately universal for pure single-qubit quantum mechanics to justify calling the ZX-calculus approximately complete, there does not currently exist any notion of approximation of operators within the calculus itself. It would be interesting to define an approximate equality relation for diagrams, which could then be used to transform arbitrary line graphs into Clifford+T diagrams, normalise, and compare them.

References

- [1] *Quantomatic*. <https://sites.google.com/site/quantomatic/>.

- [2] Miriam Backens (2013): *The ZX-calculus is complete for stabilizer quantum mechanics*. arXiv e-print 1307.7025. Available at <http://arxiv.org/abs/1307.7025>.
- [3] P. Oscar Boykin, Tal Mor, Matthew Pulver, Vwani Roychowdhury & Farrokh Vatan (1999): *On universal and fault-tolerant quantum computing: A novel basis and a new constructive proof of universality for Shor's basis*. In: *40th Annual Symposium on Foundations of Computer Science (Cat. No.99CB37039)*, IEEE, pp. 486–494, doi:10.1109/SFFCS.1999.814621.
- [4] Bob Coecke & Ross Duncan (2008): *Interacting Quantum Observables*. In: *Automata, Languages and Programming*, 5126, Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 298–310, doi:10.1007/978-3-540-70583-3_25.
- [5] Bob Coecke & Ross Duncan (2011): *Interacting quantum observables: categorical algebra and diagrammatics*. *New Journal of Physics* 13(4), p. 043016, doi:10.1088/1367-2630/13/4/043016.
- [6] Ken Matsumoto & Kazuyuki Amano (2008): *Representation of Quantum Circuits with Clifford and $\pi/8$ Gates*. arXiv:0806.3834. Available at <http://arxiv.org/abs/0806.3834>.
- [7] Vladimir Zamdzhiev (2013): *Private communication*.

Tensors, !-graphs, and non-commutative quantum structures

Aleks Kissinger

University of Oxford

aleks.kissinger@cs.ox.ac.uk

David Quick

University of Oxford

david.quick@cs.ox.ac.uk

Categorical quantum mechanics (CQM) and the theory of quantum groups rely heavily on the use of structures that have both an algebraic and co-algebraic component, making them well-suited for manipulation using diagrammatic techniques. Diagrams allow us to easily form complex compositions of (co)algebraic structures, and prove their equality via graph rewriting. One of the biggest challenges in going beyond simple rewriting-based proofs is designing a graphical language that is expressive enough to prove interesting properties (e.g. normal form results) about not just single diagrams, but entire families of diagrams. One candidate is the language of *!-graphs*, which consist of graphs with certain subgraphs marked with boxes (called *!-boxes*) that can be repeated any number of times. New *!-graph* equations can then be proved using a powerful technique called *!-box induction*. However, previously this technique only applied to commutative (or cocommutative) algebraic structures, severely limiting its applications in some parts of CQM and (especially) quantum groups. In this paper, we fix this shortcoming by offering a new semantics for *non-commutative* *!-graphs* using an enriched version of Penrose’s abstract tensor notation.

1 Introduction

Diagrammatic theories give us a way to study a wide variety of algebraic and coalgebraic structures in monoidal categories. They consist of two parts: a *signature* Σ and a set of *diagram equations* E . The signature consists of a set of objects $\{A, B, \dots\}$ along with a set of generating morphisms with input and output arities formed from combining objects with \otimes and I . For example, the signature of a Frobenius algebra consists of four morphisms: $(\mu : A \otimes A \rightarrow A, \eta : I \rightarrow A, \delta : A \rightarrow A \otimes A, \varepsilon : A \rightarrow I)$, or, written diagrammatically:

$$\Sigma = \left\{ \begin{array}{c} \text{multiplication } \mu \\ \text{comultiplication } \delta \\ \text{unit } \eta \\ \text{counit } \varepsilon \end{array} \right\}$$

Then, E is a set of equations between morphisms built from these generators, which we can picture as equations between string diagrams. For example, the theory of commutative Frobenius algebras contains the (co)associativity, (co)unit, (co)commutativity and Frobenius equations:

A *model* of (Σ, E) in a (symmetric, traced, or compact closed) monoidal category \mathcal{C} assigns a morphism to each generator in Σ such that all equations in E hold.

Remark 1.1. Many familiar algebraic constructions arise as special cases of this setup. For instance, any linear ‘term-like’ algebraic theory (i.e. where free variables occur precisely once on the LHS and RHS of every equation) can be presented this way. Also, if we restrict to equations in E that are directed acyclic, we obtain presentations of PROPs (or coloured PROPs in the multi-sorted case). In that case, models of (Σ, E) in \mathcal{C} are in 1-to-1 correspondence with strong monoidal functors from the presented PROP into \mathcal{C} .

This style of algebraic theory works well when generators have fixed, finite arity. However, it is often possible to find a much more elegant presentation of a theory if we allow the arity of our generators to vary. For instance, commutative Frobenius algebras can be alternatively presented using a single variable-arity generator sometimes called a ‘spider’, along with just two equations.

$$\Sigma = \left\{ \begin{array}{c} \cdots \\ \nearrow \quad \searrow \\ \circ \\ \nwarrow \quad \nearrow \\ \cdots \end{array} \right\} \quad E = \left\{ \begin{array}{c} \begin{array}{c} \cdots \quad \cdots \\ \nearrow \quad \searrow \quad \nearrow \quad \searrow \\ \circ \quad \circ \\ \nwarrow \quad \nearrow \quad \nwarrow \quad \nearrow \\ \cdots \quad \cdots \end{array} = \begin{array}{c} \cdots \quad \cdots \\ \nearrow \quad \searrow \\ \circ \\ \nwarrow \quad \nearrow \\ \cdots \quad \cdots \end{array}, \quad \begin{array}{c} \uparrow \\ \circ \\ \downarrow \end{array} = \begin{array}{c} \uparrow \\ \downarrow \end{array} \right\}$$

A model of such a theory is no longer just a finite set of morphisms, but rather, a set of *families* of morphisms $f_{j,k} : A^{\otimes j} \rightarrow A^{\otimes k}$, indexed by input/output arities, such that the equations in E hold for all possible arities.

Comparing this to the equations at the beginning of this section, we seem to have lost some formality. That is, the ‘concrete’ diagrammatic identities above can be formalised in such a way that proofs can be performed (and even machine-checked) via a suitable notion of diagram rewriting, as formalised in [2]. One might be tempted to think that this level of rigour is lost when we describe equations in a mathematical meta-language, making use of ellipses, for example, to represent repetition. However, in [1], the authors introduced *!-boxes* (pronounced ‘bang-boxes’) as a method for reasoning about graphs with repeated structure. As *!-box* rules, the previously informal rules can be formalised as:

$$\Sigma = \left\{ \begin{array}{c} \boxed{A} \\ \uparrow \\ \circ \\ \downarrow \\ \boxed{B} \end{array} \right\} \quad E = \left\{ \begin{array}{c} \begin{array}{c} \boxed{A} \quad \boxed{C} \\ \uparrow \quad \uparrow \\ \circ \quad \circ \\ \downarrow \quad \downarrow \\ \boxed{B} \quad \boxed{D} \end{array} = \begin{array}{c} \boxed{A} \quad \boxed{C} \\ \nearrow \quad \searrow \\ \circ \\ \nwarrow \quad \nearrow \\ \boxed{B} \quad \boxed{D} \end{array}, \quad \begin{array}{c} \uparrow \\ \circ \\ \downarrow \end{array} = \begin{array}{c} \uparrow \\ \downarrow \end{array} \right\}$$

Intuitively, marking a subgraph with a *!-box* means that subgraph (along with edges in/out of it) can be repeated any number of times to obtain an *instance* of the graph. Thus we interpret a graph with *!-boxes* as a set of all its instances.

$$\left[\begin{array}{c} \boxed{A} \\ \uparrow \\ \circ \\ \downarrow \\ \boxed{B} \end{array} \right] := \left\{ \circ, \uparrow \circ, \begin{array}{c} \nearrow \quad \searrow \\ \circ \end{array}, \dots, \begin{array}{c} \uparrow \\ \circ \end{array}, \begin{array}{c} \uparrow \\ \downarrow \end{array}, \begin{array}{c} \nearrow \quad \searrow \\ \circ \end{array}, \dots \right\}$$

Similarly, for rules with *!-boxes*, matched pairs of *!-boxes* can be repeated in the LHS and RHS to obtain instances of that rule. Thus, for our example of the commutative Frobenius algebra, we have reduced our theory of 7 equations to just 2.

!-boxes were given a formal semantics in [7], making use of adhesive categories [8]. They also come with a simple and powerful induction principle introduced by one of the authors in [6] and proven correct in [10]. But there’s a catch: note how we were careful to say that *commutative* Frobenius algebras

have an elegant presentation as above. A major drawback of the existing !-box notation is that it is only unambiguous if all of the nodes in the diagram are invariant under permuting inputs/outputs. This is severely limiting in two ways. The first and most obvious limitation is that we are forced to consider only commutative algebraic structures. The second, more subtle limitation is that we have no freedom to *definitionally* extend our theory, i.e. introduce new nodes defined as diagrams of other nodes, without making implicit assumptions about those diagrams (namely, that they are symmetric on inputs/outputs).

In order to overcome these shortcomings, we extend the !-graph notation with some extra information about how newly-created edges should be ordered when a !-box is expanded. This turns out to be fairly straightforward as soon as one shifts from a graph-based semantics for diagrams, as employed in [2], to a *tensor-based* semantics, where morphisms in the free compact closed category are represented using a version of Penrose’s abstract tensor notation [11]. This approach, recently formalised in [5], has the property that non-commutativity comes ‘for free’, where the edges connected to a single element are represented as a list of edge names. Contrast this with the graph-based semantics for string diagrams or Joyal and Street’s geometric construction [4], where one needs to add some extra structure (e.g. a total ordering or typing on adjacent edges) to break symmetries.

So, without further ado, we introduce tensor expressions for compact closed categories and extend them to accommodate !-boxes.

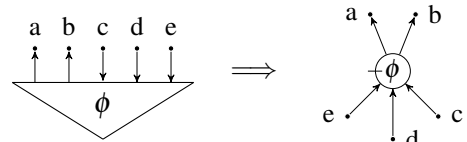
2 Tensors

Assume we are working in a compact closed category \mathcal{C} freely generated by a set of objects X, Y, Z, \dots and morphisms of the form $\phi : I \rightarrow X_1 \otimes \dots \otimes X_n$, i.e. morphisms with only non-trivial outputs. Since \mathcal{C} is compact closed, this yields no loss of generality, since we represent an input of type A as an output of type A^* . For simplicity, we’ll assume every ‘input’ is of fixed type X^* and every ‘output’ is of type X .

Since we want to distinguish inputs/outputs we label them using lower case letters. They will have a hat to illustrate being an ‘output’: $\{\hat{a}, \hat{b}, \dots\}$, or a check to illustrate being an ‘input’: $\{\check{a}, \check{b}, \dots\}$. Translating a morphism ϕ into tensor notation yields:

$$\phi : I \rightarrow X \otimes X \otimes X^* \otimes X^* \otimes X^* \implies \phi_{\hat{a}\hat{b}\check{c}\check{d}\check{e}}$$

We introduce a special graphical notation for morphisms with only outputs. We write them as circles with a tick, taking the convention that inputs/outputs are ordered clockwise from the tick.



Writing two tensors side-by-side yields a new tensor formed by taking the monoidal product and ‘contracting’ any repeated names using the compact structure on X .

$$\psi_{\check{f}\check{a}\check{b}} \phi_{\hat{a}\hat{b}\check{c}\check{d}\check{e}} := \begin{array}{c} \text{Diagram with two triangles: } \psi \text{ (left) and } \phi \text{ (right). } \psi \text{ has inputs } \check{f}, \check{a}, \check{b}. \phi \text{ has inputs } \hat{a}, \hat{b}, \check{c}, \check{d}, \check{e}. \text{ Red arcs connect } \check{a} \text{ to } \hat{a} \text{ and } \check{b} \text{ to } \hat{b}. \end{array} \implies \begin{array}{c} \text{Diagram with a stacked circle } \psi \text{ (top) and } \phi \text{ (bottom). } \psi \text{ has input } \check{f}. \phi \text{ has inputs } \check{c}, \check{d}, \check{e}. \text{ Red arcs connect } \psi \text{ to } \phi \text{ at the } \hat{a}, \hat{b} \text{ positions.} \end{array} \quad (2)$$

We say repeated edge names (e.g. a and b above) are *bound* in a tensor expression, and all other edge names are *free*. In the graph we have labelled the bound edges, though this is purely for demonstrating which edges are bound. The names of bound edges can be changed at will, provided they are replaced with new, fresh names. Hence $\psi_{\check{f}\check{a}\check{b}} \phi_{\hat{a}\hat{b}\check{c}\check{d}\check{e}}$ and $\psi_{\check{f}\check{x}\check{y}} \phi_{\hat{x}\hat{y}\check{c}\check{d}\check{e}}$ represent the same graph. As a result, we typically will not write down bound names in the graphical notation.

Definition 2.1. The set of *tensor expressions* for a signature \mathcal{S} consists of (i) the trivial tensor 1, (ii) the identity tensor $1_{\hat{a}\check{b}}$, (iii) atomic tensors $\psi_{\hat{a}\check{b}\dots}$ with the appropriate names for each $\psi \in \mathcal{S}$, (iv) GH for G, H tensor expressions, and (v) G' obtained by changing some of the names of a tensor expression G —subject to the condition that \hat{a} and \check{a} occur at most once for each name a .

Definition 2.2. Two tensor expressions G, G' are equivalent, written $G \equiv G'$ if G can be made into G' by replacing bound names or by applying one or more of the following identities:

$$(GH)K \equiv G(HK) \quad GH \equiv HG \quad G1 \equiv G$$

$$G1_{\hat{b}\check{a}} \equiv G[\check{b} \mapsto \check{a}] \quad H1_{\hat{a}\check{b}} \equiv H[\hat{b} \mapsto \hat{a}]$$

Assume for the last two identities that \check{b} and \hat{b} are free in G and H , respectively. An \equiv -equivalence class of tensor expressions is called a *tensor*.

Note that we use \equiv for syntactic equivalence of tensor expressions (and later !-tensor expressions). We reserve the normal equals sign for equality by the rules of a given theory. As such, we always assume $(G \equiv H) \implies (G = H)$, but not the converse.

Tensors are related to morphisms in the free compact closed category as follows. Suppose we fix a set of *canonical names* $\{\check{x}_1, \check{x}_2, \dots\}$ and $\{\hat{x}_1, \hat{x}_2, \dots\}$. A tensor G is said to be *canonically named* if for some N it has as a free name precisely one of \check{x}_i or \hat{x}_i for $1 \leq i \leq N$.

Theorem 2.3. *Canonically-named tensors are in 1-to-1 correspondence to morphisms in the free compact closed category generated by a signature \mathcal{S} .*

Proof. First note that adding ‘hats’ and ‘checks’ to edge names is essentially applying the Int construction (c.f. [3]) to free traced symmetric monoidal category, in the tensorial presentation given in [5]. The free compact closure of the free traced monoidal category then satisfies the appropriate universal property to make it the free compact closed category. \square

To summarise, we can interpret a tensor in a compact closed category as follows. First, we swap its free names for ‘canonical names’ (or otherwise order the outputs somehow), then interpret each atomic expression as a morphism (or one of a family of morphisms, parametrised by its arity). Finally, we construct the composed morphism by composing each of the components and contracting repeated edge names, as in (2).

Alternatively, one can study models in an existing abstract tensor system (in the sense of Penrose), in which case interpretation is trivial. These two points of view (categorical vs. ATS) are roughly equivalent, as was shown in [5].

3 Adding !-boxes to tensor expressions

We now extend the existing tensor notation with !-boxes. Graphically !-boxes are blue boxes surrounding a subgraph, labelled with a name (A, B, \dots) . We can denote this with square brackets around a subterm in a tensor expression, labelled with a superscript. Intuitively a !-box represents a portion of the graph that can be copied multiple times. For this to be well-defined in the non-commutative case we need to clarify where each new copy of the subgraph gets attached to surrounding nodes.

This is done by assigning an expansion direction (clockwise vs anticlockwise) to any group of edges from a node to a !-box. We draw these as arrows over edge groups in our !-graphs and for our tensors we

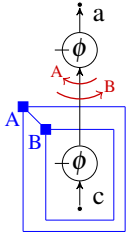
denote clockwise edge groups as $[\dots]^A$ and anticlockwise edge groups as $\langle \dots \rangle^A$. For example:

$$\phi_{\langle \hat{a} \rangle^B} [\psi_{\check{a}}]^B := \text{diagram} \quad \text{vs.} \quad \phi_{[\hat{a}]^B} [\psi_{\check{a}}]^B := \text{diagram}$$

In the next section, we will see how the arrows clarify not only which direction edges should expand, but also whether they should expand in groups or individually. For example, the following notation gives anti-clockwise expansion of $\hat{a}\check{b}$ as a group, clockwise expansion of $\hat{a}\check{b}$ as a group, and clockwise expansion of \hat{a} and \check{b} as individual edges, respectively:

$$\psi_{a'\check{b}'\langle \hat{a}\check{b} \rangle^A} \square^A \quad \text{vs.} \quad \psi_{[\hat{a}\check{b}]^A a'\check{b}'} \square^A \quad \text{vs.} \quad \psi_{[\hat{a}]^A a' [\check{b}]^A \check{b}'} \square^A$$

It is also possible for !-boxes to be nested inside other !-boxes. This means expansion of the parent box makes a new copy of the child with a new !-box name. Edge groups can correspondingly be nested if the edges enter more than one box. In the diagram to the left we have the !-graph with !-tensor expression: $\phi_{\hat{a}[\check{b}]^B \check{c}} [[\phi_{\check{b}\check{c}}]^B]^A$. We have labelled which arrow corresponds to which !-box. This is not necessary if we adopt the convention that a parent box's arrow must be drawn closer to the node than its child box's arrow. Note that the labels inside nodes are to assign a type to the node as apposed to naming the node. This means since we often have multiple nodes with the same type, we will have nodes with the same label.



We can now imagine more general generators allowing arbitrary arrangements of input and output edges. Any such node, say of type ϕ , then needs to be assigned a morphism in our category for each possible arrangement of edges. We represent an arrangement as a word over $\{\wedge, \vee\}$ where \wedge represents outputs and \vee represents inputs. For example the node has edge arrangement $\wedge \vee \wedge \wedge$ and needs to be assigned a morphism $f : I \rightarrow X \otimes X^* \otimes X \otimes X$. Hence we need $\phi : \{\wedge, \vee\}^* \rightarrow \text{Mor}(\mathcal{C})$ to model the node type ϕ .

!-tensors replace lists of edges on individual morphisms with *edgeterms* of which we now give a recursive definition.

Definition 3.1. Fix a disjoint, infinite sets \mathcal{E} and \mathcal{B} of edge names and !-box names, respectively. We denote the set of *directed edges* as $\bar{\mathcal{E}} := \{\check{a}, \hat{a} : a \in \mathcal{E}\}$. The set of *edgeterms* \mathcal{T}_e is defined recursively as follows:

- $\varepsilon \in \mathcal{T}_e$ (i.e empty)
- $\check{a}, \hat{a} \in \mathcal{T}_e$ $a \in \mathcal{E}$
- $[e]^A, \langle e \rangle^A \in \mathcal{T}_e$ $e \in \mathcal{T}_e, A \in \mathcal{B}$
- $ef \in \mathcal{T}_e$ $e, f \in \mathcal{T}_e$

Two edgeterms are equivalent if one can be transformed into the other by:

$$\varepsilon e \equiv e \equiv e \varepsilon \quad e(fg) \equiv (ef)g \quad [\varepsilon]^A \equiv \varepsilon \equiv \langle \varepsilon \rangle^A$$

Since the well-formedness conditions for !-tensor expressions are a bit more complicated than for tensor expressions, we first define the set of all !-pretensor expressions, including those that may be ill-formed.

Definition 3.2. The set of all !-pretensor expressions \mathcal{T}'_Σ for a signature Σ is defined recursively as:

$$\begin{array}{ll} \bullet 1, 1_{\check{a}\check{b}} \in \mathcal{T}'_\Sigma & a, b \in \mathcal{E} \\ \bullet \phi_e \in \mathcal{T}'_\Sigma & e \in \mathcal{T}_e, \phi \in \Sigma \\ \bullet [G]^A \in \mathcal{T}'_\Sigma & G \in \mathcal{T}'_\Sigma, A \in \mathcal{B} \\ \bullet GH \in \mathcal{T}'_\Sigma & G, H \in \mathcal{T}'_\Sigma \end{array}$$

We introduce the notion of a *context*, which lists the !-boxes in which a certain edge name occurs, from the inside-out. These come in two flavours, *edge contexts* and *node contexts*.

Definition 3.3. Given a directed edge $a \in \bar{\mathcal{E}}$ in a !-tensor G nested as $[[\phi \dots \langle a \rangle^{E_1} \dots]^{N_1} \dots]^{N_m}$.

We define the *edge context*, *node context*, and *context* of a respectively as:

$$\begin{array}{ll} \text{ectx}_G(a) := [E_1, \dots, E_n] & \text{(edge context)} \\ \text{nctx}_G(a) := [N_1, \dots, N_m] & \text{(node context)} \\ \text{ctx}_G(a) := \text{ectx}_G(a). \text{nctx}_G(a) & \text{(context)} \end{array}$$

That is, $\text{ectx}_G(a)$ lists the !-boxes containing a that occur as part of a 's edgeterm, and $\text{nctx}_G(a)$ lists the rest.

Finally, a !-tensor expression is a !-pretensor expression where !-box/edge names must be suitably unique and occur in compatible contexts.

Definition 3.4. A !-tensor expression is a !-pretensor expression satisfying the following conditions:

- F1. \check{a} and \hat{a} occur at most once for each edge name a
- F2. $[\dots]^A$ must occur at most once for each !-box name A
- C1. $\text{ectx}_G(a) \cap \text{nctx}_G(a) = \emptyset$ for all edges $a \in \mathcal{E}$ in G
- C2. If $\text{ectx}_G(a) = [B_1, \dots, B_n]$ then all $B_i \in \text{Boxes}(G)$ and $B_1 \prec_G B_2 \prec_G \dots \prec_G B_n$
- C3. For all bound pairs \check{a}, \hat{a} of edge names in G , there exist lists es, bs of !-box names such that:

$$es. \text{nctx}_G(\check{a}) = \text{ectx}_G(\hat{a}).bs \quad \text{and} \quad es. \text{nctx}_G(\hat{a}) = \text{ectx}_G(\check{a}).bs$$

where $A \prec_G B$ means that the !-box A is nested inside B in G (without other boxes nested between). We write \mathcal{T}_Σ for the set of all !-tensor expressions for a signature Σ .

The freshness conditions F1 and F2 ensure that we have not used the same name for more than one edge/box. If a node is in !-box B then any edges attached to it are already in B so it wouldn't make sense to have B in both the $\text{ectx}(a)$ and $\text{nctx}(a)$ for $a \in \bar{\mathcal{E}}$, this is enforced by C1. C2 ensures that edge contexts are compatible with the !-boxes in the rest of the !-tensor. For example $\phi_{[[\check{a}]^A]^B}$ requires A to be nested in

B so does not result in a valid $!$ -tensor when composed with e.g. $[[\psi_b]^B \xi_{\langle \hat{b} \rangle^B}]^A$. C3 ensures that edges into $!$ -boxes from the outside are decorated correctly by their edge terms. For instance, this is allowed: $\psi_{\langle \hat{a} \rangle^A} [\phi_{\hat{a}}]^A$ but this is not: $\psi_{\hat{a}} [\phi_{\hat{a}}]^A$. The freedom to pick bs, es allows bound pairs of edges to share some common context, e.g.: $[\psi_{\hat{a}} \phi_{\hat{a}}]^A$ (both nodes are inside A) or $\psi_{\langle \hat{a} \rangle^A} \phi_{\langle \hat{a} \rangle^A} []^A$ (only the edge is inside A). In the second example, A occurs in an edge term, so C2 requires the presence of $[\dots]^A$ somewhere in the $!$ -tensor, hence we append the ‘empty’ $!$ -box $[]^A$.

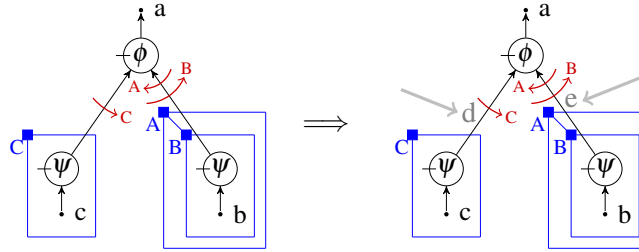
In this paper when we write a composition GH , unless stated otherwise, we will assume this forms a well defined $!$ -tensor.

Naturally, we say two $!$ -tensor expressions are equivalent, written $G \equiv H$, if one can be obtained from the other by using the usual tensor equivalences from Definition 2.2 or by using the edgeterm equivalences from Definition 3.1.

We call the graphical notation for $!$ -tensors the *non-commutative $!$ -graph notation*, or simply (non-commutative) $!$ -graphs.

Theorem 3.5. *Any $!$ -tensor can be represented unambiguously using non-commutative $!$ -graph notation.*

Proof. We show this by providing a general procedure for interpreting a $!$ -graph as a $!$ -tensor expression, and vice-versa. For the sake of clarity, we demonstrate each step on a worked example. Given a non-commutative $!$ -graph, we wish to obtain a unique equivalence class of $!$ -tensor expressions under \equiv . Begin by choosing fresh names to write on all the interior edges.



Then, write the $!$ -boxes with nesting as depicted in the diagram:

$$\dots [\dots]^C [\dots [\dots]^B]^A$$

Write each node in the diagram on the location it occurs (w.r.t. $!$ -boxes):

$$\phi_{\dots} [\psi_{\dots}]^C [[\psi_{\dots}]^B]^A$$

Finally, add the edges of each node, reading clockwise from the tick. Edges occurring under a clockwise arrow marked A should be enclosed in $[\dots]^A$, and edges under an anti-clockwise arrow should be enclosed in $\langle \dots \rangle^A$, where the outermost groups are the ones closest to the node in the picture.

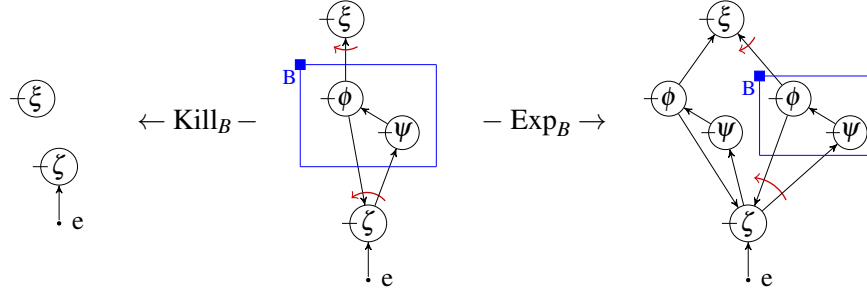
$$\phi_{\hat{a}[\langle \hat{e} \rangle^B]^A \langle \hat{d} \rangle^C} [\psi_{\hat{d}\hat{c}}]^C [[\psi_{\hat{e}\hat{b}}]^B]^A$$

The only choices we made in this process were the choice of interior edge names and the order in which to write the individual tensors. However, up to \equiv , these are irrelevant. To show that any $!$ -tensor can be represented this way, we simply run the above procedure in reverse. \square

Because of this theorem, we use the terms $!$ -tensor and $!$ -graph interchangeably, depending on whether we wish to refer to the syntactic vs. graphical notation.

4 Instantiating tensor expressions with !-boxes

The following diagram demonstrates two !-box operations we can apply to a graph: killing a !-box is the operation deleting the box B and all contents (including edges to/from B), and expanding is the operation creating a new concrete instance of the subgraph inside B (attached appropriately). We can represent the original graph in this diagram with the tensor expression $[\phi_{\hat{a}\hat{c}\hat{b}}\psi_{\hat{c}\hat{d}}]^B \xi_{[\hat{a}]^B} \zeta_{[\hat{b}\hat{d}]^B}$.



We can define both of these operations formally. Since expansion involves copying various edge/!-box names, we need a means of obtaining fresh names. Let $\text{Edges}(G) \subset \mathcal{E}$ and $\text{Boxes}(G) \subset \mathcal{B}$ be the edge names and !-box names occurring in a !-tensor G , respectively.

Definition 4.1. A *freshness function* for a !-tensor G is a pair of bijections $\mathbf{fr} : \mathcal{E} \rightarrow \mathcal{E}$ and $\mathbf{fr} : \mathcal{B} \rightarrow \mathcal{B}$ such that

$$\text{Edges}(G) \cap \mathbf{fr}(\text{Edges}(G)) = \emptyset \quad \text{and} \quad \text{Boxes}(G) \cap \mathbf{fr}(\text{Boxes}(G)) = \emptyset$$

For !-tensor expressions G or edgeterms e , we will write $\mathbf{fr}(G)$ or $\mathbf{fr}(e)$ to designate the new expression with names substituted according to the given bijections.

Definition 4.2. We define $\text{Op}_B \in \{\text{Exp}_B, \text{Kill}_B\}$ recursively over !-tensor expressions. For most cases, both operations act trivially:

$$\begin{aligned} \text{Op}_B(GH) &:= \text{Op}_B(G) \text{Op}_B(H) & \text{Op}_B(e f) &:= \text{Op}_B(e) \text{Op}_B(f) \\ \text{Op}_B([G]^A) &:= [\text{Op}_B(G)]^A & \text{Op}_B([e]^A) &:= [\text{Op}_B(e)]^A \\ \text{Op}_B(\phi_e) &:= \phi_{\text{Op}_B(e)} & \text{Op}_B(\langle e \rangle^A) &:= \langle \text{Op}_B(e) \rangle^A \\ \text{Op}_B(x) &:= x \end{aligned}$$

where $A \neq B$ and $x \in \{1, 1_{\hat{a}\hat{b}}, \hat{a}, \hat{a}, \varepsilon\}$. Then, for the final three cases:

$$\begin{aligned} \text{Exp}_B([G]^B) &:= [G]^B \mathbf{fr}(G) & \text{Kill}_B([G]^B) &:= 1 \\ \text{Exp}_B([e]^B) &:= [e]^B \mathbf{fr}(e) & \text{Kill}_B([e]^B) &:= \varepsilon \\ \text{Exp}_B(\langle e \rangle^B) &:= \mathbf{fr}(e) \langle e \rangle^B & \text{Kill}_B(\langle e \rangle^B) &:= \varepsilon \end{aligned}$$

Note that $\text{Exp}_B(G)$ implicitly takes a freshness function as input. If we wish to make this explicit, we will write $\text{Exp}_{B, \mathbf{fr}}$. The above operations can be lifted from !-tensor expressions to !-tensors, i.e. \equiv -classes of expressions, because of the following theorem.

Theorem 4.3. Let \mathbf{fr} be a freshness function for two !-tensor expressions G, H . Then $G \equiv H$ implies $\text{Exp}_{B, \mathbf{fr}}(G) \equiv \text{Exp}_{B, \mathbf{fr}}(H)$ and $\text{Kill}_B(G) \equiv \text{Kill}_B(H)$.

Proof. (Sketch) This can be shown by induction over the structure of !-tensor expressions. It is crucial that we use the *same* freshness function \mathbf{fr} for the expansions of G and H , otherwise G and H could end up with distinct free edges or !-boxes. \square

These two !-box operations give us a means to define the set of all (concrete) tensors that a single !-tensor represents.

Definition 4.4. A tensor G' is a *concrete instance* of a !-tensor G if it is obtained from G by repeatedly applying the two !-box operations Exp and Kill until G' contains no !-boxes. This sequence of operations is called the *instantiation* of G' . We write $\llbracket G \rrbracket$ for the set of all concrete instances of G .

When we fix a model in some category \mathcal{C} , concrete tensors can then be interpreted as morphisms in \mathcal{C} , just as before. We therefore interpret each !-tensor expression as a family of morphisms in \mathcal{C} , namely the interpretations of each of its concrete instances.

5 Reasoning with !-boxes

The real power of !-boxes comes from the ability to do equational reasoning using infinite families of rules. Just as it makes sense to instantiate a single !-tensor, it makes sense to instantiate an *equation* $G = H$ between two !-tensors, provided they have compatible boundaries.

Definition 5.1. A !-tensor equation ' $G = H$ ' consists of a pair of !-tensors (G, H) that have *compatible boundaries*. That is, they have identical free edge names and !-boxes, $A \prec_G B \Leftrightarrow A \prec_H B$ for all !-boxes in G and H , and $\text{ctx}_G(a) = \text{ctx}_H(a)$ for all free edge names.

Intuitively, we require that the LHS and RHS of a !-tensor equation have the same interface to attach to other graphs (same free variables and same box structure). These consistency conditions guarantee that (i) applying !-box operations to valid equations yields valid equations, and (ii) when G occurs as a sub-expression of some other !-tensor K , it can be substituted for H to yield another valid !-tensor K' .

Theorem 5.2. Let \mathbf{fr} be a freshness function for !-tensors G, H . Then, if $G = H$ is a !-tensor equation, then so too are:

$$\begin{aligned} \text{Kill}_B(G = H) &:= (\text{Kill}_B(G) = \text{Kill}_B(H)) \\ \text{Exp}_B(G = H) &:= (\text{Exp}_{B, \mathbf{fr}}(G) = \text{Exp}_{B, \mathbf{fr}}(H)) \end{aligned}$$

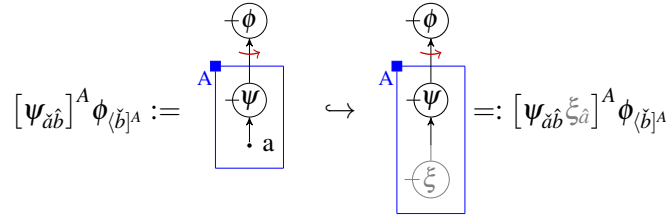
Proof. (Sketch) It is straightforward to show that killing/expanding B affects the free variables and the !-boxes in the same way on the LHS/RHS. To check the contexts, split a single free edge name into 3 cases, depending on whether the !-box B occurs in the node-context of a , the edge context of a , or neither. In all cases, a and/or $\mathbf{fr}(a)$ will have identical contexts on the LHS/RHS. \square

As in the case of !-tensors, we can define $\llbracket G = H \rrbracket$ to be the set of all concrete rules derivable from $G = H$ using the !-box operations. A valid model of a graphical theory is then one where all of the equations in $\llbracket G = H \rrbracket$ hold for each equation $G = H$. Proving that a rule holds for *all* of its instances could be a daunting task in general, however in many cases a technique called *!-box induction*—which we will meet shortly—comes to the rescue.

We obtain a notion of substitution of sub-expressions constructively, via inference rules. The first few should look familiar as congruence- and substitution-like rules for !-tensors.

$$\frac{G = H}{GK = HK} \text{ (Prod)} \quad \frac{G = H}{[G]^A = [H]^A} \text{ (Box)} \quad \frac{G = H}{G[a \rightarrow b] = H[a \rightarrow b]} \text{ (Rename)}$$

Where $G[a \rightarrow b]$ and $H[a \rightarrow b]$ are G and H with the free edge/!-box name a replaced by b . We require that K and A are chosen such that GK , HK , $[G]^A$, and $[H]^A$ are well-defined. These rules provide the conditions under which some equation $G = H$ can be unified, given some context, with a bigger equation $G' = H'$. The final inference rule (Weaken) is less intuitive from the point of view of terms, and is best understood graphically. Consider the following embedding of !-graphs:



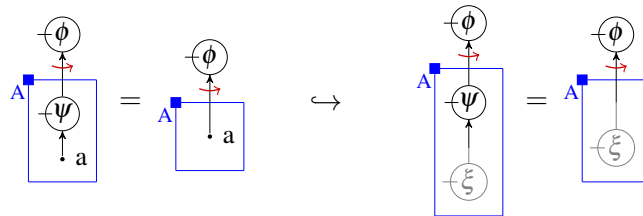
The LHS does not embed as a sub-term of the RHS, because the !-box A contains more stuff on the LHS. However, semantically, this is perfectly fine, as all of the concrete instances of the LHS will have (uniquely-determined) embeddings into all of the concrete instances of the RHS. So, we also need a rule that allows us to ‘weaken’ !-boxes by adding more nodes to them.

$$\frac{G = G'}{W_{A \rightarrow K}(G) = W_{A \rightarrow K}(G')} \text{ (Weaken)}$$

Where $W_{A \rightarrow K}(G)$ is defined recursively as:

$$\begin{aligned} W_{A \rightarrow K}([G]^A) &:= [GK]^A \\ W_{A \rightarrow K}([G]^B) &:= [W_{A \rightarrow K}(G)]^B && \text{if } A \neq B \\ W_{A \rightarrow K}(GH) &:= W_{A \rightarrow K}(G) W_{A \rightarrow K}(H) \\ W_{A \rightarrow K}(x) &:= x && x \in \{1, 1_{ab}, \phi_e\} \end{aligned}$$

These four rules give us everything we need to apply equations on !-tensors to obtain new equations. For example, the equation on the left below can be applied to delete all of the ψ -nodes occurring as input to a ϕ . An example application of this rule is shown on the right.



We can also add inference rules for each of our !-box operations i.e.

$$\frac{G = H}{\text{Exp}_B(G = H)} \text{ (Exp)} \quad \frac{G = H}{\text{Kill}_B(G = H)} \text{ (Kill)}$$

Perhaps most interestingly, we can introduce new !-boxes, where previously there were none, via *!-box induction*.

$$\frac{\text{Kill}_A(G = H) \quad G = H \Rightarrow \text{Exp}_A(G = H)}{G = H} \text{ (Induction)}$$

As mentioned in Section 1, non-commutative nodes give us the ability to make recursive definitions of variable-arity generators in terms of fixed-arity generators of our theory. This induction principle in turn gives us the means to lift rules about fixed arity generators up to more powerful $!$ -tensor rules.

We conclude by showing a simple example. Suppose we take the theory of a monoid, i.e. the pair of generators (\uparrow, \downarrow) satisfying the commutativity and unit laws from (1). Then we can recursively define, as a new generator, an n -fold tree of multiplications.

$$\begin{array}{c} \uparrow \\ \circ \end{array} := \begin{array}{c} \uparrow \\ \circ \end{array} \quad \begin{array}{c} \uparrow \\ \circ \end{array} := \begin{array}{c} \uparrow \\ \circ \end{array} \quad (3)$$

Remark 5.3. Note how non-commutative $!$ -boxes make such recursive definitions possible in the first place, without assuming *a priori* that the family of graphs generated by the definition are symmetric on their inputs/outputs. This need not be true, even in the case where all of the concrete generators are commutative. This limitation in the case of commutative $!$ -boxes was highlighted in [10], where only a partial proof of the spider theorem for commutative Frobenius algebras could be done using (commutative) $!$ -box induction.

The first property we would like to prove about such trees is that adjacent trees merge to form bigger trees. As a $!$ -box rule, it looks like this:

We can then hit this rule with the induction on B to break it into cases:

(base)

(step)

...each of which have simple rewriting proofs:

One caveat is that when we apply the induction hypothesis in step 4, the $!$ -box B must be ‘fixed’ (i.e. we’re not allowed to do any instantiation of B via Exp, Kill, etc.). This is because B occurs free on both sides of the implication $G = H \Rightarrow \text{Exp}_B(G = H)$. See [10] for details.

This style of proof is the main workhorse of soundness proofs of rules like the merging rule (a.k.a. ‘spider rule’) for commutative Frobenius algebras described in Section 1, and can be extended to the non-commutative case, proving a similar rule for e.g. *symmetric* Frobenius algebras, giving a purely diagrammatic characterisation of the normal forms described in [9].

References

- [1] L. Dixon and R. Duncan. Graphical Reasoning in Compact Closed Categories for Quantum Computation. *AMAI*, 56(1):20, 2009.
- [2] L. Dixon and A. Kissinger. Open Graphs and Monoidal Theories. arXiv:1007.3794v1 [cs.LO], 2010.
- [3] A. Joyal, R. Street, and D. Verity. Traced Monoidal Categories. *Math. Proc. Camb. Phil. Soc.*, 119(3):447–468, 1996.
- [4] A. Joyal and R. Street. The geometry of tensor calculus I. *Advances in Mathematics*, 88:55–113, 1991.
- [5] A. Kissinger. Abstract tensor systems as monoidal categories. In C. Casadio, B. Coecke, M. Moortgat, and P. Scott, editors, *Categories and Types in Logic, Language, and Physics: Festschrift on the occasion of Jim Lambek’s 90th birthday*, volume 8222 of *Lecture Notes in Computer Science*. Springer, 2014. arXiv:1308.3586.
- [6] A. Kissinger. *Pictures of Processes: Automated Graph Rewriting for Monoidal Categories and Applications to Quantum Computing*. PhD thesis, University of Oxford, 2011.
- [7] A. Kissinger, A. Merry, and M. Soloviev. Pattern graph rewrite systems. In *Proceedings of DCM 2012*, volume 143 of *EPTCS*, 2012.
- [8] S. Lack and P. Sobocinski. Adhesive and quasiadhesive categories. *Theoretical Informatics and Applications*, 39(2):522–546, 2005.
- [9] A. D. Lauda and H. Pfeiffer. Open-closed strings: Two-dimensional extended tqfts and frobenius algebras. *Topology Appl.*, 155(7):623–666, 2008.
- [10] A. Merry. *Reasoning with $!-$ Graphs*. PhD thesis, University of Oxford, 2014.
- [11] R. Penrose. Applications of negative dimensional tensors. In *Combinatorial Mathematics and its Applications*, pages 221–244. Academic Press, 1971.

On modifications of Reichenbach's principle of common cause in light of Bell's theorem

Eric Cavalcanti Raymond Lal

University of Sydney University of Oxford

e.cavalcanti@physics.usyd.edu.au - raymond.lal@cs.ox.ac.uk

This note summarises the results in:

- Eric Cavalcanti and Raymond Lal (2013) *On modifications of Reichenbach's principle of common cause in light of Bell's theorem*. Invited contribution to special issue: '50 years of Bell's theorem', Journal of Physics A. arXiv:1311.6852
-

Background. Imagine that a rare viral infection breaks out simultaneously in Australia and in Brazil. We would expect one of three possible explanations to account for this coincidence: either the virus was carried by a host travelling from Australia to Brazil, or by a host going from Brazil to Australia, or it was carried by two travellers who were in contact at some third country in the past. Reichenbach's Principle of Common Cause (R-PCC) [6] attempts to formalise this intuition. It states that when two events are correlated, either one is a direct cause of the other, or they share a common cause. This principle is considered to be fundamental for many areas of scientific practice.

Bell's theorem presents a challenge to R-PCC. Consider the situation where two correlated events A and B are space-like separated, as in the usual Bell scenario. Relativistic causal structure implies that A and B cannot be causally connected, and thus R-PCC implies they must share a common cause. Bell, like Reichenbach, considered a common cause explanation for these events to mean that there exists a sufficient specification of events (denoted by λ) in the common past of A and B such that conditioned on λ the joint probability of A and B factorises, i.e. $P(A, B|\lambda) = P(A|\lambda)P(B|\lambda)$. This assumption however leads to the Bell inequalities, which are violated by quantum theory and by numerous experiments to date [1, 7]. Thus relativistic causal structure and the R-PCC in the formulation above cannot coexist.

Faced with Bell's theorem, various lines of investigation have attempted to retain a principle similar to R-PCC by adopting a different concept of a common cause. We shall focus on the prominent approach of Hofer-Szabó and Vecsernyés (H-V) [3, 4], who define a notion of 'non-commutative common cause', which requires a factorisation of probabilities similar to the standard Reichenbach-Bell formulation. Here we analyse this programme and argue that it fails to provide a common cause explanation of the Bell correlations. The reason it fails is that it lacks an essential aspect of the notion of common cause, namely that the common cause *explains* the observed correlations. Moreover, we show that using the H-V definition, *any* quantum product state can count as a common cause for *any* quantum correlations whatsoever.

Our contribution. Our work begins by making explicit the connection between R-PCC and Bell's theorem. In particular, we show that R-PCC is the conjunction of two weaker assumptions:

1. *Principle of Common Cause (PCC)*: If two events A and B are correlated, i.e., if $P(A, B) > P(A)P(B)$, then either:

- (i) A and B are directly causally connected, i.e. either A causes B or B causes A ; or
 - (ii) A and B share a *common cause* that explains the correlation.
2. *Factorisation of probabilities (FP)*: Two events have a *common cause* if and only if there exists a sufficient specification of variables λ corresponding to events in the common causal past of A and B such that conditioned on those variables the joint probability of A and B factorises:

$$P(A, B | \lambda) = P(A | \lambda) P(B | \lambda). \quad (1)$$

Moreover, to derive Bell's theorem, we two further assumptions—less obviously related to common causes—are needed:

- 3. *Relativistic causal structure (RCS)*: Events can be embedded in a single relativistic space-time. All causes of an event are to be found in the event's past light cone.
- 4. *Law of Total Probability (LTP)*: Given a set of mutually exclusive events $\{x\}$ whose probabilities sum to unity, then the *law of total probability* is satisfied if the probability of an event y can be written as $P(y) = \sum_x P(x) P(y | x)$.

Then Bell's theorem can be seen as the following implication:

$$\text{PCC} + \text{FP} + \text{RCS} + \text{LTP} \implies \text{Bell inequalities}. \quad (2)$$

Hence, in order to accommodate a notion of common cause that is consistent with quantum theory, one must deny at least one of the conjuncts in the antecedent.

Main technical result. The preceding analysis will allow us to understand exactly how the H-V approach can provide a notion of common cause. The H-V approach is formulated in the setting of algebraic quantum field theory (AQFT). An AQFT is defined by assigning a C^* -algebra $\mathcal{A}(O)$ to each bounded region O of a given spacetime, representing the observables that can be measured in O . This assignment is used to generate the direct-limit algebra $\hat{\mathcal{A}}$, called the *quasi-local algebra*, which is used for calculating quantities of physical interest. For example, a *state* is a positive linear functional $\phi : \hat{\mathcal{A}} \rightarrow \mathbb{C}$ on the quasi-local algebra $\hat{\mathcal{A}}$. Given this framework, the authors of Ref. [4] define a non-classical notion of common cause in the following way. First, recall that the *conditional expectation* E_c is defined by $E_c(A) := \sum_{k \in K} C_k A C_k$. Let $\mathcal{P}(\hat{\mathcal{A}})$ be the set of projections in the algebra $\hat{\mathcal{A}}$. Then:

Definition 1. A partition of the unit $\{C_k\}_{k \in K} \subset \mathcal{P}(\hat{\mathcal{A}})$ is said to be the *common cause system* of the commuting events $A, B \in \mathcal{P}(\hat{\mathcal{A}})$, which correlate in the state $\phi : \hat{\mathcal{A}} \rightarrow \mathbb{C}$, if for all $k \in K$ such that $\phi(C_k) \neq 0$, the following condition holds:

$$\frac{(\phi \circ E_c)(ABC_k)}{\phi(C_k)} = \frac{(\phi \circ E_c)(AC_k)}{\phi(C_k)} \frac{(\phi \circ E_c)(BC_k)}{\phi(C_k)}. \quad (3)$$

The motivation for this definition is that, using the Lüders rule, Eq. (3) corresponds to

$$P_\phi(A, B | C_k) = P_\phi(A | C_k) P_\phi(B | C_k).$$

Hence the H-V approach formally captures the ‘screening-off’ condition, i.e. it satisfies FP. Moreover, as part of their framework of AQFT, H-V assume PCC and RCS. However, H-V do not require that a common cause system *explains* the correlations, i.e. they do not require a non-commutative analogue of the condition $P(a, b | x, y) = \sum_\lambda \mu(\lambda) P(a | x, \lambda) P(b | y, \lambda)$. Hence the H-V approach fails to satisfy LTP, from which we can prove our main result. Let $B(\mathcal{H})$ be the set of bounded linear operators on a Hilbert space \mathcal{H} .

Theorem 2. Let $\mathcal{A} \subseteq B(\mathcal{H})$. Any orthonormal set of product states for subsystems A and B forms a common cause system for the correlations of any quantum state.

Hence we see that any product state qualifies as a potential ‘non-commutative common cause’ under the proposed definition. So, for example, suppose that Alice and Bob share a Bell state and perform measurements that lead to maximal violation of a Bell-type inequality. Then we are invited to model the common cause of these correlations as a product state in the common past of Alice and Bob—and *any* product state will do. In addition, as Theorem 2 makes clear, this notion of common cause is doubly ‘degenerate’. For not only will any product state serve as a common cause, but it will do so for *any* observed quantum correlations whatsoever. There is thus no deductive or causal link between the common cause and the effects they are purported to explain.

Taxonomy. Our analysis of the relationship between the principle of common cause and Bell’s theorem allows us to categorise the responses to Bell’s theorem accordingly. We summarise the options available in Table 1, in which we can now illustrate the distinction between the H-V approach and three other

Table 1: Responses to Bell’s theorem, categorized according to assumptions satisfied

Approach	PCC	FP	RCS	LTP
H-V	✓	✓	✓	×
L-S	✓	×	✓	✓
van Fraassen	×	×	✓	✓
Bohm	✓	✓	×	✓

prominent alternatives: (i) the recent work by Leifer-Spekkens (L-S), which attempts to reformulate quantum theory as a theory of Bayesian inference [5]; (ii) the response by van Fraassen, who essentially rejected R-PCC [2]; and (iii) Bohmian mechanics. Note that although the approaches of L-S and H-V are formally similar, they provide quite different responses to Bell’s theorem. In particular, the former provides only a factorisation of *conditional quantum states*, but not of probabilities. We believe that this classification will provide a useful tool for future research, since it makes explicit the trade-offs involved.

References

- [1] Alain Aspect (1999): *Bell’s inequality test: more ideal than ever*. *Nature* 398(6724), pp. 189–190.
- [2] B. C. van Fraassen (1982): *The Charybdis of realism: epistemological implications of Bell’s inequality*. *Synthese* 52, pp. 25–38.
- [3] Gábor Hofer-Szabó & Péter Vecsernyés (2012): *Noncommuting local common causes for correlations violating the Clauser-Horne inequality*. *Journal of Mathematical Physics* 53, p. 122301.
- [4] Gábor Hofer-Szabó & Péter Vecsernyés (2013): *Bell inequality and common causal explanation in algebraic quantum field theory*. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 44(4), pp. 404 – 416.
- [5] MS Leifer & RW Spekkens (2011): *Formulating quantum theory as a causally neutral theory of Bayesian inference* ArXiv:1107.5849.
- [6] Hans Reichenbach (1956): *The direction of time*. Berkeley, University of Los Angeles Press.
- [7] Mary A Rowe, David Kielpinski, V Meyer, Charles A Sackett, Wayne M Itano, Christopher Monroe & David J Wineland (2001): *Experimental violation of a Bell’s inequality with efficient detection*. *Nature* 409(6822), pp. 791–794.

Terminality implies non-signalling

Bob Coecke

University of Oxford

bob.coecke@cs.ox.ac.uk

A ‘process theory’ is any theory of systems and processes which admits sequential and parallel composition. ‘Terminality’ unifies normalisation of pure states, trace-preservation of CP-maps, and adding up to identity of positive operators in quantum theory, and generalises this to arbitrary process theories. We show that terminality and non-signalling coincide in any process theory, provided one makes causal structure explicit. In fact, making causal structure explicit is necessary to even make sense of non-signalling in process theories. We conclude that because of its much simpler mathematical form, terminality should be taken to be a more fundamental notion than non-signalling.

1 Introduction

Causality related notions are prominent in many areas of physics, and the relationships between these are by no means obvious. Examples that are relevant to us are:

- C1** Relativistic space-time is often abstracted as a partial ordering, called *causal structure* [13].
- C2** In quantum information, for example in the context of generalised probabilistic theories [1], one often relies on the notion of *non-signalling*, by means of which one intends to implement these relativistic constraints for spatially distributed information-processing devices.
- C3** In quantum foundations, an axiom called *causality* has recently been put forward in [3, 4]. For process theories [6, 8] this axiom is equivalent to the mathematical notion of *terminality*,¹ and here we will adopt this terminology in order to avoid confusing multiple uses of the term ‘causal(ity)’.²

These three notions do not straightforwardly match each other. For example, in [3, 4] the causality axiom stands for non-signalling-from-the future (i.e. time-like), while the above mentioned notion of non-signalling concerns space-like separation. Also, in [10] it was shown that while relativistic space-like separation is invariant under time-reversal, this is by no means the case for non-signalling.

Here we investigate how these notions are related. Firstly, we argue that to even make sense of non-signalling for general process theories (cf. **C2**) one needs to make causal structure (cf. **C1**) explicit. Then, the resulting definition of non-signalling, for the particular case of two systems, becomes equivalent to terminality (cf. **C3**). In the process of doing so we also resolve the seeming contradiction in [10].

Since terminality is mathematically way simpler than non-signalling, it should be taken to be the more fundamental notion. Such a stance is already adopted in [8] where it is shown that within the context of process theories much of the relevant structure of quantum theory directly follows from terminality. Terminality also yields a covariance theorem [9] (building further on earlier work in [11, 2]).

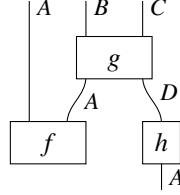
Other related work. Pearl resolved several paradoxes in probability theory [12] by making causal structure explicit within probability theory, an achievement for which he received the Turing Award. This treatment of probability theory also resulted in a fine-grained analysis of Bell’s theorem [14].

¹Process theories are symmetric monoidal categories and causality boils down to terminality of the tensor unit.

²**C1**, **C2** as well as **C3** are often referred to as causality.

2 Process theories with discarding

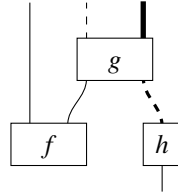
Here, by a *process theory* [6, 8] we mean a collection of *systems*, represented by *wires*, and *processes*, represented by *boxes* with wires as inputs and outputs. Moreover, when we plug these boxes together:



the resulting *diagram* should also be a process. For the purposes of this paper, outputs should be connected to inputs, and diagrams should never contain causal loops.

Remark 2.1. Equivalently, one can say that a process theory is a symmetric monoidal category.

Remark 2.2. In the above diagram we used labels to distinguish distinct systems and distinct wires, but we could also have done this by relying on different kinds of wires:



From now on we will omit labels or different kinds of wires, but all wires (within one diagram) may be interpreted as distinct.

A *state* is a process without inputs, and an *effect* is a process without outputs. We will assume that for each system there exists a *discarding effect*, which we denote as follows:



Example 2.3. When viewing probability theory as a process theory, boxes are stochastic maps, states are probability distributions, and discarding is marginalization.

Example 2.4. When viewing quantum theory as a process theory, boxes include CP-maps as well as measurements and classical data processes, states include density operators, and discarding is either the trace or deletion of classical data. A detailed description of quantum theory as a process theory is in [7].

3 Terminality vs. non-signalling

Definition 3.1. A process theory is *terminal* if for every process f we have:

$$\begin{array}{c} \text{---} \\ \text{---} \\ \boxed{f} \\ | \end{array} = \begin{array}{c} \text{---} \\ \text{---} \\ | \end{array} \quad (1)$$

Terminality has a clear operational intuition: performing an operation on a system, and then discarding the resulting system, is the same as discarding that system from the start.

Proposition 3.2. For a process theory with discarding processes, TFAE:

- (a) it is terminal,
- (b) all effects are discarding, and,
- (c) for each system there is only one effect.

Remark 3.3. When taking a process theory to be a symmetric monoidal category, by Proposition 3.2 (c) it follows that ‘terminality’ is a shorthand for ‘the tensor unit being terminal’.

Example 3.4. In the case of probability theory, terminality means stochasticity, which in the particular case of probability distributions means summing up to 1, i.e. normalisation.

Non-signalling is a bit more involved. The idea behind it is that for two spatially separated parties, say Alice and Bob, when they share a device (= a process f) with two inputs and two outputs, each having access to one input and one output, then, from one’s own input-output pair one should not be able to derive the other party’s input. Otherwise, this device would enable the parties to signal to each other while they are space-like separated, hence violating relativity. So from one of the party’s perspective, say Bob’s, who has no access to Alice’s output, something that we can represent by discarding that output, Alice’s input must not affect Bob’s input-output relationship. That is, from Bob’s perspective Alice’s input can be discarded, resulting in some process h between Bob’s input and Bob’s output. This leads to one equation with an existential quantifier. A second equation concern’s Alice’s perspective on things.

Definition 3.5. A two-input-two-output process f is *non-signalling* if we have:

$$\exists h : \begin{array}{c} \text{---} \\ \text{---} \\ \boxed{f} \\ \text{---} \end{array} = \begin{array}{c} \text{---} \\ \boxed{h} \\ \text{---} \end{array} \quad \text{and} \quad \exists h' : \begin{array}{c} \text{---} \\ \boxed{f} \\ \text{---} \end{array} = \begin{array}{c} \boxed{h'} \\ \text{---} \end{array}$$

However, it is a lot more delicate to say that a process theory is non-signalling. Simply saying ‘a process theory is non-signalling if all its processes are’ doesn’t work. For example, most process theories would include the swap process:



which evidently violates non-signalling. The obvious reason being that this process consists of two clear signalling channels. The kind of thing non-signalling aims to forbid is that without having such explicit signalling channels, signalling should not be possible. This is exactly the kind of thing that can be said in terms of a causal structure (see below), so we will need to reconsider the notion of non-signalling in the context of causal structure, in order to make sense of a non-signalling process theory.

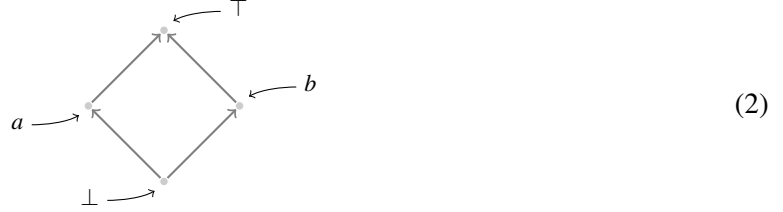
Remark 3.6. One could argue that there is no reason why in the case of non-signalling the discarding effect should be unique, something which we implicitly assumed in our notation. One could indeed weaken the definition of non-signalling for a process f to:

$$\exists e, h : \begin{array}{c} \text{---} \\ \text{---} \\ \boxed{f} \\ \text{---} \end{array} = \begin{array}{c} \triangle_e \\ \text{---} \end{array} \boxed{h} \quad \text{and} \quad \exists h', e' : \begin{array}{c} \text{---} \\ \boxed{f} \\ \text{---} \end{array} = \boxed{h'} \begin{array}{c} \triangle_{e'} \\ \text{---} \end{array}$$

This would not affect the main claims in this paper, in that we would still be able to derive non-signalling from terminality, and hence, that terminality should be taken to be the more fundamental notion.

4 Process theories with causal structure

A *causal structure* is a partial ordering, which we can represent by the corresponding Hasse-diagram. For example, if we have $\perp < a, b < \top$ then this yields the diagram:



Here, \perp can signal to a and b , which themselves can signal to \top , and by transitivity, \perp can also signal to \top . But a cannot signal to b and vice versa.

We can use such a causal structure as a support for a diagram within a process theory in the following manner, which (more or less) generalises how *causal networks* are defined by Pearl in [12]:³

- processes are positioned on nodes of the causal structure,
- wires follow the edges, from outputs to inputs, and
- we allow open inputs and open outputs at nodes.⁴

With respect to the causal structure (2), we could for example have:



where we allowed at a and b for there to be open inputs and outputs.

Remark 4.1. The manner in which a diagram of processes carries a causal structure was also considered in [9]. In brief, each diagram already carries a shadow of a causal structure in that sequential composition can be interpreted as ‘after’, i.e. time-like separated, and that parallel composition can be interpreted as ‘while’, i.e. *possibly* space-like. So the only required additional specification is to state which processes are *genuinely* space-like separated. We refer the reader to [9] for more details.

The type of scenario that we considered when defining non-signalling involves two parties that are not supposed to be able to have signalling channels to each other, just like a and b in (2).⁵ That Alice and Bob each have an input and an output is then made explicit by the *causal process network* (3).

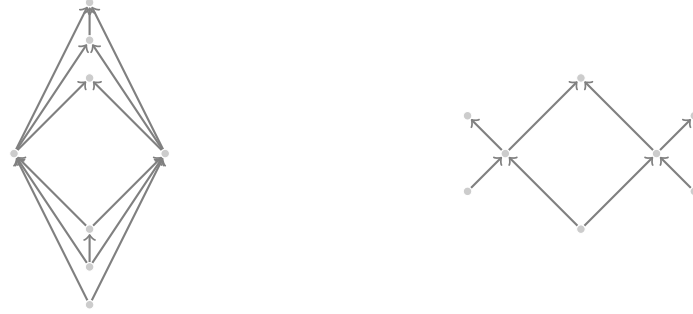
In fact, this is the most general situation that we need to consider for the purpose of evaluating

³We say ‘more or less’ since for the purposes of this paper we modify things a bit as compared to [12].

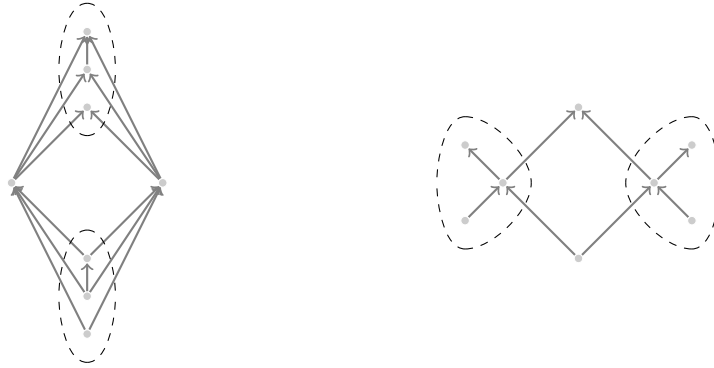
⁴It is here that we modify Pearl’s approach by allowing ‘open inputs and outputs’.

⁵So in the context of [9], these two parties have to be genuinely space-like separated.

non-signalling. Indeed, while one could of course have a causal structure like the following ones:

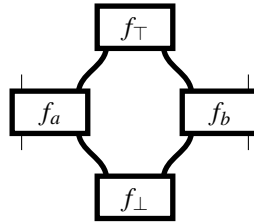


these are still covered by the diamond-shaped process network simply by treating clusters of nodes as single processes:



Definition 4.2. A process theory is non-signalling if all processes of the form (3) are.

Example 4.3 (two-sorted process theories). When one is concerned with non-signalling, then typically, there are two kinds of systems involved. One of them would correspond to probabilities, while the other one could involve some exotic physical systems, which may be existing e.g. quantum theory, or hypothetical e.g. Barrett’s box world [1] which enables one to realise correlations between systems way beyond quantum correlations, and in fact, is extremal as far as non-signalling is concerned. One treats the inputs and outputs at Alice’s and Bob’s ends as probabilities, but the ‘internal wiring’ of the box as in Definition 3.5 may involve the exotic systems. The causal process network (3) would then become:



where the bold wires and boxes mean ‘exotic’ while the normal wires mean ‘probability’.

Example 4.4 (local hidden variables). The notion of a local hidden variable theory means that such a two-sorted diagram can be replaced by a one-sorted one, only involving normal wires, and bold then standing for ‘quantum’. As we shall see below in the proof of Theorem 5.1, assuming terminality, which one does in quantum theory, the diamond shape immediately becomes a V-shape, and the definition for

a local hidden variable representation then becomes, given a quantum scenario involving f_a, f_b, f_\perp :

$$\exists h_a, h_b, h_\perp : \begin{array}{c} \boxed{h_a} \quad \boxed{h_b} \\ \diagdown \quad \diagup \\ \boxed{h_\perp} \end{array} = \begin{array}{c} \boxed{f_a} \quad \boxed{f_b} \\ \diagdown \quad \diagup \\ \boxed{f_\perp} \end{array}$$

5 Main result

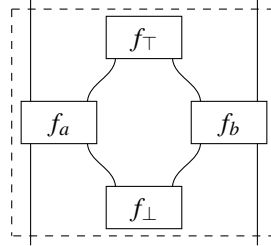
Let us first summarise what we have established so far. We want to compare the notions of terminality and non-signalling for process theories, but noted that non-signalling of a process of the form:



already requires explicit internal causal structure in order to define what it means for a process theory to be non-signalling, so that we can exclude internal signalling channels such as e.g.:



We concluded that the internal causal structure should be a diamond so that f should be of the form:



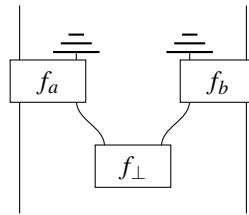
This now allows for a definition of a non-signalling process theory.

Theorem 5.1. If a process theory is terminal then it is non-signalling.

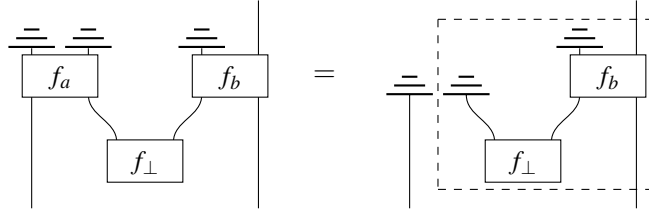
Proof. Assuming terminality, we have for f_\top in (3):

$$\begin{array}{c} \equiv \quad \equiv \\ \boxed{f_\top} \end{array} = \begin{array}{c} \equiv \quad \equiv \end{array}$$

so (3) becomes:



and applying terminality to f_a the equational requirement for non-signalling is now indeed obeyed:

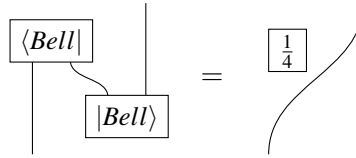


where the dashed box identifies h as in Definition 3.5. \square

To prove the converse, we need one extra assumption. What is a process with no inputs and no outputs? It interacts with nothing, so it should be independent of anything else that happens.

Definition 5.2. By (!) we mean that there is a unique diagram with no inputs nor outputs, the empty one.

Remark 5.3. The justification for there only being one box with no input and no output is that it represents *certainty*, and in this paper all processes are conceived as happening with certainty. For example, in the context of quantum theory this means that measurements are considered ‘as a whole’, that is, taking into account all possible outcomes together. This is different from, for example, a diagram for teleportation such as [5]:



where a particular measurement outcome is considered that only happen with a probability $\frac{1}{4}$.

Theorem 5.4. Assuming (!), if a process theory is non-signalling then it is terminal.

Proof. Equation (1) follows from non-signalling of processes, simply by taking the input and the output of one of the parties in Definition 3.5 to be trivial (i.e. no input nor output). Since h now has no inputs nor outputs it is the empty diagram by (!). \square

6 Discussion

So we achieved our goal and established equivalence of terminality and non-signalling for two parties, where we conceive the fact that terminality implies non-signalling as the most significant result. Terminality is both conceptually and formally the simpler and more elegant notion, and should therefore be taken to be the more fundamental one. Some other consequences are:

- Much attention has been given to the notion of non-signalling in several frameworks for theories more general than quantum theory, and maybe, these frameworks should be reconsidered.
- While in [3, 4] terminality was taken to mean non-signalling from the future, we showed here that it does much more than just that, also implying non-signalling in the usual sense.
- The reason why our result demystifies the result of [10] is that one simply should not try to match causal structure with non-signalling. Causal structure is only one ingredient when defining non-signalling, and while causal structure admits a clear notion of time-reversal, the other ingredient, the process theory does not admit such a thing, since it is governed by a manifestly time-asymmetric principle such as terminality.

Acknowledgement

Rob Spekkens has for quite a while already emphasised the importance for quantum foundations of causal structure as in causal networks. In chats last summer in Benasque he indicated that this may provide the key to demystifying [10], which, as we demonstrated here, it indeed did. In the same chat, he also strongly voiced his concerns about the the notion of non-signalling as in quantum information, which hopefully, here we (at least in part) also dethroned.

The QPL referees provided some very useful feedback on the submitted draft.

References

- [1] J. Barrett. Information processing in generalized probabilistic theories. *Physical Review A*, 75(3):032304, 2007.
- [2] R. F. Blute, I. T. Ivanov, and P. Panangaden. Discrete quantum causal dynamics. *International Journal of Theoretical Physics*, 42(9):2025–2041, 2003.
- [3] G. Chiribella, G. M. D’Ariano, and P. Perinotti. Probabilistic theories with purification. *Physical Review A*, 81(6):062348, 2010.
- [4] G. Chiribella, G. M. D’Ariano, and P. Perinotti. Informational derivation of quantum theory. *Physical Review A*, 84(1):012311, 2011.
- [5] B. Coecke. Kindergarten quantum mechanics — lecture notes. In A. Khrennikov, editor, *Quantum Theory: Reconsiderations of the Foundations III*, pages 81–98. AIP Press, 2005. arXiv:quant-ph/0510032.
- [6] B. Coecke. A universe of processes and some of its guises. In H. Halvorson, editor, *Deep Beauty: Understanding the Quantum World through Mathematical Innovation*, pages 129–186. Cambridge University Press, 2011. arXiv:1009.3786.
- [7] B. Coecke and A. Kissinger. The compositional structure of multipartite quantum entanglement. In *Automata, Languages and Programming*, Lecture Notes in Computer Science, pages 297–308. Springer, 2010. Extended version: arXiv:1002.2540.
- [8] B. Coecke and A. Kissinger. *Picturing Quantum Processes*. Cambridge University Press, 2014.
- [9] B. Coecke and R. Lal. Causal categories: relativistically interacting processes. *Foundations of Physics*, page to appear, 2012. arXiv:1107.6019.
- [10] B. Coecke and R. Lal. Time asymmetry of probabilities versus relativistic causal structure: An arrow of time. *Physical Review Letters*, 108:200403, May 2012.
- [11] F. Markopoulou. Quantum causal histories. *Classical and Quantum Gravity*, 17(10):2059, 2000.
- [12] J. Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.
- [13] R. Penrose. *Techniques of Differential Topology in Relativity*. SIAM, 1972.
- [14] C. J. Wood and R. W. Spekkens. The lesson of causal discovery algorithms for quantum correlations: causal explanations of Bell-inequality violations require fine-tuning. arXiv:1208.4119, 2012.

Contextuality and Noncommutative Geometry

Nadish de Silva

Contents

1	Summary	3
2	Background and Motivation	3
2.1	State-Observable Duality and Geometric-Algebraic Duality	3
2.2	The Noncommutative Geometry of C^* -algebras	4
2.3	The ‘Geometry’ of Noncommutative Geometry	6
3	Spatial diagrams	8
4	Extensions	9
5	K-theory, Topological to C^*-algebraic	10
6	Noncommutative Topology	10
7	Conclusions	13
7.1	Acknowledgments	13

1 Summary

We report on progress on our attempt to generalize the notion of the Gel'fand spectrum of commutative C^* -algebras to a notion of spectrum for not-necessarily-commutative C^* -algebras. Specifically, we show how to associate to each unital C^* -algebra a contravariant functor with topological spaces as the codomain in a fashion modelled on Isham and Butterfield's geometric reformulation of the Kochen-Specker theorem. We provide a framework for functorial associations of diagrams of topological spaces to operator algebras as well as an automatic method of extending any suitable functor which acts on topological spaces to one which acts on all C^* -algebras given such a functorial association of diagrams to algebras. These diagrams of spaces can be justifiably considered the Gel'fand spectra of noncommutative algebras if they lead to extensions of topological concepts which coincide with the established noncommutative geometric generalization of the concept.

These diagrams of spaces are interpreted as sample spaces featuring contextuality. We argue that the passage from classical to quantum descriptions of the core operational content of physical theories, characterized mathematically by the transition from a commutative algebra of observables to a noncommutative one, can also be understood geometrically as a shift from classical spaces to context-indexed presheaves of spaces. Alternatively, the diagrams can be thought of as a means of direct access to an imaginary noncommutative geometric space, whose algebra of continuous, complex-valued functions is a given operator algebra, via its classical quotient spaces. By synthesizing the insights of noncommutative geometers with those of physicists concerned with contextuality in quantum theory, we hope to gain deeper structural understanding of noncommutative geometry and provide automatic generalizations of classical concepts to contextual physical theories other than quantum mechanics.

After briefly reviewing a novel definition of operator K -theory given in terms of an extension of topological K -theory, we report on recent progress on a conjecture made at QPL 2013: an extension of the functor assigning to a topological space its lattice of closed sets yields the functor which assigns to a C^* -algebra its lattice of closed, two-sided ideals. This is tantamount to recovering the hull-kernel topology on the primary ideal spectrum of an arbitrary C^* -algebra as the limit of topologies coming from the algebra's contexts. Together with Rui Soares Barbosa, we prove the von Neumann algebraic analogue of the conjecture (a significant step towards proving the full conjecture) and, in doing so, answer a question raised by Heunen and Reyes by identifying precisely which partial ideals of a von Neumann algebra arise from total ideals.

2 Background and Motivation

2.1 State-Observable Duality and Geometric-Algebraic Duality

A theory of physics takes, as input, a physical system which is in the specified domain of applicability of the theory. It produces, as output, a mathematical model of the system: some mathematical structures (i.e. manifolds, vector spaces, etc.) and interpretations which allow for numerical predictions of experimental procedures to be computed. Any useful model of a system provides, at a

minimum, a representation of the multiplicity of possible states in which the system can exist as well as a description of the measurable quantitative properties of the system. The former is represented within the mathematical formalism of a theory by a collection of *states* while the latter is represented by a collection of *observables*. The predictive numerical content of the theory is summarily expressed as a map from pairings of a state with an observable to a collection of quantities which represent the possible outcomes of experiments. This map is interpreted as representing, either deterministically or probabilistically, the result of conducting the given measurement on a system in the given state.

The collections of states and observables are endowed with additional mathematical structure. The observables, being representatives of quantities which vary with states, are generally endowed with algebraic structure. The states, on the other hand, are endowed with geometric structure. Intuitively, states are close to each other when they represent configurations of a system which share similar physical properties as measured by experiments.

In models of classical systems, observables are explicitly represented as quantity-valued functions on state space. However, the fact that a pairing of a state with an observable results in a quantity means that fixing a state yields a quantity-valued function on the collection of observables. By carefully determining appropriate conditions on the quantity-valued functions on the algebra of observables, we may axiomatically select precisely those functions on observables which arise from states and thus identify the states with the functions they define. This duality hints at a fundamental conceptual symmetry between the notions of state and observable and leads to a recurring theme of modern mathematics: that of geometric-algebraic duality.

By realizing states as quantity-valued functions on observables we are identifying the idea of state of a system with the experimental outcomes prescribed by the state. When these outcomes are to be interpreted as deterministic, we are making an implicit assumption of classicality: that physical systems have simultaneously existing objective properties which the measurement process benignly extracts.

Quantum mechanical models cannot simultaneously ascribe deterministic predictions of outcomes for all possible experiments; this is, in essence, the content of the famed Kochen-Specker theorem. In the standard mathematical formalism of the theory, wherein measurable quantities are represented as self-adjoint operators on a Hilbert space, this is captured by the existence of noncommuting operators. These facts seem to place insuperable obstacles to representing quantum systems in terms of a geometric space of states on which observables might be represented as quantity-valued functions in contrast to the standard representation of observables as operators. However, reconsiderations of geometry in modern mathematics seem to give hope that such a vision could be realized.

2.2 The Noncommutative Geometry of C^* -algebras

Given a certain geometric space, i.e. a set of points together with some additional structure, we often study it by considering a commutative algebra of functions on that space. Often, these algebras retain all the geometric information about the space which leads us to establish dualities between certain categories of geometric objects and categories of commutative algebraic objects. The most important

example for our purposes is the Gel'fand duality between locally compact, Hausdorff spaces and commutative C^* -algebras. Under this duality, a space X corresponds to the commutative C^* -algebra $C(X)$ of all the continuous complex-valued functions on X . The reversal of this process—going from a commutative algebra \mathcal{A} to the topological space whose algebra of functions is \mathcal{A} —is accomplished by the Gel'fand spectrum functor Σ . In the commutative case, elements of the C^* -algebra \mathcal{A} can be thought of as continuous complex-valued functions on the space $\Sigma(\mathcal{A})$ with pointwise operations; the self-adjoint elements are the real-valued ones. This example of geometric-algebraic duality can be viewed as the mathematical statement of state-observable duality in classical physics.

Noncommutative geometry is the mathematical study of noncommutative algebras by the extension of geometric tools which have been rephrased in the language of commutative algebra [2]. Given a duality between geometric objects and commutative algebras, we can rephrase geometric concepts by expressing them algebraically in terms of functions. For example, if we wish to algebraically express the idea of an open set, we might think about the set of functions which vanish outside of it and note that this set is an ideal in $C(X)$. In fact, there is a bijective correspondence between closed ideals of $C(X)$ and open sets of X . As a more complicated example, the Serre-Swan theorem allows us to identify vector bundles over X with finitely generated projective $C(X)$ -modules. Remarkably, these algebraic descriptions of geometric concepts do not rely crucially on the commutativity of $C(X)$. This allows us to generalize geometric tools and intuition to noncommutative algebras \mathcal{A} by using these same algebraic descriptions. This justifies thinking of a noncommutative C^* -algebra as a *noncommutative topological space*. The elements of the C^* -algebra \mathcal{A} can be thought of as continuous complex-valued functions on an imaginary noncommutative space. Such a space defies explicit description by conventional mathematical ideas about what a space is; for example, it cannot be thought of as a collection of points for such an object always has a commutative algebra of functions.

One of the best examples of an extension of a topological tool to the setting of noncommutative spaces is that of K -theory. The isomorphism classes of vector bundles over a space X form a semigroup under direct sum and the Grothendieck group of this semigroup is $K(X)$. The K functor is an important cohomological invariant in the study of topology. By using the geometry-to-algebra dictionary described above, we can define an extension of K to C^* -algebras \mathcal{A} in terms of equivalence classes of finitely generated projective \mathcal{A} -modules which is called K_0 . It is an extension in the sense that when \mathcal{A} is commutative, i.e. $\mathcal{A} \simeq C(X)$ for a space X , then $K_0(\mathcal{A}) \simeq K(X)$. In this way, we obtain the most powerful invariant of C^* -algebras. We note that in the modern account of operator K_0 , one uses an equivalent formulation in terms of equivalence classes of projections in matrix algebras over \mathcal{A} .

With considerable effort, the conceptual dictionary yielded by this process of translation from geometry to algebra covers a vast terrain within mathematics. It is not just topological concepts which can be translated into the language of algebra; there exist noncommutative extensions of measure theory, differential geometry, etc. [1]

Geometry

continuous function from a space to \mathbb{C}
continuous function from a space to \mathbb{R}
range of a function
open set
vector bundle
cartesian product
disjoint union
infinitesimal
Borel probability measure
integral
1-point compactification
...

Algebra

element of the algebra (operator)
self-adjoint element of the algebra
spectrum of an operator
closed, 2-sided ideal
finite, projective module
minimal tensor product
direct sum
compact operator
state
trace
unitalization
...

Note the parallel between concepts used in the geometric, state space picture of classical mechanics and the corresponding concepts of the quantum mechanical formalism. Observable quantities of the theories are modelled by the entries of the second row. A joint classical system is described by the product of their state spaces whereas, in quantum theory, a joint system is described by the tensor product of operator algebras. We also see a hint of how noncommutative operator algebra theory generalizes measure and probability theory.

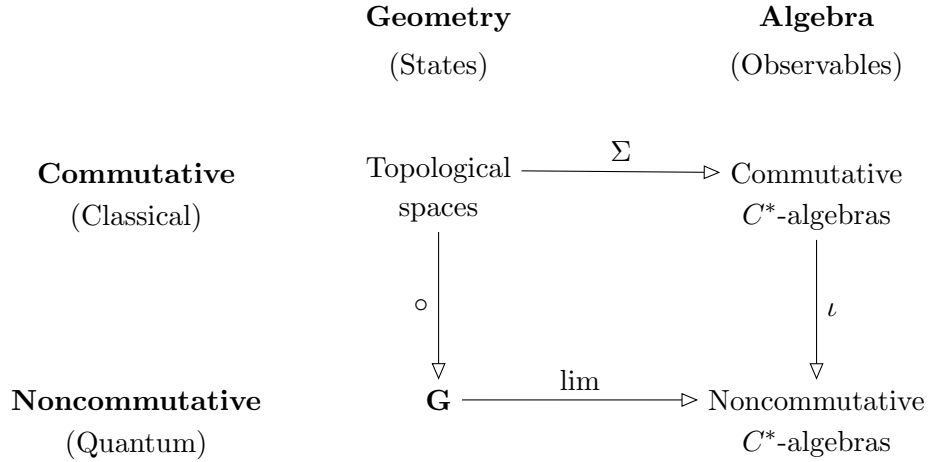
2.3 The ‘Geometry’ of Noncommutative Geometry

The unreasonable effectiveness of topological tools and intuition in the study of C^* -algebras suggests the existence of a deeper principle at work. The method of translating geometric ideas into algebra in order to generalize them is powerful but strikes one as somewhat difficult and clumsy. Ideally, one would hope for a new conception of space, of which the commutative/topological situation is a special case, which serves as the categorical dual of noncommutative C^* -algebras. That is, such a space would extend the notion of the Gel’fand spectrum of a commutative algebra to the noncommutative case and assign to an algebra \mathcal{A} an object whose set of continuous functions is, in some sense, \mathcal{A} . As pointed out above, an explicit description of (currently imaginary) noncommutative topological spaces is very difficult since such spaces defy most contemporary ideas about mathematical spaces. It is difficult to know how to begin defining such an object. However, we can imagine that equipped with such a explicit description, should it not depart too far from the commutative situation, we could find far more natural and intuitive methods of extending topological tools.

Apart from the mathematical motivation for defining an explicit spatial object to represent the spectrum of a noncommutative algebra \mathcal{A} , there is considerable physical motivation. For a quantum system whose algebra of observables is the C^* -algebra \mathcal{A} , the spectrum of the \mathcal{A} would represent the quantum analogue of the space of states of the system. The reason for this is that we would be interpreting the elements of the algebra as quantity-valued functions on a space (as opposed to operators which map a space to itself). The elements of the algebra play the role of

the observable quantities and are thus the analogue of continuous complex-valued functions on the space. This connection with physics provides a starting point for thinking about noncommutative Gel'fand duality.

The criterion for a successful spatial manifestation of noncommutative space is that it naturally leads to extensions of topological concepts which agree with well-known noncommutative geometric invariants. In effect, we aim to complete the following conceptual commutative diagram:



This informal diagram requires some explanation. The top row describes the two dually equivalent mathematical formalisms for encapsulating the operational content of a classical system: the topological picture, in which states are taken as the primitive concept, and the commutative C^* -algebra picture, in which observables are taken as primitive.

The arrows give methods for the translation and generalization of concepts. The Gel'fand spectrum functor allows for any notion or theorem phrased in terms of the topological structure of spaces to be translated into algebraic terms; i.e. open sets of a space becomes closed 2-sided ideals of an algebra. Once a concept has been phrased in terms of algebra, it can be applied without modification to the noncommutative case; i.e. finitely-generated projective modules of a commutative algebra (the equivalent of vector bundles) becomes finitely-generated projective modules of a not-necessarily-commutative algebra. Thus, the composition of the top and right arrows can be seen as the process of generating the basic entries of the noncommutative dictionary.

A topological concept can be translated in several different ways which means that intuition and judgement must be deployed when determining appropriate algebraic analogues. As a very trivial example, open sets of a space X are in correspondence with both the closed, left ideals of $C(X)$ and the closed, 2-sided ideals of $C(X)$ as these two collections are identical in the commutative case. Thus, finding a completely automatic method of translation which eliminates such ambiguities would in itself constitute an advance in the structural understanding of noncommutative geometry.

Our goal with this work is to solve the above diagram for the mathematical structure \mathbf{G} . The primary desideratum of a guess for \mathbf{G} is that it comes equipped with natural methods of generalizing notions from topology and translating them to noncommutative algebra; that is, labels for left and

bottom arrows. That the composition of these two arrows match the noncommutative dictionary is what would justify thinking of \mathbf{G} as the geometric manifestation of a noncommutative algebra. Our ansatz for \mathbf{G} , as inspired by Isham and Butterfield's insight [3] (which has already been elaborated upon in various directions by a number of authors: [4], [5], [6]), is to consider presheaves of topological spaces. In this approach, a quantum system with unital algebra of observables \mathcal{A} is analyzed in terms of a presheaf of the Gel'fand spectra of the algebra's contexts, i.e. unital, abelian subalgebras of \mathcal{A} . The elements of these spectra play the role of valuations of compatible observables, i.e. outcomes of a measurement in a context. The abelian subalgebras are connected by inclusion maps and so their spectra are connected by maps which restrict states of finer contexts to courser ones. Considered as spaces, the diagrams associated to most noncommutative von Neumann algebras have no global points; a geometric reformulation of the Kochen-Specker theorem [7]. At the same time, Borel regular probability distributions on these spaces do exist and correspond to states, i.e. positive linear functionals of norm one; a reformulation of Gleason's theorem [8]. These are two key features one would expect of the notion of quantum state space indicated above. A key observation arising from our work is that we must modify Isham and Butterfield's construction by adding morphisms which account for symmetries present only in the noncommutative case in order to find \mathbf{G} .

A parallel can be drawn between our aim and that of Akemann and Pedersen [9] who also hoped to replace the canonical translation process by finding a geometric object to represent a noncommutative topological space. Their idea was to capture the notion of open and closed sets of a space X by looking at the sorts of projections in $C(X)^{**}$ they correspond to and generalizing. Our approach is to, instead, consider the results of the translation process as, in a sense, defining the noncommutative geometry and then solve for the geometric object which regenerates the same notions.

The framework of extensions formalizes how various ways of associating diagrams of topological spaces to noncommutative algebras come with such left and bottom arrows and in this way, yield a noncommutative counterpart for every topological concept. We solve for the appropriate \mathbf{G} such that the associated extension of topological K -theory essentially matches up with the established noncommutative K -theory. As a verification of this construction of \mathbf{G} , we also use it to extend the notion of open set to the notion of closed, 2-sided ideal.

3 Spatial diagrams

To each unital C^* -algebra \mathcal{A} , we will associate a diagram of topological spaces $G(\mathcal{A})$ which is proposed as a generalization of the notion of the Gel'fand spectrum to possible noncommutative algebras such as \mathcal{A} . This diagram is expressed formally as a contravariant functor with, as its codomain, the category of compact, Hausdorff spaces. The topological spaces in this diagram can be thought of as being those which arise as quotient spaces of the hypothetical noncommutative space on which \mathcal{A} is the algebra of continuous, complex-valued functions. Not only will these spaces vary with \mathcal{A} but the shape of the diagram (the domain of the functor) will as well. This necessitates introducing the following constructions:

Definition 3.1. *For any category \mathcal{C} , $\underline{Diag}(\mathcal{C})$, the covariant category of all diagrams in \mathcal{C} , has as objects all the functors D from any small category \mathcal{S} to \mathcal{C} . Morphisms between $D_1 : \mathcal{S}_1 \rightarrow \mathcal{C}$*

and $D_2 : \mathcal{S}_2 \rightarrow \mathcal{C}$ are given by pairs (f, η) where f is a functor from $\mathcal{S}_1 \rightarrow \mathcal{S}_2$ and η is a natural transformation from D_1 to $D_2 \circ f$. The contravariant category of all diagrams $\underline{\text{Diag}}(\mathcal{C})$ has all contravariant functors to \mathcal{C} as objects; the morphisms from D_1 to D_2 are pairs (f, η) where f is a functor from $\mathcal{S}_1 \leftarrow \mathcal{S}_2$ and η is a natural transformation from $D_1 \circ f$ to D_2 .

Note that if F is a functor from \mathcal{C} to \mathcal{C}' , F naturally induces a functor from $\underline{\text{Diag}}(\mathcal{C})$ to $\underline{\text{Diag}}(\mathcal{C}')$ which we will also denote by F . Explicitly, if $D : \mathcal{A} \rightarrow \mathcal{C}$, then $F(D)$ is simply $F \circ D$. For a $\underline{\text{Diag}}(\mathcal{C})$ -morphism (f, η) , F sends (f, η) to the $\underline{\text{Diag}}(\mathcal{C}')$ -morphism $(f, F \circ \eta)$ where $(F \circ \eta)_a$ is $F(\eta_a)$. The functor F also induces, in a similar fashion, a functor from $\underline{\text{Diag}}(\mathcal{C})$ to $\underline{\text{Diag}}(\mathcal{C}')$. If F is contravariant, then it induces a contravariant functor from $\underline{\text{Diag}}(\mathcal{C})$ to $\underline{\text{Diag}}(\mathcal{C}')$ and one from $\underline{\text{Diag}}(\mathcal{C})$ to $\underline{\text{Diag}}(\mathcal{C}')$.

We can now construct the spatial diagram $G(\mathcal{A})$ associated to a unital C^* -algebra \mathcal{A} . First, we associate to \mathcal{A} a subcategory $S(\mathcal{A})$ of unital, commutative C^* -algebras. The objects of $S(\mathcal{A})$ are the unital, commutative sub- C^* -algebras of \mathcal{A} . For every inner automorphism α of \mathcal{A} , that is, one acting by conjugation by a unitary $u \in \mathcal{A}$, object $U \subset \mathcal{A}$, and any object $V \subset \mathcal{A}$ containing $\alpha(U)$, there is a restriction $\alpha|_U$ from U to V ; these are the morphisms of $S(\mathcal{A})$. Denote by $g(\mathcal{A})$ the inclusion functor from $S(\mathcal{A})$ to $uComC^*$. The diagram $G(\mathcal{A})$, an object of $\underline{\text{Diag}}(cHTop)$, is the Gel'fand spectrum functor Σ composed with $g(\mathcal{A})$. The association of diagrams to algebras is contravariantly functorial. To a unital $*$ -morphism $\phi : \mathcal{A} \rightarrow \mathcal{B}$, we define $g(\phi)$ as the $\underline{\text{Diag}}(uComC^*)$ -morphism (f, η) . The functor f maps a commutative subalgebra $U \subset \mathcal{A}$ to $\phi(U) \subset \mathcal{B}$; the action of f on the restriction of an automorphism described by a unitary u is to send it to the restriction of an automorphism described by $\phi(u)$. The component of η associated to $U \subset \mathcal{A}$ is $\phi|_U$. Finally, the $\underline{\text{Diag}}(cHTop)$ -morphism $G(\phi)$ is $\Sigma(g(\phi))$.

The spectral presheaf construction can be expressed in this framework. In this case, we consider only von Neumann algebras and their abelian sub-von Neumann algebras with only inclusions as the morphisms in the diagram. Extensions of functors, defined below to extend topological concepts to their noncommutative generalizations, can also be used with these diagrams to express Isham-Butterfield's resp. de Groote's reformulations of the Kochen-Specker resp. Gleason theorems.

4 Extensions

For a unital C^* -algebra \mathcal{A} , we have constructed the diagram of topological spaces $G(\mathcal{A})$. Any functor $F : cHTop \rightarrow \mathcal{C}$, from the category of compact, Hausdorff spaces to \mathcal{C} , can be applied directly to $G(\mathcal{A})$ to yield $F \circ G(\mathcal{A})$, a diagram in \mathcal{C} . When F is contravariant resp. covariant and \mathcal{C} is cocomplete resp. complete, we can take the colimit resp. limit of this diagram to yield a single object of \mathcal{C} . Combining all these steps gives us a new functor \tilde{F} which acts on all unital C^* -algebras. Since, when \mathcal{A} is commutative, \mathcal{A} is a terminal object in the category $S(\mathcal{A})$, it can be shown that \tilde{F} , restricted to commutative algebras, is naturally isomorphic to F composed with the Gel'fand spectrum functor.

Definition 4.1. For a contravariant functor $F : cHTop \rightarrow \mathcal{C}$ with \mathcal{C} cocomplete, $\tilde{F} : uC^* \rightarrow \mathcal{C}$ the (unitary) extension of F is $\varinjlim F \circ G$.

This definition requires the generalized colimit functor $\varinjlim : \underline{Diag}(\mathcal{C}) \rightarrow \mathcal{C}$. It assigns to a functor $F : \mathcal{A} \rightarrow \mathcal{C}$ the same object of \mathcal{C} which is assigned to F by the colimit functor of $\mathcal{C}^{\mathcal{A}}$. If η is a natural transformation between F and $G : \mathcal{A} \rightarrow \mathcal{C}$ then \varinjlim assigns to the $\underline{Diag}(\mathcal{C})$ -morphism between F and G given by $(id_{\mathcal{A}}, \eta)$ the same \mathcal{C} -morphism assigned to η by the colimit functor of $\mathcal{C}^{\mathcal{A}}$. What is novel is the ability to assign \mathcal{C} -morphisms between colimits of diagrams of different shapes. A generalized limit functor $\varprojlim : \underline{Diag}(\mathcal{C}) \rightarrow \mathcal{C}$ also exists and allows defining the extension of a covariant functor F as $\varprojlim F \circ G$.

5 K -theory, Topological to C^* -algebraic

The definition of operator K -theory which is found by the canonical method of translation to the language of commutative algebra, outlined in the Background section, is rather abstruse and free from any trace of the original geometric ideas. We aim to define operator K_0 directly in terms of topological K -theory.

The first step in computing the K_0 -group of a C^* -algebra is to take its stabilization. The stabilization functor $\mathcal{K} : C^* \rightarrow C^*$ acts on an algebra \mathcal{A} by taking it to $\mathcal{A} \otimes \mathcal{K}$, its tensor product with the C^* -algebra \mathcal{K} of compact operators on a separable Hilbert space, and sends $*$ -morphisms ϕ to $\phi \otimes id_{\mathcal{K}}$. This operation is idempotent: $\mathcal{K} \circ \mathcal{K} \simeq \mathcal{K}$. The K -theory of an algebra \mathcal{A} is the same as its stabilization $\mathcal{K}(\mathcal{A})$ —in fact, $K_0 \circ \mathcal{K} \simeq K_0$ —and, in practice, it is usually the stabilizations of algebras which are used for computations [12], [13].

Theorem 5.1. $K_0 \circ \mathcal{K} \simeq K_0 \simeq \tilde{K}_f \circ \mathcal{K}$ as functors from unital C^* -algebras to abelian groups.

Here, \tilde{K}_f is defined for unital C^* -algebras in a manner most similar to how \tilde{K} is: as $\varinjlim K \circ G_f$. The only difference in the definitions of G and G_f is that the sub- C^* -algebras of \mathcal{A} which are the objects in the full subcategory $S_f(\mathcal{A})$ of $S(\mathcal{A})$ are, in addition to being unital and commutative, finite dimensional.

6 Noncommutative Topology

The above result on extending topological K -theory to reconstruct operator K -theory indicated how Isham and Butterfield's construction must be modified to account for symmetries which arise in the noncommutative case in order to be a useful guide to reconstructing noncommutative geometric invariants, and thus to serve as a notion of noncommutative Gel'fand spectrum. The next step in using extensions to directly obtain noncommutative analogues from basic topological concepts would be to establish the following conjecture made at QPL 2013 [14] which, in essence, says that closed, two-sided ideals are the noncommutative geometric analogue of open sets:

Conjecture 6.1. *The extension of the functor which assigns to a compact, Hausdorff topological space its poset of closed sets with containment is the functor which assigns to a C^* -algebra its poset of closed, two-sided ideals.*

An equivalent way of phrasing this conjecture is to say that the lattice of the hull-kernel topology of the primary ideal space (whose points are the ideals which arise as kernels of irreducible $*$ -representations) of a C^* -algebra \mathcal{A} is the limit lattice of the topologies of the spatial diagram $G(\mathcal{A})$. It is a curious feature of this conjecture that it suggests that the proposed Gel'fand spectrum for a noncommutative algebra, which has no points in the conventional sense, has a topological lattice which matches the topology of a different notion of spectrum for noncommutative algebras which does consist of points. Since the spatial diagram seems useful for extending topological tools which other notions of spectrum do not, it seems keeping contexts separated is crucial and a natural question would be to understand precisely how; [15] seems relevant. This conjecture would establish an analogy between the functor G and the spectrum functor $Spec$ of algebraic geometry which assigns to a commutative ring the space of prime ideals with the hull-kernel topology.

There is, in our view, significant evidence to believe the conjecture holds. Our approach to this conjecture is to first prove a von Neumann algebraic analogue from which we hope to prove the general case by considering the enveloping von Neumann algebra of a given C^* -algebra. Proving this von Neumann version of the conjecture also settles a question raised by Heunen and Reyes [16, 17]: which partial ideals of von Neumann algebras arise from total ideals?

Theorem 6.2. *For a unital von Neumann algebra \mathcal{A} , the σ -weakly closed two-sided ideals of \mathcal{A} are in correspondence with choices of σ -weakly closed ideals I_V of V for every unital, commutative sub-von Neumann algebra $V \subset \mathcal{A}$ such that i) $I_V = I_{V'} \cap V$ whenever $V \subset V'$ and ii) $I_V = uI_{V'}u^*$ whenever $V = uV'u^*$ for a unitary $u \in \mathcal{A}$.*

Together with Rui Soares Barbosa, we have proved this theorem; first, in the very important case of von Neumann algebras which are factors and quite recently, in the general case of all von Neumann algebras.

Proof. For von Neumann algebras, σ -weakly closed two-sided ideals are in bijection with central projections: every such ideal is of the form $p\mathcal{A}$ for a central projection p [18, pp148].

Thus, a choice of ideals obeying the hypotheses of the theorem is the same thing as a choice of projections $\{p_V\}$ for each subalgebra V such that when $V \subset V'$, p is the largest projection in V which is less than or equal to p' and that, for any unitary $u \in \mathcal{A}$, p_{uVu^*} is up_Vu^* . We will call such a choice a *quasideal*.

A quasideal yields a central projection by simply extracting the projection $p_{\mathcal{Z}(\mathcal{A})}$ chosen at the centre of the algebra. A central projection z yields a quasideal by choosing the projections which correspond to the V -ideals $z\mathcal{A} \cap V$. When $V \subset V'$, then the projection corresponding to $z\mathcal{A} \cap V$ is indeed the inner approximation the projection corresponding to $z\mathcal{A} \cap V'$ as required. For a unitary $u \in \mathcal{A}$, the rotation by u of the projection corresponding to $z\mathcal{A} \cap V$ is the projection corresponding to $z\mathcal{A} \cap uVu^*$ and so the constructed choice of projections is indeed unitarily invariant.

A central projection z gives a quasideal whose projection at the centre corresponds to the $\mathcal{Z}(\mathcal{A})$ -ideal $z\mathcal{A} \cap \mathcal{Z}(\mathcal{A})$ which is easily seen to be z . Below, we prove that for any quasideal $\{p_V\}$, and any subalgebra W which contains the centre, the projection p_W chosen for W is simply the projection

$p_{\mathcal{Z}(\mathcal{A})}$ chosen at the centre. These facts imply that the maps defined above between central projections and quasideals are inverses.

So, suppose q is the projection p_W where W contains the centre $\mathcal{Z}(\mathcal{A})$. We claim that q is, in fact, the central carrier C of q and thus central. As $q \leq C$ is true by definition, we must show that $q \geq C$.

We define a relation called *partial orthogonality* on projections: p is partially orthogonal to q whenever there exists a central projection z such that zp and zq are orthogonal while $(1-z)p$ and $(1-q)z$ are equal. Note that partially orthogonal projections necessarily commute.

Now, let O be the unitary orbit of q . The partially orthogonal subsets of O (those such that any two projections in the subset are partially orthogonal) which contain q form a poset under inclusion. Given a chain in this poset, their union is partially orthogonal: any two projections in the union must appear together somewhere in the chain and are thus partially orthogonal. Thus, by Zorn's lemma, we can construct a maximal partially orthogonal subset M of the unitary orbit of q such that $q \in M$. Denote by S the supremum of the projections in M . The central carrier of S is the same as that of q . Denote $C - S$ by S^\perp .

By the comparison lemma for projections in a von Neumann algebra, there is a central projection z such that $zS \preceq zS^\perp$ while $(1-z)S^\perp \preceq (1-z)S$. We can assume without loss of generality that $z \leq C$ as $(1-C)S = (1-C)S^\perp = 0$. As S and S^\perp are orthogonal, there is a unitary which witnesses these order relationships. That is, there is unitary u such that $z(uSu^*) \leq zS^\perp$ and $(1-z)S^\perp \leq (1-z)(uSu^*)$. We will show that z must vanish and conclude that $S^\perp \leq (uSu^*)$.

Define v to be the unitary $zu + (1-z)1$ which acts as u within the range of z and the identity elsewhere. We establish that vqv^* and m are partially orthogonal for every $m \in M$. We have that $zvSv^*$ is simply $zuSu^*$ which, by virtue of being contained within zS^\perp , is orthogonal to S . As $zvqv^*$ and zm are contained within $zvSv^*$ and S respectively, we have that:

(i) $zvqv^*$ and zm are orthogonal.

As m is partially orthogonal to q , we have a central projection y such that yq and ym are orthogonal while $(1-y)q$ and $(1-y)m$ are equal. Cutting down by $(1-z)$, we get that $y(1-z)q$ and $y(1-z)m$ are orthogonal while $(1-y)(1-z)q$ and $(1-y)(1-z)m$ are equal. As v was constructed to act as the identity on the range of $(1-z)$, we see that in fact:

(ii) $y(1-z)vqv^*$ and $y(1-z)m$ are orthogonal

(iii) $(1-y)(1-z)vqv^*$ and $(1-y)(1-z)m$ are equal.

Summing (i) and (ii), we find that $[1-(1-y)(1-z)]vqv^*$ and $[1-(1-y)(1-z)]m$ are orthogonal. Together with (iii), we conclude that vqv^* and m are partially orthogonal as desired. As this holds for every $m \in M$, by maximality of M , it follows that $vqv^* \in M$.

As $vqv^* \in M$, we know that $vqv^* \leq S$ and so $zvqv^* \leq S$. At the same time, $zvqv^* = zuqu^* \leq zuSu^* \leq S^\perp$ is orthogonal to S . Being both contained in and orthogonal to S , $zvqv^*$ must vanish, and so zq must also vanish. As z is contained within the central carrier of q , and $zq = 0$ with q being non-zero (for else, q is central and we are already done), we are forced to conclude that z is zero. We may finally conclude that $S^\perp \leq (uSu^*)$.

We are now ready to show that q is central, that is, $q \geq C$, by showing that $q \geq S$ and $q \geq S^\perp$. Denote by $V(X)$, where X is either single projection or a commutative subset of \mathcal{A} , the unital, abelian von Neumann subalgebra of \mathcal{A} generated by X and the centre $\mathcal{Z}(\mathcal{A})$. By $P(X)$ we mean the projection $p_{V(X)}$ chosen in the quasideal at $V(X)$. We will make frequent use of two simple lemmas which follow immediately from the inner approximation property of quasideals. Whenever $V \subset V'$, $p_V \leq p_{V'}$. Further, if $p \in V$ such that $p \leq p_{V'}$ then $p \leq p_V$.

First note that $V(q) \subset W$ and of course $q \in V(q)$ and so $q \leq P(q)$. By unitary invariance of quasideals, for every $m \in M$ we have that $m \leq P(m)$. As $V(m) \subset V(M)$, $m \leq P(M)$ and so $P(M)$ is at least the supremum of M , that is, $S \leq P(M)$. From $V(S) \subset V(M)$ and $S \in V(S)$, we conclude that $S \leq P(S)$. By unitary invariance, $uSu^* \leq P(uSu^*)$. Now, $V(uSu^*) \subset V(uSu^*, S)$ and so $P(uSu^*, S) \geq uSu^* \geq S^\perp$. Since $V(S) \subset V(uSu^*, S)$, we see that $S^\perp \leq P(S)$. Having shown that $P(S)$ is at least both S and S^\perp , we see that it must be at least the central cover C of S , as S^\perp was defined at $C - S$. Now, $V(S) \subset V(M)$ and so $C \leq P(M)$. Since $V(q) \subset V(M)$, we know that $C \leq V(q)$. We may finally conclude that, as $V(q) \subset W$, that $C \leq p_W = q$. Thus, q is central. □

7 Conclusions

We have associated a geometric object $G(\mathcal{A})$ a unital C^* -algebra \mathcal{A} and gave a motivation for it as a means of keeping track of all the quotient spaces of the noncommutative space underlying \mathcal{A} which happen to be topological spaces. After showing how this association of geometric objects to algebras allows us to automatically extend topological functors, we briefly reviewed the relationship between the resulting extension of topological K -theory to a novel expression of the operator K_0 functor. We then addressed the question of how the basic topological notion of open set might be extended to the notion of closed ideal and conjectured a relationship between the geometric object $G(\mathcal{A})$ and $\text{Prim}(\mathcal{A})$, the primary ideal space of \mathcal{A} , i.e. the kernels of irreducible $*$ -representations of \mathcal{A} equipped with the hull-kernel topology. We formulated the von Neumann algebraic version of this conjecture and, together with Rui Soares Barbosa, proved it. Next, we wish to use this result to establish the original C^* -algebraic conjecture. The idea would be to, from a element of the lattice given by the extension of the topological lattice functor acting on \mathcal{A} , construct a choice of σ -weakly closed ideals as in Theorem 6.2 from \mathcal{A}^{**} , and show that the resulting central projection is also open in the sense of Akemann. As Prim is like a C^* -algebraic analogue of the Spec functor of algebraic geometry, this would strengthen the argument that G is a useful notion of noncommutative spectrum in addition to confirming its role in the automatic translation of topological concepts.

7.1 Acknowledgments

It is a pleasure to thank my supervisors, Samson Abramsky and Bob Coecke, as well as Chris Heunen and Andreas Döring for their guidance and encouragement during this project. I also wish to thank Kobi Kremnitzer and George Elliott for many crucial, constructive mathematical conversations.

References

- [1] A. Connes *Noncommutative geometry* 1994: Academic Press.
- [2] M. Khalkhali *Very basic noncommutative geometry* 2004: <http://arxiv.org/abs/math/0408416>.
- [3] J. Hamilton, C. Isham, J. Butterfield *A topos perspective on the Kochen-Specker theorem: III. Von Neumann algebras as the base category* 1999: <http://arxiv.org/abs/quant-ph/9911020>.
- [4] A. Döering, C. Isham *A topos foundation for theories of physics* 2008: Journal of Mathematical Physics.
- [5] C. Heunen, N. Landsman, B. Spitters *A topos for algebraic quantum field theory* 2009: Communications in Mathematical Physics.
- [6] S. Abramsky, A. Brandenburger *The sheaf-theoretic structure of non-locality and contextuality* 2011: New Journal of Physics.
- [7] A. Döering *Kochen-Specker theorem for von Neumann algebras* 2005: International Journal of Theoretical Physics.
- [8] H. de Groote *Observables IV: The presheaf perspective* 2007: <http://arxiv.org/abs/0708.0677>.
- [9] G. Pedersen *SAW*-algebras and corona C*-algebras, contributions to non-commutative topology* 1986: Journal of Operator Theory.
- [10] S. Mac Lane, I. Moerdijk *Sheaves in geometry and logic* 1992: Springer-Verlag.
- [11] J. Marsden, T. Ratiu *Introduction to mechanics and symmetry* 1999: Springer-Verlag.
- [12] M. Rordam, F. Larsen, N. Laustsen *An introduction to K-theory for C*-algebras* 2000: Cambridge University Press.
- [13] P. Fillmore *A user's guide to operator algebras* 1996: Wiley-Interscience.
- [14] N. de Silva *Extensions: from topological to quantum spaces* 2013: Quantum Physics and Logic 2013.
- [15] B. van den Berg, C. Heunen *No-go theorems for functorial localic spectra of noncommutative rings* 2011: Quantum Physics and Logic 2011.
- [16] M. Reyes *Obstructing extensions of the functor Spec to noncommutative rings* 2012: Israel Journal of Mathematics.
- [17] M. Reyes, C. Heunen *Private communication* 2013.
- [18] E. Alfsen, F. Shultz *State spaces of operator algebras: Basic theory, orientations, and C*-products* 2001: Birkhauser.

Complexity of Grammar Induction for Quantum Types

Antonin Delpuch

École Normale Supérieure

45 rue d'Ulm

75005 Paris, France

antonin.delpuch@ens.fr

Most categorical models of meaning use a functor from the syntactic category to the semantic category. When semantic information is available, the problem of grammar induction can therefore be defined as finding preimages of the semantic types under this forgetful functor, lifting the information flow from the semantic level to a valid reduction at the syntactic level. We study the complexity of grammar induction, and show that for a variety of type systems, including pivotal and compact closed categories, the grammar induction problem is NP-complete. Our approach could be extended to linguistic type systems such as autonomous or bi-closed categories.

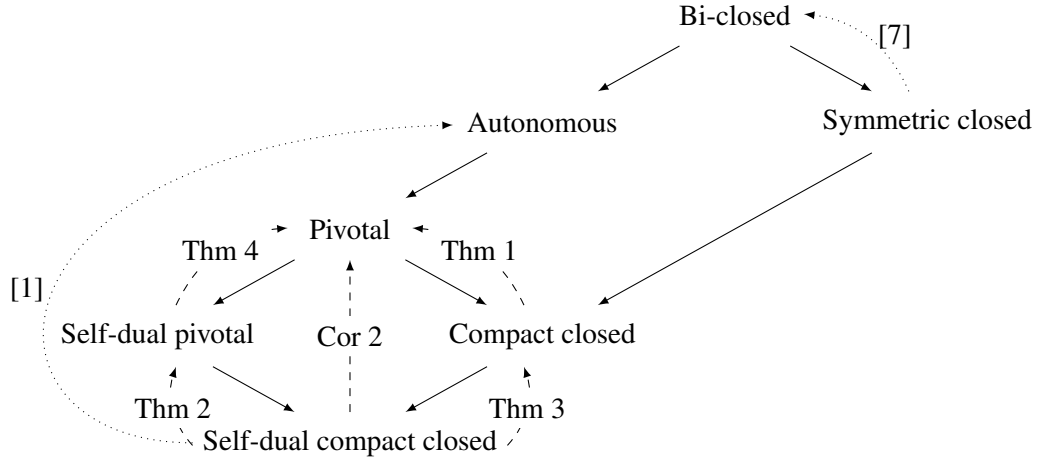
1 Introduction

1.1 Overview

Category theoretic approaches to linguistics are flourishing. They provide a convenient abstract framework for both syntax and semantics [4], and these insights enable some progress on natural language processing tasks [10]. This framework is flexible, because it allows for different types of grammars, such as the Syntactic Calculus of Lambek [15] or Compact Bilinear Logic [18], also known as pregroups [14]. It also allows for different kinds of compositional semantics, which can be distributional [4], Montagovian and extensional [17], Montagovian and intensional [6], or even hybrid models [19]. But whatever the syntax or the semantics are, these approaches rely on a functor from the syntactic category to the semantic category to give meaning to a sentence.

$$\begin{array}{ccc} \mathcal{S} & \xrightarrow{F} & \mathcal{C} \\ \text{syntax} & & \text{semantics} \end{array}$$

We propose to study the complexity of lifting the information flow at the semantic level to a valid expression at the syntactic level. In a quantum setting, this could correspond to representing a family of quantum circuits as (planar) string diagrams, for instance. In a linguistic framework, this is the task of grammar induction. Given a set of example sentences belonging to a language, the problem is to infer a grammar of this language. Originally motivated by the study of language acquisition by children [16], this task has been widely investigated in the field of formal languages [5]. If the example sentences are just raw strings, the problem is known to be intractable for most expressive classes of grammars [9]. Hence variations have been introduced, one of them consisting in adding some semantic information about the words in the example sentences. In a categorical framework, words are given syntactic types, which are objects in a monoidal category. The semantic type of a word is the image of this syntactic type under a monoidal functor to the semantic category. The categories we will use are defined in Section 2.2 and are summarised in figure 1. Our results focus on the lower part of our hierarchy of categories, which consists in quantum structures, whereas the linguistic type systems are higher up in the hierarchy.



Plain lines are functors, dashed lines are complexity results and dotted lines are existing algorithms.

Figure 1: A hierarchy of type systems

Since the grammatical correctness of a sentence is witnessed by an arrow from the product of its syntactic types to S (the type of a sentence), the problem of grammar induction can be seen as *lifting* an arrow from the semantic category to the syntactic category, as we will see in Section 4.1.

It turns out that many instances of this problem are *hard*, in the sense of computational complexity theory. This is mainly because we require that the syntactic type assigned to each word remains consistent among all the example sentences. This creates global constraints which restrict the solutions of the inference problem. In Section 4, we use this fact to reduce NP-complete problems to our grammar learning problem.

1.2 An example

Suppose we use a compact closed category for the semantics and a pivotal category for the syntax. We have to infer the possible syntactic types t_i based on their images $F(t_i)$, where F is the canonical monoidal functor from the free pivotal category to the free compact closed category on a given set of generators. In the following expressions, the tensor product \otimes is implicit.

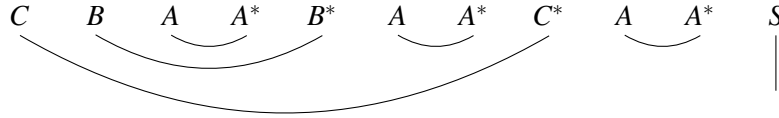
$$\begin{array}{ccccccccc} \text{Syntax} & t_1 & t_2 & t_3 & t_4 & \rightarrow S & (1) \end{array}$$

$$\begin{array}{ccccccccc} \text{Semantics} & ABC & B^*A^*A & C^*A^*A & A^*S & \rightarrow S & (2) \end{array}$$

There are many different arrows of the required domain and codomain at the semantic level. One of them is

$$\begin{array}{ccccccccccc} A & B & C & B^* & A^* & A & C^* & A^* & A & A^* & S. \\ & \searrow & \searrow & \searrow & \searrow & \searrow & \searrow & \searrow & \searrow & \searrow & | \end{array}$$

As the only difference between a free compact closed category and a free pivotal category is the symmetry, the problem bends down to finding a permutation of the basic types of each t_i such that the type reduction holds at the syntactic level. In other words, we have to find a diagrammatic reduction without crossing, such as this one:



In this particular example, one can see that it is necessary that C occurs before B in t_1 . We can add a second sentence:

$$\begin{array}{ccccccccc} \text{Syntax} & t_1 & t_5 & t_6 & t_7 & \rightarrow S & (3) \end{array}$$

$$\begin{array}{ccccccccc} \text{Semantics} & ABC & B^*C^*C & A^*C^*C & C^*S & \rightarrow S & (4) \end{array}$$

This examples forces A to occur before B in t_1 . Hence every solution of the learning problem made of these two sentences will be such that C and A occur before B in t_1 . In Section 4, this technique enables us to reduce the problem of betweenness [11] to our grammar lifting problem. This problem is known to be NP-complete.

2 A grammar hierarchy

2.1 Monoidal categories as type systems

We define how monoidal categories can be used as type systems. Both the syntactic and the semantic categories will be seen as type systems in our induction problem.

Definition 1. A *type system* (\mathcal{C}, S) is a strict monoidal category \mathcal{C} with a distinguished object S in \mathcal{C} .

When the object S is clear from the context, the type system is simply noted \mathcal{C} . The objects of this category will be used to denote types. We require the category to be monoidal, so that we can define the sentence type as the product of the types of its words. The distinguished object will play the role of the type for a grammatical sentence. The arrows in the category play the role of reductions: A reduces to B when $\mathcal{C}(A, B)$ is not empty.

The type systems we will consider are monoidal categories with some additional structure (which will be detailed in section 2.2), and freely generated by a basic category, whose objects are called **basic types** and morphisms are understood as subtyping relations: there is a morphism between two basic types A and B when A is a subtype of B .

Definition 2. A *lexicon* l over a set of words W and a type system (\mathcal{C}, S) is a function $l : W \rightarrow \mathcal{C}$.

Although it is interesting to consider the case where multiple types can be assigned to a single word, the previous definition restricts our lexicons to one type per word. We restruct ourselves to rigid grammars, according to the terminology of [1].

Definition 3. A sequence of words $w_1, \dots, w_n \in W$ is *grammatical* for a lexicon l when $\mathcal{C}(l(w_1) \otimes \dots \otimes l(w_n), S)$ is not empty.

In this definition, S is the distinguished type of the underlying type system.

Definition 4. A *functor* of type systems from (\mathcal{C}_1, S_1) to (\mathcal{C}_2, S_2) is a functor of monoidal categories $F : \mathcal{C}_1 \rightarrow \mathcal{C}_2$ such that $F(S_1) = S_2$.

From this definition, the following property follows immediately:

Proposition 1. Let $F : \mathcal{T}_1 \rightarrow \mathcal{T}_2$ be a functor of type systems. If a sentence w_1, \dots, w_n is grammatical for the lexicon \mathcal{L}_1 over \mathcal{T}_1 , then it is grammatical for the lexicon $F \circ \mathcal{L}_1$ over \mathcal{T}_2 .

This property expresses that if a sentence is correct at the syntactic level, then there is a valid reduction at the semantic level.

2.2 Various structures in monoidal categories

We now move on to the definition of the categories involved in the hierarchy of figure 1.

Definition 5. A *bi-closed category* is a monoidal category in which for all object B , the functor $_ \otimes B$ has a right adjoint $_ / B$ and the functor $B \otimes _$ has a right adjoint $B \setminus _$.

In other words, this means that for every pair of objects A, B , we have morphisms $\text{eval}_{A,B}^l : B \otimes (B \setminus A) \rightarrow A$ and $\text{eval}_{A,B}^r : (A/B) \otimes B \rightarrow A$ satisfying some coherence equations, and similarly some morphisms $A \rightarrow (A \otimes B)/B$ and $A \rightarrow B \setminus (B \otimes A)$. Type systems built on bi-closed categories correspond to grammars defined in the Syntactic Calculus of Lambek.

Definition 6. An *autonomous category*¹ is a monoidal category where for each object A , there are two objects, the left (A^l) and right (A^r) adjoints, equipped with four morphisms $\epsilon_A^l : A^l \otimes A \rightarrow 1$, $\epsilon_A^r : A \otimes A^r \rightarrow 1$, $\eta_A^l : 1 \rightarrow A \otimes A^l$ and $\eta_A^r : 1 \rightarrow A^r \otimes A$ satisfying the following equalities :

$$\begin{aligned} (\epsilon_A^r \otimes 1_A) \circ (1_A \otimes \eta_A^r) &= 1_A & (\epsilon_A^l \otimes 1_{A^l}) \circ (1_{A^l} \otimes \eta_A^l) &= 1_{A^l} \\ (1_A \otimes \epsilon_A^l) \circ (\eta_A^l \otimes 1_A) &= 1_A & (1_{A^r} \otimes \epsilon_A^r) \circ (\eta_A^r \otimes 1_{A^r}) &= 1_{A^r} \end{aligned}$$

Type systems built on a free autonomous category define pregroup grammars. For instance, let n be the type of a noun phrase and s be the distinguished type of a sentence. If we give the type n to the words *Mary* and *John*, and the type $n^r \otimes s \otimes n^l$ to *loves*, the sentence *Mary loves John* has the type $n \otimes n^r \otimes s \otimes n^l \otimes n$. This type reduces to s through the morphism

$$(\epsilon_n^r \otimes 1_s \otimes \epsilon_n^l) : n \otimes n^r \otimes s \otimes n^l \otimes n \rightarrow s$$

See [14] for a linguistic presentation of pregroup grammars and [18] for the links with category theory.

The distinction between n^l and n^r is important at a syntactical level to reject ill-formed sentences. For instance, we can give the type $s^r \otimes s$ to adverbs placed at the end of a sentence. If $s^l = s^r$, then the type $s^r \otimes s = s^l \otimes s$ reduces to 1 through ϵ_s^l , hence the adverb can be written at any place in the sentence, which does not reflect the usual rules of grammar. As one can show that for any object n , $n^{rl} \simeq n \simeq n^{lr}$, the iterated adjoints of a type n are of the form

$$\dots, n^{lll}, n^{ll}, n^l, n, n^r, n^{rr}, n^{rrr} \dots$$

so we can write $n^{ll} = n^{-2}, n^l = n^{-1}, n = n^0, n^r = n^1, n^{rr} = n^2$, and so on.

However, it makes sense to drop the distinction between left and right adjoints at the semantic level: in terms of flow of information, an adjoint is just something that can consume a resource, no matter whether it comes from the left or the right side.

Definition 7. A *pivotal category* is an autonomous category with a monoidal natural isomorphism between A^r and A^l . We set $A^* = A^l$.

Pivotal categories correspond to groups, in the sense that in a free pivotal category, two objects have an arrow between them if and only if they are equal in the corresponding free group (where $*$ plays the role of the inverse, hence $*$ will be sometimes noted $^{-1}$).

The canonical morphism between the free pregroup and the free group is defined by

$$h : t_1^{e_1} \otimes \dots \otimes t_n^{e_n} \mapsto t_1^{(-1)^{e_1}} \otimes \dots \otimes t_n^{(-1)^{e_n}}$$

where $t_1^{e_1} \otimes \dots \otimes t_n^{e_n}$ is the canonical form of a pregroup element.

¹Some authors use the name **compact closed category** instead, but this term has been used for both symmetric and planar categories. As we want to insist on the fact that these categories are not symmetric (contrary to some other ones in this article), we follow the terminology of [21].

Definition 8. A **compact closed category** is an autonomous category which is symmetric, i.e. for each objects A and B there is a monoidal natural isomorphism $s_{A,B} : A \otimes B \rightarrow B \otimes A$ such that $s_{A,B}^{-1} = s_{B,A}$.

For instance, the category of finite-dimensional vector spaces is compact closed. One can wonder why we introduced the isomorphism $A^l \simeq A^r$ before adding the symmetries $s_{A,B}$. The following fact explains our choice.

Proposition 2. *Compact closed categories are pivotal.*

This property is well known (it is stated in [4], and implicitly in [21]) but I have never seen a proof of it.

Proof. Let ϕ_A and ψ_A be the following morphisms :

$$\phi_A = \begin{array}{c} A^l \\ | \\ A^l \text{---} A^r \\ | \quad | \\ A^r \text{---} A^l \\ | \\ A^r \end{array} \quad \psi_A = \begin{array}{c} A^r \\ | \\ A \text{---} A^l \\ | \quad | \\ A^l \text{---} A \\ | \\ A^l \end{array}$$

We have $\psi_A \circ \phi_A = 1_{A^l}$ and $\phi_A \circ \psi_A = 1_{A^r}$. By symmetry, let us show the first equality only.

Moreover, one can check with similar techniques that this isomorphism is monoidal and natural. \square

Definition 9. A **self-dual compact closed category** is a compact closed category with a family of isomorphisms $h_A : A \rightarrow A^*$.

Self-dual compact closed categories have been studied in detail by Selinger in [20]. The definition we adopt here corresponds to his first option, namely self-duality without coherence. As a finite-dimensional vector space is isomorphic to its dual, the category of finite-dimensional vector spaces is self-dual. This category has been widely used as the underlying semantic category for models of meaning, such as in [4], [19] or [6]. The objects in this category have also been used in [1] as semantic types in a learning task. However, they did not introduce a whole typing system at the semantic level, as they had no notion of reduction on semantic types.

We have introduced the commutativity first and then the isomorphism between A and A^* . It is possible to swap these properties, although it requires to be more careful:

Definition 10. A **free self-dual pivotal category** is the free pivotal category generated by a category \mathcal{C} where for each object $A \in \mathcal{C}$, $A \simeq A^*$.

A self-dual pivotal category models a rewriting system where any two identical adjacent letters cancel.

It is important to notice that we require that $A \simeq A^*$ only for basic objects. If this were true for all objects, then as noted by Selinger [21], we would get the following isomorphism

$$A \otimes B \simeq (A \otimes B)^* \simeq B^* \otimes A^* \simeq B \otimes A$$

. This isomorphism is not a symmetry in general but would have the same effects on our type system.

A widespread category for semantic types in the linguistic literature is the free symmetric monoidal closed category. It has been used, among others, in [2] and [8].

Definition 11. A *symmetric closed category* is a symmetric bi-closed category. For all objects A and B , $B \backslash A \simeq A/B$, so we note $A|B = A/B$.

The objects of this category can be thought of simple types for the simply-typed λ -calculus with pairs. The object $A|B$ plays the role of the type $B \rightarrow A$ and we have a morphism $\text{eval}_{A,B} : (A|B) \otimes B \rightarrow A$ satisfying the required coherence conditions.

3 Functional types

3.1 Restricting the set of possible types

Not all types are likely to be used in a type-logical grammar. We expect types to be functional, i.e. to be built using only abstractions, the operations \backslash and $/$.

For instance, the type $n \otimes s \otimes n$ belongs to the free pregroup generated by n and s , but cannot be constructed by iterated abstractions. The type $n^r \otimes s \otimes n^l$ however can be constructed as $n \backslash (s/n)$ or $(n \backslash s)/n$.

Definition 12. Let \mathcal{L} be the free bi-closed monoidal category. The set $P \subset \text{Ob}(\mathcal{L})$ is the closure by $/$ and \backslash of the set of basic types. Given a type system (\mathcal{C}, S) and a bi-closed functor $F : \mathcal{L} \rightarrow \mathcal{C}$ the set of *functional types* in \mathcal{C} is $F(P)$.

Restricting our search of types to this form of type reduces our search space. This restriction makes sense because these types are more likely to be relevant from a linguistic point of view. For instance, [14] builds a fairly advanced grammar of English and he uses only functional types in his grammar, while not mentioning this constraint at all.

3.2 Properties of functional types

The generative power of pregroup grammars is not reduced when we require functional types: the proof given in [3] that every ε -free context free grammar is weakly equivalent to a pregroup grammar uses only functional types.

For group grammars (i.e. type systems built on pivotal categories), restricting the assignments to functional types does not harm the expressiveness either, as it is enough to multiply by $a^{-1}a$ the types that are not functional to get an equivalent grammar with functional types only. This remark will be made clear by the following proposition, which characterises functional types in pivotal categories.

Proposition 3. In a pivotal category, functional types are exactly those which are either

- basic types (generators of the free autonomous category), or
- products of basic types with exponents $t_1^{e_1} \otimes \cdots \otimes t_n^{e_n}$, where at least one e_i is -1 and at least one e_i is $+1$.

Proof. By induction on a functional Lambek type t , let us show that $F(t)$ satisfies the characterization above. If $t = a$, a basic type, then $F(t) = a$, falling into the first option. If $t = u/v$, then $F(t) = F(u)F(v)^{-1}$. By induction, there is a basic type occurring with a $+1$ exponent in $F(u)$, so it occurs again

with the same exponent in $F(t)$. Similarly, there is a basic type occurring with a $+1$ exponent in $F(v)$, so it occurs with a -1 exponent in $F(t)$.

Conversely, let us show by induction on the length of a group type $t = t_1^{e_1} \otimes \dots \otimes t_n^{e_n}$ satisfying the characterization that it is the image of a functional Lambek type. If $n = 1$, then $t = a$ where a is a basic type, so $F(a) = t$. If $n > 1$, there are several cases:

- $e_n = -1$ and e_1, \dots, e_{n-1} satisfies the characterization. Then by induction we can find a functional Lambek type u such that $F(u) = t_1^{e_1} \otimes \dots \otimes t_{n-1}^{e_{n-1}}$ and hence $F(u/t_n) = F(u)F(t_n)^{-1} = t$.
- $e_n = -1$ and $e_1, \dots, e_{n-1} = +1$: then $(-e_2), \dots, (-e_n)$ satisfies the characterization and hence there is a functional u such that $F(u)^{-1} = t_2^{e_2} \otimes \dots \otimes t_n^{e_n}$, hence $F(t_1/u) = t$.
- if $e_n = +1$ and $(-e_1), \dots, (-e_{n-1})$ satisfies the characterization. Then by induction we can find a functional Lambek type u such that $F(u)^{-1} = t_1^{e_1} \otimes \dots \otimes t_{n-1}^{e_{n-1}}$ and hence $F(u \setminus t_n) = F(u)^{-1}F(t_n) = t$.
- if $e_n = +1$ and $e_1, \dots, e_{n-1} = -1$: then e_2, \dots, e_n satisfies the characterization and hence there is a functional u such that $F(u) = t_2^{e_2} \otimes \dots \otimes t_n^{e_n}$, hence $F(t_1 \setminus u) = F(t_1)^{-1}F(u) = t$.

This completes the proof. \square

Corollary 1. *In a compact closed category, the characterization of functional types is the same. In a self-dual compact closed category, every type but 1 is functional.*

4 Complexity of the grammar induction problem

4.1 Definition of the problem

We study the complexity of learning syntactic types based on positive samples (i.e. a set of grammatical sentences) with semantic types. Each word occurrence in the samples comes with a semantic type. The nature of the syntactic and semantic types depends on the problem.

Definition 13. A **training sample** for a type system (\mathcal{C}, S) and a finite set of variables V is a finite set of sentences, where each sentence is a finite sequence of the form $(v_1, t_1), \dots, (v_n, t_n)$, where $v_i \in V$ and $t_i \in \mathcal{C}$ is functional, and such that all the sentences are grammatical for their respective type assignment.

Note that we do not require that a variable is always paired with a single type. The type of the word can depend on the context in which it appears.

In the following sections, we study the complexity of inducing a grammar, given a finite training sample. First we give a definition of the problem.

Definition 14. Let (\mathcal{C}, S) be the syntactic type system, (\mathcal{C}', S') be the semantic type system and F be a morphism from (\mathcal{C}, S) to (\mathcal{C}', S') . We call **grammar induction** the problem of, given a training sample T for the type system (\mathcal{C}', S') , find a lexicon $h : V \times \text{Ob}(\mathcal{C}') \rightarrow \mathcal{C}$ such that

$$\text{for all pair } (v_i, t_i) \in T, F(h(v_i, t_i)) \simeq t_i \text{ and } h(v_i, t_i) \text{ is functional}$$

$$\text{and for all sentence } (v_1, t_1), \dots, (v_n, t_n) \in T, \mathcal{C}(h(v_1, t_1) \otimes \dots \otimes h(v_n, t_n), S) \text{ is not empty}$$

In other words, the problem is to find functional syntactic types that are compatible with the semantic types and all the sentences are grammatical at the syntactic level. Note that we require that each pair (variable, semantic type) is associated to an unique syntactic type, following [1]. Without this restriction, the problem is trivial as the syntactic types can be chosen independently for each sentence.

4.2 Learning pivotal categories from compact closed categories

Theorem 1. *Type inference from a compact closed category to a pivotal category is NP-complete.*

Proof. We give a reduction of the betweenness problem [11] to our grammar induction problem. The betweenness problem is as follows. Given a finite set A and a set of triples $C \subset A^3$, the problem is to find a total ordering of A such that for each $(a, b, c) \in C$, either $a < b < c$ or $c < b < a$. This problem is NP-complete [?].

The compact closed category we will consider contains the objects a for each $a \in A$ and $d_{a,b,c}$ for each $(a, b, c) \in C$, with the following reduction between basic types: $a \rightarrow d_{a,b,c}$ and $c \rightarrow d_{a,b,c}$. We set $y = \prod_{x \in A} x$. The preimage of this type will define the total order satisfying the constraints induced by the sentences. For each triple $(a, b, c) \in C$, we define the following compact closed types:

$$w = \prod_{x \in A \setminus \{a,b,c\}} x \quad c_1 = d_{a,b,c}^{-1} w^{-1} w \quad c_2 = b^{-1} w^{-1} w$$

and add the following sentence to the training sample:

$$(Y, y)(W_{a,b,c,1}, c_1)(W_{a,b,c,2}, c_2)(W_{a,b,c,3}, c_1)(W_{a,b,c,4}, w^{-1})$$

where the W are words chosen to be different from any word previously seen.

This reduction is polynomial. Let us show that this grammar induction problem has a solution if and only if the corresponding betweenness problem has a solution. If there is a total ordering $<$ of A satisfying the constraints, let $A = \{x_1, \dots, x_n\}$ where $x_1 < \dots < x_n$. One can check that with the following preimages, the sample is grammatical in the pivotal category:

- The type of Y becomes $y' = \prod_{i=1}^n x_i$.
- For each $(a, b, c) \in C$, let p, q, r and s be such that

$$y' = p \cdot a \cdot q \cdot b \cdot r \cdot c \cdot s \quad \text{or} \quad y' = p \cdot c \cdot q \cdot b \cdot r \cdot a \cdot s$$

(where p, q, r and s are possibly equal to 1). y' reduces to $p \cdot d_{a,b,c} \cdot q \cdot b \cdot r \cdot d_{a,b,c} \cdot s$.

P possible type assignment for the W is:

$$W_{a,b,c,1} : s^{-1} d_{a,b,c}^{-1} s p q r (p q r)^{-1} \quad W_{a,b,c,2} : (r s)^{-1} b^{-1} r s (p q) (p q)^{-1} \quad W_{a,b,c,3} : (q r s)^{-1} d_{a,b,c}^{-1} (q r s) p p^{-1}$$

One can check that the image of this assignment is equal to the assignment from the training sample and that it makes the sentences grammatical in the pivotal category.

Conversely, if there exists a pivotal type assignment, then as the type b does not occur in the types assigned to $W_{a,b,c,1}$ and $W_{a,b,c,3}$, there is an a or a d on the right side of the occurrence of b , and similarly on the left side. But as there cannot be two occurrences of the same basic type in y' , we have either $a < b < c$ or $c < b < a$.

Hence the problem is NP-hard. As one can check a solution in polynomial time, the problem is NP-complete. \square

4.3 Learning self-dual pivotal categories from self-dual compact closed categories

Similarly, the previous proof can also be carried when $a \simeq a^{-1}$, giving the following theorem:

Theorem 2. *Type inference from a self-dual compact closed category to a self-dual pivotal category is NP-complete.*

4.4 Learning compact closed categories from self-dual compact closed categories

The problem of grammar induction from a self-dual compact closed category to a compact closed category bends down to assigning exponents to the types. It can be reduced to an integer linear programming problem where we are interested in nonnegative solutions only. This problem is NP-complete and we will show that the grammar induction problem itself is actually NP-complete.

Theorem 3. *Type inference from a self-dual compact closed category to a compact closed category is NP-complete.*

Proof. We give a polynomial reduction from 3-SAT to the problem of learning symmetric pivotal types from self-adjoint (symmetric pivotal) types. As 3-SAT is NP-complete [13], and the learning problem is in NP, this will complete the proof.

Let $\phi = c_1 \wedge \dots \wedge c_n$ be a conjunction of 3-clauses. We write $c_i = x_{i_1}^{b_{i,1}} \vee x_{i_2}^{b_{i,2}} \vee x_{i_3}^{b_{i,3}}$, where x^1 stands for x and x^0 for $\neg x$. Let p be the number of variables occurring in ϕ . We assume that they are indexed from 1 to p . For each clause c_i we define a self-adjoint type $t_i = cz_{i_1, b_{i,1}} z_{i_2, b_{i,2}} z_{i_3, b_{i,3}}$. Our idea is that we will force c to have a -1 exponent in the corresponding group type, and hence this group type will be functional if and only if one of the $z_{a,b}$ occurs with a $+1$ exponent. As a clause is true when at least one of the literals it contains is true, this will encode satisfiability.

For each literal x_i^b , $i \in \{1, \dots, p\}$, $b \in \{0, 1\}$, let $n_{i,b}$ be the number of occurrences of x_i^b in ϕ . We define a self-adjoint type $v_{i,b} = z_{i,b} y_{i,b} y_{i,b}$. Our idea is that the exponent of $z_{i,b}$ will encode the truth value of $z_{i,b}$ in the satisfying assignment. The two other types ensure that the type is functional and will be forced to have exponents $+1$ and -1 or -1 and $+1$. We define a training sample containing the following sentence (where the product denotes the concatenation of word-type pairs):

$$(S, s) \prod_{i=1}^n ((C_i, t_i)(C'_i, c)) \prod_{i=1}^p \prod_{b=0}^1 \prod_{j=1}^{n_{i,b}} (X_i^b, v_{i,b})$$

Note that the types of the words S and C'_i are basic types, so the only functional syntactic types compatible with the learning problem are these basic types. As we will show, this defines a grammar induction problem which has a solution if and only if there is a truth assignment of $x_{1,0}, x_{1,1}, x_{2,0}, x_{2,1}, \dots, x_{p,0}, x_{p,1}$ such that all the clauses are satisfied. What is missing here is that $x_{i,0} = \neg x_{i,1}$. For each $i \in \{1, \dots, p\}$ we add the following sentence:

$$(S, s)(X_i^0, v_{i,0})(D_i, z_{i,0} z_{i,1})(X_i^1, v_{i,1})$$

We will show that this ensures that x_i^0 and x_i^1 get different truth values. Let us show first that if ϕ is satisfiable, then the grammar induction problem has a solution. Let $x_1 = a_1, \dots, x_p = a_p$ be a satisfying boolean assignment. We give X_i^b the group type $z_{i,b}^e y_{i,b}^{-1} y_{i,b}$, where $e = -1$ if $b = a_i$ and $e = 1$ otherwise. This type is functional.

We give C_i the type $c^{-1} z_{i_1, b_{i,1}}^{e_1} z_{i_2, b_{i,2}}^{e_2} z_{i_3, b_{i,3}}^{e_3}$ where $e_k = 1$ if $b_{i,k} = a_{i_k}$ and $e_k = -1$ otherwise. As the clause c_i is satisfied, there is at least one $k \in \{1, 2, 3\}$ such that $b_{i,k} = a_{i_k}$, hence the type is functional. Finally, we give d_i the type $z_{i,0}^{-e} z_{i,1}^e$ where $e = 1$ if $a_i = 1$ and 0 otherwise. This type is functional.

Let us show that the main sentence is grammatical. As the exponent of $z_{i,b}$ in the type assigned to a clause only depends on a_i , there are $n_{i,b}$ occurrences of $z_{i,b}$, with the same exponent, in $\prod_{j=1}^n (c_j, t_j)$. By construction, the exponent of $z_{i,b}$ is inversed in the type assigned to x_i^b , and there are $n_{i,b}$ such occurrences in the sentence. Hence all the $z_{i,b}$ cancel. The type c assigned to c'_i cancels with c^{-1} in the type assigned to

t_i , and the $y_{i,b}$ cancel as well. Hence only s remains: the sentence is grammatical. The p other sentences are grammatical as well: with our type assignment, they become

$$s \cdot z_{i,0}^e y_{i,0}^{-1} \cdot z_{i,0}^{-e} z_{i,1}^e \cdot z_{i,1}^{-e} y_{i,1}^{-1} \rightarrow s$$

It remains to show that if the grammar induction problem has a solution, then ϕ is satisfiable. Suppose there are functional group types r_j preimage of t_j and $w_{i,b}$ preimage of $v_{i,b}$ such that the sentences are grammatical at the syntactic level. As the pregroup type c occurs with exponent $+1$ n times in the product thanks to the words c'_j , all the occurrences of c in the r_j have the exponent -1 , otherwise they would not cancel. For each r_j , it is functional so one of the $z_{i,b}$ has exponent $+1$. For each x_i^b , $z_{i,b}$ occurs $n_{i,b}$ times with the same exponent, thanks to $w_{i,b}$, and $n_{i,b}$ other times in the clauses, so the exponent assigned to $z_{i,b}$ is the same in every r_j . Finally, thanks to the p additional sentences, the exponents of $z_{i,0}$ and $z_{i,1}$ are opposite. Hence there is a satisfying assignment for ϕ . \square

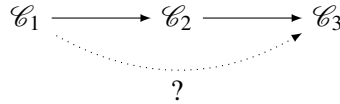
4.5 Learning pivotal categories from self-dual pivotal categories

The construction of Theorem 3 can be adapted to work in non-commutative structures, hence the following theorem:

Theorem 4. *Type inference from self-dual pivotal categories to pivotal categories is NP-complete.*

4.6 Composing complexity results

Suppose we know the complexity of the grammar induction problem between \mathcal{C}_1 and \mathcal{C}_2 , and between \mathcal{C}_2 and \mathcal{C}_3 . What can be said about grammar induction between \mathcal{C}_1 and \mathcal{C}_3 ?



Given a syntactic category \mathcal{C} and a semantic category \mathcal{C}' , we introduce the notion of exact samples.

Definition 15. *A training sample is said **exact** for some syntactic type t when it contains a word-type pair $(w, F(t))$ such that for all solutions h of this training sample, $h(w, F(t)) = t$.*

*We say that a grammar induction problem **has exact samples** when there exists exact samples for each syntactic type t .*

In other words, a grammar induction problem has exact samples when we can build sentences forcing the preimage of a particular type.

Lemma 1. *If the grammar induction problem has exact samples, then for all finite set of syntactic types $T = \{t_1, \dots, t_n\}$ there exists a training sample which is exact for t_1, \dots, t_n .*

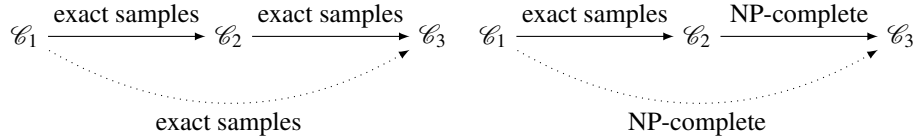
Proof. Take an exact training samples for each element of T . Make these training samples disjoint by ensuring that they use different words. The concatenation of these training samples satisfies the property claimed. \square

Lemma 2. *If grammar induction from \mathcal{C}_1 to \mathcal{C}_2 and from \mathcal{C}_2 to \mathcal{C}_3 has exact samples, then so does grammar induction from \mathcal{C}_1 to \mathcal{C}_3 .*

Proof. Take an exact sample S for t from \mathcal{C}_2 to \mathcal{C}_3 . For each type t' occurring in S , take exact samples for t' from \mathcal{C}_1 to \mathcal{C}_2 . Concatenate these samples with the image of S under the functor from \mathcal{C}_2 to \mathcal{C}_1 . \square

Proposition 4. *If grammar induction from \mathcal{C}_1 to \mathcal{C}_2 has (polynomial) exact samples and grammar induction from \mathcal{C}_2 to \mathcal{C}_3 is NP-complete, then grammar induction from \mathcal{C}_1 to \mathcal{C}_3 is NP-complete.*

Proof. Take an instance of SAT. It can be represented as an equivalent training sample from \mathcal{C}_2 to \mathcal{C}_3 . Take the image of this training sample by the functor from \mathcal{C}_2 to \mathcal{C}_1 and force this new sample to have the original preimages in \mathcal{C}_2 by adding an exact sample. This problem has a solution if and only if the instance of SAT is satisfiable. \square



Lemma 3. *The following grammar induction problems have exact samples:*

- *from self-dual compact closed to self-dual pivotal*
- *from self-dual compact closed to compact closed*
- *from self-dual pivotal to pivotal*
- *from compact closed to pivotal*

Proof. Use the same techniques as the ones we developped for our reductions. \square

Corollary 2. *Grammar induction from a self-dual compact closed category to a pivotal category is NP-complete.*

Proof. Combine Lemma 3 and Theorem 1 with Proposition 4. \square

5 Future work

A number of questions remain open. Being able to classify the complexity of the inference problem in the higher half of the hierarchy would enable us to give complexity results on the problems studied in [1] and [7].

Another issue is the expressivity of the classes of grammars defined by the categories in the lower half of the hierarchy. These grammars generate sub-classes of the context-free grammars, but it would be interesting to relate these sub-classes to known classes from the field of formal languages.

One could also use this framework to study inference problems in which the structure of a parse is known, but the types are unknown. This notion of learning with structural examples has been studied for the syntactic calculus [12].

Acknowledgements

This work has been supported by the Quantum Group in the Department of Computer Science of the University of Oxford. I am grateful to Thomas Bourgeat, Amar Hadzihasanovic and Isabelle Tellier for their help, and the reviewers for their useful comments. Special thanks go to Jamie Vicary who helped me a lot by supervising and reviewing my work.

References

- [1] Denis Béchet, Annie Foret & Isabelle Tellier (2007): *Learnability of Pregroup Grammars*. *Studia Logica* 87(2-3), pp. 225–252.
- [2] Wojciech Buszkowski (1984): *A note on the Lambek-van Benthem calculus*. *Bulletin of the Section of Logic* 13(1), pp. 31–35.
- [3] Wojciech Buszkowski & Katarzyna Moroz (2008): *Pregroup grammars and context-free grammars*. *Computational Algebraic Approaches to Natural Language*, Polimetrica, pp. 1–21.
- [4] Bob Coecke, Edward Grefenstette & Mehrnoosh Sadrzadeh (2013): *Lambek vs. Lambek: Functorial vector space semantics and string diagrams for Lambek calculus*. *Ann. Pure Appl. Logic* 164(11), pp. 1079–1100. Available at <http://dx.doi.org/10.1016/j.apal.2013.05.009>.
- [5] Colin De La Higuera (2005): *A bibliographical study of grammatical inference*. *Pattern recognition* 38(9), pp. 1332–1348.
- [6] Antonin Delpuch & Anne Preller (2014): *From Natural Language to RDF Graphs with Pregroups*. In: *Proceedings of the EACL 2014 Workshop on Type Theory and Natural Language Semantics (TTNLS)*, Association for Computational Linguistics, Gothenburg, Sweden, pp. 55–62. Available at <http://www.aclweb.org/anthology/W14-1407>.
- [7] Daniela Dudau-Sofronie, Isabelle Tellier & Marc Tommasi (2001): *From Logic to Grammars via Types*. In: *Proceedings of the 3rd Workshop on Learning Language in Logic*, pp. 35–46.
- [8] Daniela Dudau-Sofronie, Isabelle Tellier & Marc Tommasi (2003): *A learnable class of Classical Categorical Grammars from typed examples*. In: *8th Conference on Formal Grammar*, pp. 77–88.
- [9] E Mark Gold (1967): *Language identification in the limit*. *Information and control* 10(5), pp. 447–474.
- [10] Edward Grefenstette & Mehrnoosh Sadrzadeh (2011): *Experimental Support for a Categorical Compositional Distributional Model of Meaning*. In: *EMNLP*, pp. 1394–1404. Available at <http://www.aclweb.org/anthology/D11-1129>.
- [11] Walter Guttman & Markus Maucher (2006): *Variations on an ordering theme with constraints*. In: *Fourth IFIP International Conference on Theoretical Computer Science-TCS 2006*, Springer, pp. 77–90.
- [12] Makoto Kanazawa (1996): *Identification in the limit of categorial grammars*. *Journal of Logic, Language and Information* 5(2), pp. 115–155.
- [13] Richard M Karp (1972): *Reducibility among combinatorial problems*. Springer.
- [14] J Lambek (2008): *Pregroup Grammars and Chomsky's Earliest Examples*. *Journal of Logic, Language and Information* 17(2), pp. 141–160.
- [15] Joachim Lambek (1968): *Deductive systems and categories*. *Theory of Computing Systems* 2(4), pp. 287–318.
- [16] Steven Pinker (2009): *Language Learnability and Language Development, With New Commentary by the Author*. Harvard University Press.
- [17] Anne Preller (2005): *Category theoretical semantics for pregroup grammars*. In: *Logical aspects of computational linguistics*, Springer, pp. 238–254.
- [18] Anne Preller & Joachim Lambek (2007): *Free Compact 2-categories*. *Mathematical. Structures in Comp. Sci.* 17(2), pp. 309–340, doi:10.1017/S0960129506005901. Available at <http://dx.doi.org/10.1017/S0960129506005901>.
- [19] Anne Preller & Mehrnoosh Sadrzadeh (2011): *Semantic Vector Models and Functional Models for Pregroup Grammars*. *Journal of Logic, Language and Information* 20(4), pp. 419–443. Available at <http://dx.doi.org/10.1007/s10849-011-9132-2>.
- [20] Peter Selinger (2010): *Autonomous categories in which $A \simeq A^*$* .
- [21] Peter Selinger (2011): *A survey of graphical languages for monoidal categories*. In: *New Structures for Physics, Lecture Notes in Physics* 813, Springer, pp. 289–233.

Observational Equivalence Using Schedulers for Quantum Processes

Kazuya Yasuda

Takahiro Kubota

Yoshihiko Kakutani

Dept. of Computer Science
Graduate School of Information Science and Technology
The University of Tokyo
Tokyo, Japan

{kyasuda, takahiro.k11_30, kakutani}@is.s.u-tokyo.ac.jp

In the study of quantum process algebras, researchers have introduced different notions of equivalence between quantum processes like bisimulation or barbed congruence. However, there are intuitively equivalent quantum processes that these notions do not regard as equivalent. In this paper, we introduce a notion of equivalence named observational equivalence into qCCS. Since quantum processes have both probabilistic and nondeterministic transitions, we introduce schedulers that solve nondeterministic choices and obtain probability distribution of quantum processes. By definition, the restrictions of schedulers change observational equivalence. We propose some definitions of schedulers, and investigate the relation between the restrictions of schedulers and observational equivalence.

1 Introduction

Quantum communication protocols have been proposed since Bennett and Brassard [2] proposed a quantum key distribution (QKD) protocol. However, proving the correctness or security of communication protocols is very complicated and error-prone because quantum mechanical behavior is often different from our intuition based on classical mechanics. In order to analyze or verify quantum protocols successfully, quantum process calculi have been proposed, for example, QPAlg [8], CQP [6], and qCCS [4,5,12].

In quantum process calculi, it is one of the important notions whether two processes behave similarly or not, in other words, whether they are behavioral equivalent or not. One of the benefits of this notion is to provide the following technique to verify the correctness of a communication protocol. First, write a process that models the procedure of the communication protocol. Second, define a simpler process that is the specification of the protocol. Then, if these two processes are behavioral equivalent, it is proved that the protocol satisfies the specification. The notion of behavioral equivalence is defined in the process calculi mentioned above. Moreover, there is a variety of the notions of behavioral equivalence such as (weak) bisimulation and barbed congruence. For example, these notions on qCCS are defined in [3,5]. Intuitively, bisimulation is the notion that one process can simulate the other's behavior, and barbed congruence is the notion that any observers (or attackers) cannot distinguish two processes. In addition, if these two notions are defined properly, they become equivalent.

These notions of bisimulation and barbed congruence have widely been used in formal verification of processes. However, there are some processes that are not regarded as equivalent by these notions but intuitively equivalent. This problem occurs when the processes include quantum operations or communication. For example, consider the following two processes: one sends a qubit $|0\rangle$ or $|1\rangle$ with the same probability, the other sends $|+\rangle$ or $|-\rangle$ with the same probability. These two processes are not regarded as equivalent by the notion of bisimulation. However, we have intuitively regarded these two processes

as the same process because these qubits are expressed as the same density matrix. This kind of equation was used in the security proof of BB84 by Shor and Preskill [11].

The aim of this paper is to define the notion of equivalence that regards above cases as equivalent into the quantum process calculus qCCS. This notion is called observational equivalence. Intuitively, two processes are observational equivalent when they are observed the same by any attackers. Because attackers can observe their behavior only by watching the channels that they use, processes are observed the same when they use the channels with the same probability. In addition, we must consider the probability of using channels although the quantum processes of qCCS have both probabilistic and nondeterministic transitions. In order to solve this inconvenience, we define schedulers that solve nondeterministic choices and obtain probability distribution of quantum processes. By definition, the restrictions of schedulers change observational equivalence. We propose some definitions of schedulers, and investigate the relation between the restrictions of schedulers and observational equivalence.

2 Definitions of qCCS

In this section, we introduce the language qCCS proposed in [4, 5, 12].

2.1 Syntax

Three types of data are considered in qCCS: `Bool` for booleans, `Real` for real numbers and `Qbt` for qubits. Let $cVar$ be the set of classical variables, ranged over by x, y, \dots , and $qVar$ be the set of quantum variables, ranged over by q, r, \dots . We assume that $cVar$ and $qVar$ are both countably infinite and $cVar \cap qVar = \emptyset$. The indexed set $\{q_1, \dots, q_n\}$ is often abbreviated to \tilde{q} . Let Exp be the set of classical data expressions over `Real`, ranged over by e, e', \dots , which includes $cVar$ as a subset. Let $BExp$ be the set of boolean-valued expressions, ranged over by b, b', \dots .

Two types of channels are used in qCCS: $cChan$ for classical channels and $qChan$ for quantum channels. c, d, \dots range over $cChan$ and c, d, \dots range over $qChan$. We assume that $cChan \cap qChan = \emptyset$. Let $Chan$ be the set of all channels, that is, $Chan = cChan \cup qChan$. A *relabeling function* is a function $f : Chan \rightarrow Chan$ such that $f(cChan) \subset cChan$ and $f(qChan) \subset qChan$.

The set of *quantum processes* $qProc$ is defined inductively as follows:

$$\begin{aligned} qProc \ni P, Q ::= & \mathbf{nil} \mid A(\tilde{q}; \tilde{x}) \mid \tau.P \mid c?x.P \mid c!e.P \mid c?q.P \mid c!q.P \mid \mathcal{E}[\tilde{q}].P \mid M[\tilde{q}; x].P \mid \\ & P + Q \mid P \parallel Q \mid P[f] \mid P \setminus L \mid \mathbf{if } b \mathbf{ then } P \end{aligned}$$

where $c \in cChan$, $x \in cVar$, $e \in Exp$, $c \in qChan$, $b \in BExp$, $q \in qVar$, $A(\tilde{q}; \tilde{x})$ is a process constant, τ is the silent action, f is a relabeling function, $L \subset_{\text{fin}} Chan$, \mathcal{E} and M are respectively a trace-preserving super-operator and a non-degenerate projective measurement applying on the Hilbert space associated with the systems \tilde{q} . The process \mathbf{nil} may be omitted, for instance, $c!0$ is used instead of $c!0.\mathbf{nil}$.

The *free classical variable function* $fv : qProc \rightarrow 2^{cVar}$ is defined in the usual way. Note that the quantum measurement $M[\tilde{q}; x]$ binds the variable x , that is, $fv(M[\tilde{q}; x].P) = fv(P) - \{x\}$. A process P is *closed* if $fv(P) = \emptyset$. The *free quantum variable function* $qv : qProc \rightarrow 2^{qVar}$ is defined inductively as in Figure 1. For quantum processes to be legal, we require that

1. $q \notin qv(P)$ in the process $c!q.P$;
2. $qv(P) \cap qv(Q) = \emptyset$ in the process $P \parallel Q$;

$ \begin{aligned} qv(\mathbf{nil}) &= \emptyset \\ qv(A(\tilde{q}; \tilde{x})) &= \tilde{q} \\ qv(\tau.P) &= qv(P) \\ qv(c?x.P) &= qv(P) \\ qv(c!e.P) &= qv(P) \end{aligned} $	$ \begin{aligned} qv(c?q.P) &= qv(P) - \{q\} \\ qv(c!q.P) &= qv(P) \cup \{q\} \\ qv(\mathcal{E}[\tilde{q}].P) &= qv(P) \cup \tilde{q} \\ qv(M[\tilde{q}; x].P) &= qv(P) \cup \tilde{q} \\ qv(P+Q) &= qv(P) \cup qv(Q) \end{aligned} $	$ \begin{aligned} qv(P Q) &= qv(P) \cup qv(Q) \\ qv(P[f]) &= qv(P) \\ qv(P \setminus L) &= qv(P) \\ qv(\mathbf{if } b \mathbf{ then } P) &= qv(P) \end{aligned} $
--	---	--

Figure 1: Definition of qv

3. each process constant $A(\tilde{q}; \tilde{x})$ has a defining equation $A(\tilde{q}; \tilde{x}) := P$, where $P \in qProc$, $qv(P) \subset \tilde{q}$ and $fv(P) \subset \tilde{x}$.

We use $P\{v/x\}$ to denote the substitution of v for x in P . We abbreviate $P\{v_1/x_1\} \dots \{v_n/x_n\}$ to $P\{\tilde{v}/\tilde{x}\}$.

2.2 Configuration

For each $q \in qVar$, we assume a 2-dimensional Hilbert space \mathcal{H}_q to be the state space associated with the system q . Let

$$\mathcal{H}_S = \bigotimes_{q \in S} \mathcal{H}_q$$

for any $S \subset qVar$. In particular, $\mathcal{H} = \mathcal{H}_{qVar}$ is the whole state space associated with all of the quantum variables.

A *configuration* is a pair $\langle P, \rho \rangle$, where $P \in qProc$ is closed and ρ is a density operator on \mathcal{H} . Let Con be the set of all configurations, ranged over by C, D, \dots . If the state associated with the system q is $|\psi\rangle\langle\psi|$, the notation $|\psi\rangle\langle\psi|_q \otimes \rho$ or $[|\psi\rangle]_q \otimes \rho$ is used to denote this whole state, where ρ is a state associated with the systems $qVar - \{q\}$.

Let $D(Con)$ be the set of finite-support probability distribution over Con , ranged over by μ, ν, \dots . When $\mu(C) = 1$ for some $C \in Con$, we use C instead of μ to denote the distribution. We sometimes use a form $\mu = \boxplus_{i \in I} p_i \bullet C_i$ to denote the distribution μ , where C_i are distinct elements of Con and $\mu(C_i) = p_i$. For any $\mu = \boxplus_{i \in I} p_i \bullet \langle P_i, \rho_i \rangle$ and trace-preserving super-operator \mathcal{E} , the notation $\boxplus_{i \in I} p_i \bullet \langle P_i, \mathcal{E}(\rho_i) \rangle$ is often abbreviated to $\mathcal{E}(\mu)$.

2.3 Operational semantics

Let $Act = \{\tau\} \cup \{c?v, c!v \mid c \in cChan, v \in \mathbf{Real}\} \cup \{c?r, c!r \mid c \in qChan, r \in qVar\}$. For each $\alpha \in Act$, let $cn(\alpha)$ be the set of channel names used in the action α , that is, $cn(\tau) = \emptyset$, $cn(c?v) = cn(c!v) = \{c\}$ and $cn(c?r) = cn(c!r) = \{c\}$. For each $\alpha \in Act$ and relabeling function f , we use $f(\alpha)$ to denote the action of which channel is relabeled by f . For example, $f(\tau) = \tau$, $f(c?v) = f(c)?v$ and $f(c!q) = f(c)!q$.

The operational semantics of qCCS is defined by the probabilistic labeled transition system [3] $(Con, Act, \longrightarrow)$, where $\longrightarrow \subset Con \times Act \times D(Con)$ is the smallest relation satisfying the rules defined in Figure 2 (the symmetric forms for rules C-COM, Q-COM, INP-INT, OTH-INT and SUM are omitted). Here, $\llbracket e \rrbracket$ and $\llbracket b \rrbracket$ are the usual interpretations of $e \in Exp$ and $b \in BExp$ respectively, and $\mathcal{E}_{\tilde{q}}$ means that the super-operator \mathcal{E} applies on the state associated with the systems \tilde{q} . We write $C \xrightarrow{\alpha} \mu$ instead of $(C, \alpha, \mu) \in \longrightarrow$. We write $C \xrightarrow{\alpha}$ when there exists $\mu \in D(Con)$ such that $C \xrightarrow{\alpha} \mu$. We write $C \not\xrightarrow{\alpha}$ when there do not exist α and μ such that $C \xrightarrow{\alpha} \mu$.

The transition relation \longrightarrow is lifted to $D(Con) \times Act \times D(Con)$ as follows: we write $\mu \xrightarrow{\alpha} \nu$ if for any $C \in \text{supp}(\mu)$, $C \xrightarrow{\alpha} \nu_C$ for some ν_C , and $\nu = \sum_{C \in \text{supp}(\mu)} \mu(C) \nu_C$.

$\frac{}{\langle \tau.P, \rho \rangle \xrightarrow{\tau} \langle P, \rho \rangle}$	(TAU)	$\frac{\langle P_1, \rho \rangle \xrightarrow{c?r} \langle P'_1, \rho \rangle, \quad r \notin qv(P_2)}{\langle P_1 \parallel P_2, \rho \rangle \xrightarrow{c?r} \langle P'_1 \parallel P_2, \rho \rangle}$	(INP-INT)
$\frac{v \in \mathbf{Real}}{\langle c?v.P, \rho \rangle \xrightarrow{c?v} \langle P\{v/x\}, \rho \rangle}$	(C-INP)	$\frac{\langle P_1, \rho \rangle \xrightarrow{\alpha} \boxplus_{i \in I} p_i \bullet \langle P'_i, \rho_i \rangle, \quad \alpha \neq c?r}{\langle P_1 \parallel P_2, \rho \rangle \xrightarrow{\alpha} \boxplus_{i \in I} p_i \bullet \langle P'_i \parallel P_2, \rho_i \rangle}$	(OTH-INT)
$\frac{v = \llbracket e \rrbracket}{\langle c!e.P, \rho \rangle \xrightarrow{c!v} \langle P, \rho \rangle}$	(C-OUTP)	$\frac{\langle P, \rho \rangle \xrightarrow{\alpha} \mu}{\langle P + Q, \rho \rangle \xrightarrow{\alpha} \mu}$	(SUM)
$\frac{\langle P_1, \rho \rangle \xrightarrow{c?v} \langle P'_1, \rho \rangle, \quad \langle P_2, \rho \rangle \xrightarrow{c!v} \langle P'_2, \rho \rangle}{\langle P_1 \parallel P_2, \rho \rangle \xrightarrow{\tau} \langle P'_1 \parallel P'_2, \rho \rangle}$	(C-COM)	$\frac{\langle P, \rho \rangle \xrightarrow{\alpha} \boxplus_{i \in I} p_i \bullet \langle P_i, \rho_i \rangle}{\langle P[f], \rho \rangle \xrightarrow{f(\alpha)} \boxplus_{i \in I} p_i \bullet \langle P_i[f], \rho_i \rangle}$	(REL)
$\frac{r \notin qv(c?q.P)}{\langle c?q.P, \rho \rangle \xrightarrow{c?r} \langle P\{r/q\}, \rho \rangle}$	(Q-INP)	$\frac{\langle P, \rho \rangle \xrightarrow{\alpha} \boxplus_{i \in I} p_i \bullet \langle P_i, \rho_i \rangle, \quad cn(\alpha) \cap L = \emptyset}{\langle P \setminus L, \rho \rangle \xrightarrow{\alpha} \boxplus_{i \in I} p_i \bullet \langle P_i \setminus L, \rho_i \rangle}$	(RES)
$\frac{}{\langle c!q.P, \rho \rangle \xrightarrow{c!q} \langle P, \rho \rangle}$	(Q-OUTP)	$\frac{\langle P, \rho \rangle \xrightarrow{\alpha} \mu, \quad \llbracket b \rrbracket = \mathbf{true}}{\langle \mathbf{if } b \mathbf{ then } P, \rho \rangle \xrightarrow{\alpha} \mu}$	(CHO)
$\frac{\langle P_1, \rho \rangle \xrightarrow{c?r} \langle P'_1, \rho \rangle, \quad \langle P_2, \rho \rangle \xrightarrow{c!r} \langle P'_2, \rho \rangle}{\langle P_1 \parallel P_2, \rho \rangle \xrightarrow{\tau} \langle P'_1 \parallel P'_2, \rho \rangle}$	(Q-COM)	$\frac{\langle P\{\tilde{r}/\tilde{q}\}\{\tilde{v}/\tilde{x}\}, \rho \rangle \xrightarrow{\alpha} \mu, \quad A(\tilde{q}; \tilde{x}) := P}{\langle A(\tilde{r}; \tilde{v}), \rho \rangle \xrightarrow{\alpha} \mu}$	(DEF)
$\frac{\langle \mathcal{E}[\tilde{q}].P, \rho \rangle \xrightarrow{\tau} \langle P, \mathcal{E}_{\tilde{q}}(\rho) \rangle}{M = \sum_{i \in I} \lambda_i E^i \quad p_i = \text{tr}(E^i_{\tilde{q}} \rho)}$	(OPER)		
$\frac{}{\langle M[\tilde{q}; x].P, \rho \rangle \xrightarrow{\tau} \sum_{i \in I \wedge p_i \neq 0} p_i \langle P\{\lambda_i/x\}, E^i_{\tilde{q}} \rho E^i_{\tilde{q}}/p_i \rangle}$	(MEAS)		

Figure 2: Transition rules of qCCS

3 Bisimulation

In this section, we recall the relation called open bisimulation. To define it, we need to define the relation \Longrightarrow and a weight function. These definitions are introduced in [5].

Definition 1. The relation $\Longrightarrow \subset D(\text{Con}) \times D(\text{Con})$ is the smallest relation satisfying the following conditions:

1. $C \Longrightarrow C$;
2. if $C \xrightarrow{\tau} \mu$ and $\mu \Longrightarrow v$, then $C \Longrightarrow v$;
3. if $\mu = \sum_{i \in I} p_i C_i$, and for any $i \in I$, $C_i \Longrightarrow v_i$ for some v_i , then $\mu \Longrightarrow \sum_{i \in I} p_i v_i$.

For any $\mu, v \in D(\text{Con})$ and $s = \alpha_1 \dots \alpha_n \in \text{Act}^*$, we say that μ can evolve into v by a weak s -transition, denoted by $\mu \xrightarrow{s} v$, if there exist $\mu_1, \dots, \mu_{n+1}, v_1, \dots, v_n \in D(\text{Con})$, such that $\mu \Longrightarrow \mu_1$, $\mu_{n+1} = v$, and for each $i = 1, \dots, n$, $\mu_i \xrightarrow{\alpha_i} v_i$ and $v_i \Longrightarrow \mu_{i+1}$.

For any $s \in \text{Act}^*$, \hat{s} is the string obtained from s by deleting all the occurrences of τ .

Definition 2. Let $\mathcal{R} \subset \text{Con} \times \text{Con}$ and $\mu, v \in D(\text{Con})$. A weight function for (μ, v) w.r.t. \mathcal{R} is a function $\delta : \text{Con} \times \text{Con} \rightarrow [0, 1]$ that satisfies the following conditions:

1. for all $C, D \in \text{Con}$,

$$\sum_{D' \in \text{supp}(v)} \delta(C, D') = \mu(C), \quad \sum_{C' \in \text{supp}(\mu)} \delta(C', D) = v(D);$$

2. for all $C, D \in \text{Con}$, if $\delta(C, D) > 0$, then $C \mathcal{R} D$.

We write $\mu \mathcal{R} \nu$ if there exists a weight function for (μ, ν) w.r.t. \mathcal{R} .

Lemma 1. *Let $\mu, \nu \in D(\text{Con})$. Then $\mu \mathcal{R} \nu$ if and only if there exist $\{p_i\}_{i \in I}$, $\{C_i\}_{i \in I}$, and $\{D_i\}_{i \in I}$ such that $\mu = \sum_{i \in I} p_i C_i$, $\nu = \sum_{i \in I} p_i D_i$, and $C_i \mathcal{R} D_i$ for each $i \in I$. In particular, if $C \mathcal{R} \mu$ then $C \mathcal{R} D$ for each $D \in \text{supp}(\mu)$.*

Now we introduce open bisimulation on qCCS defined in [3].

Definition 3. A relation $\mathcal{R} \subset \text{Con} \times \text{Con}$ is an *open bisimulation* if $\langle P, \rho \rangle \mathcal{R} \langle Q, \sigma \rangle$ implies that $qv(P) = qv(Q)$, $\text{tr}_{qv(P)}(\rho) = \text{tr}_{qv(Q)}(\sigma)$, and for any super-operator \mathcal{E} acting on $\mathcal{H}_{qv(P)}$,

1. whenever $\langle P, \mathcal{E}(\rho) \rangle \xrightarrow{\alpha} \mu$, there exists ν such that $\langle Q, \mathcal{E}(\sigma) \rangle \xrightarrow{\hat{\alpha}} \nu$ and $\mu \mathcal{R} \nu$;
2. whenever $\langle Q, \mathcal{E}(\sigma) \rangle \xrightarrow{\alpha} \nu$, there exists μ such that $\langle P, \mathcal{E}(\rho) \rangle \xrightarrow{\hat{\alpha}} \mu$ and $\mu \mathcal{R} \nu$.

Let \approx_o be the largest open bisimulation.

There are other notions of equivalence like open bisimulation on qCCS. For example, *bisimulation* is defined in [5] and *reduction barbed congruence* is defined in [3]. According to [3], the largest open bisimulation is strictly coarser than the largest bisimulation, and the reduction barbed congruence coincides with the largest open bisimulation.

4 Observational equivalence

In this section, we introduce the notion of observational equivalence on qCCS. Intuitively, two configurations are observationally equivalent when they are observed by foreign processes in the same way, in other words, when they use the same channels with the same probability in any contexts.

First of all, we describe why we want to define the notion of observational equivalence with an example. There are two different ways to express quantum measurements in qCCS: $M[q; x]$ and $\mathcal{E}[q]$, where M is the 1-qubit projective measurement such that $M = \sum_{i=0}^1 |i\rangle \langle i|$, \mathcal{E} is the trace-preserving super-operator such that $\mathcal{E}(\rho) = \sum_{i=0}^1 |i\rangle \langle i| \rho |i\rangle \langle i|$. We intuitively want to consider that these two processes are equivalent, but they are not bisimilar. This gap is an obstacle to formalize Shor and Preskill's security proof of BB84 [10]. For simplicity, we consider the following example.

Example 1. Consider these two configurations:

$$C = \langle M[q; x].(c!0 + d!0), [!+]_q \otimes \rho \rangle, \quad D = \langle \mathcal{E}[q].(c!0 + d!0), [!+]_q \otimes \rho \rangle$$

where M and \mathcal{E} are described above. The pLTSs for these configurations are depicted as in Figure 3. It is obvious that $C \not\approx_o D$. We want to consider that C and D are equivalent.

4.1 Scheduler

Even though quantum processes on qCCS have both probabilistic and nondeterministic transitions, we have to consider a probability to use channels in order to define observational equivalence. So, we define schedulers to solve nondeterministic choices and to obtain probability distribution of configurations.

Definition 4. A function $F : \text{Con} \rightarrow (\text{Act} \times D(\text{Con})) \cup \{\perp\}$ is a *scheduler* if the following conditions are satisfied:

1. $F(C) = (\alpha, \mu)$ implies $C \xrightarrow{\alpha} \mu$,
2. $F(C) = \perp$ implies $C \not\rightarrow$.

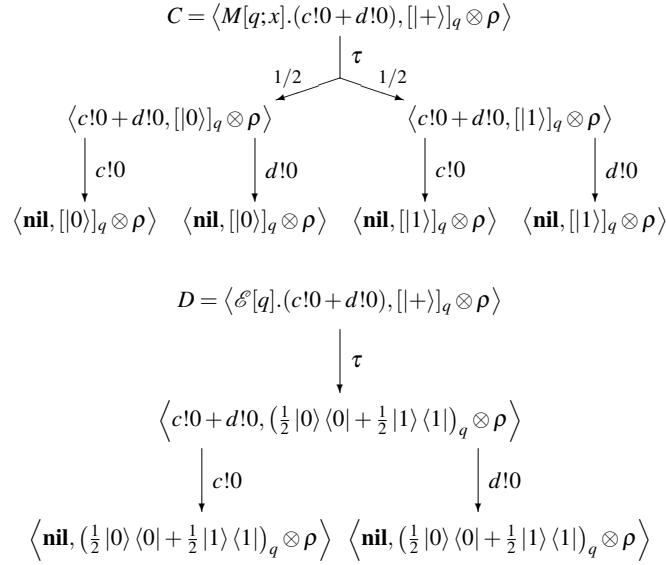


Figure 3: pLTSs for Example 1

We write $C \xrightarrow{\alpha}_F \mu$ when $F(C) = (\alpha, \mu)$. We write $C \xrightarrow{\alpha}_F$ when $F(C) = (\alpha, \mu)$ for some $\mu \in D(\text{Con})$.

The relation \Longrightarrow is limited by a scheduler as follows:

Definition 5. The relation $\Longrightarrow_F \subset D(\text{Con}) \times D(\text{Con})$ is the smallest relation satisfying the following conditions:

1. $C \Longrightarrow_F C$;
2. if $C \xrightarrow{\tau}_F \mu$ and $\mu \Longrightarrow_F \nu$, then $C \Longrightarrow_F \nu$;
3. if $\mu = \sum_{i \in I} p_i C_i$, and for any $i \in I$, $C_i \Longrightarrow_F \nu_i$ for some ν_i , then $\mu \Longrightarrow_F \sum_{i \in I} p_i \nu_i$.

4.2 Observational equivalence

We write $C \Downarrow_F^p c$ when there exists $\mu \in D(\text{Con})$ such that

- $C \Longrightarrow_F \mu$ holds;
- for each $C' \in \text{supp}(\mu)$, either $F(C') = \perp$ or $C' \xrightarrow{\lambda}_F$ holds for some $\lambda \neq \tau$; and
- the equation $\sum \{ \mu(C') \mid C' \xrightarrow{c!v}_F \text{ for some } v \} = p$ holds.

This means, intuitively, that the configuration C uses the channel c with the probability p after all internal transitions in accordance with the scheduler F .

Now, we define observational equivalence on qCCS.

Definition 6. Two configurations $\langle P, \rho \rangle, \langle Q, \sigma \rangle \in \text{Con}$ are *observationally equivalent*, we write $\langle P, \rho \rangle \approx_{oe} \langle Q, \sigma \rangle$, if $qv(P) = qv(Q)$, $\text{tr}_{qv(P)}(\rho) = \text{tr}_{qv(Q)}(\sigma)$ and for any quantum processes $R \in q\text{Proc}$,

1. for each scheduler F there exists a scheduler F' such that, for any classical channel $c \in c\text{Chan}$

$$\langle P \parallel R, \rho \rangle \Downarrow_F^p c \text{ implies that } \langle Q \parallel R, \sigma \rangle \Downarrow_{F'}^p c;$$

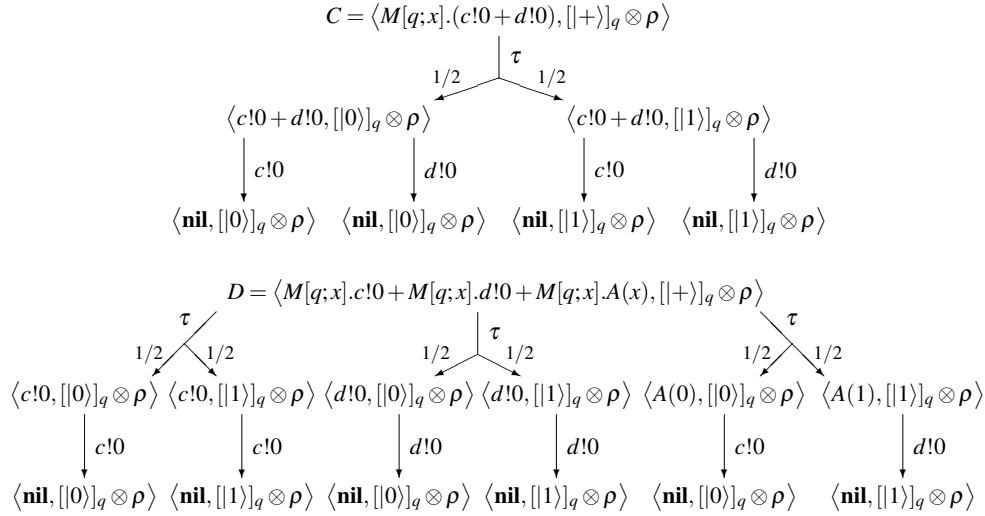


Figure 4: pLTSs for Example 2

2. for each scheduler F there exists a scheduler F' such that, for any classical channel $c \in cChan$ $\langle Q || R, \sigma \rangle \Downarrow_F^P c$ implies that $\langle P || R, \rho \rangle \Downarrow_{F'}^P c$.

We can prove that \approx_{oe} is an equivalence relation easily.

For example, we show two configurations that are not equivalent in the notion of open bisimulation but observationally equivalent.

Example 2. Consider these two configurations:

$$C = \langle M[q;x].(c!0 + d!0), [[+]]_q \otimes \rho \rangle,$$

$$D = \langle M[q;x].c!0 + M[q;x].d!0 + M[q;x].A(x), [[+]]_q \otimes \rho \rangle$$

where $A(x) := (\text{if } x = 0 \text{ then } c!0) + (\text{if } x = 1 \text{ then } d!0)$ and M is as defined in Example 1. The pLTSs for these configurations are depicted as in Figure 4. It is obvious that $C \not\approx_o D$. However, we can prove that $C \approx_{oe} D$.

Proposition 1. *Let C, D be the configurations in Example 2. Then $C \approx_{oe} D$.*

Proof. Let $P = M[q;x].(c!0 + d!0)$ and $Q = M[q;x].c!0 + M[q;x].d!0 + M[q;x].A(x)$.

We have $qv(P) = qv(Q) = \{q\}$ and $\text{tr}_{qv(P)}([+])_q \otimes \rho = \text{tr}_{qv(Q)}([+])_q \otimes \rho = \rho$.

Let R be an arbitrary quantum process. First, we need to show that, for each scheduler F , there exists a scheduler F' such that, for any classical channel c $\langle P || R, [[+]]_q \otimes \rho \rangle \Downarrow_F^P c$ implies $\langle Q || R, [[+]]_q \otimes \rho \rangle \Downarrow_{F'}^P c$. To prove it, we divide several cases of the scheduler F and construct a scheduler F' in each case.

1. The scheduler F does not choose the τ transition caused by P . In this case, we can easily construct a scheduler F' such that $\langle Q || R, [[+]]_q \otimes \rho \rangle \Downarrow_{F'}^P c$.
2. The scheduler F chooses the τ transition caused by P . In this case, we have

$$\langle P || R, [[+]]_q \otimes \rho \rangle \Rightarrow_F \boxplus_{i \in I} \left(\frac{1}{2} p_i \bullet \langle c!0 + d!0 || R_i, [[0]]_q \otimes \rho_i \rangle \boxplus \frac{1}{2} p_i \bullet \langle c!0 + d!0 || R_i, [[1]]_q \otimes \rho_i \rangle \right)$$

after all τ transitions caused by P and R independently in accordance with F . Then, there exists a scheduler F' such that

$$\langle Q || R, [[+]]_q \otimes \rho \rangle \Rightarrow_{F'} \boxplus_{i \in I} p_i \bullet \langle Q || R_i, [[+]]_q \otimes \rho_i \rangle.$$

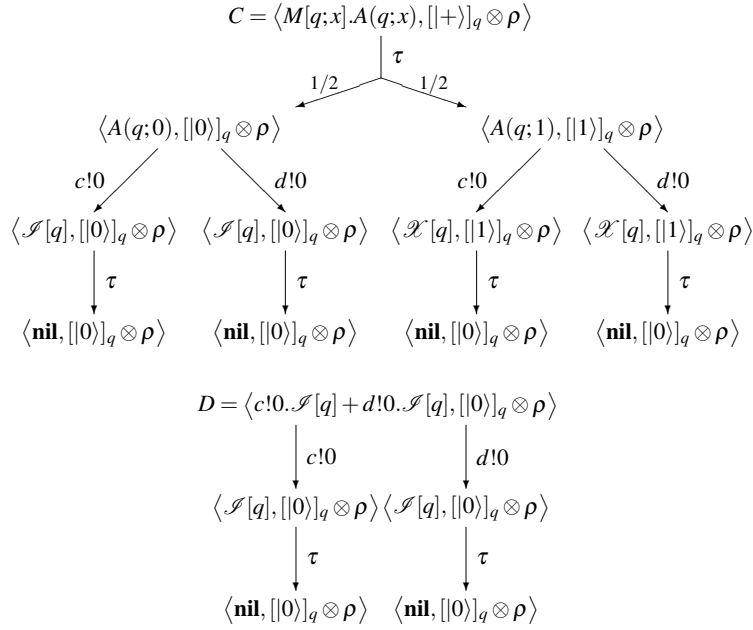


Figure 5: pLTSs for Example 3

For each $i \in I$, we again divide some cases of F and construct F' in each cases. Here we show only one case and omit the others.

When

$$\langle c!0 + d!0 \mid R_i, [0] \rangle_q \otimes \rho_i \xrightarrow{c!0}_F \langle \mathbf{nil} \mid R_i, [0] \rangle_q \otimes \rho_i,$$

$$\langle c!0 + d!0 \mid R_i, [1] \rangle_q \otimes \rho_i \xrightarrow{c!0}_F \langle \mathbf{nil} \mid R_i, [1] \rangle_q \otimes \rho_i,$$

the channel c is used with the probability 1. So, we can construct a scheduler F' such that

$$\langle Q \mid R_i, [0] \rangle_q \otimes \rho_i \xrightarrow{\tau}_{F'} \frac{1}{2} \bullet \langle c!0 \mid R_i, [0] \rangle_q \otimes \rho_i \boxplus \frac{1}{2} \bullet \langle c!0 \mid R_i, [1] \rangle_q \otimes \rho_i,$$

$$\langle c!0 \mid R_i, [0] \rangle_q \otimes \rho_i \xrightarrow{c!0}_{F'} \langle \mathbf{nil} \mid R_i, [0] \rangle_q \otimes \rho_i, \quad \langle c!0 \mid R_i, [1] \rangle_q \otimes \rho_i \xrightarrow{c!0}_{F'} \langle \mathbf{nil} \mid R_i, [1] \rangle_q \otimes \rho_i.$$

The scheduler F' satisfies the requirement.

□

We show another example that means there exist configurations C, D that $C \not\approx_{oe} D$ but $C \approx_o D$.

Example 3. Consider these two configurations:

$$C = \langle M[q; x].A(q; x), [0] \rangle_q \otimes \rho, \quad D = \langle c!0.\mathcal{J}[q] + d!0.\mathcal{J}[q], [0] \rangle_q \otimes \rho$$

where

$$A(q; x) := (\text{if } x = 0 \text{ then } (c!0.\mathcal{J}[q] + d!0.\mathcal{J}[q])) + (\text{if } x = 1 \text{ then } (c!0.\mathcal{X}[q] + d!0.\mathcal{X}[q])),$$

\mathcal{J} is an operator that does nothing, \mathcal{X} is the Pauli-X operator, and M is as defined in Example 1. The pLTSs for these configurations are depicted as in Figure 5.

We can prove that $C \approx_o D$. However, $C \not\approx_{oe} D$. Consider a scheduler F such that

$$F(\langle A(q;0), [0]_q \otimes \rho \rangle) = (c!0, \langle \mathcal{J}[q], [0]_q \otimes \rho \rangle),$$

$$F(\langle A(q;1), [1]_q \otimes \rho \rangle) = (d!0, \langle \mathcal{X}[q], [1]_q \otimes \rho \rangle).$$

Then both $C \Downarrow_F^{1/2} c$ and $C \Downarrow_F^{1/2} d$ hold. But, for any schedulers F' , neither $D \Downarrow_{F'}^{1/2} c$ nor $D \Downarrow_{F'}^{1/2} d$ holds.

Proposition 2. \approx_o and \approx_{oe} are incomparable.

4.3 Strategy: a limited scheduler

In previous section, we define schedulers and the observational equivalence. However, the processes in Example 1 are not observationally equivalent. Consider a scheduler F such that

$$F(\langle (c!0 + d!0), [0]_q \otimes \rho \rangle) = (c!0, \langle \mathbf{nil}, [0]_q \otimes \rho \rangle),$$

$$F(\langle (c!0 + d!0), [1]_q \otimes \rho \rangle) = (d!0, \langle \mathbf{nil}, [1]_q \otimes \rho \rangle).$$

Then both $C \Downarrow_F^{1/2} c$ and $C \Downarrow_F^{1/2} d$ hold. But, for any schedulers F' , neither $D \Downarrow_{F'}^{1/2} c$ nor $D \Downarrow_{F'}^{1/2} d$ holds.

This problem is due to the definition of schedulers, that is, because schedulers can choose different transitions even though the processes are the same. In order to solve this problem, we propose strategies, limited schedulers.

Definition 7. A function $F : \text{Con} \rightarrow (\text{Act} \times D(\text{Con})) \cup \{\perp\}$ is a *strategy* if the following conditions are satisfied:

1. $F(C) = (\alpha, \mu)$ implies $C \xrightarrow{\alpha} \mu$,
2. $F(C) = \perp$ implies $C \not\rightarrow$,
3. if $F(\langle P, \rho \rangle) = (\alpha, \mu)$, then there exist a set of processes $\{P_i\}_{i \in I}$, a set of super-operators $\{\mathcal{E}_i\}_{i \in I}$, acting on $\mathcal{H}_{qv(P)}$, and a set of projectors $\{E_i\}_{i \in I}$, acting on $\mathcal{H}_{qv(P)}$ and $\sum_{i \in I} E_i = I$, such that for any density operators σ ,

$$F(\langle P, \sigma \rangle) = \left(\alpha, \sum_{i \in I \wedge q_i^\sigma \neq 0} q_i^\sigma \langle P_i, \mathcal{E}_i(\sigma) / q_i^\sigma \rangle \right)$$

and

$$\mu = \sum_{i \in I \wedge q_i^\rho \neq 0} q_i^\rho \langle P_i, \mathcal{E}_i(\rho) / q_i^\rho \rangle$$

where $q_i^\sigma = \text{tr}(E_i \sigma)$.

The difference between schedulers and strategies is only the condition 3 in Definition 7. This condition means that strategies must choose the same transition for any density operators if the processes of the configurations are the same. In order to validate this condition, we use the following lemma. This lemma is stronger than Lemma 3.3 (2) in [5], but can still be easily observed from the transition rules of qCCS.

Lemma 2. If $\langle P, \rho \rangle \xrightarrow{\alpha} \mu$, then there exists a set of processes $\{P_i\}_{i \in I}$, a set of super-operators $\{\mathcal{E}_i\}_{i \in I}$, acting on $\mathcal{H}_{qv(P)}$, and a set of projectors $\{E_i\}_{i \in I}$, acting on $\mathcal{H}_{qv(P)}$ and $\sum_{i \in I} E_i = I$, such that for any density operators σ ,

$$\langle P, \sigma \rangle \xrightarrow{\alpha} \sum_{i \in I \wedge q_i^\sigma \neq 0} q_i^\sigma \langle P_i, \mathcal{E}_i(\sigma) / q_i^\sigma \rangle,$$

and

$$\mu = \sum_{i \in I \wedge q_i^\rho \neq 0} q_i^\rho \langle P_i, \mathcal{E}_i(\rho) / q_i^\rho \rangle$$

where $q_i^\sigma = \text{tr}(E_i \sigma)$.

We use the notations $C \xrightarrow{\alpha}_F \mu$, $C \xrightarrow{\alpha}_F$ and \Rightarrow_F for strategies F in the same way as schedulers.

4.4 Observational equivalence with strategies

We write $C \Downarrow_F^p c$ for strategies F in the same way as schedulers. Now, we define observational equivalence using strategies instead of schedulers.

Definition 8. Two configurations $\langle P, \rho \rangle, \langle Q, \sigma \rangle \in \text{Con}$ are *observationally equivalent with strategies*, we write $\langle P, \rho \rangle \approx_{oe}^{st} \langle Q, \sigma \rangle$, if $qv(P) = qv(Q)$, $\text{tr}_{qv(P)}(\rho) = \text{tr}_{qv(Q)}(\sigma)$ and for any quantum processes $R \in qProc$,

1. for each strategy F there exists a strategy F' such that, for any classical channel $c \in cChan$ $\langle P || R, \rho \rangle \Downarrow_F^p c$ implies that $\langle Q || R, \sigma \rangle \Downarrow_{F'}^p c$;
2. for each strategy F there exists a strategy F' such that, for any classical channel $c \in cChan$ $\langle Q || R, \sigma \rangle \Downarrow_F^p c$ implies that $\langle P || R, \rho \rangle \Downarrow_{F'}^p c$.

We can prove that \approx_{oe}^{st} is an equivalence relation easily.

Now, we can check that the two configurations in Example 1 are observationally equivalent with strategies.

Proposition 3. Let C and D be configurations in Example 1. Then $C \approx_{oe}^{st} D$.

Let us consider the relation among open bisimulation \approx_o , observational equivalence \approx_{oe} , and observational equivalence with strategies \approx_{oe}^{st} .

By Example 1 and Proposition 3, there exist some configurations C and D that $C \not\approx_{oe} D$ but $C \approx_{oe}^{st} D$. However, $\approx_{oe} \subset \approx_{oe}^{st}$ does not hold. Consider Example 2 again. The configurations in Example 2 are observationally equivalent, but they are not observationally equivalent with strategies. Consider the strategy F such that

$$F(D) = \left(\tau, \frac{1}{2} \bullet \langle A(0), [0] \rangle_q \otimes \rho \right) \boxplus \frac{1}{2} \bullet \langle A(1), [1] \rangle_q \otimes \rho \Big),$$

$$F(\langle A(0), [0] \rangle_q \otimes \rho) = (c!0, \langle \mathbf{nil}, [0] \rangle_q \otimes \rho), \quad F(\langle A(1), [1] \rangle_q \otimes \rho) = (d!0, \langle \mathbf{nil}, [1] \rangle_q \otimes \rho).$$

Then both $D \Downarrow_F^{1/2} c$ and $D \Downarrow_F^{1/2} d$ are hold. However, neither $C \Downarrow_{F'}^{1/2} c$ nor $C \Downarrow_{F'}^{1/2} d$ holds for any strategies F' . It is because, for any strategies F' , if

$$F'(\langle c!0 + d!0, [0] \rangle_q \otimes \rho) = (\alpha_0, \mu_0), \quad F'(\langle c!0 + d!0, [1] \rangle_q \otimes \rho) = (\alpha_1, \mu_1),$$

then α_0 and α_1 must be the same action by the definition of strategies.

Proposition 4. \approx_{oe} and \approx_{oe}^{st} are incomparable.

In addition, $\approx_o \subset \approx_{oe}^{st}$ does not also hold, although there exist some configurations C and D that $C \not\approx_o D$ but $C \approx_{oe}^{st} D$. Consider Example 3 again. It is proved that $C \approx_o D$, but $C \not\approx_{oe}^{st} D$. Consider the strategy F such that

$$F(C) = \left(\tau, \frac{1}{2} \bullet \langle A(q; 0), [0]_q \otimes \rho \rangle \boxplus \frac{1}{2} \bullet \langle A(q; 1), [1]_q \otimes \rho \rangle \right),$$

$$F(\langle A(q; 0), [0]_q \otimes \rho \rangle) = (c!0, \langle \mathcal{J}[q], [0]_q \otimes \rho \rangle),$$

$$F(\langle A(q; 1), [1]_q \otimes \rho \rangle) = (d!0, \langle \mathcal{X}[q], [1]_q \otimes \rho \rangle).$$

Then both $C \Downarrow_F^{1/2} c$ and $C \Downarrow_F^{1/2} d$ are hold. However, neither $D \Downarrow_{F'}^{1/2} c$ nor $D \Downarrow_{F'}^{1/2} d$ holds for any strategies F' .

Proposition 5. \approx_o and \approx_{oe}^{st} are incomparable.

5 Related work

There already exists “observational equivalence” or “observational congruence” on other process calculi such as applied pi calculus [1] and probabilistic applied pi calculus [7]. However, they are essentially the same as reduction barbed congruence because they are reduction-closed by definition. So, they are also the same as the notion of open bisimulation.

The same notion of observational equivalence in this paper is defined on quantum applied pi calculus in [9]. However, quantum applied pi calculus takes into account only pure states as quantum states, so it cannot deal with mixed states.

6 Conclusion

In this paper, we proposed the notion of observational equivalence. To define it, we used schedulers that solve nondeterministic choices. Some processes that are not bisimilar became observationally equivalent, but others remained nonequivalent. And so, we defined strategies, which are limited schedulers, and the notion of observational equivalence with strategies. Some processes that are intuitively equivalent became observationally equivalent with strategies. After that, we investigated the relation among three notions, that is, open bisimulation \approx_o , observational equivalence \approx_{oe} , and observational equivalence with strategies \approx_{oe}^{st} , and we found that it is impossible to compare these three notions. In addition, we think that \approx_{oe}^{st} is the most intuitive of the three when we consider the situation like Example 1 or the formal security proof of BB84.

However, there remains a question whether our definition of observational equivalence is really intuitive. In order to solve this question, we must formalize the “intuition” at first. And then, we can discuss whether our definition of equivalence is intuitive or not.

We should also discuss the congruence of our observational equivalences. Congruence is the property that the equivalence is preserved under process constructs. The congruence property for parallel compositions $P \parallel R$, which are the most important case, holds by definition of our observational equivalences. However, the congruence property for channel restrictions $P \setminus L$ does not holds. It remains for future work to investigate whether they are preserved under other constructs or not.

References

- [1] Martín Abadi & Cédric Fournet (2001): *Mobile Values, New Names, and Secure Communication*. In: *Proceedings of the 28th ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL '01, ACM, New York, NY, USA, pp. 104–115, doi:10.1145/360204.360213.
- [2] Charles H. Bennett & Gilles Brassard (1984): *Quantum cryptography: Public key distribution and coin tossing*. In: *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, pp. 175–179.
- [3] Yuxin Deng & Yuan Feng (2012): *Open Bisimulation for Quantum Processes*. In JosC.M. Baeten, Tom Ball & FrankS. Boer, editors: *Theoretical Computer Science, Lecture Notes in Computer Science 7604*, Springer Berlin Heidelberg, pp. 119–133, doi:10.1007/978-3-642-33475-7_9.
- [4] Yuan Feng, Runyao Duan, Zhengfeng Ji & Mingsheng Ying (2007): *Probabilistic bisimulations for quantum processes*. *Information and Computation* 205(11), pp. 1608–1639, doi:10.1016/j.ic.2007.08.001.
- [5] Yuan Feng, Runyao Duan & Mingsheng Ying (2012): *Bisimulation for Quantum Processes*. *ACM Trans. Program. Lang. Syst.* 34(4), pp. 17:1–17:43, doi:10.1145/2400676.2400680.
- [6] Simon J. Gay & Rajagopal Nagarajan (2005): *Communicating Quantum Processes*. In: *Proceedings of the 32nd ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, POPL '05, ACM, New York, NY, USA, pp. 145–157, doi:10.1145/1040305.1040318.
- [7] Jean Goubault-Larrecq, Catuscia Palamidessi & Angelo Troina (2007): *A Probabilistic Applied Pi-Calculus*. In Zhong Shao, editor: *Programming Languages and Systems, Lecture Notes in Computer Science 4807*, Springer Berlin Heidelberg, pp. 175–190, doi:10.1007/978-3-540-76637-7_12.
- [8] Philippe Jorrand & Marie Lalire (2004): *Toward a Quantum Process Algebra*. In: *Proceedings of the 1st Conference on Computing Frontiers*, CF '04, ACM, New York, NY, USA, pp. 111–119, doi:10.1145/977091.977108.
- [9] Takahiro Kubota (2011): *Formalization and automation of unconditional security proof of QKD*. Master's thesis, The University of Tokyo.
- [10] Takahiro Kubota, Yoshihiko Kakutani, Go Kato, Yasuhito Kawano & Hideki Sakurada (2012): *Application of a Process Calculus to Security Proofs of Quantum Protocols*. In: *Proceedings of Foundations of Computer Science in WORLDCOMP*, pp. 141–147.
- [11] Peter W. Shor & John Preskill (2000): *Simple Proof of Security of the BB84 Quantum Key Distribution Protocol*. *Physical Review Letters* 85(2), pp. 441–444, doi:10.1103/PhysRevLett.85.441.
- [12] Mingsheng Ying, Yuan Feng, Runyao Duan & Zhengfeng Ji (2009): *An Algebra of Quantum Processes*. *ACM Transactions on Computational Logic* 10(3), pp. 19:1–19:36, doi:10.1145/1507244.1507249.

Equivalence of wave-particle duality to entropic uncertainty

Patrick J. Coles, Jędrzej Kaniewski, and Stephanie Wehner

Centre for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117543 Singapore

(Dated: April 13, 2014)

Interferometers capture a basic mystery of quantum mechanics: a single quantum particle can exhibit wave behavior, yet that wave behavior disappears when one tries to determine the particle's path inside the interferometer. This idea has been formulated quantitatively as an inequality, e.g., by Englert and Jaeger, Shimony, and Vaidman, which upper bounds the sum of the interference visibility and the path distinguishability. Such wave-particle duality relations (WPDRs) are often thought to be conceptually inequivalent to Heisenberg's uncertainty principle, although this has been debated. Here we show that WPDRs correspond precisely to a modern formulation of the uncertainty principle in terms of entropies, namely the min- and max-entropies. This observation unifies two fundamental concepts in quantum mechanics. Furthermore, it leads to a robust framework for deriving novel WPDRs by applying entropic uncertainty relations to interferometric models. As an illustration, we derive a novel relation that captures the coherence in a quantum beam splitter.

When Feynman discussed the two-path interferometer [1], he noted that quantum systems (quantons) display the behavior of both waves and particles and that there is a competition between seeing the wave behavior versus the particle behavior. When the observer tries harder to figure out which path of the interferometer the quanton takes, the wave-like interference becomes less visible. This tradeoff is commonly called wave-particle duality (WPD). Feynman also said this is "a phenomenon which is impossible ... to explain in any classical way, and which has in it the heart of quantum mechanics. In reality, it contains the only mystery [of quantum mechanics]."

Many quantitative statements of this idea, so-called wave-particle duality relations (WPDRs), have been formulated [2–10]. Such relations typically consider the Mach-Zehnder interferometer for single photons, see Fig. 1. For example, a well-known formulation proven independently by Englert [2] and Jaeger et al. [3] quantifies the wave behavior by fringe visibility \mathcal{V} , and particle behavior by the distinguishability of the photon's path, \mathcal{D} . (See below for precise definitions; the idea is that "waves" have a definite phase, while "particles" have a definite location, hence \mathcal{V} and \mathcal{D} respectively quantify how definite the phase and location are inside the interferometer.) They found the tradeoff:

$$\mathcal{D}^2 + \mathcal{V}^2 \leq 1 \quad (1)$$

which implies $\mathcal{V} = 0$ when $\mathcal{D} = 1$ (full particle behavior means no wave behavior) and vice-versa, and also treats the intermediate case of partial distinguishability.

It has been debated [11–13] whether the WPD principle, sometimes referred to as Bohr's complementarity principle [14], is equivalent to another fundamental quantum idea with no classical analog: Heisenberg's uncertainty principle. The latter states that there are certain pairs of observables, such as position and momentum or two orthogonal components of spin, that cannot simultaneously be known or jointly measured. Likewise many uncertainty relations have been proven and modern ones

typically use entropy instead of standard deviation as the uncertainty measure, so-called *entropic* uncertainty relations (EURs) [15]. This is because the standard deviation formulation suffers from trivial bounds when applied to finite-dimensional systems, whereas the entropic formulation not only fixes this weakness but also implies Heisenberg's standard deviation relation [16] and has relevance to information-processing tasks.

At present the debate regarding wave-particle duality and uncertainty remains unresolved, to our knowledge. Yet Feynman's quote seems to suggest a belief that quantum mechanics has but one mystery and not two separate ones. In this work [17] we lend quantitative support to this belief by showing a connection between URs and WPDRs, demonstrating that URs and WPDRs capture the same underlying physics; see also [18, 19] for some partial progress along these lines. This may come as a surprise, since Englert [2] originally argued that (1) "does not make use of Heisenberg's uncertainty relation in any form". To be fair, the uncertainty relation that we show is equivalent to (1) was not known at the time of Englert's paper, and was only recently discovered [20]. Specifically, we will consider EURs, where the particular entropies that are relevant to (1) are the so-called min- and max-entropies used in cryptography [21].

We show that several WPDRs from the literature are in fact particular examples of EURs. Making this connection not only unifies two fundamental concepts in quantum mechanics, but also means that novel WPDRs can be derived simply by applying already-proven EURs. As an illustration, we derive a novel WPDR for an exotic scenario involving a "quantum beam splitter" [22–24], where testing our WPDR would allow the experimenter to verify the beam splitter's quantum coherence. Thus, in addition to unifying fundamental concepts, we provide a general framework for deriving and discussing WPDRs. We emphasize that the framework provided by EURs is highly robust, and entropies have well-characterized statistical meanings. Current approaches to deriving WP-

DRs often involve brute force calculation of the quantities one aims to bound; there is no general, elegant method currently in use. Our approach simply involves judicious application of the relevant uncertainty relation. What's more, we emphasize that uncertainty relations can be applied to interferometers in two different ways. One involves preparation uncertainty, which says that a quantum state cannot be prepared having low uncertainty for two complementary observables, and it turns out this principle is the one relevant to (1). The other involves measurement uncertainty, which says that two complementary observables cannot be jointly measured, and we discuss why this principle is actually what was tested in some recent interferometry experiments [23, 25].

RESULTS

Our unified view associates a kind of behavior with the availability of a kind of information, or lack of behavior with missing information, as follows:

lack of particle behavior: $H_{\min}(Z|J)$

lack of wave behavior: $\min_{W \in XY} H_{\max}(W|K)$

where H_{\min} and H_{\max} are the min- and max-entropies commonly used in quantum information theory, Z is the path observable identified with the standard qubit basis (see Fig. 1), W is an orthonormal basis observable in the XY plane of the Bloch sphere, and J and K are some other quantum systems that help to reveal the behavior (e.g., J could be a which-path detector and K could be the quanton's internal degree of freedom). Note that we use the same symbols (Z , W , X , etc.) for the observables as for the random variables they give rise to. We formulate our general WPDR as

$$H_{\min}(Z|J) + \min_{W \in XY} H_{\max}(W|K) \geq 1 \quad (2)$$

which states that, for a two-path interferometer for single quantons, the sum of the ignorances about the particle and wave behaviors is lower bounded by 1 (i.e., 1 bit).

To be clear, (2) is explicitly an EUR, and it has been exploited to prove the security of quantum cryptography [26]. The usefulness of (2) for cryptography is due to the clear operational meanings of the min- and max-entropies [21], which naturally express the monogamy of correlations as they give the distances to being uncorrelated (H_{\max}) and being perfectly correlated (H_{\min}). One can replace these entropies with the von Neumann entropy in (2) and the relation still holds; however, the min- and max-entropies give more refined statements about information processing since they are also applicable in the regime of finite repetitions. The fact that (2) can be thought of as a WPDR, and furthermore that it encompasses the majority of WPDRs found in the literature for two-path single-quanton interferometers, is our result.

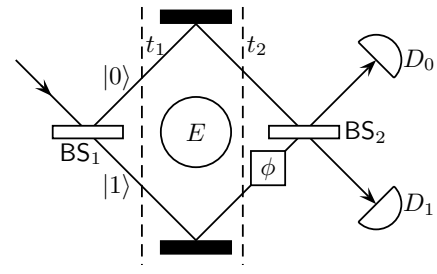


FIG. 1: Mach-Zehnder interferometer for single photons. Identifying $Z = \{|0\rangle, |1\rangle\}$ as the which-path basis, a superposition of these states is created at time t_1 . An environment E may then obtain some which-path information, e.g., E could be a gas of atoms whose internal states are affected by the presence of a photon. Finally at time t_2 a phase shift ϕ is applied to the lower arm and the two beams are recombined on a second beam splitter.

DISCUSSION

To illustrate this, we consider the celebrated Mach-Zehnder interferometer, shown in Fig. 1. In the simplest case a single photon impinges on a 50/50 beam splitter, BS_1 , resulting in the state $(|0\rangle + |1\rangle)/\sqrt{2}$, where $Z = \{|0\rangle, |1\rangle\}$ is the which-path basis, then a phase ϕ is applied to the lower arm giving the state $(|0\rangle + e^{i\phi}|1\rangle)/\sqrt{2}$. Finally the two paths are recombined on a second 50/50 beam splitter BS_2 and the output modes are detected by detectors D_0 and D_1 . Visibility is then defined as

$$\mathcal{V} := \frac{p_{\max}^0 - p_{\min}^0}{p_{\max}^0 + p_{\min}^0} \quad (3)$$

where $p_{\max}^0 = \max_{\phi} \Pr(C = 0)$ and $p_{\min}^0 = \min_{\phi} \Pr(C = 0)$, where C denotes the random variable revealing which detector D_C clicks, with $C \in \{0, 1\}$. In this trivial example, $\mathcal{V} = 1$. However many more complicated and interesting situations have been considered [2–10]; we show how these fall under the umbrella of our framework.

For example, consider the scenario in Fig. 1. Inside the interferometer the photon interacts with some environment E , which may act as a which-way detector. Most generally the interaction is given by a completely positive trace preserving (CPTP) map \mathcal{E} , with the input system being S at time t_1 and output systems being S and E at time t_2 , see Fig. 1. The final state is $\rho_{SE}^{(2)} = \mathcal{E}(\rho_S^{(1)})$, where the superscripts (1) and (2) to indicate the states at times t_1 and t_2 . The path distinguishability is defined by $\mathcal{D} := 2p_{\text{guess}}(Z|E) - 1$, where $p_{\text{guess}}(Z|E)$ is the probability for correctly guessing the path Z at time t_2 given that the experimenter performs the optimally helpful measurement on E . We find that (1) is equivalent to

$$H_{\min}(Z|E) + \min_{W \in XY} H_{\max}(W) \geq 1, \quad (4)$$

where the entropies are evaluated for the state $\rho_{SE}^{(2)}$. To

rewrite (4) as (1), we first have $H_{\min}(Z|E) = -\log \frac{1+\mathcal{V}}{2}$ from the operational meaning of the conditional min-entropy [21], and second we prove in [17] that

$$\min_{W \in XY} H_{\max}(W) = \log(1 + \sqrt{1 - \mathcal{V}^2}). \quad (5)$$

Preparation vs. Measurement Uncertainty.—The above analysis shows that (1) corresponds to applying the *preparation uncertainty relation* at time t_2 . Preparation uncertainty restricts one’s ability to predict the outcomes of *future* measurements of complementary observables. Thus, to experimentally measure \mathcal{D} , the experimenter *removes* BS_2 and sees how well he/she can guess which detector clicks. On the other hand, uncertainty relations can be applied in a conceptually different way. Instead of two complementary output measurements and a fixed input state, consider a fixed output measurement and two complementary sets of input states. Namely consider the input ensembles $Z = \{|0\rangle, |1\rangle\}$ and $W = \{|w_{\pm}\rangle\}$, where $|w_{\pm}\rangle = (|0\rangle \pm e^{i\phi}|1\rangle)/\sqrt{2}$ (see Fig. 1 for definition of $|0\rangle$ and $|1\rangle$). The two Z inputs are generated by blocking the opposite arm of the interferometer, while the W states are generated by applying a phase (either 0 or π) to the lower arm. One can imagine this as a game, where Bob controls the input and Alice has control over both E and the detectors, Bob flips a coin to determine which path he will block in the case of Z (or which phase he will apply in the case of W) and Alice’s goal is to guess the outcome of Bob’s coin flip. In [17] we discuss how this leads to a *different class* of WPDRs, which address the question of how well Alice’s apparatus can *jointly measure* Bob’s Z and W observables.

Quantum BS_2 .—As an interesting application, we consider the scenario in [22–24], where the photon’s polarisation P acts as a control system to determine whether or not BS_2 appears in the photon’s path and hence whether the interferometer is open or closed. Since P can be prepared in an arbitrary superposition, this effectively means that BS_2 is a “quantum beam splitter”, i.e., it can be in a quantum superposition of being absent or present. We argue in [17] that the only known WPDR for this scenario is insensitive to coherence and hence does not allow the experimenter to verify that their beam splitter is “quantum”. However, our framework provides a novel WPDR that is unique in that experimentally testing it will directly verify the beam splitter’s coherence.

Conclusions.—We have unified the wave-particle duality principle and the entropic uncertainty principle, showing that WPDRs are EURs in disguise. We believe the framework presented here can be applied fairly universally to the case of single quantons in two-path interferometers. Indeed we show that the main results of Refs. [7–9] (respectively concerning asymmetric beam splitters, quantum erasure, and polarization dynamics) all fall under our entropic uncertainty framework. Our

framework also makes it clear how to formulate novel WPDRs by simply applying known EURs to novel interferometer models, and these new WPDRs will likely inspire new interferometry experiments. These WPDRs may also find application in the security analysis of interferometric quantum key distribution [27].

-
- [1] R. P. Feynman, *Feynman Lectures on Physics* (Addison Wesley, Longman, 1970).
 - [2] B.-G. Englert, Phys. Rev. Lett. **77**, 2154 (1996).
 - [3] G. Jaeger, A. Shimony, and L. Vaidman, Phys. Rev. A **51**, 54 (1995).
 - [4] W. K. Wootters and W. H. Zurek, Phys. Rev. D **19**, 473 (1979).
 - [5] D. M. Greenberger and A. Yasin, Physics Letters A **128**, 391 (1988), ISSN 0375-9601.
 - [6] N.-L. Liu, L. Li, S. Yu, and Z.-B. Chen, Phys. Rev. A **79**, 052108 (2009).
 - [7] L. Li, N.-L. Liu, and S. Yu, Phys. Rev. A **85**, 054101 (2012).
 - [8] B.-G. Englert and J. A. Bergou, Optics Communications **179**, 337 (2000), ISSN 0030-4018.
 - [9] K. Banaszek, P. Horodecki, M. Karpiński, and C. Radzewicz, Nat Commun **4** (2013).
 - [10] A.-A. Jia, J.-H. Huang, W. Feng, T.-C. Zhang, and S.-Y. Zhu, Chinese Physics B **23**, 30307 (2014).
 - [11] B.-G. Englert, M. O. Scully, and H. Walther, Nature **375**, 367 (1995).
 - [12] P. Storey, S. Tan, M. Collett, and D. Walls, Nature **367**, 626 (1994).
 - [13] H. Wiseman and F. Harrison, Nature **377**, 584 (1995).
 - [14] N. Bohr, Nature **121**, 580 (1928).
 - [15] S. Wehner and A. Winter, New J. Phys. **12**, 025009 (2010).
 - [16] I. Białynicki-Birula and J. Mycielski, Communications in Mathematical Physics **44**, 129 (1975).
 - [17] P. J. Coles, J. Kaniewski, and S. Wehner, ArXiv e-prints <http://arxiv.org/abs/1403.4687> (2014), 1403.4687.
 - [18] S. Durr and G. Rempe, American Journal of Physics **68**, 1021 (2000).
 - [19] P. Busch and C. Shilladay, Physics Reports **435**, 1 (2006), ISSN 0370-1573.
 - [20] M. Tomamichel and R. Renner, Phys. Rev. Lett. **106**, 110506 (2011).
 - [21] R. König, R. Renner, and C. Schaffner, IEEE Trans. Inf. Theory **55**, 4337 (2009).
 - [22] R. Ionicioiu and D. R. Terno, Phys. Rev. Lett. **107**, 230406 (2011).
 - [23] F. Kaiser, T. Coudreau, P. Milman, D. B. Ostrowsky, and S. Tanzilli, Science **338**, 637 (2012).
 - [24] A. Peruzzo, P. Shadbolt, N. Brunner, S. Popescu, and J. L. O’Brien, Science **338**, 634 (2012).
 - [25] V. Jacques, E. Wu, F. Grosshans, F. Treussart, P. Grangier, A. Aspect, and J.-F. Roch, Phys. Rev. Lett. **100**, 220402 (2008).
 - [26] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, Nature Communications **3**, 634 (2012).
 - [27] A. K. Ekert, J. G. Rarity, P. R. Tapster, and G. Massimo Palma, Phys. Rev. Lett. **69**, 1293 (1992).

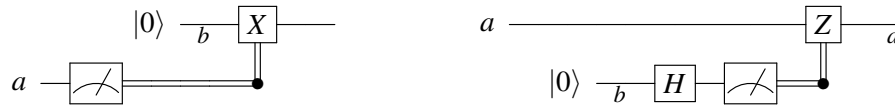
An equational characterization of quantum computation

Sam Staton
Radboud University Nijmegen

We provide an equational theory describing the algebraic structure of quantum computation. We show that our axioms are complete with respect to the interpretation as completely positive unital maps between C^* -algebras.

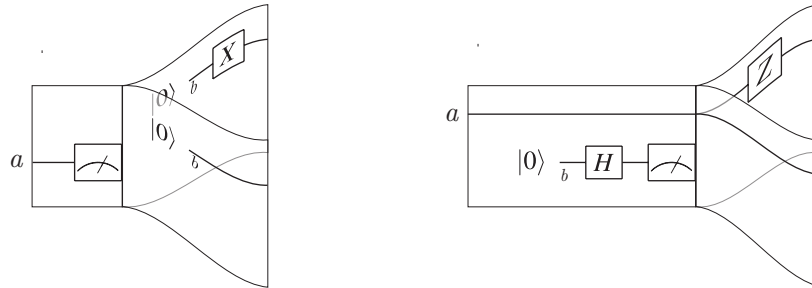
1 Introduction

In this paper we provide an axiomatization of the equations that are satisfied by quantum programs. We begin by discussing what we mean by quantum computation, beginning informally with some diagrammatic notation, and progressing to a formal algebraic syntax. For illustration we consider the following two programs, first illustrated using quantum circuit diagrams [15].



The first program collapses the state of the qubit a in the standard basis. The second program randomly flips the phase of the qubit a . The first program works by first measuring the input qubit (a) and then generating a new qubit b , initialized to $|0\rangle$ but then flipped around X depending on the outcome of the measurement. The randomness in the second program is implemented by allocating a new qubit b initialized to $|0\rangle$, but measuring it in the $(|+\rangle, |-\rangle)$ basis by applying the Hadamard unitary.

Since our aim is to analyze the essence of quantum computing, we do not want to axiomatize a theory of classical circuits, and so we rather think of the classical wires (double lines) as informal 2-dimensional notation for the idea that measurement causes a branch in the quantum circuit. This could be illustrated by 3-dimensional diagrams like



(These kinds of diagram are common in the many-worlds interpretation of quantum mechanics but are also strongly reminiscent of Melliès' notation for storage of classical bits [14].)

We are now in a position to postulate the algebraic structure of quantum computation. It supports three constructions:

- if t a computation involving a qubit a then there is a computation $\text{new}(a.t)$ that allocates that qubit, initialized to $|0\rangle$, and continues as t ;

- if t is a computation involving qubits $a_1 \dots a_n$ and U is a unitary gate over n qubits then there is a computation $\text{apply}_U(a_1 \dots a_n, a_1 \dots a_n.t)$ that first performs the unitary U and then continues as t ;
- if t and u are computations and a is a qubit then there is a computation $\text{measure}(a, t, u)$ that measures a in the standard basis and continues as either t or u depending on the result of the measurement.

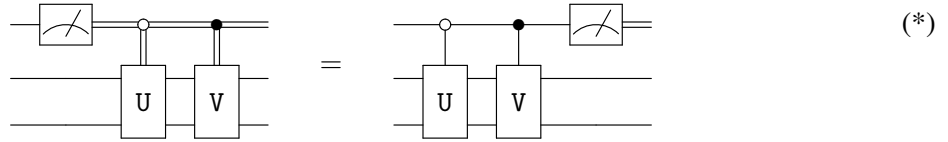
We can construct our two programs as algebraic expressions as follows:

$$\begin{aligned} &\text{measure}(a, \text{new}(b.x(b)), \text{new}(b.\text{apply}_x(b, b.x(b)))) \\ &\quad \text{new}(b.\text{apply}_{\text{had}}(b, b.\text{measure}(b, x(a), \text{apply}_z(a, a.x(a))))) \end{aligned}$$

Here x is a variable standing for the continuation of the computation (as yet unspecified) which is parameterized by the result qubit.

Different authors have considered different syntaxes and notations for quantum programs (e.g. [7, 8, 20, 28]). Our algebraic notation is not intended for practical programming, rather to describe the essence of the notions of computation. There are standard and general tools for extending an algebraic notion of computation to a more useful programming language, which we discuss briefly in Section 4.

In fact, our two programs have the same behaviour (see e.g. [15, eqns 8.66–8.70], [20, ex. 4.6]). In this paper we provide a complete system of axioms that allows us to deduce equations of this kind, explaining the interactions between the three constructions of quantum computation (measurement, unitaries, and allocating new qubits). Here is an illustration of one of our axioms, relating controlled gates with measurement:



We claim that all appropriate equations between quantum programs can be derived from our axioms.

We justify our axioms by reference to a popular model of quantum computation based on completely positive unital maps between finite dimensional C^* -algebras. We have the following theorem:

Theorem. *The following data are equivalent:*

- *An algebraic expression, modulo the equality derivable from our axioms;*
- *A completely positive unit-preserving map between finite dimensional C^* -algebras.*

(It is already well-known that a completely positive unit-preserving map can be understood in terms of allocating qubits, applying a unitary, and then measuring, and Selinger [20] phrased this in terms of programming languages; the novelty here is our axiomatization of equality.)

Conversely, our theorem provides a justification for the model of quantum computation based on C^* -algebras. Although linear algebra plays a major role in many models of quantum computation, the non-convex addition in linear algebra often has no direct physical justification. In our equational theory, vector spaces do *not* have an explicit role. The model of C^* -algebras arises (up-to categorical equivalence) from our axioms, which are all validated by basic intuitions from physics. Of course, linear algebra, representation theory and functional analysis are very useful. Our work shows that rather than jumping from physical intuitions directly to linear algebra, one can first set up a mathematical model of the physical intuitions, and then use that to justify the techniques of linear algebra.

Another motivation for our development is that it follows a strong tradition of equational reasoning about computer programs. Indeed, some equations for quantum programs are already suggested in [20, §4.9], [26], and [28, §6], and other authors have suggested equational theories for different aspects of quantum computation [4, 5, 25, 27]. More broadly across computer science it is often informative and useful to reason about computer programs in an equational way without reference to a particular denotational semantics. A profound analysis of this general phenomenon was begun by Plotkin and Power [17], who provided a categorical formalism and explained the relationship between equational reasoning and Moggi’s work on monads. Indeed, our axioms for quantum computation over qubits resemble the axioms for local store of classical bits [17, 14, 22] in several ways. The underlying general frameworks that support the axioms are closely related, and that was the starting point for this author.

In this paper we focus on quantum computation in a first-order sense. However, our analysis seems to extend to other computational features, which we briefly discuss at the end of the paper.

2 Equational theory

In this section we introduce a formalism for building quantum computations over qubits. To motivate this, we consider some simple examples. If x and y are quantum computations and a is a qubit, then $\text{measure}(a, x, y)$ is a quantum computation that first measures a , and, depending on the result, continues as x or as y . We can allocate new qubits, writing $\text{new}(a.\text{measure}(a, x, y))$ for the computation that first allocates a , then measures it, continuing as either x or y . We can also apply unitaries to qubits, writing $\text{apply}_{\text{cnot}}(a, b, a'b'.\text{measure}(a', \text{measure}(b', v, x), \text{measure}(b', y, z)))$ for the computation that first applies a controlled-not gate to qubits a and b , yielding qubits a' and b' , and then measures a' and then b' .

We make some informal remarks about this syntax, before introducing a formal system.

1. There are two kinds of variable: in the examples above, a, b stand for qubits whereas v, x, y, z stand for computations.
2. In the example with new , the a is binding, and we could just as well write $\text{new}(b.\text{measure}(b, x, y))$.
3. Since computations can involve qubit parameters, and variables x, y stand for computations, we will also allow variables with qubit parameters. When we write $x(a, b)$, the computation variable x is being passed parameters a and b . For instance we can write a computation expression $\text{new}(a.z(a))$ which allocates a new qubit a and passes it as a parameter to the continuation z ; we can substitute $\text{measure}(a, x, y)$ for $z(a)$, resulting in the term $\text{new}(a.\text{measure}(a, x, y))$. The notation means that we do not need to worry about implicit variable capture.
4. Care must be taken when using qubits. Measuring a qubit collapses it to one of the basis states, and so we adopt the convention that, in an expression $\text{measure}(a, t, u)$, the variable a should not appear free in t or u . This kind of linearity also plays a crucial role in the syntax for unitaries: in $\text{apply}_{\text{cnot}}(a, b, \dots)$ it is crucial that a and b are different qubits and not aliases for the same qubit. For simplicity we adopt the convention that apply consumes its qubit parameters, creating new ones that are passed to the continuation.

2.1 General syntactic framework

Our general syntactic framework is an algebraic framework which is not at all specific to quantum computation. Indeed, it is a substructural variation on the ‘parameterized algebraic theories’ already used to analyze various kinds of computation including local store, π -calculus-style communication [22] and logic programming [21]. There are two kinds of variable:

1. Computation variables are ranged over by x, y etc.. Computations can depend on parameters, and we write $x : p$ if x has p parameters. The number p is sometimes called a valence.
2. Parameter variables are ranged over by a, b . In the quantum situation, these stand for qubits.

The operations O have an arity $(p \mid m_1 \dots m_k)$ which is a natural number (≥ 0) followed by a list of natural numbers. This specifies that O takes p parameter arguments and k computation arguments, and the i th computation argument has m_i parameters bound in it.

We define a three-place judgement $\Gamma \mid \Delta \vdash t$ defining well-formed terms: here Γ is a list of computation variables with their valences, and Δ is a list of parameter variables. The judgement is the least one closed under the following rules, which means that if the judgements on the top of the rules are justified then so is the judgement on the bottom. (This is standard in type theory; see the example on page 5.)

$$\frac{}{\Gamma, x : p, \Gamma' \mid a_1 \dots a_p \vdash x(a_1 \dots a_p)} \quad \frac{\Gamma \mid a_1 \dots a_p \vdash t}{\Gamma \mid a_{\sigma(1)} \dots a_{\sigma(p)} \vdash t} \quad (\sigma \text{ is a permutation of } p)$$

$$\frac{\Gamma \mid \Delta, b_1 \dots b_{m_1} \vdash t_1 \quad \dots \quad \Gamma \mid \Delta, b_1 \dots b_{m_k} \vdash t_k}{\Gamma \mid \Delta, a_1 \dots a_p \vdash O(a_1, \dots, a_p, b_1 \dots b_{m_1}.t_1, \dots, b_1 \dots b_{m_k}.t_k)} \quad O : (p \mid m_1 \dots m_k)$$

In the term formation rule for operations, the b 's are binding, and we work up to renaming those bound variables, just as in predicate logic ($\forall x.P(x) = \forall y.P(y)$). The syntax admits the following simultaneous substitution law:

$$\frac{x_1 : m_1 \dots x_k : m_k \mid \Delta \vdash t \quad \Gamma \mid \Xi, a_1 \dots a_{m_1} \vdash u_1 \quad \dots \quad \Gamma \mid \Xi, a_1 \dots a_{m_k} \vdash u_k}{\Gamma \mid \Delta, \Xi \vdash t[\Xi, a_1 \dots a_{m_1} \vdash u_1 / x_1 \dots \Xi, a_1 \dots a_{m_k} \vdash u_k / x_k]}$$

2.2 Operations

Rudiments of unitaries. Recall that a complex square matrix U is *unitary* if its conjugate transpose U^* is its inverse. Unitaries of the same dimension form a group under matrix multiplication. We can thus form a groupoid whose objects are natural numbers, and where a morphism $U : n \rightarrow n$ is an $n \times n$ unitary. There are two symmetric monoidal structures on this groupoid. One monoidal structure is multiplication of dimensions: if U and V are unitaries, $m \times m$ and $n \times n$ respectively, the Kronecker product $U \otimes V$ is a unitary $(m \times n) \times (m \times n)$ matrix. The other monoidal structure is addition of dimensions: if U and V are unitaries, $m \times m$ and $n \times n$ respectively, the block diagonal $(m+n) \times (m+n)$ matrix $\begin{pmatrix} U & 0 \\ 0 & V \end{pmatrix}$ is also unitary. In general we write $D(U_1, \dots, U_k)$ for a block diagonal.

Recall some convenient unitaries. A 1×1 unitary is a complex number with unit modulus. The following 2×2 unitaries are handy: rotation by π about the X axis, $X \stackrel{\text{def}}{=} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, is the about the Z axis, $Z \stackrel{\text{def}}{=} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$, and the Hadamard rotation, $\text{had} \stackrel{\text{def}}{=} \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$. The block diagonal $CX \stackrel{\text{def}}{=} D(I_2, X)$ is often called ‘controlled not’, and $\text{swap} \stackrel{\text{def}}{=} D(1, X, 1)$ is the symmetry for the multiplication monoidal structure.

Remark: Our aim here is to study the interaction between this groupoid of unitaries and qubit allocation and measurement. There is a body of work on axiomatizing and approximating unitaries. A reviewer pointed out that we are here ignoring the topological structure of unitaries, which could perhaps be used to give an account of approximations within our framework.

Operations We consider the following operations in our algebraic theory.

$$\text{new} : (0 \mid 1) \quad \text{measure} : (1 \mid 0, 0) \quad \text{apply}_{\mathbb{U}} : (n \mid n) \quad \text{where } \mathbb{U} \text{ is a } 2^n \times 2^n \text{ unitary}$$

Explicitly, these operations induce the following term formation rules:

$$\frac{\Gamma \mid \Delta, a \vdash t}{\Gamma \mid \Delta \vdash \text{new}(a.t)} \quad \frac{\Gamma \mid \Delta \vdash t \quad \Gamma \mid \Delta \vdash u}{\Gamma \mid \Delta, a \vdash \text{measure}(a.t, u)} \quad \frac{\Gamma \mid \Delta, b_1 \dots b_n \vdash t}{\Gamma \mid \Delta, a_1, \dots, a_n \vdash \text{apply}_{\mathbb{U}}(\vec{a}, \vec{b}.t)}$$

Examples

1. We can make a new qubit initialized to 1 and continue as x .

$$x : 1 \mid - \vdash \text{new}(a.\text{apply}_x(a, b.x(b))) \quad \begin{array}{l} \text{This is justified by the} \\ \text{term formation rules:} \end{array} \quad \frac{\frac{x : 1 \mid b \vdash x(b)}{x : 1 \mid a \vdash \text{apply}_x(a, b.x(b))}}{x : 1 \mid - \vdash \text{new}(a.\text{apply}_x(a, b.x(b)))}$$

Here, the a and b are binding: this expression is equal to $\text{new}(c.\text{apply}_x(c, d.x(d)))$. The notation $x(b)$ indicates that the continuation x takes a parameter b .

A convenient shorthand: $\text{new}_0(a.t) \stackrel{\text{def}}{=} \text{new}(a.t)$, $\text{new}_1(a.t) \stackrel{\text{def}}{=} \text{new}(a.\text{apply}_x(a, a.t))$.

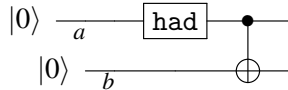
2. We can use quantum operations to make a random choice between continuations x and y .

$$x : 0, y : 0 \mid - \vdash \text{new}(a.\text{apply}_{\text{had}}(a, b.\text{measure}(b, x, y)))$$

3. We can create a Bell state, and pass it to x :

$$x : 2 \mid - \vdash \text{new}(a.\text{new}(b.\text{apply}_{\text{had}}(a, a.\text{apply}_{D(\mathbb{I}, \mathbb{X})}(a, b, ab.x(a, b)))))$$

This would be written as the following circuit diagram:



4. The linearity constraints mean that we cannot implicitly discard nor duplicate qubits. However we can explicitly discard a qubit a , by measuring it and ignoring the result.

$$x : 0 \mid a \vdash \text{measure}(a, x, x)$$

5. In our formalism, measurement consumes the qubit parameter. Another convention is that measurement retains the qubit but it is now collapsed into one of the basis states. Of course this can be simulated by immediately creating a new qubit with the result of the measurement:

$$x : 1 \mid a \vdash \text{measure}(a, \text{new}_0(a.x(a)), \text{new}_1(a.x(a)))$$

6. In fact, we will later show that (5) is the same as randomly flipping the phase of a .

$$x : 1 \mid a \vdash \text{new}_0(b.\text{apply}_{\text{had}}(b, b.\text{measure}(b, x(a), \text{apply}_z(a, a.x(a)))))$$

2.3 Equations

We subject the terms built from `new`, `apply` and `measure` to the following equations. The equations are of three main kinds: commutativity equations (all possible commutativity equations are permitted, we explicitly impose (8)–(11)); the relationship between `applyU` and the groupoid of unitaries ((1)–(4)); perhaps most interestingly, the key relationships between `apply`, `measure` and `new` ((5)–(7), (12)–(13)).

Respecting the symmetric monoidal groupoid of unitaries. The first class of equation schemes that we consider imposes the relationships between the symmetric monoidal structure of unitaries and the compositional structure of terms built from `applyU`.

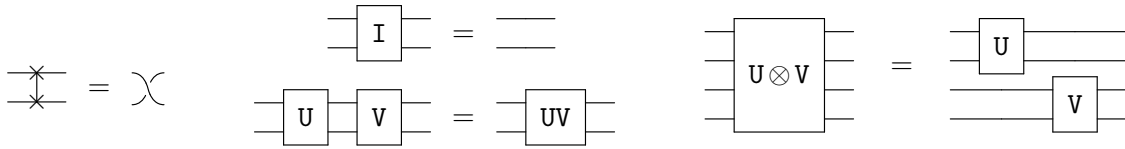
$$\text{apply}_{\text{swap}}(a, b, (a, b).x(a, b)) = x(b, a) \quad (1)$$

$$\text{apply}_{\mathbf{I}}(\vec{a}, \vec{a}.x(\vec{a})) = x(\vec{a}) \quad (2)$$

$$\text{apply}_{UV}(\vec{a}, \vec{a}.x(\vec{a})) = \text{apply}_U(\vec{a}, \vec{a}.\text{apply}_V(\vec{a}, \vec{a}.x(\vec{a}))) \quad (3)$$

$$\text{apply}_{U \otimes V}(\vec{a}, \vec{b}, (\vec{a}, \vec{b}).x(\vec{a}, \vec{b})) = \text{apply}_U(\vec{a}, \vec{a}.\text{apply}_V(\vec{b}, \vec{b}.x(\vec{a}, \vec{b}))) \quad (4)$$

Here are informal illustrations of these equations.



Interaction between unitary gates and measurement. Our next set of equations describe the interaction between the operations of unitaries and the standard basis measurement operations. We begin by setting up some shorthand. For a list of p distinct parameter variables $a_1 \dots a_p$ and a term t , define a term $\text{discard}_p(a_1 \dots a_p, t)$ by

$$\text{discard}_0(-, t) = t; \quad \text{discard}_{n+1}(a, \vec{b}, t) = \text{measure}(a, \text{discard}_n(\vec{b}, t), \text{discard}_n(\vec{b}, t)).$$

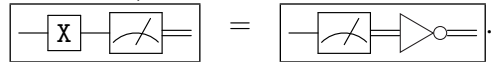
There are now three equation schemes governing the interaction between unitaries and standard basis measurement.

$$\text{apply}_U(\vec{a}, \vec{a}.\text{discard}_n(\vec{a}, x)) = \text{discard}_n(\vec{a}, x) \quad U \text{ is a } 2^n \times 2^n \text{ unitary, } n \geq 0 \quad (5)$$

$$\text{apply}_X(a, a.\text{measure}(a, x, y)) = \text{measure}(a, y, x) \quad (6)$$

$$\text{measure}(a, \text{apply}_U(\vec{b}, \vec{b}.x(\vec{b})), \text{apply}_V(\vec{b}, \vec{b}.y(\vec{b}))) = \text{apply}_{D(U,V)}(a, \vec{b}, (a, \vec{b}).\text{measure}(a, x(\vec{b}), y(\vec{b}))) \quad (7)$$

In particular, equation (5) for $n = 0$ says when global phase can be ignored. Equation (7) specifies the interaction between controlled gates and measurement, as illustrated in the introduction (*), and equation (6) could be illustrated as follows:



Commutativity. Our equational theory is commutative in the sense of [13]. The following commutativity equation scheme is derivable:

$$\text{apply}_U(\vec{b}, \vec{b}.\text{apply}_V(\vec{c}, \vec{c}.x(\vec{b}, \vec{c}))) = \text{apply}_V(\vec{c}, \vec{c}.\text{apply}_U(\vec{b}, \vec{b}.x(\vec{b}, \vec{c})))$$

This is in spite of the fact that multiplication of unitaries is not commutative, e.g. $X.\text{had} \neq \text{had}.X$: while $\text{apply}_X(a, a.\text{apply}_{\text{had}}(b, b.x(a, b))) = \text{apply}_{\text{had}}(b, b.\text{apply}_X(a, a.x(a, b)))$, this does not imply that $\text{apply}_X(a, a.\text{apply}_{\text{had}}(a, a.x(a)))$ is equal to $\text{apply}_{\text{had}}(a, a.\text{apply}_X(a, a.x(a)))$.

The following commutativity equation will also follow from our other axioms for measure:

$$\text{apply}_U(\vec{b}, \vec{b}.\text{measure}(a, x(\vec{b}), y(\vec{b}))) = \text{measure}(a, \text{apply}_U(\vec{b}, \vec{b}.x(\vec{b})), \text{apply}_U(\vec{b}, \vec{b}.y(\vec{b})))$$

but we need to explicitly require that different measurements commute:

$$\text{measure}(a, \text{measure}(b, u, v), \text{measure}(b, x, y)) = \text{measure}(b, \text{measure}(a, u, x), \text{measure}(a, v, y)) \quad (8)$$

Equations for qubit allocation: Finally we impose the following equations regarding new. Firstly commutativity:

$$\text{new}(a.\text{new}(b.x(a, b))) = \text{new}(b.\text{new}(a.x(a, b))) \quad (9)$$

$$\text{new}(a.\text{measure}(b, x(a), y(a))) = \text{measure}(b, \text{new}(a.x(a)), \text{new}(a.y(a))) \quad (10)$$

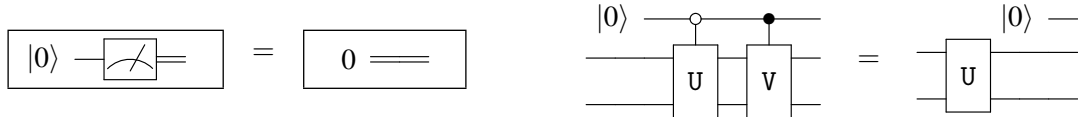
$$\text{new}(a.\text{apply}_U(\vec{b}, \vec{b}.x(a, \vec{b}))) = \text{apply}_U(\vec{b}, \vec{b}.\text{new}(a.x(a, \vec{b})))$$

(The last equation is derivable from (12).) Secondly equations governing the interaction between new and measurement and controlled unitaries, based on the idea that new qubits are initialized to 0.

$$\text{new}(a.\text{measure}(a, x, y)) = x \quad (11)$$

$$\text{new}(a.\text{apply}_{D(U,V)}(a, \vec{b}, a.\vec{b}.x(a, \vec{b}))) = \text{apply}_U(\vec{b}, \vec{b}.\text{new}(a.x(a, \vec{b}))) \quad (12)$$

These two equations can be illustrated as follows:



2.4 Examples of derivations

Rotation about Z doesn't affect standard basis measurement Let $\text{phase}\theta = \begin{pmatrix} 1 & 0 \\ 0 & e^{i\theta} \end{pmatrix} = D(1, e^{i\theta})$. Then

$$\begin{aligned} \text{apply}_{\text{phase}\theta}(a, a.\text{measure}(a, x, y)) &= \text{apply}_{D(1, e^{i\theta})}(a, a.\text{measure}(a, x, y)) = \text{measure}(a, x, \text{apply}_{e^{i\theta}}(y)) \\ &= \text{measure}(a, x, y) \end{aligned}$$

Random phase flip = measurement. Consider the following equation:

$$\text{new}_0(a.\text{apply}_{\text{had}}(a, a.\text{measure}(a, x(b), \text{apply}_Z(b, b.x(b))))) = \text{measure}(b, \text{new}_0(a.x(a)), \text{new}_1(a.x(a)))$$

where $Z = \text{phase}\pi$. It says that randomly flipping the phase of a qubit is the same as measuring it. Nielsen and Chuang discuss the equation to demonstrate the freedom in the operator-sum representation [15, eqns 8.66–8.70]. We can prove it directly in our equational theory.

First we show that measuring a qubit is the same as making an entangled copy using CX and then discarding the original:

$$\begin{aligned}
& \text{measure}(b, \text{new}_0(a.x(a)), \text{new}_1(a.x(a))) \\
&= \text{measure}(b, \text{new}_0(a.x(a)), \text{new}_0(a.\text{apply}_x(a, a.x(a)))) \\
&= \text{new}_0(a.\text{measure}(b, x(a), \text{apply}_x(a, a.x(a)))) \\
&= \text{new}_0(a.\text{apply}_{CX}(b, a, (b, a).\text{measure}(b, x(a), x(a)))) \\
&= \text{new}_0(a.\text{apply}_{CX}(b, a, (b, a).\text{discard}(b, x(a)))) \\
&= \text{new}_0(a.\text{apply}_{CX}(a, b, (a, b).\text{apply}_{CX}(b, a, (b, a).\text{discard}(a, x(b)))) \quad \text{since } CX.\text{swap}.CX.\text{swap} = \text{swap}.CX \\
&= \text{new}_0(a.\text{apply}_{CX}(b, a, (b, a).\text{discard}(a, x(b))))
\end{aligned}$$

We conclude by showing that the last line is the same as randomly flipping the phase.

$$\begin{aligned}
& \text{new}_0(a.\text{apply}_{CX}(b, a, (b, a).\text{discard}(a, x(b)))) \\
&= \text{new}_0(a.\text{apply}_{\text{had}}(a, a.\text{apply}_{CZ}(a, b, ab.\text{apply}_{\text{had}}(a, a.\text{discard}(a, x(b)))))) \quad \dagger \\
&= \text{new}_0(a.\text{apply}_{\text{had}}(a, a.\text{apply}_{CZ}(a, b, ab.\text{discard}(a, x(b)))) \\
&= \text{new}_0(a.\text{apply}_{\text{had}}(a, a.\text{measure}(a, x(b), \text{apply}_Z(b, b.x(b))))
\end{aligned}$$

\dagger : since $\text{swap}.CX.\text{swap} = (\text{had} \otimes I).CZ.(\text{had} \otimes I)$, where $CZ = D(I_2, Z)$.

Reasoning without qutrits. The following example illustrates a key point in the proof of our completeness theorem (Thm. 2). When reasoning in terms of C^* -algebras, one has access to various structures that are not definable in our syntax as it stands, such as base 3 quantum digits, ‘qutrits’. We could extend our syntax to have parameter variables of different sorts, for different bases. In fact, we do not need these structures to deduce the relevant equations between computations. Let

$$U = D(1, \text{had}, 1) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \begin{array}{c} \bullet \quad \boxed{\text{had}} \quad \bullet \\ | \quad | \quad | \\ \oplus \quad \bullet \quad \oplus \end{array}.$$

Consider the following equation:

$$\begin{aligned}
& \text{new}(b.\text{apply}_U(a, b, ab.\text{measure}(b, x(a), \text{measure}(a, y, u)))) \\
&= \text{new}(b.\text{apply}_U(a, b, ab.\text{measure}(b, x(a), \text{measure}(a, y, v)))) \tag{13}
\end{aligned}$$

It says that the pair (a, b) of qubits will never be in the state $(1, 1)$; in effect we have a qutrit, although that is not expressible in this formalism. We prove the equation by introducing an intermediate qubit c , as follows. Let $CCX = D(I_6, X)$, the 8×8 Toffoli gate, and note that if V is a block 3-1 matrix then $CCX.(V \otimes I_2) = (V \otimes I_2).CCX$.

$$\begin{aligned}
& \text{new}(b.\text{apply}_U(a, b, ab.\text{measure}(b, x(a), \text{measure}(a, y, u)))) \\
&= \text{new}(b.\text{apply}_U(a, b, ab.\text{meas}(b, \text{new}(c.\text{discard}(c, x(a))), \text{meas}(a, \text{new}(c.\text{discard}(c, y)), \text{new}(c.\text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_U(a, b, ab.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_{U \otimes I}(a, b, c, abc.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_{CCX.(U \otimes I)}(a, b, c, abc.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_{(U \otimes I).CCX}(a, b, c, abc.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_{U \otimes I}(a, b, c, abc.\text{meas}(b, \text{disc}(c, x(a)), \text{apply}_{CX}(a, c, ac.\text{meas}(a, \text{disc}(c, y)), \text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_{U \otimes I}(a, b, c, abc.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{apply}_x(c, c.\text{meas}(c, u, v)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_{U \otimes I}(a, b, c, abc.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{meas}(c, v, u)))))) \\
&= \text{new}(c.\text{new}(b.\text{apply}_U(a, b, ab.\text{meas}(b, \text{disc}(c, x(a)), \text{meas}(a, \text{disc}(c, y)), \text{meas}(c, v, u)))))) \\
&= \text{new}(b.\text{apply}_U(a, b, ab.\text{measure}(b, x(a), \text{measure}(a, y, v))))
\end{aligned}$$

3 Semantic interpretation

Our equational theory is purely syntactic. We now explain what it means to interpret terms and equations.

3.1 General interpretation in a category

Let **Bij** be the groupoid whose objects are natural numbers and whose morphisms are bijections. We will consider it as a symmetric monoidal category, where the monoidal operation is addition of numbers. In fact **Bij** is a free symmetric monoidal category on one generator (1).

Recall that an *action* of **Bij** [11] is a category \mathcal{V} together with a functor $\bullet : \mathbf{Bij} \times \mathcal{V} \rightarrow \mathcal{V}$ together with natural isomorphisms $0 \bullet X \cong X$ and $(m+n) \bullet X \cong m \bullet n \bullet X$ satisfying coherence conditions. For example, let \mathcal{V} be a monoidal category with an object A , and let $m \bullet X = A^{\otimes m} \otimes X$, where e.g. $A^{\otimes 3} = A \otimes A \otimes A$.

Let \mathcal{V} be an action of **Bij** such that the category \mathcal{V} has products and each functor $m \bullet - : \mathcal{V} \rightarrow \mathcal{V}$ preserves products.

A structure for a signature in \mathcal{V} is an object X together with, for each operation $O : (p \mid m_1 \dots m_k)$ a morphism $m_1 \bullet X \times \dots \times m_k \bullet X \rightarrow p \bullet X$.

In any structure, one can interpret each term in context $a_1 \dots a_p \mid x_1 : m_1 \dots x_k : m_k \vdash t$ as a morphism $m_1 \bullet X \times \dots \times m_k \bullet X \rightarrow p \bullet X$. This interpretation is defined by induction on the structure of terms.

3.2 Interpretation in linear algebra

Recall that a C*-algebra is a vector space over the field of complex numbers that also has multiplication, a unit, an involution, satisfying associativity laws for multiplication, idempotence laws (e.g. $x^{**} = x$, $(xy)^* = y^*x^*$) and such that the spectral radius provides a norm making it a Banach space. A *-homomorphism between C*-algebras is a linear map that preserves the multiplication, involution and unit. (NB some authors do not require the unit to be preserved.) We write **Cstar** for the category of C*-algebras and *-homomorphisms.

A key source of examples of C*-algebras are the algebras M_k of $k \times k$ complex matrices, with matrix addition and multiplication, and where involution is conjugate transpose. In particular the set $M_1 = \mathbb{C}$ of complex numbers has a C*-algebra structure.

Moreover if X is a C*-algebra then the $k \times k$ matrices valued in X form a C*-algebra, $M_k(X)$. Any linear map $f : X \rightarrow Y$ extends in the obvious way to a linear map $M_k(f) : M_k(X) \rightarrow M_k(Y)$, and $M_k(f)$ is a *-homomorphism if f is. We will consider the action of **Bij** on **Cstar** given by $m \bullet X = M_{2^m}(X)$.

The direct sum of vector spaces, given by the cartesian product, extends to C*-algebras, where it has the universal property of the categorical product, and it distributes over the action: $M_m(X \oplus Y) \cong M_m(X) \oplus M_m(Y)$.

We now investigate structures for the signatures in Section 2 whose carrier is the complex numbers. Note that $m \bullet \mathbb{C} = M_{2^m}$.

Measurement. The operations `measure` and `applyU` are interpreted using the following *-homomorphisms, `measure` : $M_1 \oplus M_1 \rightarrow M_2$ and `applyU` : $M_p \rightarrow M_p$.

$$\text{measure}(\alpha, \beta) = \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix} \quad \text{apply}_U(A) = U^*AU.$$

Allocation. To interpret allocation, we move beyond $*$ -homomorphisms. Recall that an element x of a C^* -algebra is called *positive* if there is y such that $x = y^*y$. A linear map $f : X \rightarrow Y$ is *completely positive* if for all k the map $M_k(f) : M_k(X) \rightarrow M_k(Y)$ preserves positive elements. (If either X or Y has commutative multiplication then it is sufficient to check the case $k = 1$.) We will focus on the completely positive maps that preserve units; this includes all the $*$ -homomorphisms. These form a category $\mathbf{Cstar}_{\text{CPU}}$ and the products and **Bij**-action extend from \mathbf{Cstar} ($*$ -homomorphisms) to $\mathbf{Cstar}_{\text{CPU}}$.

The operation `new` is interpreted using the following map, $\text{new} : M_2 \rightarrow M_1$.

$$\text{new} \begin{pmatrix} \alpha_{11} & \alpha_{12} \\ \alpha_{21} & \alpha_{22} \end{pmatrix} = \alpha_{11}.$$

This is not a $*$ -homomorphism, but it is completely positive and unital.

Proposition 1. *The interpretation of quantum computations is sound in $\mathbf{Cstar}_{\text{CPU}}$: If $\Gamma \mid \Delta \vdash t = u$ is derivable from the axioms via substitution and congruence then $\llbracket t \rrbracket = \llbracket u \rrbracket$ are equal linear maps.*

This is proved by induction on the derivations of equality.

3.3 Completeness theorems

Our interpretation is complete for completely positive maps.

Theorem 2. 1. *For any completely positive unital map $f : M_{2^p} \rightarrow M_{2^{m_1}} \oplus \cdots \oplus M_{2^{m_k}}$ there is a term $x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash t$ such that $f = \llbracket t \rrbracket$.*

2. *If $\Gamma \mid \Delta \vdash t, u$ and $\llbracket t \rrbracket = \llbracket u \rrbracket$ then $\Gamma \mid \Delta \vdash t = u$ is derivable.*

(NB. Part (1) of the theorem is essentially Theorem 6.14 of [20].)

We give a rough outline of our proof here, some more detail is in the Appendix. As a first step we consider the case without `new`.

Lemma 3. 1. *For any $*$ -homomorphism $f : M_{2^p} \rightarrow M_{2^{m_1}} \oplus \cdots \oplus M_{2^{m_k}}$ there is a term not involving `new`, say $x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash t$ such that $f = \llbracket t \rrbracket$.*

2. *If $\Gamma \mid \Delta \vdash t, u$ and t and u do not contain `new`, and $\llbracket t \rrbracket = \llbracket u \rrbracket$ then $\Gamma \mid \Delta \vdash t = u$ is derivable.*

The lemma is proved by noting that a term not involving `new` can always be rearranged to a term $\text{apply}_U(\vec{a}, \vec{a}.t)$ where t is built from measure and variables. We can understand a term built only from measure as a Bratelli diagram, which is a combinatorial way of understanding the $*$ -homomorphisms modulo unitaries.

To prove Theorem 2 we extend Lemma 3 to consider `new` and completely positive maps. Our proof is based around Stinespring's theorem, and we recall the following variant (e.g. [16, p. 45]). Let \tilde{p} be a natural number. If $f : A \rightarrow M_{\tilde{p}}$ is completely positive and unital then there is a natural number $q \geq \tilde{p}$ and a $*$ -homomorphism g making the following diagram commute:

$$\begin{array}{ccc} A & & \\ g \downarrow & \searrow f & \\ M_q & \longrightarrow & M_{\tilde{p}} \end{array}$$

where the unlabelled horizontal arrow is $A \mapsto A|_{\tilde{p}}$, where $A|_{\tilde{p}}$ comprises the first \tilde{p} rows and columns of the $q \times q$ matrix A . Moreover q can be chosen as the minimal such number: if $r \geq \tilde{p}$ and a $*$ -homomorphism

$h : A \rightarrow M_r$ is such that $h(-)|_{\tilde{p}} = f(-)$ then $r \geq q$ and there is an r -ary unitary U such that $g(a) = (U(ha)U^*)|_q$. As a diagram:

$$\begin{array}{ccccc}
 & & A & & \\
 & \swarrow h & \downarrow g & \searrow f & \\
 M_r & \xrightarrow{U-U^*} & M_r & \longrightarrow & M_q \longrightarrow M_{\tilde{p}}
 \end{array}$$

Now if f is as in the statement of our theorem (so $A = M_{2^{m_1}} \oplus \dots \oplus M_{2^{m_k}}$ and $\tilde{p} = 2^p$) then we can use the minimal dilation to show that f is definable, provided q is a power of 2. This is because the downward arrow can be defined using new and the $*$ -homomorphism can be defined according to Lemma 3. Moreover the minimality essentially gives us the second part of the theorem, since every term can be rearranged so that the new's are in front.

However, q need not be a power of 2. For instance, consider example (13): its minimal dilation is a $*$ -homomorphism $M_2 \oplus M_1 \oplus M_1 \rightarrow M_3$. We resolve this by generalizing our derivation of (13) from the axioms.

4 Remarks on other aspects of quantum programming

4.1 An alternative syntax for programming

The syntax of our equational theory describes quantum computation but it is not immediately amenable to practical programming because it focuses on continuing computations rather than intermediate results.

There is a standard way of moving between an equational theory like ours and a syntax more oriented towards programming. This applies to many different notions of computation [18]. In the setting of quantum computation, we can illustrate it by suggesting that a programmer might find it helpful to have a typed function measure : qubit \rightarrow bool which returns the result of a standard basis measurement. Our measurement operation can then be derived: $\text{measure}(a, x, y) = \text{if } \underline{\text{measure}}(a) \text{ then } x \text{ else } y$. Conversely, if we have a language with a bool type that is a model of our equational theory, we can derive $\underline{\text{measure}}(a) = \text{measure}(a, \text{false}, \text{true})$.

This relationship between ‘algebraic operations’ (like measure) and ‘generic effects’ (underlined, like measure) is a crucial one in the theory of computational effects [18]. Another way to understand this is in terms of the categorical duality between Lawvere theories, which are an abstract description of algebraic theories, and Freyd categories, which are an abstract description of first order programming languages [23]. The relationship allows us to pass from our algebraic syntax to a language rather like Selinger’s QPL [20]. Moreover Freyd categories are a categorical formulation of arrows, making a connection with [26].

To be more precise, we briefly demonstrate the programming language that corresponds automatically to our algebraic theory. The types are built from the grammar: $A, B ::= \text{qubit} \mid I \mid A \otimes B \mid 0 \mid A + B$. A context in this language is just an assignment of types to variables. The typed terms are built from the standard rules for a linear type theory [2], e.g.

$$\frac{}{x : A \vdash x : A} \quad \frac{\Gamma \vdash t : A \otimes B \quad \Delta, x : A, y : B \vdash u : C}{\Gamma, \Delta \vdash \text{let } (x, y) = t \text{ in } u : C}$$

To this standard linear type theory we add the following generic effects:

$$\frac{}{\vdash \underline{\text{new}}() : \text{qubit}} \quad \frac{\Gamma \vdash t : \text{qubit}^{\otimes n}}{\Gamma \vdash \underline{\text{apply}}_v(t) : \text{qubit}^{\otimes n}} \quad \frac{\Gamma \vdash t : \text{qubit}}{\Gamma \vdash \underline{\text{measure}}(t) : I + I}$$

Our equations have immediate analogues under this correspondence, as program equations. For example, the commutativity equations amount to the commutativity of `let` (e.g. [24]). Our equation (6) can be written $\text{measure}(\text{apply}_x(a)) = \neg(\text{measure}(a))$ (where $\neg(t) : I + I$ for $t : I + I$ is defined using the basic constructions for sums). Here is equation (12): $\text{measure}(\text{new}()) = 0$ (writing 0 for the left injection of $I + I$), and equation (13) can be written $\text{apply}_{D(u,v)}(\text{new}(), x) = (\text{new}(), \text{apply}_u(x))$.

4.2 Other language features

We can combine our theory of quantum computation with other notions of computation, simply by combining the equational theories.

Non-returning computations, sub-unital maps and recursion We can model computations whose results are partially undefined (e.g. they diverge or ‘fail’) by adding to our theory a constant symbol `undef` : $(0 \mid -)$. We do not need any additional equations. We interpret `undef` as the zero map $\text{undef} : M_0 \rightarrow M_1$ between C*-algebras: $\text{undef}() = 0$.

The correspondence with C*-algebras extends to ‘`undef`’ when we relax the preservation of units to the requirement that $(1 - f(1))$ is positive. This is because to give a subunital map $f : X \rightarrow Y$ is to give a unital map $f : X \oplus \mathbb{C} \rightarrow Y$.

Selinger has noted that the sub-unital maps between finite-dimensional C*-algebras form a pointed dcpo, and proposed to use this to interpret recursion in a quantum programming language [20]; this ought to be related to standard algebraic ways of analyzing iteration [3].

Non-determinism and non-unital maps Some authors drop the requirement of preserving the unit altogether (e.g. [10]). I am not aware of any attempt to justify this with physical intuitions, but I can make the following following remarks.

Recall that a non-deterministic computation is one in which the result is not fully determined. This should be thought of as an underspecified algorithm, rather than as a random result. The algebraic structure of non-deterministic computation could be described in terms of a commutative monoid, that is, a binary operation $\oplus : (0 \mid 0, 0)$ satisfying commutativity (aka medial) and unit laws

$$(u \oplus x) \oplus (y \oplus z) = (u \oplus y) \oplus (x \oplus z) \quad x \oplus \text{undef} = x \quad \text{undef} \oplus x = x$$

(although one would typically also impose idempotence, $x \oplus x = x$). We can interpret \oplus as a linear map between C*-algebras: let $\oplus : M_1 \oplus M_1 \rightarrow M_1$ be given by $\oplus(\alpha, \beta) = \alpha + \beta$. This is completely positive, but it does not preserve units.

Let us add axioms imposing that \oplus commutes with the other operations (`measure`, `apply` and `new`). I conjecture that this yields an equational characterization of not-necessarily-unital completely positive maps. For instance, one consequence of the commutativity axioms is

$$\text{measure}(a, x, y) = \text{measure}(a, x, \text{undef}) \oplus \text{measure}(a, \text{undef}, y)$$

so that all terms can be rearranged into an operator-sum, as in Choi’s theorem. However, this still does not give a proper motivation for the class of maps, because it is not clear why idempotence ($x \oplus x = x$) should be omitted in a theory of non-determinism.

Higher order computation and monads By understanding quantum computation as an algebraic effect, we are able to begin applying other techniques developed for algebraic effects in general, such as compiler optimizations and static analyses [12] and normalization by evaluation [1]. Another general method is building models of higher order computation with computational effects [19] using monads.

Indeed, the equational theory of quantum computation is not a theory in the sense of classical universal algebra, but rather a theory enriched in the functor category $[\mathbf{Bij}, \mathbf{Set}]$, which is why we used actions of \mathbf{Bij} to discuss models. Functors $(\mathbf{Bij} \rightarrow \mathbf{Set})$ are called ‘species of structure’, and have been used to analyze aspects of quantum computation including variations on the Fock space construction (e.g. [9]).

To be precise, it is a general fact that to give a parameterized algebraic theory in the style of Section 2 is to give a sifted-colimit-preserving strong monad on the symmetric monoidal closed category $[\mathbf{Bij}, \mathbf{Set}]$ (see e.g. [21, Cor. 1], [22, §VII] for details). Our theory induces a monad T on the functor category $(\mathbf{Bij} \rightarrow \mathbf{Set})$ that is determined by its action on functors $F : \mathbf{Bij} \rightarrow \mathbf{Set}$ of the form $F = \mathbf{Bij}(m_1, -) + \dots + \mathbf{Bij}(m_k, -)$, since every functor is a sifted colimit of functors of this form. Let

$$T(F)p = \{x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash t\} / \cong \cong \mathbf{Cstar}_{\text{CPU}}(M_{2^{m_1}} \oplus \dots \oplus M_{2^{m_k}}, M_{2^p}).$$

It remains to investigate the behaviour of this model at higher types.

Acknowledgements This work is supported by the ERC Grant QCLS, and I am very grateful to the QCLS team in Nijmegen for our many discussions. Thanks also to the QPL reviewers.

References

- [1] Daniel Ahman & Sam Staton (2013): *Normalization by evaluation and algebraic effects*. In: *Proc. MFPS XXIX*, pp. 51–69.
- [2] P. N. Benton, Gavin M. Bierman, Valeria de Paiva & Martin Hyland (1993): *A term calculus for intuitionistic linear logic*. In: *Proc. TLCA 1993*, pp. 75–90.
- [3] Stephen L. Bloom & Zoltan Esik (1993): *Iteration Theories: The Equational Logic of Iterative Processes*. Springer.
- [4] Bob Coecke & Ross Duncan (2008): *Interacting Quantum Observables*. In: *Proc. ICALP 2008*, pp. 298–310.
- [5] Vincent Danos, Elham Kashefi & Prakash Panangaden (2007): *The measurement calculus*. *J. ACM* 54(2).
- [6] Peter A Fillmore (1996): *A User’s Guide to Operator Algebras*. Wiley-Interscience.
- [7] Simon J Gay (2006): *Quantum programming languages: survey and bibliography*. *Mathematical Structures in Computer Science* 16(4), pp. 681–600.
- [8] Alexander S. Green, Peter LeFanu Lumsdaine, Neil J. Ross, Peter Selinger & Benoît Valiron (2013): *Quipper: a scalable quantum programming language*. In: *Proc. PLDI 2013*, pp. 333–342.
- [9] Madalin Guta & Hans Maassen (2002): *Symmetric Hilbert spaces arising from species of structures*. *Mathematische Zeitschrift* 239, pp. 477–513.
- [10] Chris Heunen, Aleks Kissinger & Peter Selinger: *Completely positive projections and biproducts*. ArXiv:1308.4557.
- [11] G. Janelidze & G.M. Kelly (2001): *A note on actions of a monoidal category*. *Theory Appl. Categ.* 9(4).
- [12] Ohad Kammar & Gordon D. Plotkin (2012): *Algebraic foundations for effect-dependent optimisations*. In: *POPL 2012*, pp. 349–360.
- [13] F E J Linton (1966): *Autonomous equational categories*. *J. Math. Mech.* 15, pp. 637–642.
- [14] Paul-André Mellies (2014): *Local stores in string diagrams*. In: *Proc. RTA-TLCA 2014*. To appear; lecture notes at tinyurl.com/mellies-itu-2011.

- [15] Michael A. Nielsen & Isaac L. Chuang (2011): *Quantum Computation and Quantum Information*. CUP.
- [16] Vern Paulsen (2003): *Completely Bounded Maps and Operator Algebras*. CUP.
- [17] Gordon D. Plotkin & J Power (2002): *Notions of computation determine monads*. In: *Proc. FOSSACS'02*.
- [18] Gordon D Plotkin & John Power (2003): *Algebraic operations and generic effects*. *Applied Categorical Structures* 11(1), pp. 69–94.
- [19] John Power (2006): *Generic models for computational effects*. *Theor. Comput. Sci.* 364(2), pp. 254–269.
- [20] Peter Selinger (2004): *Towards a quantum programming language*. *Mathematical Structures in Computer Science* 14(4), pp. 527–586.
- [21] Sam Staton: *An algebraic presentation of predicate logic*. In: *FOSSACS 2013*.
- [22] Sam Staton: *Instances of computational effects*. In: *LICS 2013*.
- [23] Sam Staton (2013): *Freyd categories are enriched Lawvere theories*. In: *Proceedings of Workshop on Algebra, Coalgebra and Topology, ENTCS 303*, pp. 197–206.
- [24] Sam Staton & Paul B Levy (2013): *Universal properties for impure programming languages*. In: *POPL 2013*.
- [25] André van Tonder (2004): *A Lambda Calculus for Quantum Computation*. *SIAM J. Comput.* 33(5), pp. 1109–1135.
- [26] Juliana Kaizer Vizzotto, Giovanni Rubert Librelotto & Amr Sabry (2009): *Reasoning about general quantum programs over mixed states*. In: *SBMF 2009*.
- [27] Mingsheng Ying & Yuan Feng (2009): *An Algebraic Language for Distributed Quantum Computing*. *IEEE Trans. Computers* 58(6), pp. 728–743.
- [28] Mingsheng Ying, Nengkun Yu & Yuan Feng (2014): *Alternation in quantum programming: from superposition of data to superposition of programs*. ArXiv:1402.5172.

A Proof of completeness

We use this appendix to give more details of our proof of the completeness theorem (2) from Section 3.

Theorem 2. 1. For any completely positive unital map $f : M_{2^p} \rightarrow M_{2^{m_1}} \oplus \cdots \oplus M_{2^{m_k}}$ there is a term $x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash t$ such that $f = \llbracket t \rrbracket$.

2. If $\Gamma \mid \Delta \vdash t, u$ and $\llbracket t \rrbracket = \llbracket u \rrbracket$ then $\Gamma \mid \Delta \vdash t = u$ is derivable.

(NB. Part (1) of the theorem is essentially Selinger’s [20, Thm. 6.14].)

As a first step we consider the case without new.

Lemma 3. 1. For any *-homomorphism $f : M_{2^p} \rightarrow M_{2^{m_1}} \oplus \cdots \oplus M_{2^{m_k}}$ there is a term $x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash t$ not involving new such that $f = \llbracket t \rrbracket$.

2. If $\Gamma \mid \Delta \vdash t, u$ and t and u do not contain new, and $\llbracket t \rrbracket = \llbracket u \rrbracket$ then $\Gamma \mid \Delta \vdash t = u$ is derivable.

Proof notes for Lemma. Recall that if $m_1 \dots m_k, p$ are non-zero natural numbers then the set of Bratelli diagrams, $\mathbf{Brat}(m_1 \dots m_k, p)$, comprises k -tuples of natural numbers $s_1 \dots s_k$ (‘partial multiplicities’) such that $\sum_{i=1}^k s_i m_i = p$. There is a function $\mu : \mathbf{Brat}(m_1 \dots m_k, p) \rightarrow \mathbf{Cstar}(M_{m_1} \oplus \cdots \oplus M_{m_k}, M_p)$ where $\mu(s_1 \dots s_k)(A_1, \dots, A_k)$ is the block diagonal matrix formed by s_1 copies of A_1 followed by s_2 copies of A_2 and so on.

The following result is standard (e.g. [6, 1.1.3]).

Proposition 4. The function $\mu : \mathbf{Brat}(m_1 \dots m_k, p) \rightarrow \mathbf{Cstar}(M_{m_1} \oplus \cdots \oplus M_{m_k}, M_p)$ has a retraction: there is a function $\rho : \mathbf{Cstar}(M_{m_1} \oplus \cdots \oplus M_{m_k}, M_p) \rightarrow \mathbf{Brat}(m_1 \dots m_k, p)$ such that $\rho\mu = \text{id}$. Moreover $\rho(f) = \rho(g)$ if and only if there is a $p \times p$ unitary U such that $f(\vec{A}) = U^*(g(\vec{A}))U$. \square

Consider a context $(m_1 \dots m_k | p)$. Suppose that the second order variables are arranged in order of increasing valence.

We now show that the following are in bijective correspondence:

1. Sequences of partial multiplicities in $\mathbf{Brat}(2^{m_1} \dots 2^{m_k}, 2^p)$.
2. Terms built only from measure and variables, such that the variables (of both kinds) appear in the same order as in the context.

This is proved by induction on the size of p . Given a Bratelli diagram $\vec{s} \in \mathbf{Brat}(2^{m_1} \dots 2^{m_k}, 2^p)$ we write $x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash \text{Measure}((\vec{s}))$ for the corresponding term.

We can thus define a section of the semantic map

$$\llbracket - \rrbracket : \{t \mid (x_1 : m_1 \dots x_k : m_k \mid a_1 \dots a_p \vdash t)\} \rightarrow \mathbf{Cstar}(M_{2^m} \oplus M_{2^{m_k}}, M_{2^p})$$

as follows. Every *-homomorphism $f : M_{2^{m_1}} \oplus \dots \oplus M_{2^{m_k}} \rightarrow M_{2^p}$, factors as $f(x) = U^* \cdot (\mu \vec{s})(x) \cdot U$, for a $2^p \times 2^p$ unitary U , so that $f = \llbracket \text{apply}_U(\vec{a}, \vec{a}.\text{Measure}(\vec{s})) \rrbracket$. Thus the first part of the lemma is proved.

We show the second part of the lemma in two stages.

1. We show that the section of the semantic map doesn't depend on the choice of U ;
2. We show that the section is a surjection, i.e. that every term is equal to one of the form $\text{apply}_U(\vec{a}, \vec{a}.\text{Measure}(\vec{s}))$.

First, then, we show that for any Bratelli diagram $\vec{s} \in \mathbf{Brat}(2^{m_1} \dots 2^{m_k}, 2^p)$ and any $2^p \times 2^p$ unitaries U, V , if

$$U^* \cdot (\mu \vec{s})(-) \cdot U = V^* \cdot (\mu \vec{s})(-) \cdot V \quad \text{as linear maps } M_{2^{m_1}} \oplus \dots \oplus M_{2^{m_k}} \rightarrow M_{2^p}. \quad (14)$$

then we can derive the equality

$$\text{apply}_U(\vec{a}, \vec{a}.\text{Measure}(\vec{s})) = \text{apply}_V(\vec{a}, \vec{a}.\text{Measure}(\vec{s})) \quad \text{between terms.} \quad (15)$$

This fact can be understood along the same lines as the freedom in the operator-sum representation of completely positive maps (e.g. [15, Ch. 8]). We have that if $(\mu \vec{s})(-) = U^* \cdot (\mu \vec{s})(-) \cdot U$ then U must be built from a tensor products of unitaries that are conditional on qubits that are being measured and that act on qubits that are being discarded.

Thus the section of the semantic map does not depend on the choice of U .

It remains for us to show that the section is a surjection, i.e. that every term is equivalent to one of the form $\text{apply}_U(\vec{a}, \vec{a}.\text{Measure}(\vec{s}))$. In brief, we use the equations to rearrange a term as follows:

1. First we push all measure operations inside apply operations.
2. Next we arrange all the second order variables to appear in the designated order by using controlled nots (possibly controlled-controlled-nots etc).
3. We arrange all the parameter variables to appear in the designated order, by using controlled swaps.
4. A sequence of apply operations can be combined into one apply operation, using tensors and composition.

This concludes our proof of Lemma 3. □

We now turn to prove Theorem 2, extending the Lemma to consider new and completely positive maps.

Recall the following variant of Stinespring's theorem (e.g. [16, p. 45]). Let \tilde{p} be a natural number. If $f : A \rightarrow M_{\tilde{p}}$ is completely positive and unital then there is a natural number $q \geq \tilde{p}$ and a *-homomorphism g making the following diagram commute:

$$\begin{array}{ccc} A & & \\ g \downarrow & \searrow f & \\ M_q & \longrightarrow & M_{\tilde{p}} \end{array}$$

where the unlabelled horizontal arrow is $A \mapsto A|_{\tilde{p}}$, where $A|_{\tilde{p}}$ comprises the first \tilde{p} rows and columns of the $q \times q$ matrix A . Moreover q can be chosen as the minimal such number: if $r \geq \tilde{p}$ and a *-homomorphism $h : A \rightarrow M_r$ is such that $h(-)|_{\tilde{p}} = f(-)$ then $r \geq q$ and there is an r -ary unitary U such that $g(a) = (U(ha)U^*)|_q$. As a diagram:

$$\begin{array}{ccccc} & & A & & \\ & \nearrow h & \downarrow g & \searrow f & \\ M_r & \xrightarrow{U-U^*} & M_r & \longrightarrow & M_q & \longrightarrow & M_{\tilde{p}} \end{array}$$

Now if f is as in the statement of our theorem (so $A = M_{2^{m_1}} \oplus \dots \oplus M_{2^{m_k}}$ and $\tilde{p} = 2^p$) then we can use the minimal dilation to show that f is definable, provided q is a power of 2. This is because the downward arrow can be defined using new and the *-homomorphism can be defined according to Lemma 3. Moreover the minimality essentially gives us the second part of the theorem.

However, q need not be a power of 2. For instance, consider example (13): its minimal dilation is a *-homomorphism $M_2 \oplus M_1 \oplus M_1 \rightarrow M_3$.

We resolve this as follows. First we note that the domain of f (A) is not 0, for if $A \cong 0$ then, since f is unital, $2^p = 0$, which is absurd. So $k \neq 0$ and we pick $i \leq k$, giving us a *-homomorphism map $\pi_i : A \rightarrow M_{2^{m_i}}$. Now the Stinespring dilation gives the following factorization of f :

$$\begin{array}{ccc} A & \xrightarrow{((2^{m_i}) \cdot g, (2^l - q) \cdot \pi_i)} & 2^{m_i} \cdot M_q \oplus (2^q - q) \cdot M_{2^{m_i}} \\ & \searrow g & \downarrow \pi_1 \\ & & M_q \\ & \searrow f & \downarrow \\ & & M_{2^p} \end{array}$$

Note that the horizontal maps are *-homomorphisms, and the vertical maps are restrictions. Thus the completely positive map f is definable: there is a term t such that $f = \llbracket t \rrbracket$.

Now suppose that t and u are terms with the same interpretation ($\llbracket t \rrbracket = \llbracket u \rrbracket$). We will show that their equality can be derived. We can assume that the terms are written with the new's on the outside, by using the commutativity equations, so that $t = \text{new}(a_1 \dots a_l.t')$ where t' doesn't contain new. Using the equation $\text{new}(a.\text{discard}(a,x)) = x$ we can add new's to t or u so that they both have the same number of new's (l , say). In summary we have *-homomorphisms $\llbracket t' \rrbracket : A \rightarrow M_{2^l}$ and $\llbracket u' \rrbracket : A \rightarrow M_{2^l}$, such that

$$\llbracket t \rrbracket = A \xrightarrow{\llbracket t' \rrbracket} M_{2^l} \rightarrow M_{2^p} \quad \text{and} \quad \llbracket u \rrbracket = A \xrightarrow{\llbracket u' \rrbracket} M_{2^l} \rightarrow M_{2^p}.$$

By Stinespring's theorem there is a natural number q and unitaries U and V on 2^l such that

$$\begin{array}{ccccccc} A & \xrightarrow{\llbracket t' \rrbracket} & M_{2^l} & \xrightarrow{U-U^*} & M_{2^l} & \xrightarrow{\quad} & M_{2^p} \\ & \searrow \llbracket u' \rrbracket & M_{2^l} & \xrightarrow{V-V^*} & M_{2^l} & \xrightarrow{\quad} & M_{2^p} \end{array}$$

where the composites of the inner square are $*$ -homomorphisms. Since U and V do not affect M_{2^p} , we have

$$\text{new}(\vec{a}.t) = \text{new}(\vec{a}.\text{apply}_U(\vec{b}, \vec{b}.t')) \quad \text{and} \quad \text{new}(\vec{a}.u) = \text{new}(\vec{a}.\text{apply}_V(\vec{b}, \vec{b}.u'))$$

This follows from the following lemma.

If there are n parameter variables in context and U is a unitary of arity 2^n then we write $\text{apply}_U(t)$ as shorthand for $\text{apply}(\vec{a}^n, \vec{a}.t)$.

Lemma 5. *Let $m < n$ and let U be a unitary on 2^n square matrices that fixes the top 2^m rows. Then*

$$x : n \mid a_1 \dots a_m \vdash \text{new}(a_{m+1} \dots a_n.\text{apply}_U(x(\vec{a}))) = \text{new}(a_{m+1} \dots a_n.x(\vec{a}))$$

is derivable.

This lemma follows from the law $\text{new}(a.\text{apply}_{D(U,V)}(a, \vec{b}, a\vec{b}.x(a, \vec{b}))) = \text{new}(a.\text{apply}_V(\vec{b}, \vec{b}.x(a, \vec{b})))$.

So we can assume wlog that U and V are identities, and

$$A \begin{array}{c} \xrightarrow{\llbracket t' \rrbracket} M_{2^l} \\ \xrightarrow{\llbracket u' \rrbracket} M_{2^l} \end{array} \begin{array}{c} \searrow \\ \searrow \end{array} M_q.$$

We now use another lemma:

Lemma 6. *If the following diagram commutes and f, g are $*$ -homomorphisms then f factors through the block diagonal map $M_p \oplus M_q \rightarrow M_{p+q}$ (measurement).*

$$A \begin{array}{c} \xrightarrow{f} M_{p+q} \\ \searrow g \\ \downarrow \\ M_p \end{array}$$

Returning to our main argument, the lemma (6) gives us f, f', g, h and the following situation:

$$A \begin{array}{c} \xrightarrow{(f,g)} M_q \oplus M_{2^l-q} \\ \xrightarrow{(f',h)} M_q \oplus M_{2^l-q} \end{array} \begin{array}{c} \xrightarrow{\llbracket t' \rrbracket} M_{2^l} \\ \xrightarrow{\llbracket u' \rrbracket} M_{2^l} \end{array} \begin{array}{c} \searrow \\ \searrow \end{array} M_q$$

We immediately have $f = f'$. Now,

$$t = \text{new}(\vec{a}.\text{new}(b.\text{measure}(b, t', u')))) = \text{new}(\vec{a}.\text{new}(b.\text{apply}_U(\text{measure}(b, t', u'))))$$

where U is the unitary that swaps the two occurrences of $(2^l - q)$ in $M_{q+(2^l-q)+q+(2^l-q)} = M_{2^l} \otimes M_2$. This equation is a consequence of Lemma 5: U fixes the upper 2^p rows.

It remains for us to show that $\text{apply}_U(\text{measure}(b, t', u')) = \text{measure}(b, u', t')$. These terms don't involve new , so by Lemma 3 the equality is derivable if and only if the corresponding $*$ -homomorphisms are equal. Indeed they are, because

$$\begin{aligned} \llbracket \text{apply}_U(\text{measure}(b, t', u')) \rrbracket(a) &= U \begin{pmatrix} \llbracket t' \rrbracket(x) & 0 \\ 0 & \llbracket u' \rrbracket(x) \end{pmatrix} U^* = U \begin{pmatrix} f(x) & 0 & 0 & 0 \\ 0 & g(x) & 0 & 0 \\ 0 & 0 & f(x) & 0 \\ 0 & 0 & 0 & h(x) \end{pmatrix} U^* \\ &= \begin{pmatrix} f(x) & 0 & 0 & 0 \\ 0 & h(x) & 0 & 0 \\ 0 & 0 & f(x) & 0 \\ 0 & 0 & 0 & g(x) \end{pmatrix} = \begin{pmatrix} \llbracket u' \rrbracket(x) & 0 \\ 0 & \llbracket t' \rrbracket(x) \end{pmatrix} = \llbracket \text{measure}(b, u', t') \rrbracket(x) \end{aligned}$$

This concludes our proof of Theorem 2.

Mixed quantum states in higher categories

Chris Heunen

Department of Computer Science,
University of Oxford
`chris.heunen@cs.ox.ac.uk` *

Jamie Vicary

Centre for Quantum Technologies,
National University of Singapore
Department of Computer Science,
University of Oxford
`jamie.vicary@cs.ox.ac.uk`

Linde Wester

Department of Computer Science,
University of Oxford
`lindewester@gmail.com`

There are two ways to describe the interaction between classical and quantum information categorically: one based on completely positive maps between Frobenius algebras, the other using symmetric monoidal 2-categories. This paper makes a first step towards combining the two. The integrated approach allows a unified description of quantum teleportation and classical encryption in a single 2-category, as well as a universal security proof applicable simultaneously to both scenarios.

1 Introduction

In the categorical approach to quantum information [1], there are two main approaches to modelling the interaction between classical and quantum data, which can be summarized as follows:

- Commutative Frobenius algebras model classical data, noncommutative Frobenius algebras model quantum data, completely positive maps model computational processes. The resulting compact category $\mathbf{CP}^*[\mathbf{FHilb}]$, reviewed in Section 1.1, incorporates both pure and mixed states in a single setting, while admitting a graphical calculus [8–11, 13, 14, 16, 21].
- Objects model classical data, 1-morphisms model quantum data, 2-morphisms model computational processes. The resulting symmetric monoidal 2-category, reviewed in Section 1.2, provides universal syntactic models that can encode entire procedures as single equations [4, 19, 22].

This article makes a first step towards combining both approaches while retaining the advantages of each. Section 2 introduces a procedure that turns a suitable symmetric monoidal category \mathbf{C} into a symmetric monoidal 2-category $2[\mathbf{C}]$. It is based on the well-known structure of bimodules and homomorphisms, but with a new definition of bimodule composition in terms of splitting of an idempotent.

Section 3 investigates basic properties of $2[\mathbf{CP}^*[\mathbf{FHilb}]]$. We show that on a large domain, which is sufficient for the intended application to quantum information, the 2-category is well-defined. We also prove the surprising result that every finite groupoid gives rise to an object in $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ in a canonical way, suggesting that the 2-category has a rich structure waiting to be explored.

Finally, Section 4 demonstrates the advantages of our combined approach. We obtain:

- An elegant abstract definition of measurement, that in $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ comes down to the usual mixed-state notion of positive operator-valued measurement.
- A single equation whose solutions in $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ simultaneously include implementations of quantum teleportation and of classical encrypted communication.
- A single proof of security that applies simultaneously to both procedures.

*Supported by the Engineering and Physical Sciences Research Council Fellowship EP/L002388/1.
We thank Aleks Kissinger for useful discussions.

There are several interesting directions for future work:

- How can objects of $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ be classified?
- Is there a direct construction $\mathbf{C} \mapsto \mathbf{Mix}[\mathbf{C}]$ of 2-categories such that $2[\mathbf{CP}^*[\mathbf{C}]] \cong \mathbf{Mix}[2[\mathbf{C}]]$?
- What are nonstandard models such as $2[\mathbf{CP}^*[\mathbf{Rel}]]$ like?
- Are there nonstandard solutions of the teleportation equation in $2[\mathbf{CP}^*[\mathbf{FHilb}]]$, which are neither pure-state quantum teleportation or encrypted communication, but some hybrid process?

1.1 The \mathbf{CP}^* -construction

Categorical quantum mechanics deals with dagger monoidal categories [1], which admit a graphical calculus; see [17]. Within this well-documented setting, let us very briefly recall the \mathbf{CP}^* construction from the perspective of [14, Lemma 1.2]; for more details we refer to [10, 14, 16]. This construction turns a dagger compact category \mathbf{C} into a new category $\mathbf{CP}^*[\mathbf{C}]$. An object in $\mathbf{CP}^*[\mathbf{C}]$ is a *special dagger Frobenius algebra* in \mathbf{C} : an object A with morphisms $\alpha: A \otimes A \rightarrow A$ and $\beta: I \rightarrow A$ satisfying the *specialness* condition $\beta \circ \alpha = \text{id}$, as well as the *dagger Frobenius algebra* laws:

$$\text{Frobenius law: } \begin{array}{c} \text{dot} \\ \text{line} \end{array} \circ \begin{array}{c} \text{line} \\ \text{dot} \end{array} = \begin{array}{c} \text{dot} \\ \text{line} \end{array} \circ \begin{array}{c} \text{line} \\ \text{dot} \end{array} \quad \text{Specialness: } \begin{array}{c} \text{dot} \\ \text{line} \end{array} \circ \begin{array}{c} \text{line} \\ \text{dot} \end{array} = \text{line} \quad \text{Dagger Frobenius law: } \begin{array}{c} \text{dot} \\ \text{line} \end{array} \circ \begin{array}{c} \text{line} \\ \text{dot} \end{array} = \begin{array}{c} \text{dot} \\ \text{line} \end{array} \circ \begin{array}{c} \text{line} \\ \text{dot} \end{array} \quad (1)$$

Commutative such objects are also called *classical structures*. A morphism $(A, \alpha, \beta) \rightarrow (B, \alpha, \beta)$ in $\mathbf{CP}^*[\mathbf{C}]$ is a morphism $f: A \rightarrow B$ in \mathbf{C} satisfying the *complete positivity* condition

$$\begin{array}{c} \text{line} \\ \text{box } f \\ \text{line} \end{array} = \begin{array}{c} \text{box } g^\dagger \\ \text{box } g \\ \text{line} \end{array} \quad (2)$$

for some morphism $g: A \otimes B^* \rightarrow X$ in \mathbf{C} . This gives a well-defined dagger compact category $\mathbf{CP}^*[\mathbf{C}]$ with the following basic interpretation:

Category theory	Geometry	Interpretation
Commutative objects	Lines with commutative dots	Classical information
Noncommutative objects	Lines with noncommutative dots	Quantum information
Morphisms	Vertices	Physical operations

The \mathbf{CP}^* -construction is of fundamental importance because it turns a category of pure states and processes into a category of mixed states: applying the \mathbf{CP}^* -construction to the category \mathbf{FHilb} of finite-dimensional Hilbert spaces and linear maps results in the category $\mathbf{CP}^*[\mathbf{FHilb}]$ of finite-dimensional C^* -algebras and completely positive maps.

1.2 Higher quantum theory

Higher quantum theory [22, 23] separates classical and quantum information by replacing monoidal categories by monoidal weak 2-categories. These also have a graphical notation [15]:

Category theory	Geometry	Interpretation
Objects	Surfaces	Classical information
1-Morphisms	Lines	Quantum systems
2-Morphisms	Vertices	Physical operations

Graphically, composition of 1-morphisms is indicated by horizontal juxtaposition, and composition of 2-morphisms by vertical juxtaposition. The tensor product is given by ‘overlying’ regions one above the other, perpendicular to the plane of the page.

Just like in the 1-categorical case, the diagrams are interpreted as describing sequences of events taking place over time, with time running from bottom to top. A dagger provides a formal time-reversal of 2-morphisms, represented graphically by reflecting a diagram about a horizontal axis.

Definition 1. A *dagger 2-category* is a 2-category equipped with an involutive operation \dagger on 2-morphisms, such that $\mu^\dagger: G \Rightarrow F$ for all $\mu: F \Rightarrow G$, in a way that is functorial and compatible with the rest of the monoidal 2-category structure.

The core theory uses the graphical components summarized below, motivated in detail in [22].



Definition 2. An object in a symmetric monoidal 2-category has a *topological boundary* when it is equipped with data (4)-(6) satisfying the following axioms, which amount to saying that the boundary of a classical system is topological and that holes can be eliminated:

$$\begin{array}{c} \text{[Diagram: Gray square with a semi-circular bump at the top]} \end{array} = \begin{array}{c} \text{[Diagram: Vertical line]} \end{array} = \begin{array}{c} \text{[Diagram: Gray square with a semi-circular hole at the bottom]} \quad \begin{array}{c} \text{[Diagram: Gray square with a semi-circular bump at the bottom]} \end{array} = \begin{array}{c} \text{[Diagram: Vertical line]} \end{array} = \begin{array}{c} \text{[Diagram: Gray square with a semi-circular hole at the top]} \quad (7)$$

$$\begin{array}{c} \text{[Diagram: Gray square with a circular hole]} \end{array} = \begin{array}{c} \text{[Diagram: Solid gray square]} \quad \begin{array}{c} \text{[Diagram: Gray square with a figure-eight hole]} \end{array} = \begin{array}{c} \text{[Diagram: Gray square with a U-shaped hole]} \quad (8)$$

Whenever we make use of the above graphical notation, it is understood that we are depicting an object with topological boundary in a symmetric monoidal dagger 2-category.

2 The $2[-]$ construction

This section introduces a construction that turns a monoidal category \mathbf{C} into a 2-category $2[\mathbf{C}]$, in such a way that $2[\mathbf{CP}^*[\mathbf{C}]]$ has the appropriate structure to express the teleportation equation solely in terms of

objects and morphisms. The idea is to adapt the well-known algebraic construction of rings, bimodules, and bimodule homomorphisms [12]. In Section 2.1 we will see how our construction is defined, and in Section 2.2 we will see that objects in $2[\mathbf{C}]$ have a topological boundary in the sense of Definition 2.

2.1 Bimodules and composition

Definition 3. Let $(C, \curvearrowright, \circlearrowleft)$ and $(D, \curvearrowright, \bullet)$ be dagger Frobenius algebras in a dagger monoidal category. A *dagger C - D -bimodule* is a morphism \mathbf{M} satisfying:

$$(9)$$

We also call the object M the bimodule, and the map \mathbf{M} its *action*, and write $\mathbf{M}_\bullet = \mathbf{M}(|\bullet\rangle)$ and $\circ\mathbf{M} = \mathbf{M}(\langle\bullet|)$. A *homomorphism* of dagger C - D -bimodules is a morphism $f: M \rightarrow M'$ that respects that actions by satisfying $f\mathbf{M} = \mathbf{M}'(\text{id}_C \otimes f \otimes \text{id}_D)$.

If M is a C - D -bimodule, and N is a D - E -bimodule, the standard algebraic construction of *tensor product* gives a C - E -bimodule $M \otimes_D N$; see [12, Section 4.5]. It is constructed by forcing the right D -action on M and the left D -action on N to cooperate. More precisely, it is the coequalizer of the two morphisms $M \otimes D \otimes N \rightarrow M \otimes N$ induced by the two D -actions.

One way to guarantee the existence of such a coequalizer is to require that some morphisms have a sensible notion of *image*, as in the following definition and lemma. Recall that an endomorphism $p: A \rightarrow A$ is a *dagger idempotent* when $p^2 = p = p^\dagger$. A dagger idempotent p *splits* when $p = ii^\dagger$ and $i^\dagger i = \text{id}$ for some morphism i , called the *image* of p . Split idempotents are a special case of dagger coequalizers [18]: a dagger idempotent $p: A \rightarrow A$ splits if and only if p and id_A have a dagger coequalizer i^\dagger .

Definition 4. A dagger monoidal category *has dagger Frobenius images* when for all classical structures $(C, \curvearrowright, \circlearrowleft)$, $(D, \curvearrowright, \bullet)$, $(E, \curvearrowright, \bullet)$, for all C - D -bimodules \mathbf{M} and all D - E -bimodules \mathbf{N} , the following dagger idempotent splits:

$$(10)$$

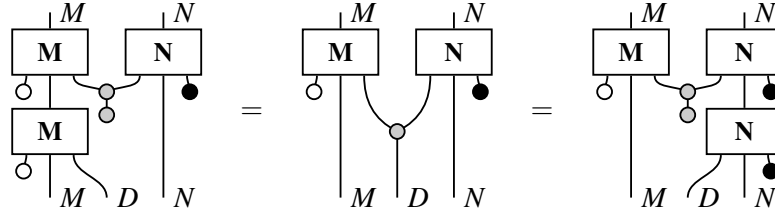
Notice that this morphism is indeed dagger idempotent by (9).

We denote the image of (10) by $i: M \otimes N \rightarrow M \otimes N$. It is a dagger C - E -bimodule:

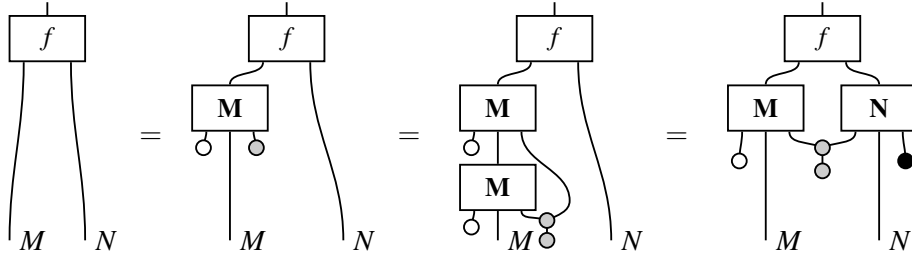
$$(11)$$

Lemma 5. If (10) splits with image i , then i^\dagger is a coequalizer of $\circ \mathbf{M} \otimes \text{id}_N$ and $\text{id}_M \otimes \mathbf{N}_\bullet$.

Proof. Observe that $i^\dagger(\circ \mathbf{M} \otimes \text{id}_N) = i^\dagger(\text{id}_M \otimes \mathbf{N}_\bullet)$ because $(\mathbf{M} \circ \mathbf{N})(\circ \mathbf{M} \otimes \text{id}_N) = (\text{id}_M \otimes \mathbf{N}_\bullet)(\circ \mathbf{M} \otimes \text{id}_N)$:



Suppose that $f(\circ \mathbf{M} \otimes \text{id}_N) = f(\text{id}_M \otimes \mathbf{N}_\bullet)$. Then f factors through i^\dagger as $f = f i i^\dagger$:



This mediating map is unique: if $f = m i^\dagger$, then $m = m i^\dagger i = f i$. □

We can use this technique to re-prove many standard results about bimodules in a graphical way, such as the following simple result, that will be useful later.

Lemma 6. For any dagger Frobenius algebra $(A, \circlearrowleft, \circlearrowright)$, the identity A - A -bimodule is \circlearrowleft . □

With this preparation we can now define our main construction.

Proposition 7. If \mathbf{C} is a dagger monoidal category that has dagger Frobenius images, then the following data define a symmetric monoidal (weak) 2-category $2[\mathbf{C}]$:

- objects are classical structures in \mathbf{C} ;
- 1-morphisms are dagger bimodules; the identity 1-morphism on $(C, \circlearrowleft, \circlearrowright)$ is \circlearrowleft ;
- 2-morphisms are dagger bimodule homomorphisms;
- horizontal composition of 1-morphisms is given by (11);
- horizontal composition of 2-morphisms follows from the universal property of Lemma 5;
- monoidal structure is inherited from \mathbf{C} .

More precisely, the horizontal composition of 2-morphisms $f: \mathbf{M} \rightarrow \mathbf{M}'$ and $g: \mathbf{N} \rightarrow \mathbf{N}'$ is the unique arrow making the following diagram commute:

$$\begin{array}{ccccc}
 M \otimes D \otimes N & \xrightarrow[\text{id}_M \otimes \mathbf{N}_\bullet]{\circ \mathbf{M} \otimes \text{id}_N} & M \otimes N & \xrightarrow{i^\dagger} & M \circ N \\
 (f \otimes \text{id}_D \otimes g) \downarrow & & \downarrow f \otimes g & & \downarrow f \circ g \\
 M' \otimes D \otimes N' & \xrightarrow[\text{id}_{M'} \otimes \mathbf{N}'_\bullet]{\circ \mathbf{M}' \otimes \text{id}_{N'}} & M' \otimes N' & \xrightarrow{i'^\dagger} & M' \circ N'
 \end{array}$$

Proof. For verification that these data indeed satisfy all the conditions required of a weak 2-category, see [24]. Verifying monoidality is a huge exercise that nevertheless seems straightforward enough. \square

We end this subsection by listing some properties of the $2[-]$ -construction; for proofs we refer to [24].

- If \mathbf{C} has a dagger, so does $2[\mathbf{C}]$.
- If \mathbf{C} is compact, so is $2[\mathbf{C}]$: 1-morphisms have duals that are both left and right adjoint.
- If \mathbf{C} has dagger biproducts, so do all hom-categories of $2[\mathbf{C}]$.
- The scalars of $2[\mathbf{C}]$ correspond to \mathbf{C} : there is an isomorphism $2[\mathbf{C}](I, I) \cong \mathbf{C}$ of categories.

2.2 Topological boundaries

We now show that objects of $2[\mathbf{C}]$ have topological boundaries in the sense of Definition 2.

Definition 8. The *boundaries* of a special dagger Frobenius algebra $(A, \lrcorner, \circlearrowleft)$ in \mathbf{C} are canonical bimodules $(A, \lrcorner, \circlearrowleft) \xrightarrow{L} (I, \lambda_I, \text{id}_I)$ and $(I, \lambda_I, \text{id}_I) \xrightarrow{R} (A, \lrcorner, \circlearrowleft)$ induced by the multiplication map \lrcorner :

$$\begin{array}{ccc} \boxed{\text{L}} & := & \text{diagram} \end{array} \quad \begin{array}{ccc} \boxed{\text{R}} & := & \text{diagram} \end{array} \quad (12)$$

The diagrams show the boxes L and R defined by their graphical representations. Box L is a rectangle with a solid line entering from the top and two dashed lines exiting from the bottom. Box R is a rectangle with a solid line entering from the top and two dashed lines exiting from the bottom. The first diagram shows box L equal to a cup shape with a solid line entering from the top and two dashed lines exiting from the bottom. The second diagram shows box R equal to a cap shape with a solid line entering from the top and two dashed lines exiting from the bottom.

The dashed lines indicate the monoidal unit object; we will typically omit these from now on. These boundaries are depicted as lines that bound solid regions, as shown in the diagrams (4).

Lemma 9. For a special dagger Frobenius algebra $(A, \lrcorner, \circlearrowleft)$, the composite bimodule

$$(I, \lambda_I, \text{id}_I) \xrightarrow{R} (A, \lrcorner, \circlearrowleft) \xrightarrow{L} (I, \lambda_I, \text{id}_I)$$

is isomorphic to the object A .

Proof. By Lemma 5, we must find the dagger splitting of the left-hand diagram below:

$$\text{diagram} = \text{diagram} \quad (13)$$

The diagram shows a complex shape on the left, which is a cup shape with a solid line entering from the top and two dashed lines exiting from the bottom. This is equal to a simpler cup shape on the right, also with a solid line entering from the top and two dashed lines exiting from the bottom.

By the dagger Frobenius axioms it equals the right-hand diagram. But the dagger specialness axiom makes this a dagger splitting via the object A , so the object A gives the composite of the bimodules. \square

Lemma 10. The boundaries of a special dagger Frobenius monoid $(A, \lrcorner, \circlearrowleft)$ in \mathbf{C} can be equipped with data (5) and (6) satisfying equations (7) and (8) as follows:

$$\begin{array}{cccc} \begin{array}{c} \text{L} \quad \text{R} \\ \text{diagram} \\ \text{id}_A \end{array} & \begin{array}{c} \text{id}_A \\ \text{diagram} \\ \text{L} \quad \text{R} \end{array} & \begin{array}{c} \text{L} \quad \text{R} \\ \text{diagram} \\ \text{id}_I \end{array} & \begin{array}{c} \text{id}_I \\ \text{diagram} \\ \text{L} \quad \text{R} \end{array} \\ \lrcorner : A \rightarrow A \otimes A & \lrcorner : A \otimes A \rightarrow A & \circlearrowleft : I \rightarrow A & \circlearrowleft : A \rightarrow I \end{array} \quad (14)$$

The diagrams show the boundaries L and R equipped with data. The first diagram shows L and R with a cup shape and id_A. The second diagram shows id_A with a cup shape and L and R. The third diagram shows L and R with a cap shape and id_I. The fourth diagram shows id_I with a cap shape and L and R. Below each diagram is a label: lrcorner : A -> A tensor A, lrcorner : A tensor A -> A, circlearrowleft : I -> A, and circlearrowleft : A -> I.

Note that we are relying on Lemmas 6 and 9 for these definitions to make sense.

Proof. Equations (7) follow immediately from the (co)unit equation for a dagger Frobenius algebra. Equations (8) follow immediately from specialness and commutativity. \square

3 The case of Hilbert spaces

This section discusses $2[\mathbf{CP}^*[\mathbf{FHilb}]]$. We show that a substantial portion is well-defined, which we characterize in concrete terms: it consists of natural numbers, matrices of finite-dimensional C^* -algebras, and matrices of completely positive maps. Thus this is completely analogous to the case of $2[\mathbf{FHilb}]$, which is equivalent to the 2-category of 2-Hilbert spaces that consists of natural numbers, matrices of Hilbert spaces, and matrices of linear maps [3, 24]. The difficulty of establishing that $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ is well-defined in general arises because $\mathbf{CP}^*[\mathbf{FHilb}]$ does not have good completeness properties.

Lemma 11. *Not all coequalizers in the category $\mathbf{CP}^*[\mathbf{FHilb}]$ are split epimorphisms.*

Proof. Due to the dagger we may equivalently show that not all equalizers split. Suppose the completely positive maps $f = \begin{pmatrix} 1 & 1 & 0 & 0 \end{pmatrix}$ and $g = \begin{pmatrix} 0 & 0 & 1 & 1 \end{pmatrix}: \mathbb{C}^4 \rightarrow \mathbb{C}$ had an equalizer $e: A \rightarrow \mathbb{C}^4$ in $\mathbf{CP}^*[\mathbf{FHilb}]$. Then $fe = ge$, so e factors through the equalizer of f and g in \mathbf{FHilb} :

$$\begin{array}{ccccc}
 \mathbb{C}^3 & \xrightarrow{\begin{pmatrix} -1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}} & \mathbb{C}^4 & \xrightleftharpoons[g]{f} & \mathbb{C} \\
 \uparrow m & \nearrow e & & & \\
 A & & & &
 \end{array}$$

We show below that the function e is injective¹. Then the linear map m is injective, and so $\dim(A) \leq 3$. It follows that A must be a commutative C^* -algebra (as 2-by-2 matrices already have dimension 4).

Suppose $e(a) = 0$ with $a = x + iy$ for self-adjoint $x, y \in A$. Then $e(x) = e(y) = 0$ because positive maps preserve adjoints [20, page 2]. Say $x = x_+ - x_-$ for positive $x_+, x_- \in A$; then $e(x_+) = e(x_-)$. But the completely positive maps $h_{\pm}: \mathbb{C} \rightarrow A$ defined by $h_{\pm}(1) = x_{\pm}$ satisfy $eh_+ = eh_-$. So $h_+ = h_-$ since e is monic, whence $x = 0$. Similarly $y = 0$. So $\ker(e) = \{0\}$, and e is injective.

On the other hand, there are at least four completely positive maps $\mathbb{C} \rightarrow \mathbb{C}^4$, given by $x_1 = (1, 0, 1, 0)$, $x_2 = (1, 1, 0, 0)$, $x_3 = (0, 1, 0, 1)$, $x_4 = (0, 0, 1, 1)$, which satisfy $fx_i = gx_i$. No x_i is a linear combination of the others with nonnegative coefficients. If d is a retraction of e , therefore none of $dx_i \in A$ is a linear combination of the others with nonnegative coefficients, as $edx_i = x_i$. Moreover $dx_i \geq 0$ by complete positivity of d . But this contradicts $\dim(A) \leq 3$ as A is commutative. \square

It follows immediately that $\mathbf{CP}^*[\mathbf{FHilb}]$ does not have dagger coequalizers. The point is that there are nevertheless enough coequalizers for our purposes, as we show below.

3.1 Analysis

A subcollection of the objects in $\mathbf{CP}^*[\mathbf{FHilb}]$ are classical structures $C = (A, \curvearrowright, \circ)$ in \mathbf{FHilb} . Since the morphisms \curvearrowright and \circ are completely positive with respect to this algebra structure, this also gives rise to an algebra $C' = (C, \curvearrowright, \circ)$ in $\mathbf{CP}^*[\mathbf{FHilb}]$. We call this a *classical structure over itself*. Note that, up to isomorphism $C \cong \mathbb{C}^n$, such structures are just natural numbers. In this section, except for the last theorem, we restrict consideration to objects of $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ which are classical structures over themselves.

Lemma 12. *There is a one-to-one correspondence between dagger bimodules on classical structures over themselves in $\mathbf{CP}^*[\mathbf{FHilb}]$, and matrices of finite-dimensional C^* -algebras.*

¹We thank Narutaka Ozawa for this observation.

Proof. Let $\mathbf{M}: \mathbb{C}^m \otimes M \otimes \mathbb{C}^n \rightarrow M$ be a dagger \mathbb{C}^m - \mathbb{C}^n -bimodule in $\text{CP}^*[\mathbf{FHilb}]$, between classical structures over themselves. Then M is a finite-dimensional C^* -algebra, and \mathbf{M} is a completely positive map. By 16, the units of \mathbb{C}^m and \mathbb{C}^n are given by the sum over the standard basis vectors $|i\rangle$ and $|j\rangle$ of \mathbb{C}^m and \mathbb{C}^n , respectively. Set $p_{ij} = (|i\rangle\langle i|) \otimes \text{id}_M \otimes (|j\rangle\langle j|)$; this is a completely positive dagger idempotent. Hence its image $M_{ij} = p_{ij}(M)$ is a finite-dimensional C^* -algebra by a classic theorem of Choi and Effros [6]; see [14, Proposition 2.4]. Thus the bimodule \mathbf{M} gives rise to a matrix (M_{ij}) of finite-dimensional C^* -algebras.

Conversely, let (M_{ij}) be an m -by- n matrix of finite-dimensional C^* -algebras. Set $M = \bigoplus_{i,j} M_{ij}$, and define $\mathbf{M}: \mathbb{C}^m \otimes M \otimes \mathbb{C}^n \rightarrow M$ by mapping $|i\rangle \otimes a \otimes |j\rangle$ to $1_{ij} \cdot a$, where 1_{ij} is the unit of M_{ij} . In other words, $\mathbf{M}(|i\rangle \otimes a \otimes |j\rangle)$ is the projection of $a \in A$ onto the summand M_{ij} . This is a $*$ -homomorphism, and hence a completely positive map [10, Lemma 3.8]. To verify that it is a bimodule, we need to check equation (9). The first two equalities are easily verified, the third equality uses that $\mathbf{M}^\dagger: M \rightarrow \mathbb{C}^m \otimes M \otimes \mathbb{C}^n$ maps $b \in M$ to $\sum_{i,j} |i\rangle \otimes (1_{ij}b) \otimes |j\rangle$. Hence these two constructions, which are inverse to each other, are well-defined. \square

It follows that an important part of $2[\text{CP}^*[\mathbf{FHilb}]]$ is well-defined, which will be sufficient for our applications in Section 4 to quantum information and encryption.

Proposition 13. *Let $(C, \curvearrowright, \circ)$, $(D, \curvearrowright, \circ)$, $(E, \curvearrowright, \bullet)$ be classical structures over themselves in $\text{CP}^*[\mathbf{FHilb}]$, and let \mathbf{M} and \mathbf{N} be a C - D -bimodule and a D - E -bimodule. The idempotent (10) splits.*

Proof. Write $|i\rangle, |j\rangle, |k\rangle$ for the standard bases of $\mathbb{C}^l, \mathbb{C}^m, \mathbb{C}^n$. Let \mathbf{M} be a dagger \mathbb{C}^l - \mathbb{C}^m -bimodule, and let \mathbf{N} be a dagger \mathbb{C}^m - \mathbb{C}^n -bimodule in $\text{CP}^*[\mathbf{FHilb}]$. Then (10) maps $m \otimes n$ to $\sum_{i,j,k} \mathbf{M}(|i\rangle \otimes m \otimes |j\rangle) \otimes \mathbf{N}(|j\rangle \otimes n \otimes |k\rangle)$. This morphism is a sum of orthogonal projections, and hence a projection itself. As in the proof of Lemma 12, this means that it has a well-defined dagger image in $\text{CP}^*[\mathbf{FHilb}]$. The proof is finished by noticing that any classical structure in \mathbf{FHilb} is isomorphic to the commutative C^* -algebra \mathbb{C}^n for some n . \square

Lemma 14. *There is a one-to-one correspondence between homomorphisms of dagger bimodules between classical structures over themselves in $\text{CP}^*[\mathbf{FHilb}]$, and matrices of completely positive maps between finite-dimensional C^* -algebras.*

Proof. Let $f: \mathbf{M} \rightarrow \mathbf{N}$ be a homomorphism of dagger \mathbb{C}^m - \mathbb{C}^n -bimodules in $\text{CP}^*[\mathbf{FHilb}]$. Write $|i\rangle$ and $|j\rangle$ for the standard bases of \mathbb{C}^m and \mathbb{C}^n . According to the proof of Lemma 12, let $p_{ij}: M \rightarrow M_{ij}$ and $q_{ij}: N \rightarrow N_{ij}$ be the completely positive maps implementing the biproduct decompositions $M = \bigoplus_{i,j} M_{ij}$ and $N = \bigoplus_{i,j} N_{ij}$. Then $f_{ij} = q_{ij} f p_{ij}^\dagger: M_{ij} \rightarrow N_{ij}$ is an m -by- n matrix of completely positive maps. Conversely, let (f_{ij}) be an m -by- n matrix of completely positive maps $f_{ij}: M_{ij} \rightarrow N_{ij}$. According to Lemma 12 we have to find a map $f: M \rightarrow N$ for $M = \bigoplus_{i,j} M_{ij}$ and $N = \bigoplus_{i,j} N_{ij}$. Just take $f = \bigoplus_{i,j} f_{ij}$; this is well-defined because $\text{CP}^*[\mathbf{C}]$ inherits biproducts from \mathbf{C} [14, Theorem 3.2]. We have to verify that this is a well-defined homomorphism of dagger bimodules:

$$\begin{aligned} f\mathbf{M}(|i_0\rangle \otimes a \otimes |j_0\rangle) &= f(1_{i_0 j_0} a) = \bigoplus_{i,j} f_{ij}(1_{i_0 j_0} a) = f_{i_0 j_0}(1_{i_0 j_0} a) \\ &= 1_{i_0 j_0} \bigoplus_{i,j} f_{ij}(1_{ij} a) \\ &= 1_{i_0 j_0} f(a) = \mathbf{N}(|i_0\rangle \otimes f(a) \otimes |j_0\rangle) \end{aligned}$$

These two constructions are clearly inverse to each other. \square

We can now characterize a well-defined part of $2[\text{CP}^*[\mathbf{FHilb}]]$.

Theorem 15. *The following full sub-2-category is well-defined within $2[\mathbf{CP}^*[\mathbf{FHilb}]]$:*

- *objects are natural numbers m ;*
- *1-morphisms $m \rightarrow n$ are m -by- n matrices (M_{ij}) of finite-dimensional C^* -algebras;*
- *2-morphisms $(M_{ij}) \rightarrow (N_{ij})$ are m -by- n matrices (f_{ij}) of completely positive maps;*
- *horizontal composition of 1-morphisms is given by $(\bigoplus_j M_{ij} \otimes N_{jk})$;*
- *horizontal composition of 2-morphisms is given by $(\bigoplus_j f_{ij} \otimes g_{jk})$;*
- *vertical composition of 2-morphisms is given by $(g_{ij} f_{ij})$.*

Proof. It suffices to show that the correspondences of Lemmas 12 and 14 turn the compositions of Proposition 7 into the ones of the statement. Let \mathbf{M} and \mathbf{M}' be \mathbb{C}^l - \mathbb{C}^m -bimodules, and let \mathbf{N} and \mathbf{N}' be \mathbb{C}^m - \mathbb{C}^n -bimodules. These correspond to matrices of C^* -algebras, where M_{ij} is the image of $|i\rangle\langle i| \otimes \text{id}_M \otimes |j\rangle\langle j|$. Let $f: \mathbf{M} \rightarrow \mathbf{M}'$ and $g: \mathbf{N} \rightarrow \mathbf{N}'$ be bimodule homomorphisms. These correspond to matrices of completely positive maps $f_{ij}: M_{ij} \rightarrow M'_{ij}$ and $g_{ij}: N_{ij} \rightarrow N'_{ij}$. Now, by definition $M \circ N$ is the image of the map $\sum_{i,j,k} \mathbf{M}(|i\rangle \otimes [-] \otimes |j\rangle) \otimes \mathbf{N}(|j\rangle \otimes [-] \otimes |k\rangle)$. But this is just $\bigoplus_{i,j,k} M_{ij} \otimes N_{jk}$. Similarly, horizontal composition of f and g corresponds to $(\bigoplus_j f_{ij} \otimes g_{jk})$. \square

In future work we would of course like to show that $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ is completely well-defined. The first task will be to characterize its objects up to isomorphism. We offer the following theorem, which generalizes [7, Corollary 3.10], as evidence that this is a nontrivial question. Recall that a state $x \in C$ of a classical structure $(C, \blacktriangleright, \blacktriangleleft)$ in \mathbf{FHilb} is *copyable* when $\blacktriangledown(x) = x \otimes x$.

Theorem 16. *Consider a classical structure C in \mathbf{FHilb} as an object of $\mathbf{CP}^*[\mathbf{FHilb}]$. There is a one-to-one correspondence between dagger special Frobenius algebras on C in $\mathbf{CP}^*[\mathbf{FHilb}]$, and finite groupoids whose morphisms are the copyable states of C .*

Proof. Let $(\mathbb{C}^n, \blacktriangleright, \blacktriangleleft)$ be a dagger special Frobenius algebra on \mathbb{C}^n in $\mathbf{CP}^*[\mathbf{FHilb}]$. That is, it is a dagger special Frobenius algebra in \mathbf{FHilb} —i.e. a finite-dimensional C^* -algebra [21]—satisfying the extra condition that \blacktriangleright and \blacktriangleleft are completely positive maps. Since they are maps between commutative C^* -algebras, saying that \blacktriangleright and \blacktriangleleft are completely positive is the same as saying that they are linear maps that preserve positive elements [20, Theorem 1.2.4]. Write \blacktriangleright and \blacktriangleleft as a matrix using the standard basis $|i\rangle$ of \mathbb{C}^n . Then all matrix entries $\langle i | \blacktriangleright | jk \rangle$ and $\langle i | \blacktriangleleft | 1 \rangle$ are nonnegative real numbers, and conversely, if all the matrix entries are nonnegative, then the linear maps \blacktriangleright and \blacktriangleleft certainly preserve positive elements. Thus $(\mathbb{C}^n, \blacktriangleright, \blacktriangleleft)$ is a dagger special Frobenius algebra in $\mathbf{CP}^*[\mathbf{FHilb}]$ if and only if it is a C^* -algebra whose multiplication and unit have nonnegative matrix entries on the standard basis $|i\rangle$ of \mathbb{C}^n .

But then, by [2, Proposition 34], the matrix entries of \blacktriangleright must in fact be either 0 or 1 (see also [10, Section 5.2].) So we may equally think of the matrix of \blacktriangleright as a morphism in the category \mathbf{Rel} of sets and relations, where it still is a special dagger Frobenius algebra. Hence it encodes the multiplication of a groupoid whose arrows are the row indices $|i\rangle$ [13]. As units for a monoid are unique, also the matrix of \blacktriangleleft must take values in $\{0, 1\}$, and encode the identities of the groupoid. Finally, any classical structure C in \mathbf{FHilb} is isomorphic to \mathbb{C}^n for some n , with the standard basis of \mathbb{C}^n corresponding to the copyable states of C . Similarly, a $*$ -isomorphism between classical structures in \mathbf{Rel} corresponds to an isomorphism of groupoids [13, Theorem 19]. \square

We leave open the interesting question of whether isomorphism between these objects in $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ (so-called *Morita equivalence*) corresponds to *equivalence* of groupoids.

4 Applications

We now consider applications to quantum information of the well-defined part of $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ constructed in Theorem 15. We give an abstract 2-categorical definition of *measurement*, and show it recovers the ordinary notion positive operator-valued measure. We then analyze the 2-categorical equation for quantum teleportation, and show that it has solutions in our 2-category given by both encrypted communication and quantum teleportation. We then give a proof of a security property, which applies simultaneously to both types of solution.

4.1 Measurement

Earlier work on the 2-categorical syntax for pure-state quantum theory [22] demonstrated that a projective quantum measurement corresponds to a *unitary* 2-morphism which converts a local system into an extended system. Since our measurements in general are mixed, unitarity is not appropriate; instead we impose a counit-preservation condition.

Definition 17. In $2[\mathbf{CP}^*[\mathbf{FHilb}]]$, a *measurement* is a counit-preserving 2-morphism of type:

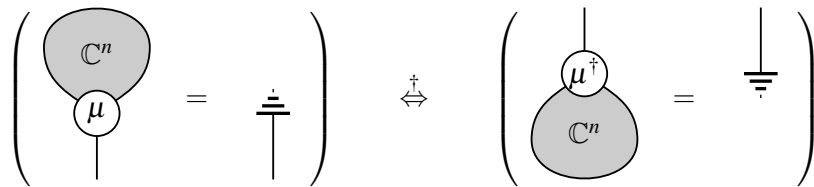

(15)

It is not ideal that we must modify the definition of a measurement in this way. The situation is analogous to the work in [19], where measurements were required to be kernel-free. That requirement can be replaced with the more elegant unitarity condition [4]. With further work we hope to show the same in the current setting, a task which is likely to require making use of a larger part of $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ than we have so-far shown to be well-defined. However, Definition 17 elegantly captures precisely the desired notion, as we now show.

Theorem 18. *Restricting to the part of $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ defined in Theorem 15, measurements on matrix algebras are exactly positive operator-valued measures.*

Proof. The 2-morphism μ is a trace-preserving completely positive map from a matrix algebra to a classical structure. Its adjoint μ^\dagger is therefore a completely positive map out of a classical structure. Such a map is completely defined by its action on the n copyable states of the classical structure, which must be sent to positive elements of $H \otimes H^*$. Thus μ^\dagger is defined by a family of n positive operators $P_i : H \rightarrow H$.

The counit-preservation condition is given by the left-hand condition below:


(16)

On the right-hand side we take the adjoint of this condition. We use the ‘earth’ symbol to represent the counit of a matrix algebra, which is just the trace map, following previous work [11]. The second equation says precisely that $\sum_i P_i = \text{id}_H$, which is exactly the condition for the family of positive operators P_i to define a positive operator-valued measurement. \square

4.2 Unification of quantum teleportation and classical encrypted communication

Definition 19. In a symmetric monoidal 2-category containing an object with a topological boundary, *teleportation* is a solution to the following equation with μ a measurement and v unitary:

$$(17)$$

Note that this definition relies on our earlier Definition 17 of a measurement.

Theorem 20. When the nontrivial region is labelled by a discrete groupoid, solutions to the teleportation equation in $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ can be obtained as follows:

1. when the incoming system is a classical structure, by implementations of classical encrypted communication using a one-time pad;
2. when the incoming system is a matrix algebra, by implementations of quantum teleportation.

Proof. We can only give a sketch here. It is already established separately that both classical encrypted communication via a one-time pad [19] and quantum teleportation [22] can be characterized exactly as solutions to this equation, in **2Rel** and **2Hilb** respectively. Both families of solutions can be embedded into $2[\mathbf{CP}^*[\mathbf{FHilb}]]$ in an appropriate fashion. \square

It is an interesting open question whether these are the only solutions, or whether solutions exist which somehow *mix* the encryption and teleportation aspects.

4.3 Security of teleportation

In both quantum teleportation and classical encrypted communication with a one-time pad, it is true that if you throw away the second half of the cryptographic resource—the entangled state or the secret key, respectively—all information about the message is lost. An abstract proof of this has already been given in the 2-categorical setup for the case of encrypted communication [19]. We now give a general proof that applies simultaneously to quantum teleportation and encrypted communication.

Theorem 21. For any solution to the teleportation equation (17), destroying the second half of the shared resource is equivalent to destroying the original message:

$$(18)$$

Proof. Adjoin a trace map to the final system on both sides of the teleportation equation (17). The map v is a family of invertible completely positive maps by Lemma 14, and thus is necessarily trace-preserving [5, Theorem 3.3]; the left-hand side therefore simplifies, giving equation (18). \square

References

- [1] S. Abramsky & B. Coecke (2004): *Categorical Semantics of Quantum Protocols*. In: *Proceedings of the 19th ACM/IEEE Symposium on Logic in Computer Science*, IEEE, pp. 415–425.
- [2] S. Abramsky & C. Heunen (2012): *H^* -Algebras and Nonunital Frobenius Algebras*. In: *Clifford Lectures, Proceedings of Symposia in Applied Mathematics* 71, American Mathematical Society, pp. 1–24.
- [3] J. C. Baez (1997): *Higher-Dimensional Algebra II: 2-Hilbert Spaces*. *Adv. Math.* 127, pp. 125–189.
- [4] K. Bar & J. Vicary (2014): *Groupoid Semantics for Thermal Computing*. <http://arxiv.org/abs/1401.3280>.
- [5] D. Cariello (2012): *An Elementary Description of the Positive Maps with Positive Inverse*. In: *National Congress of Applied Mathematics and Computation* 34. http://www.sbmec.org.br/eventos/cnmac/xxxiv_cnmac/pdf/541.pdf.
- [6] M.-D. Choi & E. G. Effros (1977): *Injectivity and Operator Spaces*. *J. Func. Anal.* 24, pp. 156–209.
- [7] B. Coecke, R. Duncan, A. Kissinger & Q. Wang (2012): *Strong Complementarity and Non-Locality in Categorical Quantum Mechanics*. In: *Logic in Computer Science* 27, IEEE, pp. 245–254.
- [8] B. Coecke & C. Heunen (2012): *Pictures of Quantum Processes in Arbitrary Dimension*. In: *Quantum Physics and Logic IX, Electronic Proceedings in Theoretical Computer Science* 95, pp. 27–35.
- [9] B. Coecke, C. Heunen & A. Kissinger (2013): *Compositional Quantum Logic*. In: *Computation, Logic, Games, and Quantum Foundations*, Springer, pp. 21–36.
- [10] B. Coecke, C. Heunen & A. Kissinger (2014): *Categories of Quantum and Classical Channels*. *Q. Inf. Processing*.
- [11] B. Coecke & S. Perdrix (2012): *Environment and Classical Channels in Categorical Quantum Mechanics*. *Logical Methods in Computer Science* 8(4), pp. 1–24.
- [12] M. Hazewinkel, N. Gubareni & V. V. Kirichenko (2004): *Algebras, Rings and Modules*. 1, Kluwer.
- [13] C. Heunen, I. Contreras & A. Cattaneo (2013): *Relative Frobenius Algebras are Groupoids*. *Journal of Pure and Applied Algebra* 217, pp. 114–124.
- [14] C. Heunen, A. Kissinger & P. Selinger (2013): *Completely Positive Projections and Biproducts*. In: *Quantum Physics and Logic X, Electronic Proceedings in Theoretical Computer Science*.
- [15] A. D. Lauda (2006): *Frobenius Algebras and Ambidextrous Adjunctions*. *Theory Appl. Categories* 16(4), pp. 84–122.
- [16] P. Selinger (2005): *Dagger Compact Categories and Completely Positive Maps*. In: *Quantum Physics and Logic III, Electronic Notes in Theoretical Computer Science* 170, pp. 139–163.
- [17] P. Selinger (2011): *A survey of graphical languages for monoidal categories*. In: *New Structures for Physics, Lecture Notes in Physics* 813, Springer, pp. 289–355.
- [18] Peter Selinger (2006): *Idempotents in Dagger Categories*. *Electronic Notes in Theoretical Computer Science* 210, pp. 107–122. *Proceedings of the 4th International Workshop on Quantum Programming Languages*.
- [19] M. Stay & J. Vicary (2013): *Bicategorical Semantics of Nondeterministic Computation*. In: *Mathematical Foundations of Programming Semantics* 29, *Elec. Notes Theor. Comp. Sci.* 298, pp. 367–382.
- [20] E. Størmer (2013): *Positive Linear Maps of Operator Algebras*. Springer.
- [21] J. Vicary (2011): *Categorical Formulation of Quantum Algebras*. *Comm. Math. Phys.* 304(3), pp. 765–796.
- [22] J. Vicary (2012): *Higher Semantics of Quantum Protocols*. In: *Proceedings of the 27th ACM/IEEE Symposium on Logic in Computer Science*, pp. 606–615.
- [23] J. Vicary (2013): *Topological Structure of Quantum Algorithms*. In: *Proceedings of the 28th ACM/IEEE Symposium on Logic in Computer Science*, pp. 93–102.
- [24] L. Wester (2013): *Categorical Models for Quantum Computing*. Master’s thesis, University of Oxford.

A 2-Categorical Analysis of Complementary Families, Quantum Key Distribution and the Mean King Problem

Krzysztof Bar

Department of Computer Science, University of Oxford
krzysztof.bar@cs.ox.ac.uk

Jamie Vicary

Centre for Quantum Technologies, University of Singapore
and Department of Computer Science, University of Oxford
jamie.vicary@cs.ox.ac.uk

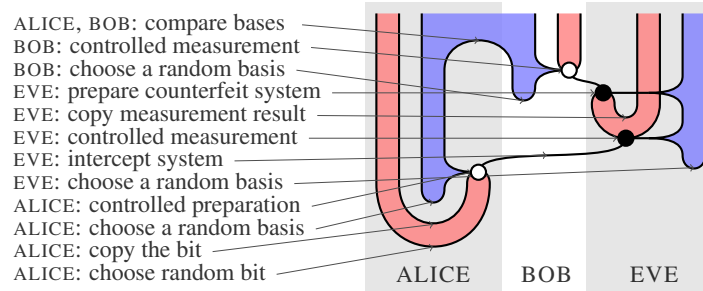
This paper explores the use of 2-categorical technology for describing and reasoning about complex quantum procedures. We give syntactic definitions of a family of complementary measurements, and of quantum key distribution, and show that they are equivalent. We then show abstractly that either structure gives a solution to the Mean King problem, which we also formulate 2-categorically.

1 Introduction

The 2-categorical approach to quantum information is now well-developed in its basic aspects [4, 12, 14]. The central aim is to use the structure of a symmetric monoidal 2-category to describe quantum procedures in an abstract way, such that the ordinary versions of these procedures are recovered when we apply the formalism in **2Hilb**, the symmetric monoidal 2-category of 2-Hilbert spaces [3]. This builds on the highly successful categorical quantum mechanics research programme of Abramsky, Coecke and collaborators [1, 2], in which quantum procedures are axiomatized in terms of monoidal 1-categories.

The key advantage of the 2-categorical setup is that many important structures, such as teleportation, dense coding and complementary observables, can be defined by single 2-categorical equations. These typically have direct physical interpretations, with the defining equation for a structure following immediately from a careful physical description of its required properties. A computer algebra system *TwoVect* [10, 11] allows these equations to be directly evaluated computationally.

In this paper, we show that the formalism can be applied successfully to more sophisticated quantum procedures: measurements in a complementary family of bases, quantum key distribution (QKD), and the Mean King problem. For each scenario, we write down a 2-categorical equation that defines the entire procedure in a precise way. For example, here is the defining diagram for BB84 QKD:



We have annotated this diagram with text to show how the different parts are to be interpreted, but we emphasize that this annotation is superfluous: everything about the flow of quantum and classical information is captured by the diagram itself. To complete the abstract definition of BB84 quantum key

distribution, we require that this diagram is equal to a second diagram which encodes the intended result of the procedure.

The main contributions of this paper are as follows:

- Definitions 15, 20, 21 and 29 give 2-categorical equations whose solutions in **2Hilb** correspond exactly to implementations of a family of complementary observables, BB84 QKD, E91 QKD, and solutions of the Mean King problem respectively.
- In Theorem 27 we show that the 2-categorical definition for a family of complementary measurements is equivalent to that for QKD. While an equivalence between these notions seems generally expected in the community, we are not able to find an existing crisp proof in the literature.
- In Theorem 37 we give a graphical proof of correctness of Klappenecker and Rottler's solution [9] to the Mean King problem. This is roughly the same complexity as the original proof, but quite different in nature. The graphical proof makes clear the role played by complementarity.

A significant result on the categorical basis of quantum key distribution was given by Coecke and Perdrix in [7, Proposition 7.4], which demonstrates the correctness of QKD based on a pair of complementary observables. Our work goes beyond this result, as we work with arbitrary families of complementary observables rather than a single pair, and we further show that every implementation of QKD gives rise to a family of complementary observables.

A primary avenue of future work arising from our results will be investigating the existence of nonstandard models. It has been shown that a category of groupoids, profunctors and spans admits combinatorial ‘toy models’ of teleportation, as solutions to a 2-categorical equation, from which ordinary quantum teleportation can be recovered by applying a 2-functor into **2Hilb** [4]. It will be interesting to use the results of this paper to investigate whether combinatorial toy models of quantum key distribution can also be built in that setting.

Remark on colour and transparency. The diagrams in this paper make essential use of colour and transparency. We therefore recommend reading this paper on a screen, or as a colour printout. For printing we recommend Adobe Reader, as some other PDF viewers do not correctly handle transparency.

1.1 The 2-category **2Hilb** of 2-Hilbert spaces

It has been argued in [14] that the 2-category of 2-Hilbert spaces [3] is the correct 2-categorical setting in which to analyze quantum informatic procedures. We recall the following construction of **2Hilb**, which is most useful for calculation. For more details, see the papers cited above.

Definition 1. The symmetric monoidal 2-category **2Hilb** has *objects* given by natural numbers, *1-morphisms* given by matrices of finite-dimensional Hilbert spaces, and *2-morphisms* given by matrices of linear maps. Details of the compositional structure of **2Hilb** are available in the references given.

This gives the formal categorical semantics that forms the primary model of our abstract syntax, introduced in Section 1.2.

1.2 The topological formalism

The basic 2-categorical structures on which the theory is built have simple graphical representations [14], thanks to the graphical notation for monoidal 2-categories. This graphical formalism involves surfaces, lines and vertices. Their basic interpretation is as follows:

Category theory	Geometry	Interpretation
Objects	Surfaces	Classical information
1-Morphisms	Lines	Quantum systems
2-Morphisms	Vertices	Physical operations



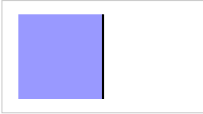


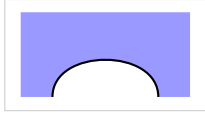
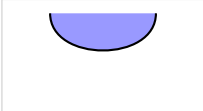
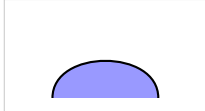
Composite diagrams involving many vertices are interpreted as a series of actions that place over time, with time flowing from bottom to top. In the graphical calculus, composition of 1-morphisms is given by horizontal juxtaposition, and composition of 2-morphisms by vertical juxtaposition. The tensor product is given by ‘overlying’ regions one above the other, perpendicular to the plane of the page and the tensor unit is expressed by an unlabelled, empty region.

We desire the ability to take the formal adjoint of 2-cells, represented graphically by flipping a diagram about a horizontal axis.

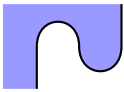


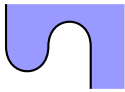

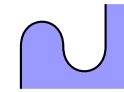


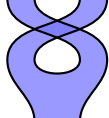
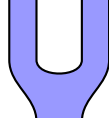
Definition 2. A *dagger 2-category* is a 2-category equipped with an involutive operation \dagger on 2-cells, such that for all $\mu : F \Rightarrow G$ we have $\mu^\dagger : G \Rightarrow F$, which is functorial and compatible with the rest of the monoidal 2-category structure.

Definition 3. A 2-cell μ is *unitary* when $\mu \circ \mu^\dagger = \text{id}$ and $\mu^\dagger \circ \mu = \text{id}$.

The core graphical theory makes use of only a small number of graphical components. They give the formal syntax for our theory. We summarize them here, along with their interpretations, which are motivated in detail in [14].

	Quantum system		Classical system	(1)
	Right-hand boundary of classical system		Left-hand boundary of classical system	(2)
	Copy classical information		Compare classical information	(3)
	Create uniform classical information		Delete classical information	(4)

These components are required to satisfy a set of axioms, which amount to saying that the boundary of a region is topological, and that holes can be eliminated:

	=		=				=		=		(5)
	=				=		(6)				

All rotations and mirrored versions of the last of these axioms are also imposed. The net effect of these axioms is that any two connected networks of copying, comparison, creation and deletion operations, with the same number of inputs and the same number of outputs, will be equal. It follows that every such region carries the structure of a commutative dagger-Frobenius algebra in a canonical way. Note that the symmetric monoidal 2-category structure is used crucially in the last equation here, allowing one region to pass above another.

Definition 4. In a symmetric monoidal 2-category, an object has a *topological boundary* if it is equipped with the data (2)–(4) satisfying equations (5)–(6).

We will assume throughout that we are working with dagger 2-categories whose objects are equipped with topological boundaries.

1.3 Controlled operations

In this paper, a key role will be played by the concept of a *controlled family of measurements* which we define here in a new way. This has the following definition in the 2-categorical formalism.

Definition 5. A *controlled family of measurements* is a unitary 2-cell of the following type:



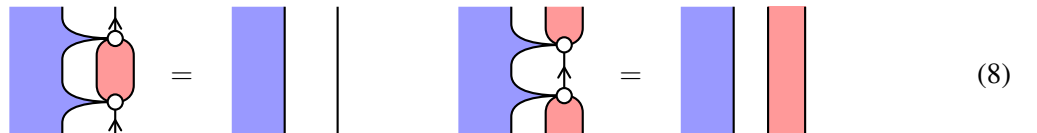
The left-hand pool of classical information represents classical data that will determine the measurement basis, which we will always draw in blue. The line at the bottom-right of the diagram represents the quantum system to be measured. The upper-right pool of classical information represents the classical result of the measurement, which we will always draw in red.

We interpret these measurements as perfectly fine-grained (that is, non-degenerate), and projective. The motivation for the definition above is made clear by analyzing its models in **2Hilb**.

Lemma 6. In **2Hilb**, a controlled family of measurements corresponds precisely to a Hilbert space equipped with a list of orthonormal bases.

Proof. Given a 2-cell ζ of the type (7) in **2Hilb**, write n for the dimension of the blue object, and m for the dimension of the red object. Then it is immediate that ζ constitutes a list of length n , whose entries are m -by- m matrices [14]. For ζ to be unitary means exactly that each m -by- m matrix is unitary. So we have a list of n unitary operators. However, the red region comes equipped with a canonical commutative dagger-Frobenius algebra structure, and hence a canonical orthonormal basis. Writing the unitaries in terms of this basis, it is clear that the data of ζ is canonically equivalent to a list of n orthonormal bases for the incoming m -dimensional Hilbert space. \square

The unitarity property of (7) takes the following graphical form:



Definition 7 (Conjugate measurement bases). Following the standard conventions, a controlled measurement with respect to the *conjugate* set of bases is represented by mirroring the diagram about a vertical axis:

$$\begin{array}{c} \text{Red box} \\ \downarrow \end{array} \quad := \quad \left(\begin{array}{c} \text{Blue box} \\ \uparrow \end{array} \right)^* \quad \equiv \quad \begin{array}{c} \text{Red box} \\ \downarrow \end{array} \quad (9)$$

Here we decompose the conjugation operation into a composition of adjoint and transpose operations. The blue classical data controlling the choice of basis is now naturally on the right-hand side.

However, we may want to change the side of the classical data controlling the choice of basis *without* passing to the conjugate set of bases. To do this, we use the symmetric monoidal 2-category structure to directly move the blue classical region to the other side. In order to distinguish this from the conjugate controlled measurement (9), we represent it as a black vertex.

Definition 8 (Control from the other side). We use a black vertex to indicate control of the measurement and encoding vertices from the other side:

$$\begin{array}{c} \text{Blue box} \\ \uparrow \end{array} \quad := \quad \begin{array}{c} \text{Blue box} \\ \downarrow \end{array} \quad \begin{array}{c} \text{Red box} \\ \downarrow \end{array} \quad := \quad \begin{array}{c} \text{Red box} \\ \uparrow \end{array} \quad (10)$$

Arrows indicating dual objects will generally be omitted for simplicity.

1.4 Projectors

We will often need to constrain the value held by a pool of classical data, which we do with projectors of different kinds.

Definition 9. Given an object $\mathbf{C} \equiv \mathbf{Hilb}^n$ in $\mathbf{2Hilb}$, a *classical data projector* is an element of the canonical n -element basis for the vector space $\text{Hom}_{\mathbf{2Hilb}}(\text{id}_{\mathbf{C}}, \text{id}_{\mathbf{C}})$.

These projectors act to constrain the classical data stored in a region to a particular value. We write them as floating labels that decorate our regions. The following lemma establishes some of their key properties.

Lemma 10 (Properties of classical data projectors). *For diagrams in $\mathbf{2Hilb}$, we can use classical data projectors to decompose the identity, and two adjacent projectors annihilate unless they are identical:*

$$\begin{array}{c} \text{Blue box} \end{array} = \sum_{a=1}^n \begin{array}{c} a \\ \text{Blue box} \end{array} \quad \begin{array}{c} a \quad b \\ \text{Blue box} \end{array} = \delta_{a,b} \begin{array}{c} a \\ \text{Blue box} \end{array} \quad (11)$$

The projectors can move freely around within regions, much like scalars in the theory of monoidal categories. Furthermore, labelled regions can be connected and disconnected arbitrarily:

$$\begin{array}{c} \text{Blue box} \\ a \end{array} = \begin{array}{c} a \\ \text{Blue box} \end{array} \quad \begin{array}{c} a \\ \text{Blue box} \end{array} = \begin{array}{c} a \\ \text{Blue box} \end{array} \quad (12)$$

Proof. Straightforward, but omitted for reasons of space. □

We can also define a different type of projector, which constrains the values of two separate regions of classical data to be the same, or to be different. We will always apply these projectors to regions that are coloured blue in our notation. There will always be exactly 2 blue regions in every diagram where we use the projectors, so it will be unambiguous to which regions they ‘attach’.

Definition 11 (Same-value and different-value projectors). In a symmetric monoidal 2-category whose hom-categories are **Ab**-enriched, for an object with topological boundary, the *same-value projector* P_s and *different-value projector* P_d are defined as follows:

$$P_s := \text{[Diagram: A blue rectangle with two semi-circular indentations on the top and bottom edges, facing each other.]} \quad (13)$$

$$P_d := \text{[Diagram: Two vertical blue rectangles side-by-side.]} - \text{[Diagram: A blue rectangle with two semi-circular indentations on the top and bottom edges, facing each other.]} \quad (14)$$

The main 2-category we are concerned with is **2Hilb**, in which hom-categories are indeed **Ab**-enriched.

Lemma 12. The projectors P_s and P_d satisfy $P_s^2 = P_s$, $P_d^2 = P_d$, $P_s \circ P_d = P_d \circ P_s = 0$ and $P_s + P_d = \text{id}$.

Proof. Straightforward graphical proof, omitted for reasons of space. \square

We will assume throughout that we are working in an **Ab**-enriched 2-category, and so these projectors are well-defined. The tensor product, vertical and horizontal composition in the 2-category all distribute over the additive structure introduced by these projectors. Distributivity of 2-cell composition with respect to addition is illustrated by the following. Note, that we can apply the projectors P_s , P_d whenever the appropriate regions have any open boundary:

$$P_d \left(\text{[Diagram: Two vertical blue rectangles side-by-side.]} \right) = \left(\text{[Diagram: Two vertical blue rectangles side-by-side.]} - \text{[Diagram: A blue rectangle with two semi-circular indentations on the top and bottom edges, facing each other.]} \right) \circ \text{[Diagram: Two vertical blue rectangles side-by-side.]} = \text{[Diagram: Two vertical blue rectangles side-by-side.]} - \text{[Diagram: A blue rectangle with two semi-circular indentations on the top and bottom edges, facing each other.]} \quad (15)$$

1.5 Attaching controlled phases

Definition 13. A *controlled phase* ϕ is an unitary endomorphism of a family of boundaries:

$$\text{[Diagram: A white rectangular box labeled } \phi \text{ with four small white circles on its top and bottom edges. It is surrounded by blue regions. Two vertical red rectangles are positioned behind the box, one on the left and one on the right.]} \quad (16)$$

The white nodes decorating the 2-cell ϕ indicate the boundaries to which it attaches.

In **2Hilb**, such a structure gives a controlled phase in the ordinary sense: a family of unit complex numbers, indexed by the values of the classical information of the regions to which the phase is connected. The result of such a controlled phase is to render the overall wavefunction of the system entangled, without introducing any classical statistical correlation between local measurement results [14].

2 Complementary families of measurements

In Definition 5 we introduced a 2-categorical axiomatization of a controlled family of measurements. In this Section, we add the extra requirement that any two distinct measurements in the family are *complementary*. In this case, we say that we have a *complementary family* of measurements. These play an essential role in quantum key distribution and the Mean King problem, which we study in Sections 3 and 4.

For a single pair of nondegenerate measurements to be complementary is a standard condition in quantum information, sometimes also known as *unbiasedness*.

Definition 14. Two bases $\{|a_i\rangle\}$, $\{|b_j\rangle\}$ of a finite-dimensional Hilbert space H are *complementary*, or *unbiased*, when for all i, j we have $|\langle a_i | b_j \rangle|^2 = \dim(H)^{-1}$.

A first characterization of this property in terms of monoidal categories was given by Coecke and Duncan [6], and a 2-categorical definition was given in [14].

2.1 Basic definition

Definition 15 (Complementary family). A *complementary family of measurements*, or simply a *complementary family*, is an ordinary family of measurements as given in Definition 5, such that there exists some unitary 2-cell ϕ satisfying the following equation:

$$\begin{array}{l}
 \text{Measure in right basis} \\
 \text{Encode in left basis} \\
 \text{Copy result} \\
 \text{Measure in left basis}
 \end{array}
 \begin{array}{c}
 \text{---} \\
 \text{---} \\
 \text{---} \\
 \text{---}
 \end{array}
 P_d
 \begin{array}{c}
 \text{---} \\
 \text{---} \\
 \text{---} \\
 \text{---}
 \end{array}
 = \frac{P_d}{n}
 \begin{array}{c}
 \text{---} \\
 \text{---} \\
 \text{---} \\
 \text{---}
 \end{array}
 \begin{array}{l}
 \text{Controlled phase } \phi \\
 \text{Measure in left basis} \\
 \text{Create random data}
 \end{array}
 \quad (17)$$

The black measurement vertex is as defined in 8. Attenuation of the region controlling the measurement choice is a notation allowing to avoid obstructing the rest of the diagram. This definition has an immediate physical motivation. On the left-hand side, we measure a quantum system in some particular basis, copy the result, and then re-encode the result back into a quantum state using the same basis. Then, according to a second basis guaranteed to be different to the first thanks to the projector P_d , we make a new measurement, represented by the black vertex, on the re-encoded state. The right-hand side of the equation says that this entire procedure must be equivalent to doing the original measurement with respect to the original basis, but then choosing the second measurement result uniformly at random, up to the application of some overall phase that allows the wavefunctions to be entangled without introducing any classical correlation.

That this definition is correct for ordinary quantum theory is immediate from previous results on the 2-categorical characterization of complementary measurements.

Lemma 16. In 2Hilb , the complementary families are exactly Hilbert spaces equipped with a collection of pairwise-complementary orthonormal bases.

Proof. Labelling the left- and right-hand blue regions on each side of equation (17) with distinct projectors a and b respectively, we obtain the ordinary 2-categorical condition for a complementary pair of orthonormal bases [14]. Conversely, suppose we have a family of orthonormal bases; then by writing the identity as a sum of projectors using Lemma 10, equation (17) follows. \square

2.2 Alternative characterizations

Here we examine alternative characterizations of the complementary family definition.

Lemma 17 (Complementarity through unitarity). *A controlled family of measurements is complementary if and only if the following 2-cell is unitary on the support of the projector P_d :*

$$\alpha := \text{[Diagram]} \quad (18)$$

Proof. See Appendix. □

The following alternative formulations of complementarity will prove useful in a later section.

Lemma 18 (Complementarity condition under horizontal reflection). *A family of controlled measurement operations is complementary if and only if the following equation is satisfied:*

$$P_d \text{ [Diagram]} = \frac{P_d}{\sqrt{n}} \text{ [Diagram]} \quad (19)$$

Proof. By Lemma 17, both α and α^\dagger are unitary on the support of the projector P_d . The condition 19 can be obtained by elementary 2-cell operations from the unitarity of α^\dagger . □

Lemma 19 (Alternative formulation of complementarity). *A controlled family of measurement operations is complementary if and only if the following condition is satisfied:*

$$P_d \text{ [Diagram]} = \frac{P_d}{n} \text{ [Diagram]} \quad (20)$$

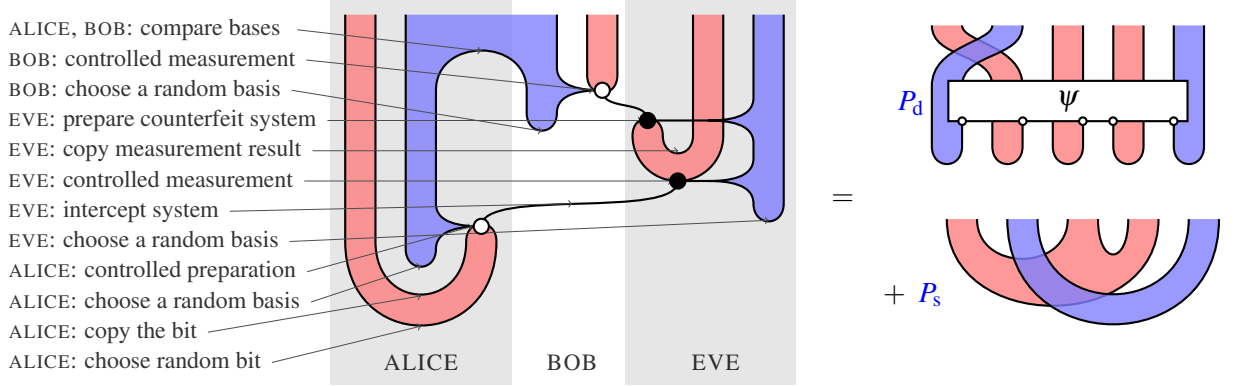
Proof. The result is proved using Lemma 17 and by performing topological manipulations. □

3 Quantum key distribution

In this Section we give 2-categorical equations defining quantum key distribution (QKD), in both its BB84 [5] and E91 [8] forms. In Theorem 22 we show that these forms are equivalent. Our main result is Theorem 27, in which we demonstrate that these quantum key distribution equations are equivalent to Definition 15 of a complementary family of measurements.

3.1 Abstract definitions

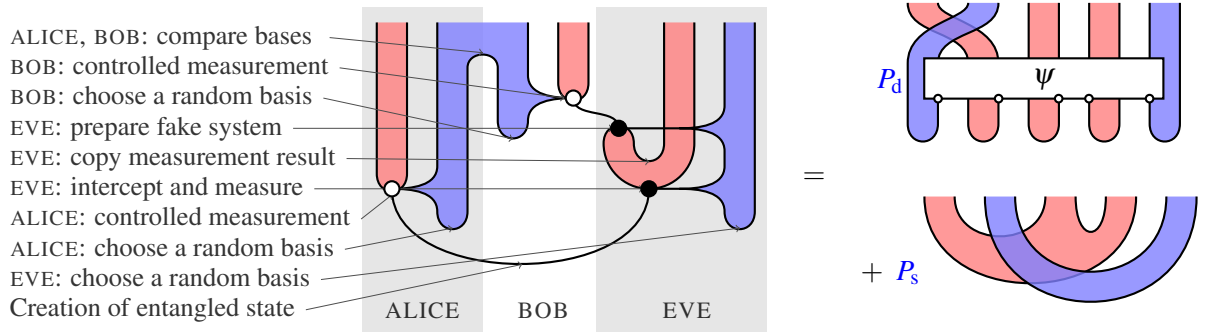
Definition 20 (BB84 QKD). A controlled family of measurements satisfies *BB84 quantum key distribution* if there exists a unitary 2-cell ψ satisfying the following equation:



Each of the diagrams on the right-hand side corresponds to the desired behaviour depending on whether Eve guessed the basis correctly. If Eve guesses incorrectly, then the P_d term says that both basis choices and all 3 measurement results are classically uncorrelated. If Eve guesses correctly, then the P_s term says that Eve shares Alice and Bob's basis, and that all 3 agents share the same classical data.

A different equation can be obtained from consideration of the E91 QKD protocol.

Definition 21 (E91 QKD). A controlled family of measurements satisfies *E91 quantum key distribution* if there exists a unitary 2-cell ψ satisfying the following equation:



Theorem 22. *The equations for BB84 and E91 QKD are equivalent.*

Proof. Elementary topological manipulation. □

Lemma 23 (Eve's successful interference). *On the support of projector P_s , the quantum key distribution specification is satisfied for any controlled family of measurements.*

Proof. See Appendix. □

3.2 Quantum key distribution from a complementary family

Lemma 24. *If a controlled family of measurements is complementary, then it satisfies the quantum key distribution specification with:*

Diagrammatic equation (21) showing the decomposition of a vertex labeled ψ into two vertices labeled ϕ and ϕ^\dagger . The left side shows a blue square labeled P_d containing a white rectangle labeled ψ with four red vertical lines passing through it. The right side shows the same blue square labeled P_d containing two stacked white rectangles labeled ϕ and ϕ^\dagger , each with two red vertical lines passing through them. The equation is labeled (21) on the right.

Proof. See Appendix.

3.3 A complementary family from quantum key distribution

Lemma 25. *If a controlled family of measurements allows quantum key distribution with a phase ψ , then:*

$$\alpha^\dagger \circ \alpha = P_d = P_d \psi = P_d \quad (22)$$

Proof. See Appendix.

Lemma 26. *If a controlled family of measurements allows quantum key distribution, then the family is complementary.*

Proof. By Lemma 25 the following map is unitary:

By Lemma 17, we can conclude that the controlled family of measurements is complementary.

Theorem 27. *A controlled family of measurements satisfies quantum key distribution if and only if it is complementary.*

Proof. Immediate by Lemmas 24 and 26. \square

Lemma 28. *If a controlled family of measurements allows quantum key distribution with phase ψ , then we can decompose ψ in the following way:*

$$P_d \text{ (with } \psi \text{)} = P_d \text{ (with } \phi \text{ and } \phi^\dagger \text{)} \quad (23)$$

Proof. Immediate by Lemmas 24 and 26. \square

4 The Mean King problem

The Mean King problem is defined as follows [13, 9]. There are two agents, Alice and the King, who take part in the following procedure.

1. Alice hands a quantum state to the King.
2. The King measures the state in one of n mutually unbiased bases, keeping both the basis and the outcome secret, and returns the state to Alice.
3. Alice performs any quantum measurement she wishes.
4. The King reveals his measurement basis to Alice.
5. Using only classical processes, Alice must calculate the King's earlier measurement outcome.

With some thought, it becomes clear that for Alice to have the best chance of succeeding, she should retain an entangled partner to the system initially passed to the King. Alice should then apply a predetermined measurement procedure to the entangled state that will reveal the King's measurement result every time, regardless of his basis choice. In other words, she should perform some nondegenerate PVM μ on both systems in step 3, and prepare a lookup table f that tells her, depending on the King's measurement basis choice and her own measurement result, what King's result was.

The key results of this Section is a graphical definition of a solution of the Mean King problem, and a graphical proof of the correctness of Klappenecker and Rötteler's solution [9] to the Mean King problem.

4.1 Abstract definition

We begin with an abstract definition of the Mean King problem. Recall that in categorical quantum mechanics, a *classical function* is defined as a morphism between classical data which satisfies the comonoid homomorphism property, and that regions with topological boundary carry a canonical comonoid structure.

Definition 29 (Mean King scheme). Given a complementary family of measurements \bullet , a bipartite measurement μ , and a classical function f , a *Mean King scheme* $\text{MK}_{\bullet, \mu, f}$ is defined as the following composite:

$$\text{MK}_{\bullet, \mu, f} := \text{Diagram} \quad (24)$$

Definition 30 (Mean King solution). A Mean King scheme $\text{MK}_{\bullet, \mu, f}$ solves the Mean King problem if the following equation holds:

$$\text{MK}_{\bullet, \mu, f} = \text{Diagram} \quad (25)$$

This says exactly that, after carrying out the procedure, Alice and the King carry the same measurement result information. It will be satisfied precisely if the Mean King scheme $\text{MK}_{\bullet, \mu, f}$ is correct.

4.2 Solving the Mean King problem

Our solution to the Mean King problem is presented entirely graphically. It is based on a solution due to Klappenecker and Rötteler [9]. Giving this solution graphically is an interesting exercise in the graphical formalism, and demonstrates that it is capable of reasoning about sophisticated schemes. Our presentation is roughly comparable in complexity to Klappenecker and Rötteler's. One advantage of our presentation is that it is perhaps clearer exactly how complementarity is being used.

We begin by giving a scheme to construct a bipartite state from a classical function.

Definition 31. Given a classical function $f_i : n \rightarrow m$, and an n -fold controlled family of measurements on \mathbb{C}^n , the associated bipartite state $|\mu_f\rangle \in \mathbb{C}^n \otimes \mathbb{C}^n$ is defined as follows:

$$\mu_f := \frac{1}{\sqrt{n}} \text{ (diagram of a blue U-shaped wire with a red box labeled } f_i \text{ inside)} - \text{ (diagram of a simple U-shaped wire)}$$

Definition 32 (Collisions). Given classical functions $f, g : n \rightarrow m$, let $f \diamond g := |\{a | f(a) = g(a)\}|$ be the number of *collisions* between them.

Lemma 33. Let $f, g : [n+1] \rightarrow [n]$ be functions. Then $n\langle\mu_f|\mu_g\rangle + 1 = f \diamond g$.

Proof. A straightforward graphical proof is possible, which we omit. □

Lemma 34. Given a family of n^2 classical functions $f_i : [n+1] \rightarrow [n]$ with $i \neq j \Rightarrow f_i \diamond f_j = 1$, the states $|\mu_{f_i}\rangle$ form an orthonormal basis.

Proof. Rearranging the result of Lemma 33, we see that $\langle\mu_f|\mu_g\rangle = ((f \diamond g) - 1)/n$, and the conclusion follows. □

Lemma 35. For any prime power $n = p^k$, the following structures exist:

1. a family of n^2 functions $f_i : [n+1] \rightarrow [n]$, such that for $i \neq j$ we have $f_i \diamond f_j = 1$;
2. a family of $n+1$ mutually complementary bases.

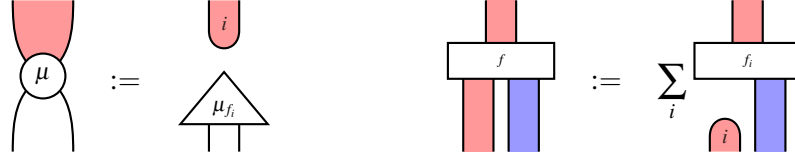
Proof. See [9, Section 2]. □

Lemma 36. For a complementary family of controlled measurements and a classical function g , the following holds:

$$\text{(diagram of a blue U-shaped wire with a red box labeled } g \text{ inside)} = \text{(diagram of a red box labeled } g \text{ inside a blue U-shaped wire)} + 1 \quad (26)$$

Proof. See Appendix. □

Theorem 37 (Solution to the Mean King problem). *For a family of functions $f_i : [n+1] \rightarrow [n]$ such that $|\mu_{f_i}\rangle$ form a basis, and a family \bullet of $n+1$ complementary bases of \mathbb{C}^n , the following assignments give a Mean King solution:*



Proof. See Appendix. □

References

- [1] Samson Abramsky and Bob Coecke. A categorical semantics of quantum protocols. *Proceedings of the 19th Annual IEEE Symposium on Logic in Computer Science*, pages 415–425, 2004. IEEE Computer Science Press.
- [2] Samson Abramsky and Bob Coecke. *Handbook of Quantum Logic and Quantum Structures*, volume 2, chapter Categorical Quantum Mechanics. Elsevier, 2008.
- [3] John C Baez. Higher-dimensional algebra II: 2-Hilbert spaces. *Advances in Mathematics*, 127:125–189, 1997.
- [4] Krzysztof Bar and Jamie Vicary. Groupoid semantics for thermal computing. 2014. Available online at <http://arxiv.org/abs/1401.3280>.
- [5] C.H. Bennett and G. Brassard. Quantum public key distribution. 1985. IBM Technical Disclosure Bulletin 28, 3153–3163.
- [6] Bob Coecke and Ross Duncan. Interacting Quantum Observables: Categorical Algebra and Diagrammatics. *New J. Phys.*, 13(043016), Jun 2011. arXiv:0906.4725v3.
- [7] Bob Coecke and Simon Perdrix. Environment and classical channels in categorical quantum mechanics. *Logical Methods in Computer Science*, 8(4):1–24, 2012.
- [8] Artur Ekert. Quantum cryptography based on Bell’s theorem. *Physical Review Letters*, 67:661, 1991.
- [9] Andreas Klappenecker and Martin Rötteler. New Tales of the Mean King. 2005. arxiv.org/abs/quant-ph/0502138.
- [10] Dan Roberts. Representing modular tensor categories: A computer algebra system for topological quantum computing. Master’s thesis, Department of Computer Science, University of Oxford, 2011.
- [11] Dan Roberts and Jamie Vicary. The TwoVect package. A computer algebra system for higher linear algebra. <http://ncatlab.org/nlab/show/TwoVect>.
- [12] Mike Stay and Jamie Vicary. Bicategorical semantics of nondeterministic computation. *Electronic Notes in Theoretical Computer Science*, 298:367–382, 2013. Proceedings of the 29th Conference on the Mathematical Foundations of Programming Semantics.
- [13] Lev Vaidman, Yakir Aharonov, and David Z. Albert. How to ascertain the values of σ_x , σ_y and σ_z of a spin- $\frac{1}{2}$ particle. 1987. *Phys. Rev. Lett.* 58, 1385–1387.
- [14] Jamie Vicary. Higher semantics of quantum protocols. *Proceedings of the 27th Annual IEEE Symposium on Logic in Computer Science*, pages 606–615, 2012. Expanded version at <http://arxiv.org/abs/1207.4563>.

Appendix

Proof of Lemma 17. We consider the following chain of equivalences:

$$\begin{aligned}
 \left[P_d \text{ (diagram)} \right] &= \frac{P_d}{\sqrt{n}} \left[\text{diagram with } \phi \right] \Leftrightarrow \left[P_d \text{ (diagram)} \right] = \frac{P_d}{\sqrt{n}} \left[\text{diagram with } \phi \right] \\
 &\Leftrightarrow \left[P_d \text{ (diagram)} \right] = \frac{P_d}{\sqrt{n}} \left[\text{diagram with } \phi \right]
 \end{aligned}$$

For the first equivalence we compose at the bottom with the inverse of the controlled measurement vertex; for the second we perform a topological manipulation. Since ϕ is an arbitrary unitary 2-cell, it is clear that the last condition is exactly that given in the statement of the lemma. \square

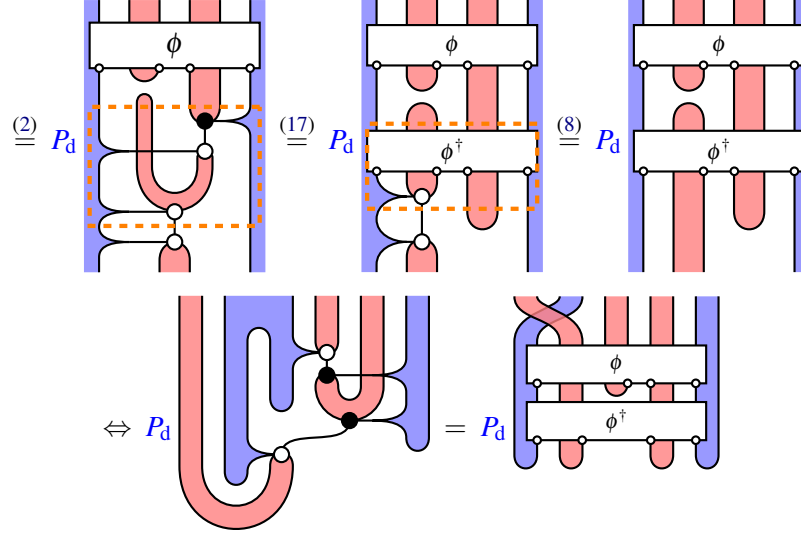
Proof of Lemma 23. We investigate this scenario by applying the projector P_s on both sides of the specification. In this case the right-hand side only retains the P_s component, and the left hand side simplifies as follows:

$$P_s \text{ (diagram)} = P_s \text{ (diagram)} = P_s \text{ (diagram)}$$

Vertex colour changes are justified by changing the side from which the operations are controlled in accordance with Definition 8. By this, we can conclude that after application of P_s the QKD specification becomes a tautology. \square

Proof of Lemma 24. Suppose the controlled complementarity condition 15 is satisfied. Then we make the following argument:

$$P_d \text{ (diagram)} \stackrel{(19)}{=} P_d \text{ (diagram)} \stackrel{(8)}{=} P_d \text{ (diagram)}$$



The final equality follows from the first chain of equalities by topological deformation. This final equality is equivalent to the statement of BB84 quantum key distribution as given in Definition 20, since by Lemma 23 the P_s component is trivially satisfied. \square

Proof of Lemma 25. If Eve picks the wrong basis but does not influence the communication between Alice and Bob, their key information is still the same. We post-select on this scenario by applying a projector P_d to pools of classical information corresponding to Alice's, Bob's and Eve's basis information and by applying the comparison operation to pools corresponding to Bob's and Alice's key information:

(27)

Using topology preserving elementary 2-cell operations, the first equality in 22 is obtained. For the second equality, up to application of P_d on the outer pools of classical information, the middle 2-cell in equation 22 is a unitary, since ψ is a unitary. Also, $\alpha^\dagger \circ \alpha$ is a positive map. The only positive unitary is the identity, hence the result is established. \square

Proof of Lemma 36. We use the fact that these controlled measurements form a family of complementary controlled operations. Hence equation (20) holds, and we combine it with the classical

function g to obtain the left-hand side given below:

$$P_d \left[\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right] = \frac{P_d}{n} \left[\begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right] \Leftrightarrow \left[\begin{array}{c} \text{Diagram 5} \\ \text{Diagram 6} \end{array} \right] - \left[\begin{array}{c} \text{Diagram 7} \\ \text{Diagram 8} \end{array} \right] = \frac{1}{n} \left[\begin{array}{c} \text{Diagram 9} \\ \text{Diagram 10} \end{array} \right]$$

The right-hand side is obtained by expanding out the action of the projectors P_d . We next assign specific values a, b to pools of classical information and perform elementary 2-cell operations. We can replace black vertices with white, as long as we switch the side from which the vertex is controlled. Since pools of classical information exhibit topological behaviour, we can reposition them freely.

$$\left[\begin{array}{c} \text{Diagram 11} \\ \text{Diagram 12} \end{array} \right] - \left[\begin{array}{c} \text{Diagram 13} \\ \text{Diagram 14} \end{array} \right] = \frac{1}{n} \left[\begin{array}{c} \text{Diagram 15} \\ \text{Diagram 16} \end{array} \right] \Leftrightarrow \left[\begin{array}{c} \text{Diagram 17} \\ \text{Diagram 18} \end{array} \right] - \left[\begin{array}{c} \text{Diagram 19} \\ \text{Diagram 20} \end{array} \right] = \frac{1}{n} \left[\begin{array}{c} \text{Diagram 21} \\ \text{Diagram 22} \end{array} \right]$$

After we cancel out measurement and encoding operations controlled by the same pools of classical information, the equation is simplified to:

$$\begin{aligned} & \left[\begin{array}{c} \text{Diagram 23} \end{array} \right] = \left[\begin{array}{c} b \\ s \\ a \end{array} \right] + \frac{1}{n} \left(\left[\begin{array}{c} b \\ s \\ a \end{array} \right] \text{ (blue circle)} \text{ (red circle)} - \left[\begin{array}{c} b \\ s \\ a \end{array} \right] \text{ (red circle)} \text{ (blue circle)} \right) \\ & = \left[\begin{array}{c} b \\ s \\ a \end{array} \right] + \frac{1}{n}(n+1) - \frac{1}{n} = \left[\begin{array}{c} b \\ s \\ a \end{array} \right] + 1 \end{aligned} \quad (28)$$

Proof of Theorem 37. By Lemma 35 a suitable family of n^2 functions $f_i : [n+1] \rightarrow [n]$ and a complementary family of controlled measurements in $n+1$ bases exist. The latter by defining a controlled operation to pick one of the $n+1$ complementary bases to measure in. For each f_i we define a state μ_{f_i} in accordance with Lemma 31. By Lemma 34 states $|\mu_{f_i}\rangle$ form an orthonormal basis μ that we

use to solve the problem. The scheme $\text{MK}_{\bullet,\mu,f}$ then simplifies to:

$$\begin{aligned}
 & \text{MK}_{\bullet,\mu,f} \\
 &= \sum_{i,a,b} \left[\begin{array}{c} \text{Diagram 1} \\ \text{Diagram 2} \end{array} \right] \\
 &= \sum_{i,a,b} \left[\begin{array}{c} \text{Diagram 3} \\ \text{Diagram 4} \end{array} \right]
 \end{aligned}$$

The diagrams are as follows:

- Diagram 1:** A box labeled $\text{MK}_{\bullet,\mu,f}$ with three vertical lines (two red, one blue) entering from the top.
- Diagram 2:** A difference of two diagrams. The first shows a red loop labeled f_i and a blue loop labeled f with a box f_i on the red loop. The second shows a similar configuration with a box f on the blue loop. Both diagrams have a red line labeled b and a blue line labeled a .
- Diagram 3:** A diagram with a red loop labeled f_i and a blue loop labeled f_i with a box f_i on the red loop. It also has a red line labeled b and a blue line labeled a .
- Diagram 4:** A diagram with a red loop labeled b and a blue loop labeled a with a box f_i on the red loop. It also has a red line labeled b and a blue line labeled a .

By Lemma 36, this simplifies as follows:

$$\begin{aligned}
 & \sum_{i,a,b} \left[\left(\begin{array}{c} b \\ f_i \\ a \end{array} \right) + 1 \right] \left(\begin{array}{c} f_i \\ b \\ a \end{array} \right) - \begin{array}{c} f \\ b \\ a \end{array} \right] = \sum_{i,a,b} \left[\begin{array}{c} b \\ f_i \\ a \end{array} \right] \left[\begin{array}{c} f_i \\ b \\ a \end{array} \right] - \begin{array}{c} f \\ b \\ a \end{array} \right] \\
 &= \sum_{i,a,b} \left[\begin{array}{c} b \\ f_i \\ a \end{array} \right] \left[\begin{array}{c} f_i \\ b \\ a \end{array} \right] = \sum_{i,a,b} \begin{array}{c} b \\ f_i \\ f_i \\ a \end{array} = \sum_i \begin{array}{c} b \\ f_i \\ f_i \\ a \end{array} \\
 &= \sum_i \begin{array}{c} b \\ f_i \\ f_i \\ a \end{array} = \sum_i \begin{array}{c} b \\ f \\ i \\ a \end{array} = \begin{array}{c} b \\ f \\ a \end{array}
 \end{aligned}$$

The diagrams are as follows:

- Diagram 1:** A diagram with a red loop labeled f_i and a blue loop labeled f_i with a box f_i on the red loop. It also has a red line labeled b and a blue line labeled a .
- Diagram 2:** A diagram with a red loop labeled b and a blue loop labeled a with a box f_i on the red loop. It also has a red line labeled b and a blue line labeled a .
- Diagram 3:** A diagram with a red loop labeled f_i and a blue loop labeled f_i with a box f_i on the red loop. It also has a red line labeled b and a blue line labeled a .
- Diagram 4:** A diagram with a red loop labeled f and a blue loop labeled f with a box f on the red loop. It also has a red line labeled b and a blue line labeled a .

The final diagram clearly remains unchanged under application of the projector as per Definition 30, hence the result is established. \square

Reflections on the PBR Theorem: Reality Criteria & Preparation Independence

Shane Mansfield

Quantum Group
Department of Computer Science
University of Oxford

shane.mansfield@cs.ox.ac.uk

This paper contains initial work on attempting to bring recent developments in the foundations of quantum mechanics concerning the nature of the wavefunction within the scope of more logical and structural methods. A first step involves dualising a criterion for the reality of the wavefunction proposed by Harrigan & Spekkens, which was central to the Pusey-Barrett-Rudolph theorem. The resulting criterion has several advantages, including the avoidance of certain technical difficulties relating to sets of measure zero. By considering the ‘reality’ not of the wavefunction but of the observable properties of any ontological physical theory a new characterisation of non-locality and contextuality is found. Secondly, a careful analysis of preparation independence, one of the key assumptions of the PBR theorem, leads to a precise analogy with the kind of locality prohibited by Bell’s theorem. Motivated by this, we propose a weakening of the assumption to something analogous to no-signalling. This amounts to allowing global or non-local correlations in the joint ontic state, which nevertheless do not allow for superluminal signalling. This is, at least, consistent with the Bell and Kochen-Specker theorems. We find a counter-example to the PBR argument, which violates preparation independence, but does satisfy this physically motivated assumption. The question of whether the PBR result can be strengthened to hold under the relaxed assumption is therefore posed.

1 Introduction

The issue of the reality of the wavefunction has received a lot of attention recently (see especially [31, 15]). In this paper, we show that insights may also be gained by taking a similar approach to considering the ‘reality’ of objects and properties in physical theories more generally, and in particular that such an approach can provide a new perspective on non-locality and contextuality. The first step will be to formalise a suitably general criterion for ‘reality’ inspired by the Harrigan-Spekkens criterion for the reality of the wavefunction [22], which was the subject of the Pusey-Barrett-Rudolph theorem [31].

The aim is to formulate the ideas in a manner that can allow for a deeper, structural understanding of what is at play. Indeed, the initial motivation was to bring considerations of this kind within the scope of the methods of the unified sheaf-theoretic approach to non-locality and contextuality [2, 4, 29]. The resulting criterion has several advantages. It avoids certain technical difficulties, and due to its generality it can be applied within any ontological physical theory: e.g. generalised probabilistic theories [6], or classical mechanics.

The initial investigations here also show how such considerations can provide an alternative perspective on foundational questions more generally. We find an alternative characterisation of both local and non-contextual correlations, as those that can arise from observations or measurements of properties that can be considered ‘real’ in this sense. This ties together the notions of locality and reality, bring-

ing to light another link between the Bell and Pusey-Barrett-Rudolph (PBR) theorems [31], which deal, respectively, with these properties.

We begin, in section 2, by presenting our general, reformulated criterion for reality, which requires minimal background. Much of the literature on the foundations of quantum mechanics, including that concerning recent developments on the reality of the wavefunction, deals with hidden variable or ontological models. Therefore, we will provide a brief review of this framework in section 3, which readers familiar with the material may wish to skim over, paying attention to the notation used. In section 4, we apply the criterion not to the wavefunction but to observable properties, leading to a characterisation of locality akin to that of the unified sheaf-theoretic approach to non-locality and contextuality [2] or to Kochen-Specker contextuality [25]. We demonstrate how this may be used to arrive at treatments of the fact that local hidden variable models can be subsumed by the sheaf-theoretic framework [12, 13], and the EPR argument [16].

Finally, in section 5, we give a detailed consideration of preparation independence, which first appeared as one of the assumptions of the PBR theorem. We show that the assumption, which is crucial to the theorem, is analogous in a precise sense to Bell locality. Aside from this being another link between the Bell and PBR theorems, the analogy would also suggest that the assumption may be too strong, and that it could be weakened to something analogous to no-signalling [18]. A counter-example to the PBR result was constructed by Lewis et al. [26] by dropping the assumption of preparation independence entirely, with the caveat that for compound systems it would necessarily introduce superluminal signalling. Here, we relax preparation independence to an independence assumption that is still well motivated and rules out signalling, and construct a counter-example which avoids this caveat. It is not clear, however, if it is possible to strengthen the PBR theorem so that its result still holds under the weaker assumption. We will mention, too, that by assuming preparation independence one can very easily prove a weakened form of Bell's theorem, a fact that may cast further suspicion on the strength of the stricter assumption.

2 A Criterion for Reality

In this section we will use the terminology of Harrigan & Spekkens [22], which has been established in the literature. We begin by reviewing their criterion for the reality, or *onticity*, of the wavefunction, which we then dualise and re-cast. We note that a dual view was suggested in [22], though it was not formalised. For this, we need only postulate, for each system, a space Λ of *ontic states*. These can be considered to correspond to real, physical states of the system. The idea will be that objects or properties that are determined with certainty by the ontic state can themselves be considered ontic. The term ontic is chosen deliberately, and it is supposed that such objects, properties, or states, have a real objective existence as opposed to having a merely phenomenal existence. We do not, however, propose to get into a discussion of the suitability of terminology here. Similarly, objects or properties that are not determined with certainty are said to be *epistemic*, recalling that the literal meaning of the term is that which relates to knowledge or to its degree of validation. The use of the term in [22] can be taken to reflect the fact that objects and properties of this kind are necessarily probabilistic and could thus be assumed to represent a degree of knowledge about some underlying ontic object or property. It should be borne in mind, of course, that results relating to these definitions will hold regardless of the physical significance attached to them.

As well as the existence of an ontic state space, the authors of [22] also posit the assumption that the preparation of any quantum state $|\psi\rangle$ induces a distribution $\mu_{|\psi\rangle}$ over the ontic state space Λ for that system, specifying the probabilities for the system to be in each ontic state given that it has been prepared

in this way.

Definition 2.1 (Harrigan & Spekkens [22]). If, for all wavefunctions $|\psi\rangle \neq |\phi\rangle$ of each system, the induced distributions $\mu_{|\psi\rangle}$ and $\mu_{|\phi\rangle}$ have non-overlapping supports, the wavefunction is said to be *ontic*. Otherwise, there exist some $|\psi\rangle \neq |\phi\rangle$ such that $\mu_{|\psi\rangle}(\lambda) > 0$ and $\mu_{|\phi\rangle}(\lambda) > 0$ for some $\lambda \in \Lambda$, and the wavefunction is said to be *epistemic*.

We now formalise a more general, dual version of the definition. We thereby shift from thinking of values of properties as giving probabilistic information about ontic states, to thinking of ontic states as giving probabilistic information about values of properties. As we will see, the definition can be applied to any object or property. Though the wavefunction would more usually be considered to be (at least) a mathematical object rather than a property of a system, for simplicity we only refer to properties from now on. The terms ontic and epistemic apply to the relationship or specification of values for each ontic state, and it is this that we refer to as a property.

Definition 2.2. A \mathcal{V} -valued property over Λ is a function $f : \Lambda \rightarrow \mathcal{D}(\mathcal{V})$, where $\mathcal{D}(\mathcal{V})$ is the set of probability distributions over \mathcal{V} . The property is said to be *ontic* in the special case that, for all $\lambda \in \Lambda$, the distribution $f(\lambda)$ over \mathcal{V} is a delta function. Otherwise, it is said to be *epistemic*.

Another way of stating this is that ontic properties are generated by functions $\hat{f} : \Lambda \rightarrow \mathcal{V}$; i.e. they map each ontic state to a unique value. For epistemic properties, however, there is at least one ontic state that is compatible with two or more distinct values in \mathcal{V} .

We now set about showing how these definitions relate, which may not be immediately clear. Any \mathcal{V} -valued property f specifies probability distributions over \mathcal{V} , conditioned on Λ . Bayesian inversion can be used to obtain probability distributions over Λ , conditioned on \mathcal{V} , which we (suggestively) label $\{\mu_v\}_{v \in \mathcal{V}}$. Explicitly,

$$\mu_v(\lambda) := \frac{f(\lambda)(v) \cdot p(\lambda)}{\int_{\Lambda} f(\lambda')(v) \cdot p(\lambda') d\lambda'}, \quad (1)$$

assuming a uniform prior distribution $p(\lambda)$ on Λ . Note that this is only well-defined for finite Λ , and that a more careful measure theoretic treatment, which will not be provided here, is required for the infinite case.

Proposition 2.3. A \mathcal{V} -valued property over finite Λ is ontic (definition 2.2) if and only if the distributions $\{\mu_v\}_{v \in \mathcal{V}}$ have non-overlapping supports.

Proof. Suppose the property f is ontic according to definition 2.2, let $\lambda \in \Lambda$, and let $v, v' \in \mathcal{V}$ such that $v \neq v'$. Assume for a contradiction that $\mu_v(\lambda) > 0$ and $\mu_{v'}(\lambda) > 0$. Then, by (1), $f(\lambda)(v) > 0$ and $f(\lambda)(v') > 0$; but since f is ontic,

$$v_{\lambda} = v \neq v' = v_{\lambda},$$

where $v_{\lambda} := \hat{f}(\lambda)$.

Conversely, suppose that the distributions $\{\mu_v\}_{v \in \mathcal{V}}$ have non-overlapping supports and assume for a contradiction that $f(\lambda)(v) > 0$ and $f(\lambda)(v') > 0$. Then, by (1), $\mu_v(\lambda) > 0$ and $\mu_{v'}(\lambda) > 0$. \square

One way of thinking about this correspondence could be as a kind of Stone duality, or as a special case of the dual equivalence between the category of von Neumann algebras and $*$ -homomorphisms and the category of measure spaces and measurable functions [23].

To illustrate, we provide a couple of simple examples of ontic and epistemic properties.

Example 2.4 (Classical Mechanics). The phase space of a system is taken to be the ontic state space. Classical mechanical observables (energy, momentum, etc.) are represented by real-valued functions on phase space, and are therefore ontic.

Example 2.5 (Fuzzy Measurement). Consider an experiment in which a bag is prepared containing two coins, which can each be green or white, with equal probability, but are otherwise identical. We claim that the process of removing one and checking its colour measures an epistemic property. If the ontic states are $\Lambda = \{GG, GW, WG, WW\}$, the property cannot be represented by a $\{G, W\}$ -valued function on Λ . Given the ontic state GW , for example, both G and W are compatible, and can arise with equal probability.

In this second example, the property that is being measured is, according to the present definition, epistemic with respect to the state of the bag; it might also be said the example describes a fuzzy measurement on the state of the bag.

Definition 2.2 has some advantages.

- It is fully general and can be applied to any object or property in any ontological theory.
- It avoids measure theoretic problems relating to sets of measure zero that are inherent to the original [19].
- It is also mathematically straightforward and conceptually transparent.
- We will see in section 4 that, while generalisation would be possible in the original formulation, generalising the criterion in the present form avoids the need to postulate additional non-specified distributions.

3 Ontological Models

We are concerned with theories that give operational predictions for outcomes to measurements; we refer to sets of such predictions as empirical models. Quantum mechanics is one such theory, which might be described operationally by saying that we associate a density matrix ρ^p with each preparation p , a POVM $\{E_o^m\}_{o \in O}$ with each measurement m , and prescribe the probability of the outcome o given preparation p and measurement m by

$$p(o \mid m, p) = \text{tr}(\rho^p E_o^m).$$

We wish, more generally, to consider theories with the same kind of operational structure. In order to do so, we will use some notation that is similar to that of the sheaf-theoretic approach. For each system we assume spaces P of preparations, X of measurements, and O of outcomes. There may be some compatibility structure on the space of measurements, say $\mathcal{M} \subseteq \mathcal{P}(X)$, specifying which sets of measurements can be made jointly (in quantum mechanics, this is specified by the commutative subalgebras of the algebra of observables). This information encodes which kind of measurement scenario we are working in: e.g. the Bell-CHSH model [14, 8], Hardy model [20, 21], and PR correlations [30] all deal with two-party scenarios in which each party can choose freely between two binary-outcome measurements¹. Again, we additionally assume a space Λ of *ontic states*, over which each preparation induces a probability distribution.

In an effort to simplify notation, we will use an overline to denote a joint measurement

$$\overline{m} = \{m_A, m_B, \dots\} \in \mathcal{M}$$

and also to denote joint outcomes $\overline{o} \in \mathcal{E}(\overline{m})$ to a joint measurement; here $\mathcal{E}(\overline{m})$ is the set of functions $\overline{o} : \overline{m} \rightarrow O$. Readers familiar with the sheaf-theoretic approach will recall that $\mathcal{E} : X \rightarrow O^X$ is the event

¹This measurement scenario is referred to as the (2,2,2) Bell scenario.

sheaf. On the other hand, $m \in X$ and $o \in O$, without overlines, denote individual measurements and outcomes, respectively. It may be the case that a particular individual measurement can belong to several allowed sets of joint measurements, etc. Joint preparations and joint ontic states will be treated similarly in section 5.

Definition 3.1. An *ontological* or *hidden variable model* h over Λ specifies:

1. A distribution

$$h(\lambda \mid p)$$

over the ontic states Λ for each preparation $p \in P$;

2. A distribution

$$h(\bar{o} \mid \bar{m}, \lambda) \tag{2}$$

over joint outcomes $\mathcal{E}(\bar{m})$, for each ontic state $\lambda \in \Lambda$ and joint measurement $\bar{m} \in \mathcal{M}$.

The *operational probabilities* are then prescribed by

$$h(\bar{o} \mid \bar{m}, p) = \int_{\Lambda} d\lambda \, h(\bar{o} \mid \bar{m}, \lambda) \, h(\lambda \mid p). \tag{3}$$

The terms ontological model and hidden variable model are both used in the literature, but recently the term ontological model has gained some popularity. It may be a more suitable term in the sense that the ‘hidden’ variable need not necessarily be hidden at all: it could be directly observable. In Bohmian mechanics [9, 10], for example, position and momentum play the role of the hidden variable. It also carries the connotation that such a model is an attempt to describe some underlying ontological reality.

Definition 3.2. A theory which predicts the measurement statistics for the ontic states (2) will be referred to as an *ontological theory* over Λ .

We are especially interested in ontological models and theories that can reproduce quantum mechanical predictions. Trivially, the simplest such theory is quantum mechanics itself, regarded as an ontological theory.

Example 3.3 (ψ -complete Quantum Mechanics). The ontic state is identified with the quantum state. A preparation produces a density matrix, which is regarded as a distribution over the projective Hilbert space associated with the system. By construction, the operational probabilities are those given by the Born rule.

Of course, quantum mechanics, treated as an ontological theory in itself in this way, has certain non-intuitive features (Einstein, Podolsky & Rosen provided one early discussion of this [16]) but later results such as Bell’s theorem [7] and the Kochen-Specker theorem [25] clarified the fact that non-locality and contextuality are necessary features of any theory that can account for quantum mechanical predictions. In order to address these issues, we point out some relevant properties that ontological models may have.

Definition 3.4. An ontological model is λ -*independent* if and only if the distributions over Λ induced by each preparation $p \in P$ do not depend on the joint measurement $\bar{m} \in \mathcal{M}$ to be performed.

We have already implicitly assumed this in definition 3.1, but it is worth making it clear since it is a crucial assumption in all of the familiar no-go theorems. In a λ -*dependent* model, on the other hand, the probabilities of being in the various ontic states would depend on both the preparation of the system and the joint measurement being performed, and we would have $h(\lambda \mid p, \bar{m})$ rather than $h(\lambda \mid p)$ in equation (3).

Definition 3.5. An ontological model is *deterministic* if and only if for each $\lambda \in \Lambda$ and set of compatible measurements $\bar{m} \in \mathcal{M}$ there exists some joint outcome $\bar{o} \in \mathcal{E}(\bar{m})$ such that $h(\bar{o} | \bar{m}, \lambda) = 1$.

In such a model, the outcome to any measurement that can be performed on an ontic state is determined with certainty.

For any distribution $h(\bar{o} | \bar{m}, \lambda)$ over joint outcomes $\bar{o} \in \mathcal{E}(\bar{m})$ to the joint measurement $\bar{m} \in \mathcal{M}$ on the ontic state $\lambda \in \Lambda$, we can find a distribution $h(o | m, \lambda)$ over outcomes $o \in O$ to any individual measurement $m \in \bar{m}$ by marginalisation.

Definition 3.6. An ontological model is *parameter-independent* if and only if the probability distribution $h(o | m, \lambda)$ over O is well-defined for each $m \in X$ and $\lambda \in \Lambda$.

By well-definedness we mean that the same marginal distribution $h(o | m, \lambda)$ is obtained regardless of which set of joint of measurements we marginalise from (in the case that $m \in \bar{m}$ and $m \in \bar{m}'$ for example). Parameter independence thus asserts that the probabilities of outcomes to a particular measurement do not depend on the other measurements being performed. Essentially, it imposes no-signalling [18] with respect to the ontic states.

Definition 3.7. An ontological model is *local* (or *non-contextual*) if and only if it is both deterministic and parameter-independent; empirical correlations are *local* (*non-contextual*) if and only if they can be realised by a local (non-contextual) model.

This says that for each ontic state there is a certain outcome to any measurement that can be performed, and that this does not depend on which other measurements are made. The term local is generally only used when the system being modelled is spatially distributed; where such an arrangement is not assumed, the model is said to be non-contextual.

We draw attention to the fact that another definition of locality that is common in the literature concerns the factorisability of the distributions

$$h(\bar{o} | \bar{m}, \lambda) = \prod_{m \in \bar{m}} h(o | m, \lambda). \quad (4)$$

While the present definition may be a less familiar means of presenting non-locality, it is important to note that these definitions were shown to be equivalent, in the sense that they generate the same sets of empirical models, in [2], which built on work by Fine [17] that was specific to the (2, 2, 2) Bell scenario.

4 Observable Properties

If we are to assume that the outcomes of measurements provide the values of properties of a system, then we require that for each measurement $m \in X$ there must exist an O -valued property $f_m : \Lambda \rightarrow \mathcal{D}(O)$ such that $f_m(\lambda)(o) = h(o | m, \lambda)$ for all $\lambda \in \Lambda$ and $o \in O$.

Definition 4.1. The *observable properties* of an ontological model h over Λ are the O -valued properties $f_m : \Lambda \rightarrow \mathcal{D}(O)$ given by

$$f_m(\lambda)(o) := h(o | m, \lambda) \quad (5)$$

for each $m \in X$ such that the marginal $h(o | m, \lambda)$ is well-defined.

By generalising in the present dualised formulation, we avoid postulating that particular values of properties induce non-specified distributions μ_v over the space of ontic states and reasoning in terms of these, in favour of the more palatable postulate that outcomes of measurements correspond to the values of properties of a system.

Theorem 4.2. *An ontological model is local (or non-contextual) if and only if all measurements are of ontic observable properties.*

Proof. First, we claim that a model is deterministic if and only if its observable properties are ontic. This holds since, by (5),

$$h(o \mid m, \lambda) = 1 \quad \Leftrightarrow \quad f_m(\lambda)(o) = 1.$$

Next, we claim that a model is parameter independent if and only if all measurements are of observable properties. This holds since, by definition 4.1, all measurements are of observable properties if and only if all marginals $h(o \mid m, \lambda)$ are well-defined. The result follows. \square

This characterisation of locality, which falls out easily from the definitions, is similar to the Kochen-Specker [25] or topos approach [24] treatments of non-contextuality. It can provide an alternative and sometimes simpler approach to certain results. The first result we mention shows that local ontological models have a canonical form. In fact, it shows that local ontological or hidden variable models can equivalently be expressed as distributions over the set of global assignments. (In this sense it shows how local ontological models are subsumed by the sheaf-theoretic approach.) It has been proved in measure theoretic generality in [13], and can also be seen to relate to the work of Fine [17]. An interesting, related point is that, by allowing for negative probabilities, these canonical models can also generate all no-signalling correlations [2, 3, 27].

Theorem 4.3. *Local models can be expressed in a canonical form, with an ontic state space $\Omega := \mathcal{E}(X)$, and probabilities*

$$h(\bar{o} \mid \bar{m}, \omega) = \prod_{m \in \bar{m}} \delta(\omega(m), \bar{o}(m))$$

for all $\bar{m} \in \mathcal{M}$, $\bar{o} \in \mathcal{E}(\bar{m})$, and $\omega \in \Omega$.

Proof. See [27]. \square

The next proposition will not be surprising in light of the EPR argument [16]. It shows that if one were to take the view that quantum mechanics is ψ -complete then all non-trivial observables are epistemic or inherently probabilistic. Indeed, we can obtain a re-statement of the EPR result as a corollary.

Proposition 4.4. *Any non-trivial quantum mechanical observable is epistemic with respect to ψ -complete quantum mechanics.*

Proof. Any observable $\hat{A} \neq \mathbf{I}$ has eigenvectors, say $|v_1\rangle$ and $|v_2\rangle$, corresponding to distinct eigenvalues, say o_1 and o_2 . Consider any state $|\psi\rangle$ such that $\langle v_1 | \psi \rangle > 0$ and $\langle v_2 | \psi \rangle > 0$. In a ψ -complete model, the wavefunction is the ontic state, so $\lambda = |\psi\rangle$. Then

$$f_{\hat{A}}(\lambda)(o_1) = h(o_1 \mid \hat{A}, \lambda) = |\langle v_1 | \psi \rangle|^2 > 0,$$

and similarly $f_{\hat{A}}(\lambda)(o_2) > 0$. Therefore $f_{\hat{A}}$ is epistemic. \square

Corollary 4.5 (EPR). *Under the assumption of locality, quantum mechanics cannot be ψ -complete.*

Proof. By proposition 4.4, any non-trivial quantum observable is epistemic with respect to ψ -complete quantum mechanics. Therefore, by theorem 4.2, ψ -complete quantum mechanics is not local. \square

This is the same result that was argued for by EPR, though this proof has more in common with an earlier argument by Einstein at the 1927 Solvay conference [5], and also with a more recent, general treatment found in [1] and [11].

Table 1: Analogy between measurement and preparation scenarios, up to a sheaf-theoretic description of preparation models (c.f. [2]).

Measurement Scenario		Preparation Scenario	
Measurements	X	Preparations	P
Outcomes	O	Ontic states	Λ
Non-locality		Preparation independence	
No-signalling		No-preparation-signalling	
Measurement compatibility	\mathcal{M}	Preparation compatibility	\mathcal{P}
Measurement events	$\mathcal{E}(\bar{m}) := O^{\bar{m}}$	Preparation events	$\mathcal{E}(\bar{p}) := \Lambda^{\bar{p}}$
Empirical model	$\{e_{\bar{m}}\}_{\bar{m} \in \mathcal{M}}$	Preparation model	$\{e_{\bar{p}}\}_{\bar{p} \in \mathcal{P}}$
$(\forall \bar{m} \in \mathcal{M}. e_{\bar{m}} \in \mathcal{D}\mathcal{E}(\bar{m}))$		$(\forall \bar{p} \in \mathcal{P}. e_{\bar{p}} \in \mathcal{D}\mathcal{E}(\bar{p}))$	

5 The PBR Theorem

In this section we briefly make some observations relating to the PBR theorem, which deals with the reality (i.e. onticity in the sense of definitions 2.1 and 2.2) of the wavefunction. One of the assumptions for this result is *preparation independence* [31]:

systems that are prepared independently have independent physical states.

The other assumptions are implicit in the present framework.

Theorem 5.1 (PBR). *For any preparation independent theory that reproduces (a certain set of) quantum correlations, the wavefunction is ontic.*

The preparation independence assumption is concerned with the composition of systems and has not appeared in previous no-go results. We will attempt to give this a more careful treatment. First of all, the PBR theorem describes a *preparation scenario*. More generally, we might think of preparation scenarios as an analogue of measurement scenarios, in which the preparations P play the role of measurements and the ontic states Λ play the role of outcomes; see table 1. Just as we had a compatibility structure \mathcal{M} for measurements, which in Bell scenarios allowed us to choose one measurement from each site, we should in general allow for a compatibility structure \mathcal{P} for preparations, which in the case of the PBR result allows us to choose one preparation per site. We should allow for joint ontic states $\bar{\lambda}$, just as we allowed for joint outcomes. Similarly to before, we will take \bar{p} to denote a tuple of joint preparations, one for each site, and $\bar{\lambda} : \bar{p} \rightarrow \Lambda$ to denote a tuple of joint hidden variables. The definitions of an ontological model and the properties from section 3 can be modified in the obvious way to account for this additional structure.

We are now in a position to give a more careful definition of preparation independence.

Definition 5.2. An ontological theory h over Λ is *preparation independent* if and only if we can factor

$$h(\bar{\lambda} \mid \bar{p}) = \prod_{p \in \bar{p}} h(\lambda_p \mid p) \quad (6)$$

for all $\bar{p} \in \mathcal{P}$, where $\lambda_p := \bar{\lambda}|_p$.

Presented in this way, preparation independence (6) in a preparation scenario is clearly seen to be analogous to non-contextuality or Bell locality (4) in a measurement scenario. An intriguing question is what happens if this is relaxed to an assumption analogous to no-signalling, in which we only assume that the marginal distributions $h(\lambda_p | p)$ are well-defined. Such a ‘no-preparation-signalling’ assumption would ensure that the preparation at one site cannot affect the probabilities of various ontic states at another site. Preparation independence would trivially imply no-preparation-signalling, but not vice versa. It is true that it would allow for global or non-local correlations in the joint ontic state $\bar{\lambda}$; but perhaps in light of the Bell and Kochen-Specker theorems this is to be expected. We therefore propose this as a more reasonable independence condition.

Definition 5.3. An ontological theory h over Λ is *no-preparation-signalling* if and only if the marginal probabilities $h(\lambda_p, \bar{p})$ are well-defined.

If we weaken the assumption of preparation independence to that of no-preparation-signalling, we will show it is possible to avoid the conclusion of PBR. Under the modified assumptions we will show how to construct a counter-example to the *argument* given by PBR for the onticity of the wavefunction [31] (proposition 5.4). The important question that will remain to be answered, therefore, is whether, with the weaker no-preparation-independence assumption, a result similar to, or indeed counter to, that of PBR can be proved.

Proposition 5.4. *The PBR argument for the onticity of the wavefunction breaks down for ontological theories which satisfy no-preparation-signalling but not preparation independence.*

Proof. We begin by summarising the PBR argument up to the point at which we can find a counter-example. It is assumed that a quantum system may be prepared in states $|\psi_0\rangle$ or $|\psi_1\rangle$, inducing distributions $\mu_0(\lambda)$ and $\mu_1(\lambda)$, respectively, over the space Λ of ontic states. Furthermore, it is assumed for a contradiction that the supports of these distributions overlap on a region $\Delta \subseteq \Lambda$, and that

$$q := \min \left\{ \int_{\Delta} d\lambda \mu_0(\lambda), \int_{\Delta} d\lambda \mu_1(\lambda) \right\} > 0.$$

The argument proceeds by considering two such systems, each of which is prepared independently in either $|\psi_0\rangle$ or $|\psi_1\rangle$. Given that the systems are prepared independently, then with probability $q^2 > 0$ both systems have ontic states in the region Δ . It therefore follows that with this probability q^2 the joint ontic state is compatible with each of $|\psi_0\rangle \otimes |\psi_0\rangle$, $|\psi_0\rangle \otimes |\psi_1\rangle$, $|\psi_1\rangle \otimes |\psi_0\rangle$ and $|\psi_1\rangle \otimes |\psi_1\rangle$ (in the sense that it lies in the support of the distributions on the joint ontic state space that are induced by these quantum states). However, if the systems are only required to obey no-preparation-signalling, rather than preparation independence, it is possible to find a counter-example, which we present in table 2.

Notice that in this hypothetical preparation model, the individual ontic states never both lie in the overlap region. Hence, a joint ontic state which is compatible with all of the aforementioned quantum states can never arise. Nevertheless, the preparation model is no-preparation signalling. For either subsystem, given that the quantum state prepared is $|\psi_0\rangle$, the probability of the ontic state being in the overlap region Δ is q and the probability of being outside is $1 - q$, and similarly for $|\psi_1\rangle$. So the choice of quantum state prepared in one system does not affect the ontic state in the other. \square

This improves on a counter-example to the PBR argument given by Lewis, et al. [26], which completely dropped the assumption of preparation independence. Here, we have provided a counter-example

Table 2: A preparation model that satisfies no-preparation-signalling but not preparation independence, and provides a counter-example to the PBR argument. The tabular representation is analogous to the framework for Bell-type measurement scenarios introduced in [28]. We read the table as saying, if the first system is prepared in the quantum state $|\psi_0\rangle$ and the second system is prepared in state $|\psi_0\rangle$ (i.e. joint quantum state $|\psi_0\rangle \otimes |\psi_0\rangle$), the probability of both ontic states being in the respective regions Δ is 0, etc.

		System 2			
		$ \psi_0\rangle$		$ \psi_1\rangle$	
		Δ	$\Lambda - \Delta$	Δ	$\Lambda - \Delta$
System 1	$ \psi_0\rangle$	Δ	0 q	0 q	
		$\Lambda - \Delta$	q $1 - 2q$	q $1 - 2q$	
	$ \psi_1\rangle$	Δ	0 q	0 q	
		$\Lambda - \Delta$	q $1 - 2q$	q $1 - 2q$	

that can apply to compound systems without invoking superluminal signalling, due the fact that we have maintained an assumption of no-preparation-signalling.

Another observation, which is also pointed out in [22], is that onticity of the wavefunction is actually inconsistent with locality. This can be demonstrated as a consequence of what Schrödinger called *steering* [32]. If a local measurement in the basis $\{|0\rangle, |1\rangle\}$ is made on the first qubit of the state

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$$

then this can be considered as a remote preparation of the second qubit in one of the states $|0\rangle$ or $|1\rangle$, and similarly for a measurement in the basis $\{|+\rangle, |-\rangle\}$. If the second sub-system has an ontic state λ that is independent of measurements made elsewhere, then λ must be consistent with one state from each of the sets $\{|0\rangle, |1\rangle\}$ and $\{|+\rangle, |-\rangle\}$, but this contradicts the onticity of the wavefunction.

The following theorem, which we propose to think of as a weak Bell theorem, since it draws the same conclusion as Bell's theorem [7] but with the extra assumption of preparation independence, is an obvious consequence of this.

Theorem 5.5. *Quantum mechanics is not realisable by any preparation independent, local ontological theory.*

Proof. This follows from the PBR theorem and the occurrence of steering in quantum mechanics (see discussion above). \square

The ease at which this result falls out may lead us to be cautious of the strength of the preparation independence assumption.

6 Discussion

We have presented a more general, dualised version of the criterion for the reality or onticity of the wavefunction proposed by Harrigan & Spekkens. Recasting the criterion in this form has been seen to give certain advantages; it avoids measure theoretic technicalities relating to sets of measure zero, is general enough to apply to any object or property in any ontological theory, and is also mathematically and conceptually straightforward. Furthermore, generalising in the present formulation avoids the need to postulate that particular values of any property induce non-specified distributions over the space of ontic states.

The obvious application of the criterion to an object or property other than the wavefunction is to the observable properties of a system. This led to a characterisation of locality and non-contextuality in terms of the nature of the observed properties. This may provide a useful tool for looking at foundational results: we have used it to obtain a short proof that local ontological models have a canonical form and to gain another perspective on the EPR argument. The characterisation is similar to the Kochen-Specker [25] or topos approach [24] treatments of non-contextuality.

It is worth mentioning that the characterisation draws a connection between locality and onticity: these are the properties that are dealt with by the Bell and PBR theorems, respectively. A further connection was found in theorem 5.5, which showed that a weakened version of Bell's result can be obtained by an argument that combines the PBR result with the incompatibility that arises between steering and the onticity of the wavefunction.

In relation to the PBR result itself, we have attempted to give a more careful treatment of the assumption of preparation independence, and made a concrete analogy between this property and locality/non-contextuality. It is possible to relax the assumption to one that is analogous to no-signalling, and which may still be well motivated. In this case we have provided a counter-example to the PBR argument. It improves on the counter-example provided by Lewis, et al. [26] in that it applies to compound systems, while still employing a reasonable independence condition that rules out superluminal influences. This amounts to introducing global or non-local correlations in the joint ontic state, which at least is consistent with the Bell and Kochen-Specker theorems. An open question is whether by another argument the result can be shown to hold with the relaxed assumption of no-preparation-signalling.

Acknowledgements

The author thanks Samson Abramsky, Clare Horsman, Nadish de Silva and Rui Soares Barbosa for comments and discussions.

References

- [1] Samson Abramsky (2013): *Relational hidden variables and non-locality*. *Studia Logica* 101(2), pp. 411–452, doi:10.1007/s11225-013-9477-4.
- [2] Samson Abramsky & Adam Brandenburger (2011): *The sheaf-theoretic structure of non-locality and contextuality*. *New Journal of Physics* 13(11), p. 113036.
- [3] Samson Abramsky & Adam Brandenburger (2014): *An Operational Interpretation of Negative Probabilities and No-Signalling Models*. *arXiv:1401.2561* ArXiv preprint.
- [4] Samson Abramsky, Shane Mansfield & Rui Soares Barbosa (2012): *The cohomology of non-locality and contextuality*. In: *Electronic Proceedings in Theoretical Computer Science*, 95, pp. 1–14.

- [5] Guido Bacciagaluppi, Antony Valentini & Martin Jähnert (2010): *Quantum Theory at the Crossroads: Reconsidering the 1927 Solvay Conference*. *Physics Today* 63(10), p. 53.
- [6] J. Barrett (2007): *Information processing in generalized probabilistic theories*. *Physical Review A* 75(3), p. 032304.
- [7] John S. Bell (1964): *On the Einstein-Podolsky-Rosen paradox*. *Physics* 1(3), pp. 195–200.
- [8] John S. Bell (1987): *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy*. Cambridge University Press.
- [9] David Bohm (1952): *A suggested interpretation of the quantum theory in terms of “hidden” variables. I*. *Physical Review* 85(2), p. 166.
- [10] David Bohm (1952): *A suggested interpretation of the quantum theory in terms of “hidden” variables. II*. *Physical Review* 85(2), p. 180.
- [11] A. Brandenburger & N. Yanofsky (2008): *A classification of hidden-variable properties*. *Journal of Physics A: Mathematical and Theoretical* 41, p. 425302.
- [12] Adam Brandenburger & H. Jerome Keisler (2011): *What does a hidden variable look like*.
- [13] Adam Brandenburger & H. Jerome Keisler (2013): *Use of a canonical hidden-variable space in quantum mechanics*. In: *Computation, Logic, Games, and Quantum Foundations. The Many Facets of Samson Abramsky*, Springer, pp. 1–6.
- [14] John F. Clauser, Michael A. Horne, Abner Shimony & Richard A. Holt (1969): *Proposed experiment to test local hidden-variable theories*. *Physical Review Letters* 23(15), pp. 880–884.
- [15] Roger Colbeck & Renato Renner (2012): *Is a System’s Wave Function in One-to-One Correspondence with Its Elements of Reality?* *Physical Review Letters* 108(15), p. 150402.
- [16] Albert Einstein, Boris Podolsky & Nathan Rosen (1935): *Can quantum-mechanical description of physical reality be considered complete?* *Physical Review* 47(10), p. 777.
- [17] Arthur Fine (1982): *Hidden variables, joint probability, and the Bell inequalities*. *Physical Review Letters* 48(5), pp. 291–295.
- [18] Gian-Carlo Ghirardi, Alberto Rimini & Tullio Weber (1980): *A general argument against superluminal transmission through the quantum mechanical measurement process*. *Lettere Al Nuovo Cimento* (1971–1985) 27(10), pp. 293–298.
- [19] Michael J.W. Hall (2011): *Generalisations of the recent Pusey-Barrett-Rudolph theorem for statistical models of quantum phenomena*. arXiv:1111.6304 ArXiv preprint.
- [20] Lucien Hardy (1992): *Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories*. *Physical Review Letters* 68(20), pp. 2981–2984.
- [21] Lucien Hardy (1993): *Nonlocality for two particles without inequalities for almost all entangled states*. *Physical Review Letters* 71(11), pp. 1665–1668.
- [22] Nicholas Harrigan & Robert W Spekkens (2010): *Einstein, incompleteness, and the epistemic view of quantum states*. *Foundations of Physics* 40(2), pp. 125–157.
- [23] Chris Heunen (2013): Private communication.
- [24] CJ Isham & Jeremy Butterfield (1998): *Topos perspective on the Kochen-Specker theorem: I. Quantum states as generalized valuations*. *International Journal of Theoretical Physics* 37(11), pp. 2669–2733.
- [25] Simon Kochen & Ernst P. Specker (1975): *Logical structures arising in quantum theory*. In: *The Logico-Algebraic Approach to Quantum Mechanics*, Springer, pp. 263–276.
- [26] Peter G Lewis, David Jennings, Jonathan Barrett & Terry Rudolph (2012): *Distinct quantum states can be compatible with a single state of reality*. *Physical Review Letters* 109(15), p. 150404.
- [27] Shane Mansfield (2013): *The mathematical structure of non-locality & contextuality*. D.Phil. thesis, Oxford University. Available at <http://ora.ox.ac.uk/objects/uuid:394bb375-db3f-4a12-bdd8-cd1ab5809573>.

- [28] Shane Mansfield & Tobias Fritz (2012): *Hardy's non-locality paradox and possibilistic conditions for non-locality*. *Foundations of Physics* 42, pp. 709–719. Available at <http://dx.doi.org/10.1007/s10701-012-9640-1>. 10.1007/s10701-012-9640-1.
- [29] Shane Mansfield & Rui Soares Barbosa (2013): *Extendability in the Sheaf-theoretic Approach: Construction of Bell Models from Kochen-Specker Models*. In: *Proceedings of Quantum Physics & Logic X*.
- [30] Sandu Popescu & Daniel Rohrlich (1994): *Quantum nonlocality as an axiom*. *Foundations of Physics* 24(3), pp. 379–385.
- [31] Matthew F Pusey, Jonathan Barrett & Terry Rudolph (2012): *On the reality of the quantum state*. *Nature Physics* 8(6), pp. 476–479.
- [32] Erwin Schrödinger (1936): *Probability relations between separated systems*. In: *Mathematical Proceedings of the Cambridge Philosophical Society*, 32, Cambridge University Press, pp. 446–452.

Abstract structure of unitary oracles for quantum algorithms

William Zeng

Department of Computer Science, University of Oxford

william.zeng@cs.ox.ac.uk

Jamie Vicary

Centre for Quantum Technologies, University of Singapore
and Department of Computer Science, University of Oxford

jamie.vicary@cs.ox.ac.uk

We show that a pair of complementary dagger-Frobenius algebras, equipped with a self-conjugate comonoid homomorphism onto one of the algebras, produce a nontrivial unitary morphism on the product of the algebras. This gives an abstract understanding of the structure of an oracle in a quantum computation, and we apply this understanding to develop a new algorithm for the deterministic identification of group homomorphisms into abelian groups. We also discuss an application to the categorical theory of signal-flow networks.

1 Introduction

1.1 Overview

Pairs of complementary dagger-Frobenius algebras play an important role in the high-level characterization of quantum phenomena [7, 12]. In Section 2, we show that if a such a pair is equipped with a self-conjugate comonoid homomorphism onto one of the algebras, a *unitary* map can be constructed that has the same abstract structure as an *oracle* in the theory of quantum algorithms. This gives insight into the logical structure of quantum algorithms, and opens up a new avenue for their generalization.

Most known quantum algorithms are constructed using these black-box quantum oracles, whose structure can be depicted graphically in the following way:



Here we read the diagram from bottom to top, defining a map of type $\mathbb{C}^n \otimes \mathbb{C}^m \rightarrow \mathbb{C}^n \otimes \mathbb{C}^m$ that acts as $|x\rangle \otimes |y\rangle \mapsto |x\rangle \otimes |y \oplus f(x)\rangle$ for a group product \oplus . Section 2 contains a full abstract description. Oracle-based algorithms include the Deutsch-Jozsa and Grover algorithms, where the oracle implements a function $f : S \rightarrow \{0, 1\}$ where S is a finite set, and the hidden subgroup problem, where the oracle implements a function $f : G \rightarrow S$ where G is a finite group and S is a finite set. In [12] it was shown that the unitary oracle described in Section 2 characterizes the structure of these well-known algorithms.

For these oracles to be physically implementable, they must be *unitary operators*. In this paper we give an abstract proof of unitarity using categorical algebra. In Section 3 we apply this result to develop a new quantum algorithm for the identification of group homomorphisms into an abelian group, in a number of queries which is equal to the number of simple factors of the target group. The graphical

approach provides a simple proof of correctness of the algorithm, and leads to an algorithm which is more general than existing work in the literature [10]. The graphical approach also emphasizes that this algorithm can be seen as an improvement of the usual hidden subgroup algorithm, by taking direct control of more parameters.

In Section 4 we investigate an application to the theory of signal-flow networks due to Baez, Erlebe and Fong [2, 9]. We show that their formalism contains dagger-Frobenius algebras equipped with self-conjugate homomorphisms, and that, as a consequence, the network representing a single resistor is unitary.

Acknowledgements. We are grateful to John Baez for useful discussions about signal-flow networks.

1.2 Frobenius monoids and complementarity

In this Section we collect some standard results from the literature [7]. We assume some familiarity with the graphical calculus for symmetric monoidal dagger-categories [11].

Definition 1. In a monoidal category, a *monoid* is a triple (A, m, u) of an object A , a morphism $A \otimes A \xrightarrow{m} A$, and a point $I \xrightarrow{u} A$, satisfying associativity and unitality equations:

$$(2)$$

The dagger operation applied to these equations produces an associated comonoid, which is represented graphically by flipping the diagrams about a horizontal axis.

Definition 2. In a monoidal dagger-category, a monoid is *dagger-Frobenius* when the following equation holds:

$$(3)$$

Definition 3. In a symmetric monoidal dagger-category, a *classical structure* is a commutative dagger-Frobenius monoid satisfying the *specialness* condition:

$$(4)$$

Definition 4. In a symmetric monoidal dagger-category, the *dimension* $d(A)$ of an object A equipped

with a dagger-Frobenius algebra is given by the following composite:

$$d(A) := \text{diagram of a vertical line with two loops} \quad (5)$$

When the algebra is commutative and special, equation (5) can be simplified to the composition of the unit and counit.

Definition 5 (Complementarity). In a symmetric monoidal category, two dagger-Frobenius algebras on the same object are *complementary* when the following property holds:

$$d(A) = \text{diagram of a vertical line with a dot} \quad (6)$$

Definition 6. A comonoid homomorphism $f : (A, \bullet, \bullet) \rightarrow (B, \bullet, \bullet)$ between classical structures is *self-conjugate* when the following property holds:

$$\text{diagram of } f \text{ with loops} = \text{diagram of } f \text{ as a box} \quad (7)$$

Lemma 7. In **Hilb**, comonoid homomorphisms $f : (A, \bullet, \bullet) \rightarrow (B, \bullet, \bullet)$ of classical structures are self-conjugate.

Proof. Recall that comonoid homomorphisms between classical structures in **Hilb** are exactly classical functions between the copyable points [8]. The linear maps in (7) will be the same if their matrix elements are the same, obtained by composing with $|i\rangle$ at the bottom and $\langle j|$ at the top. On the left-hand side, this gives the following result:

$$\text{diagram of } f \text{ with triangles } i \text{ and } j = \text{diagram of } f \text{ as a box with triangles } i \text{ and } j = \begin{cases} 1 & \text{if } i = f(j), \\ 0 & \text{if } i \neq f(j). \end{cases} \quad (8)$$

On the right we can do this calculation:

$$\begin{array}{c} \triangleup_j \\ | \\ \text{---} f \text{---} \\ | \\ \triangleleft_i \end{array} = \left(\begin{array}{c} \triangleup_i \\ | \\ \text{---} f \text{---} \\ | \\ \triangleleft_j \end{array} \right)^\dagger = \begin{Bmatrix} 1 & \text{if } i = f(j) \\ 0 & \text{if } i \neq f(j) \end{Bmatrix}^\dagger = \begin{Bmatrix} 1 & \text{if } i = f(j), \\ 0 & \text{if } i \neq f(j). \end{Bmatrix} \quad (9)$$

This is the same result as for the left-hand side, and so expression (7) holds. \square

2 Unitary oracles

2.1 Complementarity via unitarity

A pair of dagger-Frobenius algebras can be used to build a linear map in the following way:

$$\sqrt{d(A)} \quad \begin{array}{c} \text{---} \triangleleft \text{---} \\ | \\ \text{---} \triangleup \text{---} \end{array} \quad (10)$$

Here we have assumed that we operate in a category where square roots of scalars exist. The two algebras are complementary exactly when this composite is unitary, as we show in the following theorem.

Theorem 8 (Complementarity via a unitary). *In a dagger symmetric monoidal category, two dagger Frobenius algebras are complementary if and only if the composite (10) is unitary.*

Proof. Composing (10) with its adjoint in one order, we obtain the following:

$$d(A) \quad \begin{array}{c} \text{---} \triangleleft \text{---} \\ | \\ \text{---} \triangleup \text{---} \end{array} = d(A) \quad \begin{array}{c} \text{---} \triangleleft \text{---} \\ | \\ \text{---} \triangleup \text{---} \end{array} = d(A) \quad \begin{array}{c} \text{---} \triangleleft \text{---} \\ | \\ \text{---} \triangleup \text{---} \end{array} \quad (11)$$

If the complementarity condition (6) holds then this is clearly the identity on $A \otimes A$. The other composite can be shown to be the identity in a similar way, and so the composite (10) is unitary.

Conversely, suppose (10) is unitary. Then the final expression of (11) certainly equals the identity on $A \otimes A$. Composing with the black counit at the top-left and the white unit at the bottom-right then gives

back complementarity condition (6) as required:

$$\left[\begin{array}{c} \text{Diagram 1} \end{array} \right] = d(A) \left[\begin{array}{c} \text{Diagram 2} \end{array} \right] \Rightarrow \left[\begin{array}{c} \text{Diagram 3} \end{array} \right] = d(A) \left[\begin{array}{c} \text{Diagram 4} \end{array} \right] = d(A) \left[\begin{array}{c} \text{Diagram 5} \end{array} \right] \quad (12)$$

This completes the proof. \square

2.2 Families of unitary oracles

This pair of complementary observables automatically gives rise to a much larger family of unitaries, one for each self-conjugate comonoid homomorphism onto one of the classical structures in the pair. See equation (7) for the definition of the self-conjugacy property. Lemma 7 demonstrated that in **FHilb**, every comonoid homomorphism of classical structures is self-conjugate.

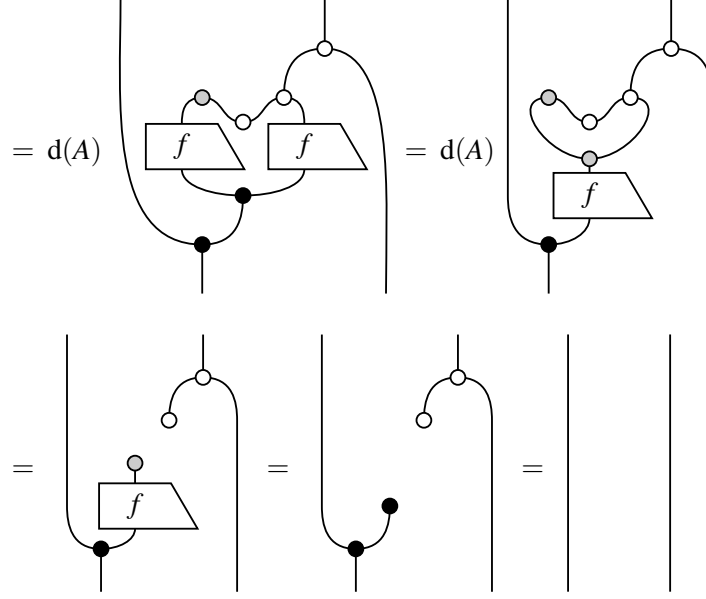
Definition 9 (Oracle). In a symmetric monoidal dagger-category, given a pair of complementary dagger-Frobenius algebras $(B, \triangleleft, \triangleright)$, $(B, \triangleright, \triangleleft)$ and a self-conjugate comonoid homomorphism $f : (A, \triangleleft, \triangleright) \rightarrow (B, \triangleleft, \triangleright)$ between dagger-Frobenius algebras, the *oracle* associated to f is defined as the following composite:

$$\sqrt{d(A)} \left[\begin{array}{c} \text{Diagram 6} \end{array} \right] \quad (13)$$

Theorem 10. *Oracles are unitary.*

Proof. To demonstrate that the oracle (13) is unitary, we must compose it with its adjoint on both sides and show that we get the identity in each case. In one case, we obtain the following, making use of the Frobenius laws, self-conjugacy of f , associativity and coassociativity, the fact that f preserves comultiplication, the complementarity condition, the fact that f preserves the counit, and the unit and counit laws:

$$d(A) \left[\begin{array}{c} \text{Diagram 7} \end{array} \right] = d(A) \left[\begin{array}{c} \text{Diagram 8} \end{array} \right] = d(A) \left[\begin{array}{c} \text{Diagram 9} \end{array} \right]$$



There is a similar argument that the other composite also gives the identity. \square

3 Identifying group homomorphisms into abelian groups

3.1 Introduction

In this Section we construct a new deterministic quantum algorithm to identify group homomorphisms.

Definition 11 (Group homomorphism identification problem). Given finite groups G and A where A is abelian, and a blackbox function $f : G \rightarrow A$ that is promised to be a group homomorphism, identify the homomorphism f .

We will define a quantum algorithm that solves the group homomorphism identification problem with a number of queries equal to the number of simple factors of the abelian group A .

For comparison, we can consider the obvious classical algorithm for this problem.

Lemma 12. *Given finite groups G and A where A is abelian and G has a generating set of order m , and a blackbox function $f : G \rightarrow A$ that is promised to be a group homomorphism, a classical algorithm can determine f with m oracle queries.*

Proof. Once we have evaluated f classically on generating set of G , we have fully characterized f . \square

We are unable to prove optimality in either the quantum or classical case. However, we note that the query complexities of these quantum and classical algorithms depend of different and unrelated parameters of the problem. Instances where the order of the generating set of G is larger than the number of factors in the target group A will demonstrate a quantum advantage.

In the simpler case where G is an abelian group this quantum algorithm was previously described by Høyer [10], though his algebraic presentation differs significantly from ours. Høyer also notes that the algorithm by Bernstein and Vazirani in [3] is an instance of the abelian group identification problem where $G = \mathbb{Z}_n^n$ and $A = \mathbb{Z}_2$. Independently, Cleve et. al. [5] also presented an algorithm for the abelian case where $G = \mathbb{Z}_2^n$ and $A = \mathbb{Z}_2^m$.

We will proceed using the abstract structure defined earlier, but will now work in the dagger-symmetric monoidal category **FHilb**. Any choice of orthonormal basis for an object A in **FHilb** endows it with a dagger Frobenius algebra (A, \bullet, \circ) , whose copying map $d : A \rightarrow A \otimes A$ is defined as the linear extension of $d(|i\rangle) = |i\rangle \otimes |i\rangle$. Any finite group G induces a different dagger-Frobenius algebra on an object $A = \mathbb{C}[G]$, the Hilbert space with orthonormal basis given by the elements G , with multiplication given by the linear extension of the group multiplication; we represent this structure as (A, \bullet, \circ) . These two Frobenius algebras are complementary.

In the case that G is finite, its representations can be characterized as the homomorphisms $G \xrightarrow{\rho} \text{Mat}(n)$. The homomorphism conditions take the following form [12, Section A.7]:

$$\begin{array}{c} \text{Mat}(n) \\ \downarrow \\ \boxed{\rho} \\ \downarrow \\ \bigcirc \\ \swarrow \quad \searrow \\ G \quad G \end{array} = \begin{array}{c} \text{Mat}(n) \\ \swarrow \quad \searrow \\ \boxed{\rho} \quad \boxed{\rho} \\ \downarrow \quad \downarrow \\ G \quad G \end{array} \qquad \begin{array}{c} \text{Mat}(n) \quad \text{Mat}(n) \\ \downarrow \quad \downarrow \\ \boxed{\rho} \quad \downarrow \\ \downarrow \quad \downarrow \\ \bigcirc \quad \downarrow \end{array} = \begin{array}{c} \text{Mat}(n) \\ \downarrow \\ \downarrow \end{array} \quad (14)$$

These will be essential for our proofs below.

3.2 The algorithm

The structure of the quantum algorithm that solves the group homomorphism identification problem is given by the topological diagram (15) below. Here $\sigma : G \rightarrow \mathbb{C}$ is a normalized irreducible representation of G , representing the result of the measurement, and $\rho : A \rightarrow \mathbb{C}$ is a normalized irreducible representation of A . The representation ρ is one-dimensional as A is an abelian group. Physically, we are able to produce the input state ρ efficiently, using $O(\log n)$ time steps, via the quantum Fourier transform for any finite abelian group [6]. The measurement result σ arises from a measurement in the Fourier basis, which can, by a similar procedure for any finite group [4], also be implemented efficiently.

$$\begin{array}{c} \text{Measure the left system} \\ \hline \begin{array}{c} \sigma \\ \downarrow \\ \sqrt{|G|} \end{array} \quad \begin{array}{c} \downarrow \\ \bigcirc \\ \downarrow \\ f \end{array} \quad \begin{array}{c} \downarrow \\ \downarrow \\ \rho^\dagger \end{array} \\ \hline \text{Apply a unitary map} \\ \hline \begin{array}{c} \downarrow \\ \frac{1}{\sqrt{|G|}} \end{array} \quad \begin{array}{c} \downarrow \\ \rho^\dagger \end{array} \\ \hline \text{Prepare initial states} \end{array} \quad (15)$$

We can compare the structure of this algorithm to that of the standard quantum algorithm for the hidden subgroup problem. There, the second system is prepared in a state given by the identity element of the group, corresponding to a uniform linear combination of the irreducible representations. A later measurement of this second system—which is not a part of the standard hidden subgroup algorithm, but can be done without changing the result of the procedure—would collapse this combination to a classical mixture of these representations. The hidden subgroup algorithm therefore contains an amount

of classical nondeterminism in its initial setup. In principle removing this, and selecting the input representation strategically, can only improve performance, and we take advantage of this here.

We analyze the effect of our new algorithm as follows.

Lemma 13. *The algorithm defined by (15) gives output σ with probability given by the square norm of $\sigma \circ f^* \circ \rho^*$.*

Proof. Using that ρ is a group homomorphism and simple diagrammatic rewrites defined in [12, Section A.9], we show the following:

$$\text{Diagram 1} = \text{Diagram 2} = \text{Diagram 3} \quad (16)$$

The left hand system is thus in the state $\sigma \circ f^* \circ \rho^*$, and using the Born rule, the squared norm of this state gives the probability of this experimental outcome. \square

Lemma 14. *The composite $\rho \circ f$ is an irreducible representation of G .*

Proof. The map f is a homomorphism, so $\rho \circ f : G \rightarrow \mathbb{C}$ is a one-dimensional representation of G . All one-dimensional representations are irreducible, so $\rho \circ f$ is an irreducible representation. \square

Lemma 15. *One-dimensional representations are equivalent only if they are equal.*

Proof. Let $\rho_1, \rho_2 : G \rightarrow \mathbb{C}$ be irreducible representations of G . If they are isomorphic, then there exists a linear map $\mathcal{L} : \mathbb{C} \rightarrow \mathbb{C}$, i.e. some complex number, such that $\forall g \in G$

$$\mathcal{L} \rho_1(g) = \rho_2(g) \mathcal{L}.$$

Hence we see that $\forall g \in G, \rho_1(g) = \rho_2(g)$. \square

Theorem 16 (Structure theorem for finite abelian groups). *Every finite abelian group is isomorphic to a direct product of cyclic groups of prime power order.*

Proof. See [1, Theorem 6.4] for a proof of this standard result. \square

Theorem 17. *For a finite group G and cyclic group of prime power order \mathbb{Z}_{p^n} , the algorithm (15) identifies a group homomorphism $f : G \rightarrow \mathbb{Z}_{p^n}$ in a single query.*

Proof. Choose the input representation ρ to be the fundamental representation of \mathbb{Z}_{p^n} . This representation is faithful. This means exactly that

$$\rho \circ f = \rho \circ f' \quad \Leftrightarrow \quad f = f'.$$

Thus $\rho \circ f$ and $\rho \circ f'$ are different irreducible representations if and only if f and f' are different group homomorphisms. The single measurement on the state $(\rho \circ f)^*$ is performed by the algorithm in the

representation basis of G , allowing us to determine $\rho \circ f$ up to isomorphism. Due to Lemma 15 we know that each equivalence class contains only one representative, and thus we can determine f with a single query. \square

Theorem 18. *For any two finite groups G and A , where A is abelian with n simple factors, the quantum algorithm (15) can identify a group homomorphism $f : G \rightarrow A$ with n oracle queries.*

Proof. We prove the result by induction.

Base case. When $A = \mathbb{Z}_{p^n}$ is simple, then by Theorem 18 we can identify the homomorphism with a single query.

Inductive step. If A is not simple, then we must have $A = H_1 \times H_2$ by Theorem 16, where the following hold:

1. the product \times is the direct product whose projectors (p_1, p_2) are homomorphisms
2. H_1 and H_2 are groups with n_1 and n_2 factors respectively such that the theorem holds, i.e. homomorphisms of the type $f_1 : G \rightarrow H_1$ and $f_2 : G \rightarrow H_2$ can be identified in n_1 and n_2 queries respectively

Since $p_1 \circ f$ and $p_2 \circ f$ are homomorphisms, we can run subroutines of the algorithm to determine them. Hence we recover f as

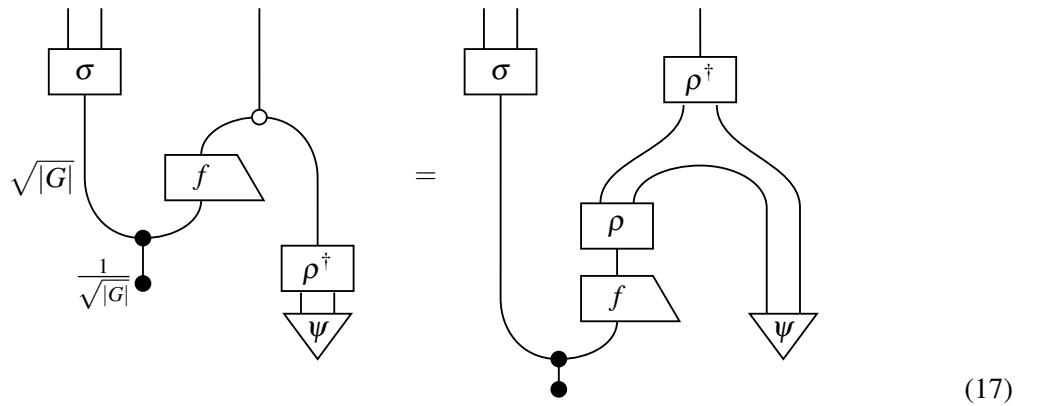
$$f(x) = ((p_1 \circ f)(x), (p_2 \circ f)(x)).$$

The first subroutine will require n_1 queries and the second will require n_2 queries, so the total number of queries will be $n_1 + n_2$, which is the number of factors of $H_1 \times H_2$. \square

3.3 Extension to the non-abelian case

We now consider the more general case where A is non-abelian. We do not know how to extend the algorithm described above to this case. Nevertheless, it is instructive to analyze this scenario in our graphical approach.

Irreducible representations of a non-abelian group A are not necessarily one dimensional, though we are still able to compute them via Fourier transform efficiently [4]. In this case the algorithm has the following structure, where ψ represents the initial state in the representation space:



We notice two additional features in this case. First, it is clear that the left and right systems are no longer in a product state, as they were in the final diagram of (16). Second, we now have an additional choice when preparing the input representation ρ ; in order to construct a state from a representation ρ we also must choose the state ψ .

While this provides a clear description of the algorithm in this more general setting, it is not clear that it would identify homomorphisms into non-abelian groups. Complications include the lack of a structure theorem that satisfies the conditions for Theorem 18, and that Lemma 14 no longer applies. In this setting it may be useful to make the problem easier by restricting to the identification of homomorphisms up to *natural isomorphism*, i.e. where two homomorphisms $f_1, f_2 : G \rightarrow H$ are considered equivalent when there exists some $\eta \in H$ such that, for all $g \in G$, we have $\eta f_1(g) \eta^{-1} = f_2(g)$.

4 Application to signal-flow calculus

4.1 Introduction

Signal-flow diagrams are a notation in electrical engineering that describe the flow of information in electrical circuits, including rich phenomena such as feedback. In recent work, Baez, Erbele and Fong [2, 9] have developed a categorical approach to modelling signal-flow diagrams, based on a category of linear relations on vector spaces over a field k . We show in this Section that unitary oracles exist in their setup, in the sense of our Definition 9, and discuss the consequences of this.

We begin with a brief introduction to their theory.

Definition 19. The category **FinRel** $_k$ of *linear relations* is defined in the following way, for any field k :

- **Objects** are finite dimensional k -vector spaces
- A **morphism** $f : V \rightsquigarrow W$ is a *linear relation*, defined as a subspace $S_f \hookrightarrow V \oplus W$
- **Composition** of linear relations $f : U \rightsquigarrow V$ and $g : V \rightsquigarrow W$ is defined as the following subspace of $U \oplus W$:

$$\{(u, w) \mid \exists v \in V \text{ with } (u, v) \in S_f \text{ and } (v, w) \in S_g\} \quad (18)$$

It can be verified that this defines a linear subspace of $U \oplus W$.

Note that a linear relation is in particular an ordinary relation, and that composition of linear relations is the same as for ordinary relations. The category **FinRel** $_k$ can be given a monoidal structure in a natural way, using the direct sum of vector spaces.

For every linear relation, we can define a converse as follows.

Definition 20. Given a linear relation $f : U \rightsquigarrow V$ defined as the subspace $S_f \hookrightarrow U \oplus V$, its *converse* is the linear relation $f^\dagger : V \rightsquigarrow U$ defined as the subspace $S_f \hookrightarrow U \oplus V \xrightarrow{\text{swap}} V \oplus U$.

This makes **FinRel** $_k$ into a monoidal dagger-category. Following the usual convention [11], we depict the dagger of a linear relation as the original morphism flipped about a horizontal axis.

Certain canonical linear relations play an important role in the theory. We define them here, along with the graphical symbol we will use to denote them.

Definition 21. The *addition*, *zero*, *copying*, *deletion* and *multiplier* linear relations are defined as follows, where the definitions in the last line are valid for all $a, b \in k$, and where the multiplier relation takes a

parameter given by some $r \in k$:

$$\begin{array}{ccccc}
 \begin{array}{c} \text{Addition} \\ \blacktriangle : k \oplus k \rightsquigarrow k \\ (a, b, a + b) \in \blacktriangle \end{array} &
 \begin{array}{c} \text{Zero} \\ \bullet : \{0\} \rightsquigarrow k \\ (0, 0) \in \bullet \end{array} &
 \begin{array}{c} \text{Copying} \\ \nabla : k \rightsquigarrow k \oplus k \\ (a, a, a) \in \nabla \end{array} &
 \begin{array}{c} \text{Deletion} \\ \circ : k \rightsquigarrow \{0\} \\ (a, 0) \in \circ \end{array} &
 \begin{array}{c} \text{Multiplier} \\ \mathbf{r} : k \rightsquigarrow k \\ (a, ra) \in \mathbf{r} \end{array}
 \end{array} \tag{19}$$

They use their theory to model resistors in electrical circuits, using the following network:

$$\tag{20}$$

The left-hand wire represents the current variable, and the right-hand wire represents the voltage variable. The initial current-voltage pair (i, v) is mapped to the output current-voltage pair $(i, v + ir)$. This respects the usual law for resistors in electrical circuits, whereby if δv is the change in voltage over a resistor, i is the current through the resistor, and the value of the resistance is r , then $\delta v = ir$.

It has been recognized in [2] that the linear relations given in Definition 21 satisfy many interesting relationships, which we summarize here without proof:

Lemma 22. *In \mathbf{FinRel}_k , the following relationships hold between the addition, zero, copying, deletion and multiplier linear relations:*

1. *Addition and zero together form a commutative monoid.*
2. *Copying and deletion together form a commutative comonoid.*
3. *This monoid and comonoid together form a bialgebra.*
4. *The multiplier relation is a monoid homomorphism for addition, and a comonoid homomorphism for copying.*

4.2 Complementary dagger-Frobenius structure

In this section we prove new results about the structures introduced in Section 4.1. We begin by establishing the existence of dagger-Frobenius properties of the addition and copying operations.

Lemma 23. *In \mathbf{FinRel}_k , the addition and copying linear relations separately form commutative dagger-Frobenius algebras.*

Proof. That addition and zero forms a commutative monoid, and copying and deletion forms a commutative comonoid, is established in Lemma 22. It remains to demonstrate that the dagger-Frobenius conditions hold for each of these structures.

We first evaluate the action of the following composite linear relation, which is one side of the dagger-Frobenius condition for the copying linear relation:

$$\begin{array}{c}
 \text{(if } a = b) \quad a \quad b \\
 \quad \quad \quad \diagup \quad \diagdown \\
 \quad \quad \quad \triangle \\
 \quad \quad \quad \diagdown \quad \diagup \\
 \quad \quad \quad b \quad \triangle \\
 \quad \quad \quad \diagup \quad \diagdown \\
 a \quad \quad b
 \end{array}
 \tag{21}$$

We see that this composite relation can be defined as $\forall a, (a, a) \curvearrowright (a, a)$, and similarly it can be shown that $\forall a, (a, a) \curvearrowleft (a, a)$. Hence we have demonstrated the dagger-Frobenius condition $\curvearrowright = \curvearrowleft$.

For the addition linear relation, we calculate the left side of the dagger-Frobenius condition as follows:

$$\begin{array}{c}
 \forall c, \quad a + c \quad b - c \\
 \quad \quad \quad \diagup \quad \diagdown \\
 \quad \quad \quad \blacktriangle \\
 \quad \quad \quad \diagdown \quad \diagup \\
 \quad \quad \quad \forall c, c \quad \blacktriangle \\
 \quad \quad \quad \diagup \quad \diagdown \\
 a \quad \quad b
 \end{array}
 \tag{22}$$

We can write this action succinctly as $\forall c, (a, b) \curvearrowright (a + c, b - c)$. Similarly, the other composite can be shown to have action $\forall c, (a, b) \curvearrowleft (a - c, b + c)$. Making the substitution $c' := -c$, we can rewrite this second definition as $\forall c', (a, b) \curvearrowleft (a + c', b - c')$. This demonstrates that $\curvearrowright = \curvearrowleft$ as linear relations, verifying the dagger-Frobenius condition for the addition linear relation. \square

Furthermore, these Frobenius algebras interact as complementary structures.

Lemma 24. *In \mathbf{FinRel}_k , the addition and copying linear relations form complementary dagger-Frobenius algebras.*

Proof. We have already established the Frobenius properties in Lemma 23. It remains to demonstrate the complementarity condition.

We evaluate the action of the following composite relation:

$$\begin{array}{c}
 a + b \quad b \\
 \quad \quad \quad \diagup \quad \diagdown \\
 \quad \quad \quad \blacktriangle \\
 \quad \quad \quad \diagdown \quad \diagup \\
 \quad \quad \quad b \quad \triangle \\
 \quad \quad \quad \diagup \quad \diagdown \\
 a \quad \quad b
 \end{array}
 \tag{23}$$

Writing K for this linear relation, we see that K is given by $\forall a, b \in k, (a, b)K(a + b, b)$. By Definition 20 of the converse relation, we see that K^\dagger is defined as $\forall a, b \in k, (a + b, b)K^\dagger(a, b)$, or equivalently $\forall a, b \in k, (a, b)K^\dagger(a - b, b)$. Since K is single-valued and total, it is clear that K and K^\dagger are inverse, as can be shown by explicit calculation. By Theorem 8, it follows that addition and copying are complementary. \square

The final property that we establish is that multipliers are self-conjugate.

Lemma 25. *In \mathbf{FinRel}_k , a multiplier $r : k \rightsquigarrow k$ is a self-conjugate morphism.*

Proof. We must verify that r is equal to the transpose of its dagger:

(24)

On the right-hand side we see that a is related to $-b$, with the constraint that $a + b/r = 0$, i.e. that $-b = ra$. This is equal as a linear relation to that of r itself, given on the left-hand side. This establishes the result. \square

Given these results, we are motivated to make the following definitions which generalize the motivating example of the theory of signal-flow diagrams in \mathbf{FinRel}_k .

Definition 26. In a symmetric monoidal dagger-category, a *signal-flow structure* is an object A equipped with a pair of commutative dagger-Frobenius algebras, which interact as a bialgebra. A *multiplier* for this signal-flow structure is a self-conjugate morphism $r : A \rightarrow A$ which is a monoid and comonoid homomorphism for both structures.

Definition 27. Given a signal-flow structure equipped with a multiplier r , the *resistor* associated to r is the composite given by diagram (20).

We then apply our earlier result to show that resistors are always unitary.

Corollary 28. *Given a signal-flow structure equipped with a multiplier, its associated resistor is unitary.*

Proof. An immediate application of Theorem 10. \square

The appearance of this unitary structure in the quantum algorithms of \mathbf{FHilb} and the signal-flow calculus of \mathbf{FinRel}_k highlights the role that their abstract structure plays in process theories in different settings.

References

- [1] M. Artin (1991): *Algebra*. Prentice Hall.
- [2] John Baez, Jason Erlebe & Brendan Fong (2014): *Electrical Circuits and Signal-Flow Graphs*. Seminar given at the Department of Computer Science, University of Oxford. Video and slides available online at math.ucr.edu/home/baez/networks_oxford/.
- [3] E. Bernstein & U. Vazirani (1997): *Quantum Complexity Theory*. *SIAM J. on Computing* 26(5), pp. 1411–1473.
- [4] Andrew M. Childs & Wim van Dam (2010): *Quantum Algorithms for Algebraic Problems*. *Reviews of Modern Physics* 82, pp. 1–52.
- [5] R. Cleve, A. Ekert, C. Macchiavello & M. Mosca (1998): *Quantum Algorithms Revisited*. *Proc. R. Soc. London* 454(339).
- [6] R. Cleve & J. Watrous (2000): *Fast Parallel Circuits for the Quantum Fourier Transform*. *Proceedings of the 41st IEEE Symposium on Foundations of Computer Science*, pp. 526–536. Available at <http://arxiv.org/abs/0006004>.
- [7] Bob Coecke & Ross Duncan (2011): *Interacting Quantum Observables: Categorical Algebra and Diagrammatics*. *New Journal of Physics* 13.
- [8] Bob Coecke, Dusko Pavlovic & Jamie Vicary (2008): *A New Description of Orthogonal Bases*. ArXiv:0810.0812.
- [9] Brendan Fong (2013): *A Compositional Approach to Control Theory*. Transfer of status report. Available at http://math.ucr.edu/home/baez/networks_oxford/.
- [10] Peter Høyer (1999): *Conjugated Operators in Quantum Algorithms*. *Physical Review A* 59(5), pp. 3280–3289.
- [11] Peter Selinger (2011): *A Survey of Graphical Languages for Monoidal Categories*. *Springer Lecture Notes in Physics* (813), pp. 289–355. Available at arxiv.org/abs/0908.3347.
- [12] Jamie Vicary (2013): *Topological Structure of Quantum Algorithms*. *Proceedings of 28th Annual ACM/IEEE Symposium on Logic in Computer Science*, pp. 93–102. Available at <http://arxiv.org/abs/1209.3917>.

Dichromatic and Trichromatic Calculus for Qutrit Systems

Quanlong Wang

Xiaoning Bian

School of Mathematics and Systems Science
Beihang University
Beijing, China

qlwang@buaa.edu.cn

bianxiaoning@smss.buaa.edu.cn

We introduce a dichromatic calculus (RG) and a trichromatic calculus (RGB) for qutrit systems. Using the dichromatic calculus, we depict a quantum algorithm with a single qutrit. To investigate the relation between the dichromatic calculus and the trichromatic calculus, a translation functor from RG to RGB is given. We also show that the decomposition of the qutrit Hadamard gate is not derivable from the dichromatic calculus.

1 Introduction

In [1], Coecke and Duncan developed dichromatic ZX-calculus for qubit systems. Then Lang and Coecke [2] introduced a trichromatic graphical calculus for qubit computing. To extend the graphical calculus to higher dimensions, Ranchin considered qudit graphical calculus [3]. At almost the same time, the authors of this paper investigated the theory and application of qutrit ZX-calculus [4]. Unlike in [3] and [4], we introduce two new rules P1 and P2 in this paper. The necessity of these two rules is demonstrated by depicting in dichromatic calculus the simplest quantum speed-up algorithm with a single qutrit [6].

For the qubit case, Duncan and Perdrix [5] proved that the Euler decomposition is not derivable from ZX calculus. In this paper, we also prove that the decomposition of the qutrit Hadamard gate is not derivable from a dichromatic qutrit ZX-calculus. Furthermore, we extend the dichromatic qutrit ZX-calculus to trichromatic qutrit calculus in a similar fashion as Lang and Coecke did in [2]. This trichromatic calculus expresses the presence of 3 complementary observables in qutrits, and also how they relate to each other. We also give a translation functor from the category of dichromatic graphs to the category of trichromatic graphs.

Although dichromatic qudit ZX-calculus is shown to be universal for quantum mechanics [3], we hope the trichromatic qutrit calculus can be helpful in understanding qutrits and deriving quantum protocols using three complementary qutrit observables.

2 Red and Green Graphs

We fix some notations here. Let \mathbf{FdHilb} be the symmetric monoidal \dagger -category (SM \dagger -category) of finite-dimensional complex Hilbert spaces and linear maps between them. Let $\mathbf{FdHilb}_{\mathbf{p}}$ be The SM \dagger -category of finite-dimensional complex Hilbert spaces and linear maps modulo the relation $f \equiv g$ if $\exists z \in \mathbb{C}, z \neq 0 : f = zg$. $\mathbf{FdHilb}_{\mathbf{Q}}$ is defined as the full subcategory of $\mathbf{FdHilb}_{\mathbf{p}}$ generated by the objects $\underbrace{Q \otimes \cdots \otimes Q}_n \mid n \geq 0$, where $Q := \mathbb{C}^3$. This is essentially the category of qutrits.

2.1 RG category

We define a category **RG** where the objects are n -fold monoidal products of an object $*$, denoted $*^n (n \geq 0)$. In **RG**, a morphism from $*^m$ to $*^n$ is a finite undirected open graph from m wires to n wires, built from

$$\begin{array}{cccccc} \delta_Z = \text{green circle with 3 incoming wires} & \delta_Z^\dagger = \text{green circle with 3 outgoing wires} & \epsilon_Z = \text{green circle with 1 incoming and 1 outgoing wire} & \epsilon_Z^\dagger = \text{green circle with 1 incoming and 1 outgoing wire} & P_Z(\alpha, \beta) = \text{green circle with } \alpha \text{ and } \beta \text{ on a wire} & H = \boxed{H} \\ \delta_X = \text{red circle with 3 incoming wires} & \delta_X^\dagger = \text{red circle with 3 outgoing wires} & \epsilon_X = \text{red circle with 1 incoming and 1 outgoing wire} & \epsilon_X^\dagger = \text{red circle with 1 incoming and 1 outgoing wire} & P_X(\alpha, \beta) = \text{red circle with } \alpha \text{ and } \beta \text{ on a wire} & H^\dagger = \boxed{H^\dagger} \end{array}$$

where $\alpha, \beta \in [0, 2\pi)$. For convenience, we denote the frequently used angles $\frac{2\pi}{3}$ and $\frac{4\pi}{3}$ by 1 and 2 respectively. The generator H is called a Hadamard gate. Additionally, the identity morphism on $*$ is represented as the straight wire. Composition is connecting up the edges, while tensor is simply putting two diagrams side by side. We also mention here that we ignore connected components of a graph which are connected to neither input nor output. This is in order to not have to deal with scalars.

RG morphisms are also subject to the equations depicted below.

1. Equations in Figure 1.
2. All equations hold under flip of graphs, negation of angles, and exchange of H and H^\dagger .
3. All equations hold under flip of colours (except for rules $K2$ and $H2$).

The equations below can be derived from the rules of **RG** given above. These equations can often be useful when wanting to demonstrate some more complex equalities in describing quantum protocols [4] and algorithms [6].

$$\begin{array}{lcl} \text{Diagram 1} & = & \text{Diagram 2} \quad (1) \\ \text{Diagram 3} & = & \text{Diagram 4} \quad (2) \\ \text{Diagram 5} & = & \text{Diagram 6} \quad (3) \\ \text{Diagram 7} & = & \text{Diagram 8} \quad (4) \end{array}$$

It is worth noting that there are some remarkable differences between qutrit rules and qubit rules. First, in qubit case we have $\text{green circle} = \text{red circle}$, while in qutrit case we have $\text{green circle} \neq \text{red circle}$. Second, the dualizer of the two observables Z and X is an even permutation, i.e., the identical permutation. And there is only one odd permutation red circle in qubit case such that

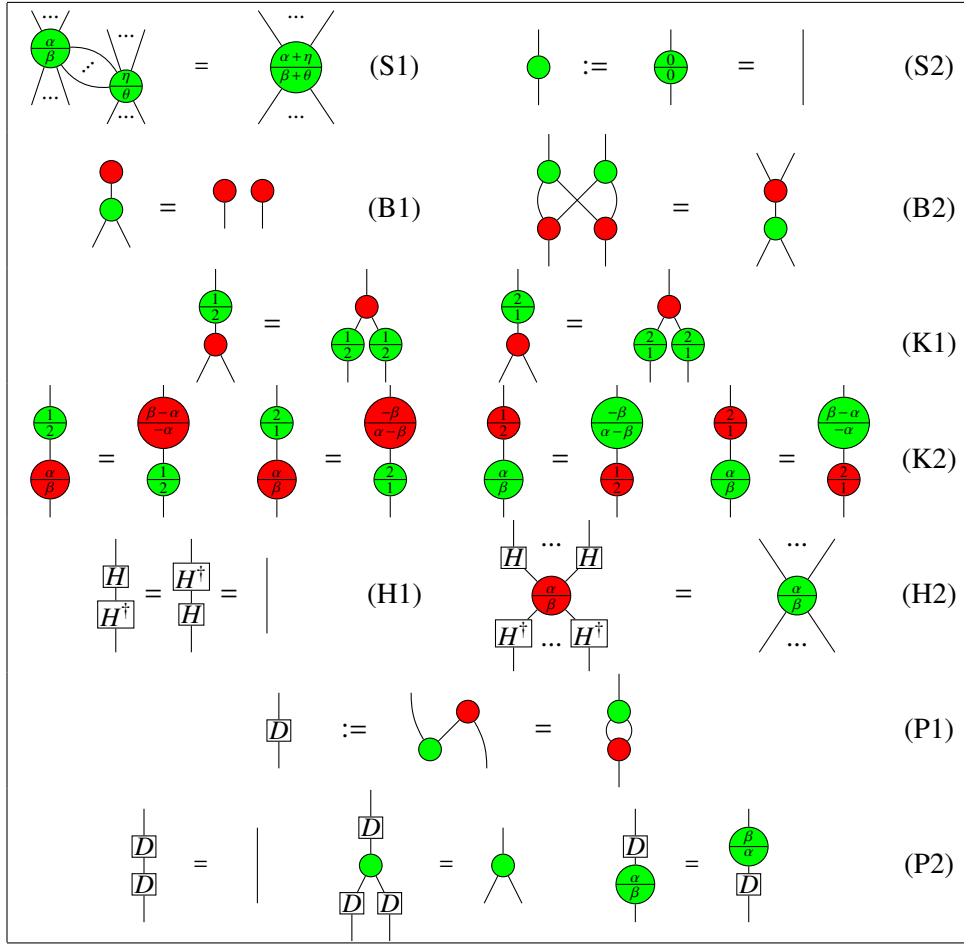
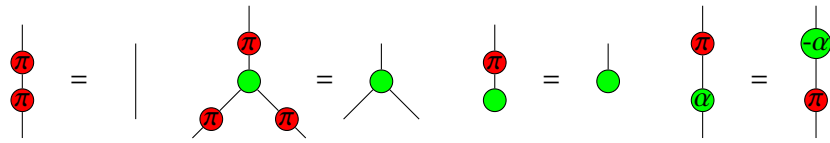


Figure 1: RG rules



While in qutrit case, the dualizer of Z and X is an odd permutation which satisfies rule *P2*. Third, in qubit case the K2 rule still holds when flipping the colours, while it doesn't hold under flip of colours in qutrit case.

Now **RG** is a symmetric monoidal category, which can further be made into a \dagger -SMC by having \dagger act on the generators like so:

$$\left(\begin{array}{c} \bullet \\ | \end{array} \right)^\dagger = \bullet \quad \left(\begin{array}{c} \bullet \\ | \end{array} \right)^\dagger = \bullet \quad \left(\begin{array}{c} \bullet \\ \diagup \diagdown \end{array} \right)^\dagger = \begin{array}{c} \bullet \\ \diagdown \diagup \end{array} \quad \left(\begin{array}{c} \bullet \\ \diagup \diagdown \end{array} \right)^\dagger = \begin{array}{c} \bullet \\ \diagdown \diagup \end{array} \quad \left(\frac{\alpha}{\beta} \right)^\dagger = \frac{-\alpha}{-\beta} \quad \left(\boxed{H} \right)^\dagger = \boxed{H^\dagger}$$

$$\left(\begin{array}{c} | \\ \bullet \\ | \end{array} \right)^\dagger = \begin{array}{c} | \\ \bullet \\ | \end{array} \quad \left(\begin{array}{c} | \\ \bullet \\ | \end{array} \right)^\dagger = \begin{array}{c} | \\ \bullet \\ | \end{array} \quad \left(\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right)^\dagger = \begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \quad \left(\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right)^\dagger = \begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \quad \left(\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right)^\dagger = \begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \quad \left(\begin{array}{c} | \\ \boxed{H^\dagger} \\ | \end{array} \right)^\dagger = \begin{array}{c} | \\ \boxed{H} \\ | \end{array}$$

where functoriality of $(\cdot)^\dagger$ is guaranteed by Rule 2.

2.2 RG interpretation

Here, we give an interpretation for these graphs by describing a monoidal functor $[\cdot]_{RG} : \mathbf{RG} \rightarrow \mathbf{FdHilb}_Q$, mapping the morphisms like so (as expressed in Dirac notation):

$$\begin{aligned} \left[\begin{array}{c} | \\ \bullet \\ | \end{array} \right]_{RG} &= |+\rangle & \left[\begin{array}{c} | \\ \bullet \\ | \end{array} \right]_{RG} &= \langle +| & \left[\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right]_{RG} &= |00\rangle\langle 0| + |11\rangle\langle 1| + |22\rangle\langle 2| \\ \left[\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right]_{RG} &= |0\rangle\langle 00| + |1\rangle\langle 11| + |2\rangle\langle 22| & \left[\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right]_{RG} &= |0\rangle\langle 0| + e^{i\alpha}|1\rangle\langle 1| + e^{i\beta}|2\rangle\langle 2| \\ \left[\begin{array}{c} | \\ \bullet \\ | \end{array} \right]_{RG} &= |0\rangle & \left[\begin{array}{c} | \\ \bullet \\ | \end{array} \right]_{RG} &= \langle 0| & \left[\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right]_{RG} &= |++\rangle\langle +| + |\omega\omega\rangle\langle \omega| + |\bar{\omega}\bar{\omega}\rangle\langle \bar{\omega}| \\ \left[\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right]_{RG} &= |+\rangle\langle ++| + |\omega\rangle\langle \omega\omega| + |\bar{\omega}\rangle\langle \bar{\omega}\bar{\omega}| & \left[\begin{array}{c} | \\ \diagup \diagdown \\ \bullet \\ \diagdown \diagup \\ | \end{array} \right]_{RG} &= |+\rangle\langle +| + e^{i\alpha}|\omega\rangle\langle \omega| + e^{i\beta}|\bar{\omega}\rangle\langle \bar{\omega}| \\ \left[\begin{array}{c} | \\ \boxed{H} \\ | \end{array} \right]_{RG} &= |+\rangle\langle 0| + |\omega\rangle\langle 1| + |\bar{\omega}\rangle\langle 2| & \left[\begin{array}{c} | \\ \boxed{H^\dagger} \\ | \end{array} \right]_{RG} &= |0\rangle\langle +| + |1\rangle\langle \omega| + |2\rangle\langle \bar{\omega}| \end{aligned}$$

where $\omega = e^{\frac{2}{3}\pi i}$, $\bar{\omega} = e^{\frac{4}{3}\pi i}$, and

$$\begin{cases} |+\rangle &= |0\rangle + |1\rangle + |2\rangle \\ |\omega\rangle &= |0\rangle + \omega|1\rangle + \bar{\omega}|2\rangle \\ |\bar{\omega}\rangle &= |0\rangle + \bar{\omega}|1\rangle + \omega|2\rangle \end{cases}$$

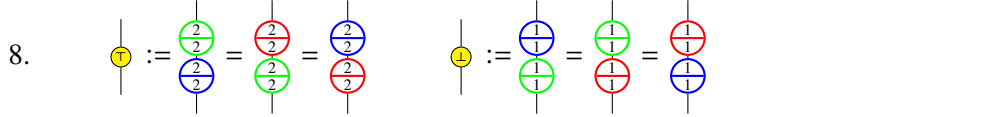
Proposition 2.1 $[\cdot]_{RG}$ is a symmetric monoidal \dagger -functor.

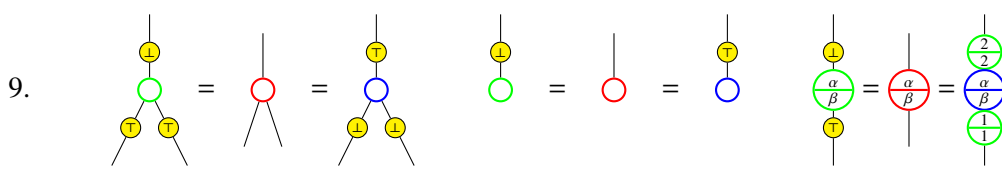
Proof: This involves checking for each rule $f = g$ in \mathbf{RG} that $[f]_{RG} = [g]_{RG}$, that $[\cdot]_{RG}$ respects the symmetric monoidal structure on the generators, and for each generator f , we have $[f]_{RG}^\dagger = [f^\dagger]_{RG}$. \square

3 Red, Green and Blue Graphs

In this section, we introduce a theory of trichromatic graphs as a category \mathbf{RGB} , which speaks of three complementary observable structures.

[illegible]

8. 

9. 

In a similar fashion as in **RG**, the symmetric monoidal category **RGB** can further be made into a \dagger -SMC by having \dagger act on the generators as follows:

$$\begin{array}{ccccc}
 \left(\begin{array}{c} | \\ \text{green circle} \end{array} \right)^\dagger = \begin{array}{c} \text{green circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{red circle} \end{array} \right)^\dagger = \begin{array}{c} \text{red circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{blue circle} \end{array} \right)^\dagger = \begin{array}{c} \text{blue circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{green circle} \end{array} \right)^\dagger = \begin{array}{c} \text{green circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{green circle with } \frac{\alpha}{\beta} \end{array} \right)^\dagger = \begin{array}{c} \text{green circle with } \frac{-\alpha}{-\beta} \\ | \end{array} \\
 \left(\begin{array}{c} | \\ \text{red circle} \end{array} \right)^\dagger = \begin{array}{c} \text{red circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{red circle} \end{array} \right)^\dagger = \begin{array}{c} \text{red circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{red circle} \end{array} \right)^\dagger = \begin{array}{c} \text{red circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{red circle} \end{array} \right)^\dagger = \begin{array}{c} \text{red circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{red circle with } \frac{\alpha}{\beta} \end{array} \right)^\dagger = \begin{array}{c} \text{red circle with } \frac{-\alpha}{-\beta} \\ | \end{array} \\
 \left(\begin{array}{c} | \\ \text{blue circle} \end{array} \right)^\dagger = \begin{array}{c} \text{blue circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{blue circle} \end{array} \right)^\dagger = \begin{array}{c} \text{blue circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{blue circle} \end{array} \right)^\dagger = \begin{array}{c} \text{blue circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{blue circle} \end{array} \right)^\dagger = \begin{array}{c} \text{blue circle} \\ | \end{array} & \left(\begin{array}{c} | \\ \text{blue circle with } \frac{\alpha}{\beta} \end{array} \right)^\dagger = \begin{array}{c} \text{blue circle with } \frac{-\alpha}{-\beta} \\ | \end{array}
 \end{array}$$

3.2 RGB interpretation

In a similar fashion as in **RG**, we give an interpretation $[\cdot]_{RGB} : \mathbf{RGB} \rightarrow \mathbf{FdHilb}_Q$ of the morphisms in **RGB**:

$$\begin{array}{lll}
 \left[\begin{array}{c} | \\ \text{green circle} \end{array} \right]_{RGB} = |+\rangle & \left[\begin{array}{c} | \\ \text{green circle} \end{array} \right]_{RGB} = \langle +| & \left[\begin{array}{c} | \\ \text{green circle} \end{array} \right]_{RGB} = |00\rangle\langle 0| + |11\rangle\langle 1| + |22\rangle\langle 2| \\
 \left[\begin{array}{c} | \\ \text{green circle} \end{array} \right]_{RGB} = |0\rangle\langle 00| + |1\rangle\langle 11| + |2\rangle\langle 22| & \left[\begin{array}{c} | \\ \text{green circle with } \frac{\alpha}{\beta} \end{array} \right]_{RGB} = |0\rangle\langle 0| + e^{i\alpha}|1\rangle\langle 1| + e^{i\beta}|2\rangle\langle 2| & \\
 \left[\begin{array}{c} | \\ \text{red circle} \end{array} \right]_{RGB} = |u\rangle & \left[\begin{array}{c} | \\ \text{red circle} \end{array} \right]_{RGB} = \langle u| & \left[\begin{array}{c} | \\ \text{red circle} \end{array} \right]_{RGB} = |++\rangle\langle +| + \omega|\omega\omega\rangle\langle \omega| + \omega|\bar{\omega}\bar{\omega}\rangle\langle \bar{\omega}| \\
 \left[\begin{array}{c} | \\ \text{red circle} \end{array} \right]_{RGB} = |+\rangle\langle ++| + \bar{\omega}|\omega\rangle\langle \omega\omega| + \bar{\omega}|\bar{\omega}\rangle\langle \bar{\omega}\bar{\omega}| & \left[\begin{array}{c} | \\ \text{red circle with } \frac{\alpha}{\beta} \end{array} \right]_{RGB} = |+\rangle\langle +| + e^{i\alpha}|\omega\rangle\langle \omega| + e^{i\beta}|\bar{\omega}\rangle\langle \bar{\omega}| & \\
 \left[\begin{array}{c} | \\ \text{blue circle} \end{array} \right]_{RGB} = |0\rangle & \left[\begin{array}{c} | \\ \text{blue circle} \end{array} \right]_{RGB} = \langle 0| & \left[\begin{array}{c} | \\ \text{blue circle} \end{array} \right]_{RGB} = |uu\rangle\langle u| + |tt\rangle\langle t| + |vv\rangle\langle v| \\
 \left[\begin{array}{c} | \\ \text{blue circle} \end{array} \right]_{RGB} = |u\rangle\langle uu| + |t\rangle\langle tt| + |v\rangle\langle vv| & \left[\begin{array}{c} | \\ \text{blue circle with } \frac{\alpha}{\beta} \end{array} \right]_{RGB} = |u\rangle\langle u| + e^{i\alpha}|t\rangle\langle t| + e^{i\beta}|v\rangle\langle v| &
 \end{array}$$

where

$$\begin{cases} |u\rangle &= |0\rangle + \bar{\omega}|1\rangle + \bar{\omega}|2\rangle \\ |t\rangle &= |0\rangle + |1\rangle + \omega|2\rangle \\ |v\rangle &= |0\rangle + \omega|1\rangle + |2\rangle \end{cases}$$

Proposition 3.1 $[\cdot]_{RGB}$ is a symmetric monoidal \dagger -functor.

The proof is similar to that of Proposition 2.1.

4 RG to RGB Translation

We define a translation from dichromatic diagrams to trichromatic diagrams as a functor $\mathfrak{T} : \mathbf{RG} \rightarrow \mathbf{RGB}$ by first defining \mathfrak{T} by its value on the generators of \mathbf{RG} and then checking that equal diagrams of \mathbf{RG} are equal under translation.

Proposition 4.1 \mathfrak{T} is a functor.

Proof: It suffices to do a routine check that for all rules $f = g$ in \mathbf{RG} , we can prove $\mathfrak{T}f = \mathfrak{T}g$ in \mathbf{RGB} .

□

Proposition 4.2 The following diagram commutes

$$\begin{array}{ccc} RG & \xrightarrow{\mathfrak{T}} & RGB \\ & \searrow [\cdot]_{RG} & \swarrow [\cdot]_{RGB} \\ & FdHilb_Q & \end{array}$$

Proof: This follows from a routine check that for each generator ϕ in \mathbf{RG} , $\mathfrak{T}[\phi]_{RGB} = [\phi]_{RG}$

□

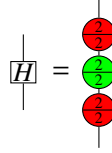
Since dichromatic qutrit ZX-calculus is universal for quantum mechanics[3], it follows from proposition 4.2 that the trichromatic calculus we introduced here is also universal for quantum mechanics.

5 Decomposition of the Hadamard Gate

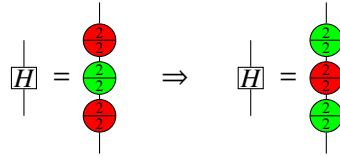
It can be directly checked that in \mathbf{FdHilb}_Q , we have

$$[H]_{RG} = [P_X(\frac{4\pi}{3}, \frac{4\pi}{3})]_{RG} \circ [P_Z(\frac{4\pi}{3}, \frac{4\pi}{3})]_{RG} \circ [P_X(\frac{4\pi}{3}, \frac{4\pi}{3})]_{RG}$$

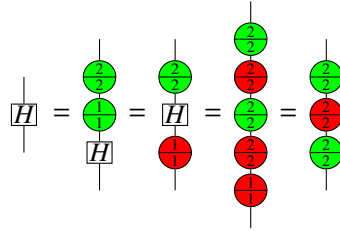
We call the following graph an Euler decomposition of the Hadamard gate:



Proposition 5.1 *The Euler decomposition is not unique:*



Proof:



□

In the qubit case, Duncan and Perdrix [5] proved that the Euler decomposition is not derivable from ZX calculus. Similarly, we have

Proposition 5.2 *The Euler decomposition is not derivable from RG.*

Proof: We define an alternative interpretation functor $[\cdot]_0 : \mathbf{RG} \rightarrow \mathbf{FdHilb}_Q$ exactly as $[\cdot]_{RG}$ with the following change:

$$[P_X(\alpha, \beta)]_0 = [P_X(0, 0)]_{RG} \quad [P_Z(\alpha, \beta)]_0 = [P_Z(0, 0)]_{RG}$$

This functor preserves all the rules introduced in Figure 1, so its image is indeed a valid model of the theory. However we have the following inequality

$$[H]_0 \neq [P_X(\frac{4\pi}{3}, \frac{4\pi}{3})]_0 \circ [P_Z(\frac{4\pi}{3}, \frac{4\pi}{3})]_0 \circ [P_X(\frac{4\pi}{3}, \frac{4\pi}{3})]_0$$

hence the Euler decomposition is not derivable from \mathbf{RG} .

□

6 Quantum Algorithm with a Single Qutrit

Recently, Gedik[6] introduces a simple algorithm using only a single qutrit to determine the parity of permutations of a set of three objects. As in the case of Deutsch's algorithm, a speed-up relative to corresponding classical algorithms is obtained.



















Consider the six permutations of the set $\{0, 1, 2\}$. Each permutation can be treated as a function $f(x)$ defined on the set $x \in \{0, 1, 2\}$. Then the task is to determine its parity. The problem could be solved by evaluating $f(x)$ for two different values of x .

The function f has a domain and range of three values. These three values correspond to the three states of a qutrit $|m\rangle$ where $m = 0, 1, 2$. The unitary U_f corresponding to the function f is a simple transposition of orthonormal states $|m\rangle$. Applying U_f to the eigenstate $|\omega\rangle$ of the X observable we obtain

$$\begin{cases} U_f|\omega\rangle = |\omega\rangle(\text{up to a phase}) & \text{if } f \text{ is an even permutation;} \\ U_f|\omega\rangle = |\bar{\omega}\rangle(\text{up to a phase}) & \text{if } f \text{ is an odd permutation.} \end{cases}$$

Thus, a single evaluation of the function is enough to determine its parity.

The above algorithm can be depicted by the dichromatic calculus as follows:

f	(0)	(1 2)(0 1)	(1 2)(0 2)	(1 2)	(0 1)	(0 2)
U_f					 	 
$U_f w\rangle$		 	 	 	 	 
Parity		Even			Odd	

7 Future Work

There are many issues requiring further exploration. Here we just list a few of them as follows. First what is the structure of the group generated by three phase gates in **RGB** with both angles being $\frac{4\pi}{3}$? In addition, does there exist an inverse functor of \mathfrak{J} ? Second, since there is a maximum of four observable structures that are mutually complementary in 3 dimensional Hilbert space, whether there exists a good graphical calculus with four colours for qutrit systems? Finally, is the dichromatic ZX calculus or the trichromatic ZXY calculus complete for qutrit stabilizer quantum mechanics?

References

- [1] B. Coecke, R. Duncan (2011): Interacting quantum observables: Categorical algebra and diagrammatics. New Journal of Physics 13, p. 043016.
- [2] A. Lang and B. Coecke. Trichromatic open digraphs for understanding qubits. In Proceedings of the 8th International Workshop on Quantum Physics and Logic (QPL), volume 95 of Electronic Proceedings in Theoretical Computer Science, pages 193-209, October 2011.
- [3] André Ranchin, Depicting qudit quantum mechanics and mutually unbiased qudit theories, QPL2014.
- [4] Xiaoning Bian, Quanlong Wang, Graphical calculus for qutrit systems, 2013.

- [5] R. Duncan, S. Perdrix: Pivoting makes the zx-calculus complete for real stabilizers, QPL2013, arXiv:1307.7048
- [6] Z. Gedik: Computational Speed-up with a Single Qutrit, arXiv:1403.5861

General probabilistic theories on arbitrary causal structures

Joe Henson

Imperial College London

Raymond Lal

University of Oxford

Matthew F. Pusey

Perimeter Institute

j.henson@imperial.ac.uk - raymond.lal@cs.ox.ac.uk - mpusey@perimeterinstitute.ca

Bayesian networks provide a powerful tool for reasoning about probabilistic causation, used in many areas of science. They are, however, intrinsically classical. In particular, Bayesian networks naturally yield the Bell inequalities. Inspired by this connection, we generalise the formalism of classical Bayesian networks in order to investigate non-classical correlations in arbitrary causal structures. Our framework of ‘generalised Bayesian networks’ replaces latent variables with the resources of any generalised probabilistic theory, most importantly quantum theory, but also, for example, Popescu-Rohrlich boxes. We obtain three main sets of results. Firstly, we prove that all of the observable conditional independences required by the classical theory also hold in our generalisation; to obtain this, we extend the classical d-separation theorem to our setting. Secondly, we find that theory-independent constraints on probabilities can go beyond these conditional independences. For example, we find that no probabilistic theory predicts perfect correlation between three parties using only bipartite common causes. Finally, we begin a classification of those causal structures, such as the Bell scenario, that yield a separation between classical, quantum and general-probabilistic correlations.

1 Introduction

Bell’s theorem [3] is a central result in the foundations of quantum mechanics. It reveals that certain quantum correlations are stronger than those obtainable in any locally causal model as defined by Bell. Recently, new results have been obtained by using variations of the scenario that Bell originally considered [13, 4]. For example, Fritz [8] showed that the ‘free will’ assumption of Bell’s theorem can be redefined by replacing the measurement settings of the Bell scenario with additional sources. The common theme in these results is the consideration of more complicated causal structures than the one usually assumed in the Bell scenario. This leads to new insights into how quantum theory deviates from classical physics: by considering arbitrary causal structures, these examples expose a rich structure to quantum correlations. However, to clarify and unify these results, it would be helpful to have a *general* framework that formalises the connection between causal structure and observable correlations.

A framework that achieves this for classical correlations is that of *Bayesian networks*, which has been pioneered in particular by Pearl [12]. When this framework is applied to a Bell-type experiment, and the causal structure implied by special relativity and independence of settings is assumed, one obtains exactly Bell’s notion of local causality [16]. The significance of this is two-fold: firstly, Bayesian networks are the natural setting for generalising Bell scenarios; secondly, a new formalism—but structurally similar to Bayesian networks—will be needed to describe the behaviour of quantum theory and other probabilistic theories on arbitrary causal structures.

Our contribution. In this paper, we propose a generalisation of Bayesian networks which incorporates the framework of generalised probabilistic theories, i.e. which generalises the ‘no-signalling’ condition of the Bell scenario to arbitrary causal structures. After introducing classical Bayesian networks in Section 2, we present our three main results.

Our first result, in Section 3, shows that all the observable conditional independences that follow from a classical Bayesian network still follow in our generalisation. The conditional independences mandated by a DAG are characterised graphically by the ‘d-separation criterion’. Technically our result is that this condition is still sound in our generalisation.

Secondly, in Section 4.1, we explore what constraints further than conditional independences can be derived for a given causal structure, even in the most general theories. We examine two quantitative limits on classical correlations that have appeared in the literature, finding that both carry over to any generalised probabilistic theory. For example we show that perfect correlation between three parties cannot be explained by bipartite common causes alone, regardless of which physical theory is used. We also show that any generalised probabilistic theory obeys the ‘instrumental inequality’, a close cousin of the Bell inequalities that applies to a simple four-node DAG.

Finally, in Section 4.2, we identify and address an important classification problem: which are the causal structures that, even classically, have no observable consequences beyond conditional independences? Structures not in this class will certainly be the focus of attention in quantum foundations, but we believe this classification will also be of interest in other applications of the classical causality framework. We make progress on this problem by providing a sufficient condition for a DAG to imply only the observable conditional independences as constraints on probability distributions, and not e.g. Bell-type inequalities.

Related work. Our work extends Pearl’s research programme [12] to the study of non-classical correlations. In this respect we build upon the work of Wood and Spekkens [16], who showed that such a connection can be made. We also build upon the circuit framework developed by Chiribella, D’Ariano and Perinotti (CDP) [7]. There are several other similar lines of investigation, e.g. Fritz [8] and Leifer and Spekkens [10].

2 Classical Bayesian networks

We often have reasons to assume a given set of causal relations between random variables. For example, consider the Bell scenario, in which we have probability distributions $P(a, b|x, y)$. The causal structure of the underlying spacetime leads to the ‘no-signalling’ conditions such as $P(a|x, y) = P(a|x)$, which restrict the possible probability distributions [14]. In general, how do we characterise the set of allowed probability distributions given a certain causal structure? In the case where we only consider causal relations between classical random variables, this question is answered by the theory of Bayesian networks.

2.1 Probabilities on graphs

Recall that a *directed graph* G is a pair (V, E) , where V is a set of nodes, and $E \subseteq V \times V$ is a set of directed edges. It is often useful to label the nodes with an index, so that we can write $V = \{X^{(i)}\}_i$. A directed graph may have a *directed cycle*, viz. a sequence of edges $X^{(1)} \rightarrow X^{(2)} \rightarrow \dots \rightarrow X^{(n)} \rightarrow X^{(1)}$. A *directed acyclic graph* (DAG) is a directed graph which has no directed cycles. In our work, DAGs will represent causal structure: more specifically, an edge $X \rightarrow Y$ will represent the possibility of direct causal influence from X to Y , where ‘direct causal influence’ will be defined in terms of probabilistic conditional dependence. The nodes that can directly influence Y are all nodes X for which there is an edge $X \rightarrow Y$; these are the *parents* of Y , and the set of all parents of Y is denoted $\text{PA}Y$. Similarly, if $X \rightarrow Y$ then Y is a *child* of X ; the set of all children of X is denoted $\text{CH}X$. A *directed path* is a sequence of nodes

$X^{(1)}, X^{(2)}, \dots, X^{(n)}$ such that $X^{(i)} \rightarrow X^{(i+1)}$ for $1 \leq i \leq n-1$. In keeping with familial terminology, we say that Y is a *descendant* of X , and X is an *ancestor* of Y , if there is a directed path from X to Y . The set of all descendants of X is denoted DEX . We also define the following two functions on sets of nodes:

- (i) we define $m(U) := U \cup (\bigcup_{X \in U} \text{CH}X)$, i.e. $m(U)$ is the union of the set of nodes U with all the children of each of the nodes in U ;
- (ii) we define $J^{-1}(U)$ to be the union of U with the set of all ancestors of nodes in U (the entire ‘past’ of U).

Now, consider the Bell scenario in which a common cause is assumed to exist. We can represent DAGs graphically, and the DAG for this scenario is depicted in Fig. 1, where the A and B nodes represent

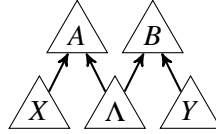


Figure 1: The Bayesian network representing a Bell-type experiment, with a classical hidden variable Λ .

experimental outcomes for the two parties, X and Y are the respective measurement settings, and Λ is the common cause (i.e. the ‘hidden variable’). Writing down such a DAG incorporates various causal assumptions, for example: (i) that the settings are ‘free’, i.e. there are no edges $\Lambda \rightarrow X$ or $\Lambda \rightarrow Y$; and (ii) that the two parties are causally disconnected from each other (which could arise from spacelike separation between Alice and Bob), e.g. there is no edge $X \rightarrow B$.

Let us now consider random variables associated to the nodes of the DAG. Only certain probability distributions will be consistent with the causal structure, if it is to have the intended meaning. As in other treatments, $X^{(1)}$ will denote a random variable, while $x^{(1)}$ denotes the value of the random variable, and the same label will also be used for the node in the graph associated to this variable (it will be clear from the context which is meant). Sometimes capital letters will also be used to signify sets of random variables, and the lowercase letter a value for each of these variables. It is convenient to extend this notation to the parents in the following way:

- $\text{PA}X^{(i)}$ is set of random variables associated to the parents of the node $X^{(i)}$;
- $\text{pa}x^{(i)}$ is a value of the random variable $\text{PA}X^{(i)}$.

The basic objects of interest will be probability distributions P over all the nodes. The notion of causality that we now apply is as follows. Given a random variable X , once direct causal influence of the parents has been taken into account by conditioning, then X should be independent of every other node, except for its descendants. This is known as the Markov condition.

Definition 1 (Markov condition). Let G be a DAG. A probability distribution P is *Markov relative to G* if P satisfies

$$P(x^{(1)}, \dots, x^{(n)}) = \prod_i P(x^{(i)} | \text{pa}x^{(i)})$$

relative to G .

A simple example is given by a probability distribution P that is Markov with respect to the chain $X \rightarrow Y \rightarrow Z$: this means that Y ‘screens off’ the influence of X from Z , i.e. $P(z|x,y) = P(z|y)$.

Definition 2. A (classical) *Bayesian network* is a pair (P, G) where G is a DAG, and P is a probability distribution that is Markov relative to G .

Often, only a subset of the nodes in a Bayesian network represent observable outcomes. These are called *observation* nodes, whereas the other nodes are referred to as *latent* or *hidden* nodes. Latent nodes are usually added by hypothesis in an attempt to explain observed correlations.

We can describe the Bell scenario in this language [16]. Consider a Bayesian network with G given by the DAG of the Bell scenario, as in Fig. 1. In the context of quantum nonlocality, the ‘hidden variable’ Λ is introduced in order to explain certain quantum correlations, with the hypothesis of the existence of common cause in the past of both A and B (the outcomes for Alice and Bob respectively). Hence, in the language of Bayesian networks, the variable Λ is a latent node. Moreover, the fact that P is Markov relative to G is simply a restatement of the local causality condition, since Definition 1 applied to this DAG implies:

$$P(a, b, x, y, \lambda) = P(a|x, \lambda)P(b|y, \lambda)P(x)P(y)P(\lambda).$$

After marginalising over λ , we obtain Bell’s locality condition.

2.2 A graphical criterion for independence: d-separation

A Bayesian network will imply conditional independences of the form $P(x, y|z) = P(x|z)P(y|z)$, which we denote by $X \perp\!\!\!\perp Y \mid Z$ (where X, Y and Z can also denote sets of nodes). In general, further independences will be derivable from those given directly by the fact that P is Markov with respect to G . Deriving all such independences using probability theory can be impractical, especially in complicated DAGs. The condition of d-separation, developed by Verma and Pearl [15], provides a way to ‘read off’ these conditional independences from the structure of the graph.

We shall use the form of d-separation originally developed by Lauritzen [9]. Let G be a DAG with disjoint subsets X, Y and Z . Then we define the set $W := G \setminus J^-(X \cup Y \cup Z)$. In words, the set W is every node in G that is not in the past of any node in X, Y or Z . Now define a *pseudo-path* from node $P^{(1)}$ to node $P^{(n)}$ to be a sequence of nodes $(P^{(1)}, P^{(2)}, \dots, P^{(p)})$ such that, for all $i = \{1, \dots, p\}$, $P^{(i)} \notin W$, and $m(P^{(i)}) \cap m(P^{(i+1)}) \not\subseteq W$. In other words, a pseudo-path does not intersect W , and two sequential elements in a pseudo-path must be related by an edge or share a common child that is not in W .

Definition 3. Let G be a DAG G with disjoint subsets X, Y and Z . We say that X and Y are *d-separated* by Z , written $X \perp\!\!\!\perp Y \mid Z$, if, for all nodes $A \in X$ and $B \in Y$, all pseudo-paths from A to B are non-trivially intersected by Z .

The following theorem shows that the d-separation condition is both sound and complete for conditional independences in a Bayesian network.

Theorem 4 (Verma and Pearl [15]). *Let G be a DAG with disjoint subsets X, Y and Z . Then:*

- (i) *If P is Markov with respect to G , then $X \perp\!\!\!\perp Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z$.*
- (ii) *If $X \perp\!\!\!\perp Y \mid Z$ holds for all P which are Markov with respect to G , then $X \perp\!\!\!\perp Y \mid Z$.*

In words, item (i) says that d-separation is a sound criterion for conditional independence, and item (ii) says that d-separation is complete, i.e. *all* conditional independences arise through applying the d-separation condition to the underlying DAG. Theorem 4 is of central importance to classical Bayesian networks, e.g. used in many causal discovery algorithms. In our setting, it is easily shown that the standard no-signalling conditions in the Bell scenario follow from the d-separation criterion.

3 Generalised Bayesian networks

We will now extend classical Bayesian networks to include general probabilistic theories [2], by building on the circuit framework developed by Chiribella, D’Ariano and Perinotti (CDP) [7].

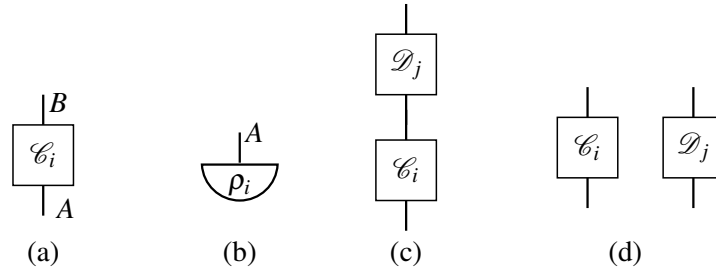


Figure 2: Graphical notation for tests

3.1 The Chiribella-D'Ariano-Perinotti framework

The CDP framework provides an abstract description of ‘circuits’ consisting of operations (which include preparations, transformations and observations) connected by propagating systems. These will be used to describe sources of correlations in our generalised Bayesian networks. First, the way in which elements of the circuits compose will be specified (the ‘operational’ part); then the way in which probabilities are attached to circuits will be described. Together these parts constitute an *operational-probabilistic theory*.

The operational part. To specify this, we consider a collection of systems A, B, C, \dots , including a *trivial* system I (note that our notion of system is that of type, e.g. qubit; CDP have slightly different usage, and would call this an operational equivalence class of systems). Systems are the inputs and outputs of *tests* $\{\mathcal{C}_i\}_{i \in X}$, which represent a single use of some physical device, e.g. a Stern-Gerlach device. The input and output systems need not be identical. The elements of tests, \mathcal{C}_i , represent operationally distinguishable outcomes of the test. They are referred to as *events*, and are indexed by outcomes $i \in X$. For example, for the test corresponding to the use of a Stern-Gerlach device with a spin-half particle, the outcome set would have two elements, corresponding to the two different spin outcomes. If a test $\{\mathcal{C}_i\}_{i \in X}$ is a singleton, i.e. if there is only one outcome $i = i_0$, then we say that this is a *deterministic test*.

CDP use a graphical notation inspired by the work of Abramsky and Coecke [1]. An event \mathcal{C}_i with input system A and output system B is depicted in Fig. 2(a). A *preparation-event* has the trivial system as input; this is depicted in Fig. 2(b). *Observation-events* are the dual notion, for which the output is the trivial system. When the output system of $\{\mathcal{C}_i\}$ is the same as the input system of $\{\mathcal{D}_j\}$, they can be composed *in sequence*, and any two tests can be composed *in parallel*. These are respectively depicted (for events) in Fig. 2(c) and (d). Each type of composition of tests yields another test, whose outcomes (i, j) are ordered pairs formed by the outcomes i and j of each factor. In the case of parallel composition, if \mathcal{C}_i has input system A and output B , and \mathcal{D}_j has input C and output D , then their parallel composition has the *composite systems*, AC and BD , as the input and output types respectively.

The graphical notation keeps track of how the tests in a circuit are connected. Below we will find it useful to also do this symbolically, using an index-based symbolic notation to explicitly refer to the input and output wires of a test. An event with input wires α and output wires β will be represented as $\mathcal{C}_{i\alpha}^\beta$ with the index left blank if there are no wires, i.e. the input or output is the trivial system.¹ Sequential

¹ Note that the superscripts and subscripts do *not* refer to input and output systems. Instead they merely carry information on how tests are connected. For example, in the parallel composition $\mathcal{C}_{i\alpha}^\beta \mathcal{D}_{j\gamma}^\delta$, it may be that $\mathcal{C}_{i\alpha}^\beta$ has the same output type as the input type of $\mathcal{D}_{j\gamma}^\delta$ or vice versa. Our notation is just an alternative to the \otimes vs. \circ notation used for monoidal categories.

composition is then denoted $\mathcal{C}_{i\alpha}^\beta \mathcal{D}_{j\beta}^\gamma$, and parallel composition denoted $\mathcal{C}_{i\alpha}^\beta \mathcal{D}_{j\gamma}^\delta$.

The probabilistic part. An operational-probabilistic theory is defined as one in which every test from the trivial system to itself (pictorially, a diagram with no input or output wires) is a probability distribution over the outcome set, and where the composition of such tests is given by the corresponding product distribution. Two tests are called *operationally equivalent* if substituting one for the other never affects a probability distribution. An operationally equivalent class of observation-events is called an *effect*.

To complete this framework we shall assume the existence of a *unique* deterministic effect \top_A for each system A . This assumption is referred to by CDP as *causality*. In particular, ignoring the outcome of any observation-test always corresponds to this unique deterministic effect. This assumption is necessary for the comparison to Bayesian networks below to make sense: CDP show that it is equivalent to the assumption that the probability of an outcome at time t_1 does not depend on which operation is performed at time t_2 , where $t_2 > t_1$. Hence the causality assumption can also be thought of as ‘no-signalling from the future to the past’. We can now give some examples of operational-probabilistic theories.

Example 5 (Quantum theory). The systems A, B, C, \dots are given by complex Hilbert spaces $\mathcal{H}_A, \mathcal{H}_B, \mathcal{H}_C, \dots$; and in particular, the trivial system is given by the one-dimensional space $I = \mathbb{C}$. Tests are quantum instruments, i.e. sets of complete positive linear maps that sum to a trace preserving map. In particular, deterministic preparation-tests are unit trace positive operators, and observation-tests are of the form $\text{Tr}(E_i \cdot)$ where $\{E_i\}$ is a POVM. Tests compose in sequence by ordinary composition of maps, and in parallel by the vector tensor product. The unique deterministic effect is Tr .

Example 6 (Classical probability theory). We obtain a *classical operational-probabilistic theory* by letting all systems be classical systems, i.e. sample spaces for probability distributions (with finite support). More specifically, the state space of a classical system is a simplex in \mathbb{R}^d , i.e. the convex hull of $d + 1$ affinely independent points. For example, for $d = 1$, the classical state space is a line, i.e. a bit.

A natural question is whether a classical operational-probabilistic theory is, in fact, a Bayesian network. However, there are two reasons why this is not the case. Firstly, there is no classical conditioning in an operational-probabilistic probabilistic theory. That is, a test $\{\mathcal{C}_i\}_i$ should be thought of as a device which can indicate classical outcome, e.g. a light that flashes red or green depending on whether spin up or down is detected. However, in general a physical device will have ‘dials’, which can be used to control which operation will take place (as in the Bell setup). This corresponds to allowing a test $\{\mathcal{C}_i\}_i$ to be a function of a classical input.

Secondly, an operational-probabilistic theory carries *two* types of information in each circuit element: the *systems* that ‘travel’ along the wires, and the *classical outcomes*. For example, consider the following sequence of tests where each system is classical, as in Fig. 3(a). For example, suppose that ρ is the preparation of a coin, which can have either heads or tails facing up, and can be in one of two colours. The test \mathcal{C} could repaint the colour of the coin, but for simplicity let us suppose that each outcome leaves the state of the coin unchanged. The classical outputs are as follows: $x^{(1)}$ is a bit representing the colour of the coin at t_1 , $x^{(2)}$ is a bit representing heads facing up at t_2 , and $x^{(3)}$ is a bit representing the colour of the coin at t_3 . This yields a classical probability distribution $P(x^{(1)}, x^{(2)}, x^{(3)})$. Now, suppose that we try to interpret this circuit as a classical Bayesian network, as in Fig. 3(b). The Markov condition implies that $X^{(3)} \perp\!\!\!\perp X^{(1)} \mid X^{(2)}$. But if ρ is the preparation of a coin with heads facing up, and in each colour with uniform probability, then $X^{(3)}$ is perfectly correlated with $X^{(1)}$, even conditioning on $X^{(2)}$. Hence the Markov condition fails to hold. The root of this problem is that, in the CDP formalism, a test’s outcome doesn’t always fix the output state, even when the systems are assumed to be classical.

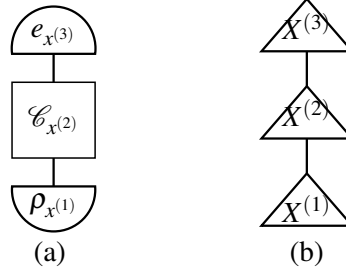


Figure 3: (a) A sequence of classical tests. (b) The putative translation to a Bayesian network.

In the next subsection we shall define a framework for general probabilistic theories which overcomes these two problems.

3.2 Generalised Bayesian networks

Our aim in this subsection will be to generalise Bayesian networks in a way that can allow non-classical resources. For example, the generalisation in the Bell scenario is depicted in Figure 4, where the circular node represents the new ‘general’ resources that we shall introduce. We begin by labelling the nodes in

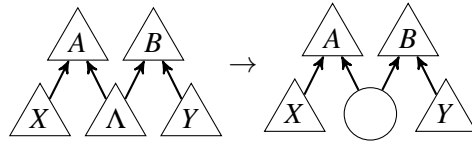


Figure 4: Schematic depiction of our generalisation of Bayesian networks.

a DAG accordingly.

Definition 7. Let G be a DAG with nodes $V = \{X^{(1)}, X^{(2)}, \dots, X^{(m)}\}$. We shall say that G is a *generalised DAG* if V can be partitioned into two sets of nodes:

1. the *classical nodes* $\{X^{(1)}, X^{(2)}, \dots, X^{(n)}\}$, and
2. the *general nodes* $\{X^{(n+1)}, \dots, X^{(m)}\}$.

We choose this terminology because all ‘classical data’, e.g. the observations from measurements, will be associated to classical nodes. On the other hand, the general nodes will replace ‘latent’ random variables with ‘general resources’, e.g. replacing the source λ in the Bell DAG with a general node will allow Alice and Bob to share a quantum state or a PR box state, as in Figure 4.

We shall use the CDP framework to assign tests to each node. However, in the previous subsection, we discussed that a circuit element in the CDP framework carries *two* types of data: the classical data associated with an outcome, and the system. We noted that this makes it problematic to interpret a CDP circuit as a Bayesian network. Our framework will address this problem by using generalised DAGs. In particular we shall define the *outputs* of classical and general nodes in distinct ways:

1. *Classical nodes*: these will *not* output a system, but *will* have a classical outcome (i.e. random variable) X assigned to it. In CDP language, a classical node has the trivial system as output.

2. *General nodes*: these will *always* output a system, and but will *not* have any non-trivial outcomes assigned to it. For convenience of notation we shall associate a classical random variable with every node² $X^{(i)}$. However, the random variable associated with general nodes will be trivial, taking only one value with probability one.

Accordingly, we shall associate a non-trivial probability distribution $P(x^{(1)}, x^{(2)}, \dots, x^{(n)})$ only with the classical nodes.

However, although classical and general nodes will have different definitions for outputs, they will have the same definition for inputs. That is, both types of node can have inputs which are either classical nodes or general nodes or both. The upshot of this is that:

- All nodes will be assigned CDP tests, $\{\mathcal{T}_{x^{(i)}}\}_{x^{(i)}}$, where the test referred to is implied by the name of the classical outcome $x^{(i)}$ (which in turn is associated to particular node $X^{(i)}$).
- General nodes will be assigned deterministic tests that have a non-trivial output system, while classical nodes will be assigned observation tests (tests with the trivial output system). Hence a classical node can have a general node as a parent (a ‘general parent’). E.g. the classical nodes A and B in the right-hand diagram of Fig. 4 have the same general parent (e.g. a Bell state).

For a node $X^{(i)}$ in a generalised DAG, we denote the ingoing edges that come from general parents as $\text{IN}X^{(i)}$, and the outgoing edges (to all children) as $\text{OUT}X^{(i)}$. Then $\mathcal{T}_{\text{IN}X^{(i)}}^{\text{OUT}X^{(i)}}$ denotes a test assigned to node $X^{(i)}$ (we briefly suppress the index $x^{(i)}$), with an input wire for each ingoing edge from a general parent, and an output wire for each outgoing edge. Then we can write $\mathcal{C}_{\text{IN}X^{(1)}}^{\text{OUT}X^{(1)}} \mathcal{D}_{\text{IN}X^{(2)}}^{\text{OUT}X^{(2)}}$ using the notation defined in Section 3.1, e.g. if there is an outgoing edge from $X^{(1)}$ to $X^{(2)}$ in the underlying DAG, then the tests are composed in sequence, \mathcal{C} followed by \mathcal{D} .

Definition 8. Let G be a generalised DAG. A probability distribution is *generalised Markov* with respect to G if it satisfies³:

$$P(x^{(1)}, x^{(2)}, \dots, x^{(n)}) = \prod_{i=1}^m \mathcal{T}_{x^{(i)}}(\text{cpa}x^{(i)})_{\text{IN}X^{(i)}}^{\text{OUT}X^{(i)}}$$

where $\mathcal{T}_{x^{(i)}}(\text{cpa}x^{(i)})$ is the outcome $x^{(i)}$ of a test assigned to node $X^{(i)}$, and this test is a function of the outcome $\text{cpa}x^{(i)}$ of the classical parents of node $X^{(i)}$.

Informally, the generalised Markov condition holds when a probability distribution is equal to a CDP closed circuit whose diagram matches the structure of the given DAG, i.e. circuit boxes are assigned to nodes. Moreover, tests now have classical ‘dials’, i.e. classical parents, which can be used e.g. to control which measurement occurs in a Bell scenario. We can now formalise the graphical notation depicted in Fig. 4. Outgoing edges from general nodes represent general systems, and the output type of the test will be the parallel composition of the systems on all the outgoing edges. On the other hand, outgoing edges from a classical node X do *not* represent systems, but instead represent the fact that the test implemented at the child node is a function of the outcome at X .

Example 9 (Prepare and measure). A preparation followed by a measurement can be depicted as Fig. 5(a). This generalised DAG has one general node, which has only a classical child and a classical parent. A probability distribution $P(x, y)$ which is generalised Markov with respect to this generalised DAG is given

²As is the case for classical Bayesian networks, we shall use the same symbol $X^{(i)}$ to denote both the node and the random variable associated with the node; context will determine which is being referred to.

³The product denoted by \prod that is used here follows the notation defined earlier, e.g. we write $\mathcal{C}_{i\alpha}^{\beta} \mathcal{D}_{j\beta}^{\gamma}$ for sequential composition of $\mathcal{C}_{i\alpha}^{\beta}$ followed by $\mathcal{D}_{j\beta}^{\gamma}$.

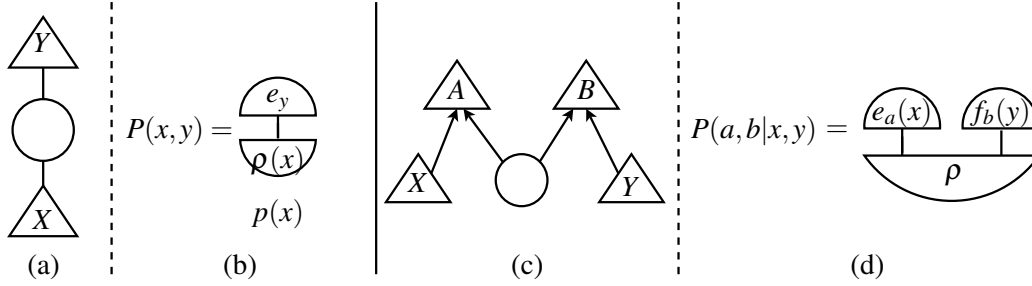


Figure 5: Prepare and measure: (a) in our framework and (b) the corresponding family of CDP circuits; the Bell setup: (c) in our framework and (d) the corresponding family of CDP circuits.

by the family of CDP circuits depicted in Fig. 5(b). Here $\rho(x)$ denotes a state prepared according to the value of the classical control variable x , whose probability distribution is $p(x)$ (which is not part of the CDP graphical notation, and is hence put in ‘by hand’ in the equation).

Definition 10. A *generalised Bayesian network* is a pair (P, G) , such that G is a generalised DAG, and P is generalised Markov with respect to G .

The definition of a generalised Bayesian network is therefore exactly analogous to that of a classical Bayesian network.

Example 11 (Bell setup). We can define a generalised Bayesian network depicted in Fig. 5(c), corresponding to the Bell scenario. A probability distribution $P(a, b|x, y)$ that is generalised Markov for this generalised DAG is given by the family of CDP circuits depicted in Fig. 5(d). In this case, the measurements are controlled by classical inputs, e.g. Alice’s measurement $\{e_a\}$ is a function of the classical variable x .

A generalised Bayesian network will allow us to explore the consequences of using non-classical resources in place of classical latent variables. However, we recover classical Bayesian networks if we restrict to classical nodes.

Proposition 12. If all nodes are classical, then a generalised Bayesian network is a classical Bayesian network.

For a given DAG, we will use \mathcal{G} to denote the set of probabilities that are generalized Markov. If the theory is restricted to be quantum theory, we will use \mathcal{Q} . If the theory is restricted to be classical probability theory, we will use \mathcal{C} . Since classical probability theory can be embedded into quantum theory by using diagonal operators, we have $\mathcal{C} \subseteq \mathcal{Q} \subseteq \mathcal{G}$. Finally, we will use \mathcal{I} to denote the set of probabilities that satisfy all of the conditional independences that follow from d -separation. In this notation, the first part of Theorem 4 states that $\mathcal{C} \subseteq \mathcal{I}$ for all DAGs. We will now strengthen this to $\mathcal{G} \subseteq \mathcal{I}$, which is implied by the following theorem.

Theorem 13. Let G be a generalised DAG with disjoint classical subsets X, Y and Z . Then

- (i) If P is generalised Markov with respect to G , then $X \perp Y \mid Z \Rightarrow X \perp\!\!\!\perp Y \mid Z$.
- (ii) If $X \perp\!\!\!\perp Y \mid Z$ holds for all P which are generalised Markov with respect to G , then $X \perp Y \mid Z$.

The importance of this theorem is that: (a) it generalises a powerful tool from the classical theory (and so e.g. the relevant causal discovery algorithms will still apply); (b) for a given DAG, generalised Bayesian networks satisfy all the observable conditional independences that a classical Bayesian network does. This suggests that our framework captures a similar notion of causality as the classical one, and it hence provides an alternative to the Leifer-Spekkens approach [10], which also aims at achieving this.

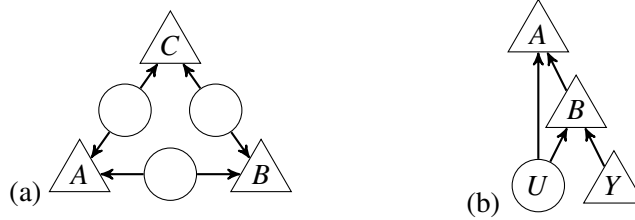


Figure 6: (a) The ‘triangle’ DAG and (b) the ‘instrumental’ DAG.

4 Beyond conditional independence

4.1 Quantitative bounds on correlations

In the Bell scenario, Bell inequalities limit the classical correlations, and Tsirelson inequalities limit the quantum correlations. What limits the correlations in a general probabilistic theory? In the Bell scenario, a general probabilistic theory is limited *only* by the no-signalling principle (see for example, [2]). In our notation, this means that $\mathcal{G} = \mathcal{I}$ for Bell DAGs. Here we show that this fact does not extend to every scenario, i.e. we show that $\mathcal{G} \subsetneq \mathcal{I}$ holds for two specific scenarios, which we call the ‘triangle’ DAG and the ‘instrumental’ DAG. In other words, causal structure can impose quantitative limits *independently* of the precise physical theory under consideration.

1. The triangle DAG. The triangle scenario, shown in Fig. 6(a), has already received some interest in quantum foundations [5, 8, 6] and the causality literature. Branciard *et al.* initially introduced the scenario with definitions matching our \mathcal{C} and \mathcal{Q} [5]. It was noted that understanding the classical correlations \mathcal{C} in this scenario is much more mathematically challenging than in the Bell scenario. Nevertheless, Fritz showed that there exist quantum correlations for this scenario which cannot be reproduced using classical sources, i.e. $\mathcal{C} \subsetneq \mathcal{Q}$ [8]. A key part of this proof was showing that any $P \in \mathcal{C}$ satisfies a ‘monogamy’ inequality, defined in terms of the mutual information I and the Shannon entropy H :

$$I(A : B) + I(B : C) \leq H(B). \quad (1)$$

That is, the stronger the correlations between A and B , the weaker the correlations between B and C .

This has some interesting consequences. For example, note that there are no independences between observable nodes for this DAG. Hence the ‘perfectly correlated bits’ distribution $P(0, 0, 0) = P(1, 1, 1) = \frac{1}{2}$ is in \mathcal{I} . However, this perfect correlation violates Eq. (1), and hence cannot be produced using classical sources. We generalise this as follows.

Theorem 14. *Equation (1) holds for any $P \in \mathcal{G}$.*

Hence perfect correlation cannot be produced in this DAG using any generalised probabilistic theory. In other words, $\mathcal{G} \subsetneq \mathcal{I}$.

2. The instrumental DAG. The fact that $\mathcal{C} \subsetneq \mathcal{I}$ for the DAG in Fig. 6(b) has already been noted in the causality literature [11]. The original interest in this DAG arose in the study of cases of imperfect compliance in a controlled trial. For example Y might be a randomly assigned treatment, B the treatment the patient actually follows, and A recovery. There could be factors U that influence both the chance of recovery under each treatment, and also the chance of compliance with a particular treatment (e.g.

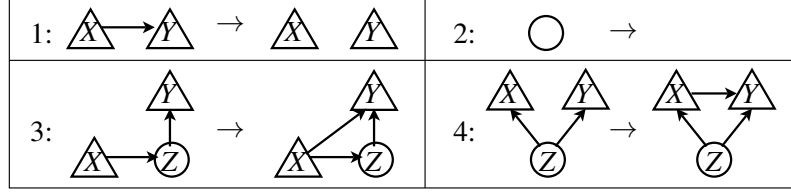


Figure 7: Illustrations of the allowed transformations for the sufficient condition.

due to side effects). This model does not imply any conditional independences on the observable nodes $\{A, B, Y\}$, and so \mathcal{I} is just the set of all probability distributions. However, Pearl [11] showed that the model can still be tested because for any $P \in \mathcal{C}$:

$$\max_b \sum_a \max_y P(a, b|y) \leq 1. \quad (2)$$

This is known as the *instrumental inequality*. Here we strengthen this result to:

Theorem 15. *Equation (2) holds for any $P \in \mathcal{G}$.*

This result implies that $\mathcal{G} \subsetneq \mathcal{I}$ for this DAG. Our results are surprising because the states of generalised probabilistic theories are not usually thought of as providing a ‘cause’ of the correlations. But our results suggest otherwise: they show that even if an arbitrary probabilistic theory is allowed (e.g. box-world), the causal structure that states are assigned to *still* restricts the possible observed correlations.

4.2 Towards a classification of ‘interesting’ DAGs

It is known that a Bell scenario where only one party has a choice of measurement is not ‘interesting’, in the sense that $\mathcal{C} = \mathcal{I}$. But for any DAG $\mathcal{C} \subseteq \mathcal{Q} \subseteq \mathcal{G} \subseteq \mathcal{I}$, so DAGs in which $\mathcal{C} = \mathcal{I}$ must have $\mathcal{C} = \mathcal{Q} = \mathcal{G} = \mathcal{I}$. It is therefore of interest to classify which DAGs have $\mathcal{C} = \mathcal{I}$ and which do not. The DAGs that are interesting are candidates for quantum advantages in ‘black-box’ information processing.

Here we make significant progress towards such a classification by providing a sufficient condition for $\mathcal{C} = \mathcal{I}$. Let X and Y be classical nodes. We write $X \rightsquigarrow Y$ to denote the existence of a directed path from X to Y , for which any intermediate nodes are unobserved.

Theorem 16. *Consider the set of classical correlations \mathcal{C}_G for a generalised DAG G . Suppose that one of the following transformations (depicted in Fig. 7) is performed on G , producing a generalised DAG H : (1) removal of an edge (between nodes of either type); (2) removal of an isolated unobserved node; (3) addition of an edge $X \rightarrow Y$ where previously $X \rightsquigarrow Y$; (4) addition of an edge $X \rightarrow Y$ where previously $PAX \subseteq PAY$ and PAX contained at least one unobserved node. Then $\mathcal{C}_H \subseteq \mathcal{C}_G$.*

The sufficient condition for $\mathcal{I} = \mathcal{C}$ is: If starting with a given DAG one can apply a sequence of transformations, such that each transformation is one from Theorem 16, and produce a DAG with: (1) no circular nodes; and (2) requiring no more conditional independences on the triangular nodes than the original DAG did; then $\mathcal{I} = \mathcal{C}$ for the original DAG. To see this, start with some probability distribution $P \in \mathcal{I}$. Recalling that the conditional independences are the only restrictions on DAGs with no latent variables, the above two properties ensure that the distribution P is Markov for the new DAG H , and hence classical, i.e. $P \in \mathcal{C}_H$. But then by Theorem 16 there is a classical model for the original DAG with the same probabilities for the triangular nodes, and we are done.

For example, this condition establishes that the Bell scenario where only one party has a measurement choice has $\mathcal{I} = \mathcal{C}$, as shown in Fig. 8 (numbers indicate the relevant transformation from Theorem 16).

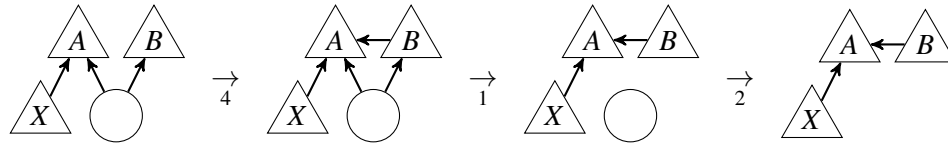


Figure 8: Repeated applications of Theorem 16 transform the one-setting Bell DAG into a new DAG without enlarging \mathcal{C} . Allowed distributions on the final DAG are constrained only by $X \perp\!\!\!\perp B$, which held for \mathcal{I} in the initial DAG, and so $\mathcal{C} = \mathcal{I}$ in the initial DAG.

Acknowledgements. We are grateful for useful discussions with Jonathan Barrett, Giulio Chiribella, Tobias Fritz and Rob Spekkens.

References

- [1] S. Abramsky & B. Coecke (2004): *A categorical semantics of quantum protocols*. In: *Proc. 19th Annual IEEE Symposium on Logic in Computer Science*, pp. 415–425, doi:10.1109/LICS.2004.1319636.
- [2] Jonathan Barrett (2007): *Information processing in generalized probabilistic theories*. *Phys. Rev. A* 75, p. 032304, doi:10.1103/PhysRevA.75.032304.
- [3] John S. Bell (1964): *On the Einstein-Podolsky-Rosen paradox*. *Physics* 1(3), pp. 195–200.
- [4] C. Branciard, N. Gisin & S. Pironio (2010): *Characterizing the Nonlocal Correlations Created via Entanglement Swapping*. *Phys. Rev. Lett.* 104, p. 170401.
- [5] Cyril Branciard, Denis Rosset, Nicolas Gisin & Stefano Pironio (2012): *Bilocal versus nonbilocal correlations in entanglement-swapping experiments*. *Phys. Rev. A* 85, p. 032119.
- [6] R. Chaves, L. Luft & D. Gross (2013): *Causal structures from entropic information: Geometry and novel scenarios*.
- [7] G. Chiribella, G.M. D’Ariano & P. Perinotti (2010): *Probabilistic theories with purification*. *Phys. Rev. A* 81, p. 062348, doi:10.1103/PhysRevA.81.062348.
- [8] Tobias Fritz (2012): *Beyond Bell’s theorem: correlation scenarios*. *New J. Phys.* 14, p. 103001.
- [9] Steffen Lauritzen (1996): *Graphical models*. Clarendon, Oxford.
- [10] Matthew. S. Leifer & Robert W. Spekkens (2013): *Towards a formulation of quantum theory as a causally neutral theory of Bayesian inference*. *Phys. Rev. A* 88, p. 052130, doi:10.1103/PhysRevA.88.052130.
- [11] Judea Pearl (1995): *On the Testability of Causal Models with Latent and Instrumental Variables*. In: *Proceedings of the Eleventh Conference Annual Conference on Uncertainty in Artificial Intelligence (UAI-95)*, Morgan Kaufmann, San Francisco, CA, pp. 435–443.
- [12] Judea Pearl (2009): *Causality*, 2 edition. Cambridge University Press.
- [13] Sandu Popescu (1995): *Bell’s Inequalities and Density Matrices: Revealing Hidden Nonlocality*. *Phys. Rev. Lett.* 74, pp. 2619–2622, doi:10.1103/PhysRevLett.74.2619.
- [14] Sandu Popescu & Daniel Rohrlich (1994): *Quantum nonlocality as an axiom*. *Found. Phys.* 24(3), pp. 379–385, doi:10.1007/BF02058098.
- [15] Thomas Verma & Judea Pearl (1988): *Causal Networks: Semantics and Expressiveness*. In: *Proceedings of the 4th Workshop on Uncertainty in Artificial Intelligence*, Minneapolis, MN, pp. 352–359.
- [16] C. J. Wood & R. W. Spekkens (2012): *The lesson of causal discovery algorithms for quantum correlations: Causal explanations of Bell-inequality violations require fine-tuning*.

A Study of Entanglement in a Categorical Framework of Natural Language

Dimitri Kartsaklis

University of Oxford
Department of Computer Science
Oxford, UK

dimitri.kartsaklis@cs.ox.ac.uk

Mehrnoosh Sadrzadeh

Queen Mary University of London
School of Electronic Engineering and Computer Science
London, UK

mehrnoosh.sadrzadeh@qmul.ac.uk

In both quantum mechanics and corpus linguistics based on vector spaces, the notion of entanglement provides a means for the various subsystems to communicate with each other. In this paper we examine a number of implementations of the categorical framework of Coecke et al. [4] for natural language, from an entanglement perspective. Specifically, our goal is to better understand in what way the level of entanglement of the relational tensors (or the lack of it) affects the compositional structures in practical situations. Our findings reveal that a number of proposals for verb construction lead to almost separable tensors, a fact that considerably simplifies the interactions between the words. We examine the ramifications of this fact, and we show that the use of Frobenius algebras mitigates the potential problems to a great extent. Finally, we briefly examine a machine learning method that creates verb tensors exhibiting a sufficient level of entanglement.

1 Introduction

Category theory in general and compact closed categories in particular provide a high level framework to identify and study universal properties of mathematical and physical structures. Abramsky and Coecke [1], for example, use the latter to provide a structural proof for a class of quantum protocols, essentially recasting the vector space semantics of quantum mechanics in a more abstract way. This and similar kinds of abstraction have made compact closed categories applicable to other fields with vector space semantics, for the case of this paper, corpus linguistics. Here, Coecke et al.[4] used them to unify two seemingly orthogonal semantic models of natural language: a syntax-driven compositional approach as expressed by Lambek [15] and distributional models of meaning based on vector spaces. The latter approach is capable of providing a concrete representation of the meaning of a word, by creating a vector with co-occurrence counts of that word in a corpus of text with all other words in the vocabulary. Distributional models of this form have been proved useful in many natural language processing tasks [22, 17, 16], but in general they do not scale up to larger text constituents such as phrases and sentences. On the other hand, the type-logical approaches to language as introduced in [15] are compositional but unable to provide a convincing model of word meaning.

The unification of the two semantics paradigms is based on the fact that both a type logic expressed as a pregroup [15] and finite dimensional vector spaces share a compact closed structure; so in principle there exists a way to express a grammatical derivation as a morphism that defines mathematical manipulations between vector spaces, resulting in a sentence vector. In [4], the solution was based on a Cartesian product between the pregroup category and the category of finite dimensional vector spaces; later this was recasted in a functorial passage from the former to the latter [19, 3, 10]. The general idea behind any of these frameworks is that the grammatical type of each word determines the vector space where the corresponding vector lives. Words with atomic types, such as nouns, are simple vectors living in N . On the other hand, words with relational types, such as adjectives or verbs, live in tensor product spaces of higher order. For instance, an intransitive verb will be an element of an order-2 space such as

$N \otimes S$, whereas a transitive verb will live in $N \otimes S \otimes N$. These tensors act on their arguments by *tensor contraction*, a generalization of the familiar notion of matrix multiplication to higher order tensors.

Since every relational word is represented by a tensor, naturally *entanglement* becomes an important issue in these models. Informally speaking, elements of tensor spaces which represent meanings of relational words should be entangled to allow for a so called ‘flow of information’ (a terminology borrowed from categorical quantum mechanics [1]) among the meanings of words in a phrase or sentence. Otherwise, parts of the meaning of these words become isolated from the rest, leading to unwanted consequences. An example would be that all sentences that have the same verb end up to get the same meaning regardless of the rest of the context, and this is obviously not the case in language. Whereas at least intuitively the above argument makes sense, in some of the language tasks we have been experimenting with, non-entangled tensors have provided very good results. For example, in [8] Grefenstette and Sadrzadeh provide results for verbs that are built from the outer product of their context vectors. These results beat the state of the art of that time (obtained by the same authors in a previous paper [7]) by a considerable difference.

The purpose of the current paper is to provide a preliminary study of the entanglement in corpus linguistics and to offer some explanation why phenomena such as the above have been the case: is this a by-product of the task or the corpus or the specific concrete model? We work with a number of concrete instantiations of the framework in sentence similarity tasks and observe their performances experimentally from an entanglement point of view. Specifically, we investigate a number of models based on the weighted relations method of [7], where a verb matrix is computed as the structural mixing of all subject/object pairs with which it appears in the training corpus. We also test a model trained using linear regression [2]. Our findings for the first case have been surprising. It turns out that, contrary to intuition and despite the fact that the construction method should yield entangled matrices, the results are very close to their rank-1 approximations, that is, they are in effect separable. We further investigate the ramifications of this observation and try to explain the good practical predictions. We then experiment with the linear regression model of [2] and show that the level of entanglement is much higher in the verbs of this model. Finally, we look at a number of Frobenius variations of the weighted relation models, such as the ones presented in [13] and a few new constructions exclusive to this paper. The conclusions here are also surprising, but in a positive way. It seems that Frobenius models are able to overcome the unwanted “no-flow” collapses of the separable verbs by generating a partial flow between the verb and either its subject or its object, depending which dimension they are copying.

2 Quantizing the grammar

The purpose of the categorical framework is to map a grammatical derivation to some appropriate manipulation between vector spaces. In this section we will shortly review how this goal is achieved. Our basic type logic is a *pregroup grammar* [15], built on the basis of a pregroup algebra. This is a partially ordered monoid with unit 1, whose each element p has a left adjoint p^l and a right adjoint p^r . This means that they satisfy the following inequalities:

$$p^l \cdot p \leq 1 \quad p \cdot p^r \leq 1 \quad \text{and} \quad 1 \leq p \cdot p^l \quad 1 \leq p^r \cdot p \quad (1)$$

A pregroup grammar is the pregroup freely generated over a set of atomic types, which for this paper will be $\{n, s\}$. Here, type n refers to nouns and noun phrases, and type s to sentences. The atomic types and their adjoints can be combined to create types for *relational words*. The type of an adjective, for example, is $n \cdot n^l$, representing something that inputs a noun (from the right) and outputs another noun. Similarly, the type of a transitive verb $n^r \cdot s \cdot n^l$ reflects the fact that verbs of this kind expect two inputs, one noun at each side. A grammatical reduction then follows from the properties of pregroups

and specifically the inequalities in (1) above. The derivation for the sentence ‘Happy kids play games’ has the following form:

$$(n \cdot n^l) \cdot n \cdot (n^r \cdot s \cdot n^l) \cdot n = n \cdot (n^l \cdot n) \cdot n^r \cdot s \cdot (n^l \cdot n) \leq n \cdot 1 \cdot n^r \cdot s \cdot 1 = n \cdot n^r \cdot s \leq 1 \cdot s = s$$

Categorically, a pregroup grammar conforms to the definition of a non-symmetric *compact closed category* (to which we will refer as **Preg_F**). Specifically, the inequalities in (1) correspond to the ε and η morphisms of a compact closed category, given as follows:

$$\varepsilon^l : A^l \otimes A \rightarrow I \quad \varepsilon^r : A \otimes A^r \rightarrow I \quad (2)$$

$$\eta^l : I \rightarrow A \otimes A^l \quad \eta^r : I \rightarrow A^r \otimes A \quad (3)$$

Hence the above grammatical reduction becomes the following morphism:

$$(\varepsilon_n^r \otimes 1_s) \circ (1_n \otimes \varepsilon_n^l \otimes 1_{n^r \cdot s} \otimes \varepsilon_n^l) \quad (4)$$

Let us now refer to the category of finite-dimensional vector spaces and linear maps over \mathbb{R} as **FVect_W**, where W is our basic distributional vector space with an orthonormal basis $\{w_i\}_i$. This category is again compact closed (although a symmetric one, since $W \cong W^*$), with the ε and η maps given as follows:

$$\varepsilon^l = \varepsilon^r : W \otimes W \rightarrow \mathbb{R} :: \sum_{ij} c_{ij}(\vec{w}_i \otimes \vec{w}_j) \mapsto \sum_{ij} c_{ij} \langle \vec{w}_i | \vec{w}_j \rangle \quad (5)$$

$$\eta^l = \eta^r : \mathbb{R} \rightarrow W \otimes W :: 1 \mapsto \sum_i \vec{w}_i \otimes \vec{w}_i \quad (6)$$

The transition from a pregroup reduction to a morphism between vector spaces is achieved through a *strongly monoidal functor* $\mathcal{F} : \mathbf{Preg}_F \rightarrow \mathbf{FVect}_W$ which preserves the compact structure between the two categories, that is we have $\mathcal{F}(A^l) = \mathcal{F}(A)^l$ and $\mathcal{F}(A^r) = \mathcal{F}(A)^r$. Further, since **FVect_W** is symmetric and W has a fixed basis, we have that $\mathcal{F}(A)^r = \mathcal{F}(A)^l \cong \mathcal{F}(A)$. As motivated in previous work [13], we assume that \mathcal{F} assigns the basic vector space W to both of the atomic types, that is we have:

$$\mathcal{F}(n) = \mathcal{F}(s) = W \quad (7)$$

The partial orders between the atomic types are mapped to linear maps from W to W by functoriality. The adjoints of atomic types are also mapped to W , whereas the complex types are mapped to tensor products of vector spaces:

$$\mathcal{F}(n \cdot n^l) = \mathcal{F}(n^r \cdot s) = W \otimes W \quad \mathcal{F}(n^r \cdot s \cdot n^l) = W \otimes W \otimes W \quad (8)$$

We are now in position to define the meaning of a sentence $w_1 w_2 \dots w_n$ with type reduction α as follows:

$$\mathcal{F}(\alpha)(\vec{w}_1 \otimes \vec{w}_2 \otimes \dots \otimes \vec{w}_n) \quad (9)$$

For example, the meaning of the sentence ‘happy kids play games’, which has the grammatical reduction (4), is computed as follows:

$$\begin{aligned} & \mathcal{F} \left[(\varepsilon_n^r \otimes 1_s) \circ (1_n \otimes \varepsilon_n^l \otimes 1_{n^r \cdot s} \otimes \varepsilon_n^l) \right] \left(\overrightarrow{\text{happy}} \otimes \overrightarrow{\text{kids}} \otimes \overrightarrow{\text{play}} \otimes \overrightarrow{\text{games}} \right) = \\ & (\varepsilon_W \otimes 1_W) \circ (1_W \otimes \varepsilon_W \otimes 1_{W \otimes W} \otimes \varepsilon_W) \left(\overrightarrow{\text{happy}} \otimes \overrightarrow{\text{kids}} \otimes \overrightarrow{\text{play}} \otimes \overrightarrow{\text{games}} \right) \end{aligned}$$

The above categorical computations simplify to the following form:

$$(\overrightarrow{\text{happy}} \times \overrightarrow{\text{kids}})^T \times \overrightarrow{\text{play}} \times \overrightarrow{\text{games}} \quad (10)$$

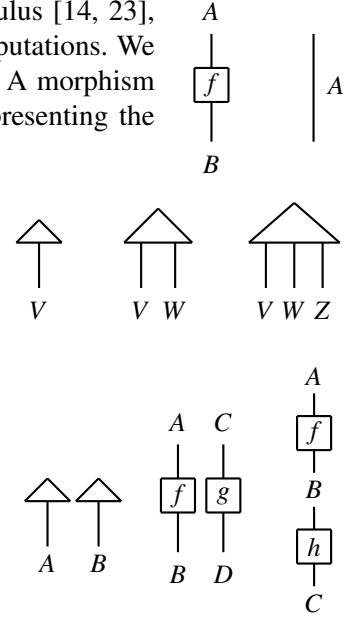
where symbol \times denotes tensor contraction and the above is a vector living in our basic vector space W .

3 Pictorial calculus

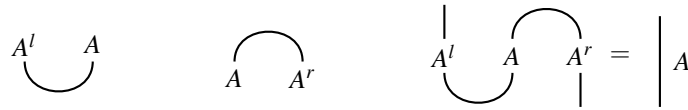
Compact closed categories are complete with regard to a pictorial calculus [14, 23], which can be used for visualizing the derivations and simplifying the computations. We introduce the fragment of calculus that is relevant to the current paper. A morphism $f : A \rightarrow B$ is depicted as a box with incoming and outgoing wires representing the objects; the identity morphism $1_A : A \rightarrow A$ is a straight line.

Recall that the objects of \mathbf{FVect}_W are vector spaces. However, for our purposes it is also important to access individual vectors within a vector space. In order to do that, we represent a vector $\vec{v} \in V$ as a morphism $\vec{v} : I \rightarrow V$. The unit object is depicted as a triangle, while the number of wires emanating from it denotes the order of the corresponding tensor.

Tensor products of objects and morphisms are depicted by juxtaposing the corresponding diagrams side by side. Composition, on the other hand, is represented as a vertical superposition. For example, from left to right, here are the pictorial representations of the tensor of a vector in A with a vector in B , a tensor of morphisms $f \otimes g : A \otimes C \rightarrow B \otimes D$, and a composition of morphisms $h \circ f$ for $f : A \rightarrow B$ and $h : B \rightarrow C$.

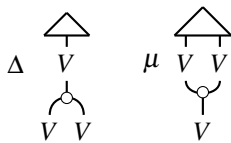
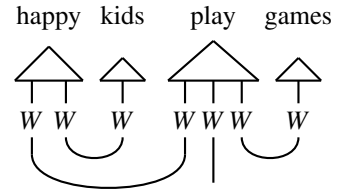


The ε -maps are represented as cups (\cup) and the η -maps as caps (\cap). Equations such as $(\varepsilon_A^l \otimes 1_{A^r}) \circ (1_{A^l} \otimes \eta_A^r) = 1_A$ now get an intuitive visual justification:



We are now in position to provide a diagram for the meaning of the sentence ‘happy kids play games’ (*right*).

We conclude this section with one more addition to our calculus. As in most quantum protocols, some times the flow of information in linguistics requires elements of classical processing; specifically, we will want the ability to *copy* and *delete* information, which can be provided by introducing *Frobenius algebras*. In \mathbf{FVect} , any vector space V with a fixed basis $\{\vec{v}_i\}$ has a Frobenius algebra over it given by Eqs. 11 below.



$$\begin{aligned} \Delta :: \vec{v}_i &\mapsto \vec{v}_i \otimes \vec{v}_i \\ \mu :: \vec{v}_i \otimes \vec{v}_j &\mapsto \delta_{ij} \vec{v}_i := \begin{cases} \vec{v}_i & i = j \\ 0 & i \neq j \end{cases} \end{aligned}$$

$$\begin{aligned} \iota :: \vec{v}_i &\mapsto 1 \\ \zeta :: 1 &\mapsto \sum_i \vec{v}_i \end{aligned} \quad (11)$$

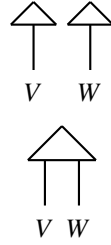
4 Entanglement in quantum mechanics and linguistics

Given two non-interacting quantum systems A and B , where A is in state $|\psi\rangle_A$ and B in state $|\psi\rangle_B$, we denote the state of the composite system $A \otimes B$ by $|\psi\rangle_A \otimes |\psi\rangle_B$. States of this form that can be expressed as the tensor product of two state vectors are called *product* states, and they constitute a special case of separable states. In general, however, the state of a composite system is not necessarily a product state or even a separable one. Fixing bases $\{|i\rangle_A\}$ and $\{|j\rangle_B\}$ for the vector spaces of the two states, a general composite state (separable or not) is denoted as follows:

$$|\psi\rangle_{AB} = \sum_{ij} c_{ij} |i\rangle_A \otimes |j\rangle_B \quad (12)$$

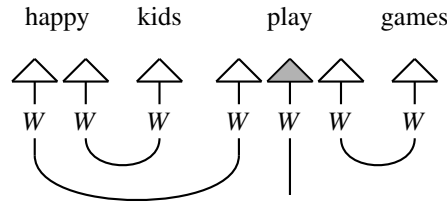
In the case of a pure quantum state, $|\psi\rangle_{AB}$ is separable only if it can be expressed as the tensor product of two vectors; otherwise it is *entangled*. In a similar way, the tensor of a relational word is separable if it is equal to the tensor product of two vectors. In our graphical calculus, these objects are depicted by the juxtaposition of two or more triangles.

In general, a tensor is not separable if it is a linear combination of many separable tensors. The number of separable tensors needed to express the original tensor is equal to the *tensor rank*. Graphically, a tensor of this form is shown as a single triangle with two or more legs.



5 Consequences of separability

In categorical quantum mechanics terms, entangled states are necessary to allow the flow of information between the different subsystems. In this section we show that the same is true for linguistics. Consider the diagram of our example derivation, where all relational words are now represented by separable tensors (in other words, no entanglement is present):



In this version, the ε -maps are completely detached from the components of the relational tensors that carry the results (left-hand wire of the adjective and middle wire of the verb); as a consequence, flow of information is obstructed, all compositional interactions have been eliminated, and the meaning of the sentence is reduced to the middle component of the verb (shaded vector) multiplied by a scalar, as follows (superscripts denote the left-hand, middle, and right-hand components of separable tensors):

$$\langle \overrightarrow{happy}^{(r)} | \overrightarrow{kids} \rangle \langle \overrightarrow{happy}^{(l)} | \overrightarrow{play}^{(l)} \rangle \langle \overrightarrow{play}^{(r)} | \overrightarrow{games} \rangle \overrightarrow{play}^{(m)}$$

Depending on how one measures the distance between two sentences, this is a very unwelcome effect, to say the least. When using cosine distance, the meaning of all sentences with ‘play’ as the verb will be exactly the same and equal to the middle component of the ‘play’ tensor. For example, the sentence “trembling shadows play hide-and-seek” will have the same meaning as our example sentence. Similarly, the comparison of two arbitrary transitive sentences will be reduced to comparing just the middle components of their verb tensors, completely ignoring any surrounding context. The use of

Structure	Simplification	Cos-measured result
adjective-noun	$\overrightarrow{adj} \times \overrightarrow{noun} = (\overrightarrow{adj}^{(l)} \otimes \overrightarrow{adj}^{(r)}) \times \overrightarrow{noun} = \langle \overrightarrow{adj}^{(r)} \overrightarrow{noun} \rangle \cdot \overrightarrow{adj}^{(l)}$	$\overrightarrow{adj}^{(l)}$
intrans. sentence	$\overrightarrow{subj} \times \overrightarrow{verb} = \overrightarrow{subj} \times (\overrightarrow{verb}^{(l)} \otimes \overrightarrow{verb}^{(r)}) = \langle \overrightarrow{subj} \overrightarrow{verb}^{(l)} \rangle \cdot \overrightarrow{verb}^{(r)}$	$\overrightarrow{verb}^{(r)}$
verb-object	$\overrightarrow{verb} \times \overrightarrow{obj} = (\overrightarrow{verb}^{(l)} \otimes \overrightarrow{verb}^{(r)}) \times \overrightarrow{obj} = \langle \overrightarrow{verb}^{(r)} \overrightarrow{obj} \rangle \cdot \overrightarrow{verb}^{(l)}$	$\overrightarrow{verb}^{(l)}$
transitive sentence	$\overrightarrow{subj} \times \overrightarrow{verb} \times \overrightarrow{obj} = \overrightarrow{subj} \times (\overrightarrow{verb}^{(l)} \otimes \overrightarrow{verb}^{(m)} \otimes \overrightarrow{verb}^{(r)}) \times \overrightarrow{obj} = \langle \overrightarrow{subj} \overrightarrow{verb}^{(l)} \rangle \cdot \langle \overrightarrow{verb}^{(r)} \overrightarrow{obj} \rangle \cdot \overrightarrow{verb}^{(m)}$	$\overrightarrow{verb}^{(m)}$

Table 1: Consequences of separability in various grammatical structures. Superscripts (l) , (m) and (r) refer to left-hand, middle, and right-hand component of a separable tensor

Euclidean distance instead of cosine would slightly improve things, since now we would be at least able to also detect differences in the magnitude between the two middle components. Unfortunately, this metric has been proved not very appropriate for distributional models of meaning, since in the vastness of a highly dimensional space every point ends up to be almost equidistant from all the others. As a result, most implementations of distributional models prefer the more relaxed metric of cosine distance which is length-invariant. Table 1 presents the consequences of separability in a number of grammatical constructs.

6 Concrete models for verb tensors

Whereas for the vector representations of atomic words of language one can use the much-experimented-with methods of distributional semantics, the tensor representations of relational words is a by-product of the categorical framework whose concrete instantiations are still being investigated. A number of concrete implementations have been proposed so far, e.g. see [7, 13, 9, 12]. These constructions vary from corpus-based methods to machine learning techniques. One problem that researchers have had to address is that tensors of order higher than 2 are difficult to create and manipulate. A transitive verb, for example, is represented by a cuboid living in $W^{\otimes 3}$; if the cardinality of our basic vector space is 1000 (and assuming a standard floating-point representation of 8 bytes per real number), the space required for just a single verb becomes 8 gigabytes. A workaround to this issue is to initially create the verb as a matrix, and then expand it to a tensor of higher order by applying Frobenius Δ operators—that is, leaving one or more dimensions of the resulting tensor empty (filled with zeros).

A simple and intuitive way to create a matrix for a relational word is to structurally mix the arguments with which this word appears in the training corpus [7]. For a transitive verb, this would be given us:

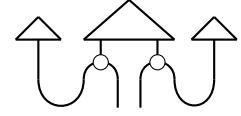
$$\overrightarrow{verb} = \sum_i (\overrightarrow{subject_i} \otimes \overrightarrow{object_i}) \quad (13)$$

where $\overrightarrow{subject_i}$ and $\overrightarrow{object_i}$ are the vectors of the subject/object pair for the i th occurrence of the verb in the corpus. The above technique seems to naturally result in an entangled matrix, assuming that the family of subject vectors exhibit a sufficient degree of linear independence, and the same is true for the family of object vectors. Compare this to a straightforward variation which naturally results in a separable matrix, as follows:

$$\overrightarrow{verb} = \left(\sum_i \overrightarrow{subject_i} \right) \otimes \left(\sum_i \overrightarrow{object_i} \right) \quad (14)$$

In what follows, we present a number of methods to embed the above verbs from tensors of order 2 to tensors of higher order, as required by the categorical framework.

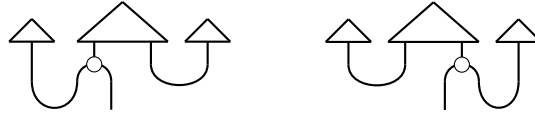
Relational In [7], the order of a sentence space depends on the arity of the verb of the sentence; for a transitive sentence the result will be a matrix, for an intransitive one it will be a vector, and so on. For the transitive case, the authors expand the original verb matrix to a tensor of order 4 (since now $S = N \otimes N$, the original $N \otimes S \otimes N$ space becomes $N^{\otimes 4}$) by copying both dimensions using Frobenius Δ operators in the fashion shown in the diagram (right). Linear-algebraically, the meaning of a transitive sentence is a matrix living in $W \otimes W$ obtained by the following equation:



$$\overrightarrow{subj\ verb\ obj} = (subj \otimes obj) \odot \overrightarrow{verb} \quad (15)$$

where the symbol \odot denotes element-wise multiplication.

Frobenius The above method has the limitation that sentences of different structures live in spaces of different tensor orders, so a direct comparison thereof is not possible. As a solution, Kartsaklis et al. [13] propose the copying of only one dimension of the original matrix, which leads to the following two possibilities:



The result is now a vector, computed in the following way, respectively for each case:

$$\text{Copy-subject:} \quad \overrightarrow{subj\ verb\ obj} = \overrightarrow{subj} \odot (\overrightarrow{verb} \times \overrightarrow{obj}) \quad (16)$$

$$\text{Copy-object:} \quad \overrightarrow{subj\ verb\ obj} = \overrightarrow{obj} \odot (\overrightarrow{verb}^T \times \overrightarrow{subj}) \quad (17)$$

Each one of the vectors obtained from Eqs. 16 and 17 above addresses a partial interaction of the verb with each argument. It is reasonable then to further combine them in order to get a more complete representation of the verb meaning (and hence the sentence meaning). We therefore define three more models, in which this combination is achieved through vector addition (**Frobenius additive**), element-wise multiplication (**Frobenius multiplicative**), and tensor product (**Frobenius tensored**) of the above.

We conclude this section with two important comments. First, although the use of a matrix for representing a transitive verb might originally seem as a violation of the functorial relation with a pre-group grammar, this is not the case in practice; the functorial relation is restored through the use of the Frobenius operators, which produce a tensor of the correct order, as required by the grammatical type. Furthermore, this notion of “inflation” has the additional advantage that can also work from a reversed perspective: a matrix created by Eq. 13 can be seen as an order-3 tensor originally in $N \otimes S \otimes N$ where the S dimension has been discarded by a ζ Frobenius map. Using this approach, Sadrzadeh and colleagues provide intuitive analyses for wh-movement phenomena and discuss compositional treatments of constructions containing relative pronouns [20, 21].

Finally, we would like to stress out the fact that, despite of the actual level of entanglement in our original verb matrix created by Eq. 13, the use of Frobenius operators as described above equips the inflated verb tensors with an extra level of entanglement in any case. As we will see in Sect. 8 when discussing the results of the experimental work, this detail will be proven very important in practice.

7 Experiments

7.1 Creating a semantic space

Our basic vector space is trained from the ukWaC corpus [5], originally using as a basis the 2,000 content words with the highest frequency (but excluding a list of stop words as well as the 50 most frequent content words since they exhibit low information content). As context we considered a 5-word window from either side of the target word, whereas for our weighting scheme we used local mutual information (i.e. point-wise mutual information multiplied by raw counts). The vector space was normalized and projected onto a 300-dimensional space using singular value decomposition (SVD). These choices are based on our best results in a number of previous experiments [12, 11].

7.2 Detecting sentence similarity

In this section we test the various compositional models of Sect. 6 in two similarity tasks involving pairs of transitive sentences; for each pair, we construct composite vectors for the two sentences, and then we measure their semantic similarity using cosine distance and Euclidean distance. We then evaluate the correlation of each model’s performance with human judgements, using Spearman’s ρ . In the first task [7], the sentences to be compared are constructed using the same subject and object and semantically correlated verbs, such as ‘spell’ and ‘write’; for example, ‘pupils write letters’ is compared with ‘pupils spell letters’. The dataset consists of 200 sentence pairs.

We are especially interested in measuring the level of entanglement in our verb matrices as these are created by Eq. 13. In order to achieve that, we compute the *rank-1 approximation* of all verbs in our dataset. Given a verb matrix \overline{verb} , we first compute its SVD so that $\overline{verb} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$, and then we approximate this matrix by using only the highest eigenvalue and the related left and right singular vectors, so that $\overline{verb}_{R1} = \mathbf{U}_1\mathbf{\Sigma}_1\mathbf{V}_1^T$. We compare the composite vectors created by the original matrix (Eq. 13), their rank-1 approximations, and the results of the separable model of Eq. 14. We also use a number of baselines: in the ‘verbs-only’ model, we compare only the verbs (without composing them with the context), while in the additive and multiplicative models we construct the sentence vectors by simply adding and element-wise multiplying the distributional vectors of their words.

The results (Table 2) revealed a striking similarity in the performances of the entangled and separable versions. Using cosine distance, all three models (relational, rank-1 approximation, separable model) have essentially the same behaviour; with Euclidean distance, the relational model performs again the

Model	ρ with cos	ρ with Eucl.
Verbs only	0.329	0.138
Additive	0.234	0.142
Multiplicative	0.095	0.024
Relational	0.400	0.149
Rank-1 approx. of relational	0.402	0.149
Separable	0.401	0.090
Copy-subject	0.379	0.115
Copy-object	0.381	0.094
Frobenius additive	0.405	0.125
Frobenius multiplicative	0.338	0.034
Frobenius tensored	0.415	0.010
Human agreement	0.60	

Table 2: Results for the first dataset (same subjects/objects, semantically related verbs)

same as its rank-1 approximation, while this time the separable model is lower.

The inevitable conclusion that Eq. 13 actually produces a separable matrix was further confirmed by an additional experiment: we calculated the average cosine similarity of the original matrices with their rank-1 approximations, a computation that revealed a similarity as high as 0.99. Since this result might obviously depend on the form of the noun vectors used for creating the verb matrix, this last experiment was repeated with a number of variations of our basic vector space, getting in every case similarities between verb matrices and their rank-1 approximations higher than 0.97. The observed behaviour can only be explained with the presence of a very high level of linear dependence between the subject vectors and between the object vectors. If every subject vector can be expressed as a linear combination of a small number of other vectors (and the same is true for the family of object vectors), then this would drastically reduce the entanglement of the matrix to the level that it is in effect separable.

Our observations are also confirmed in the second sentence similarity task. Here, we use a variation of one of the datasets in [12], consisting of 108 pairs of transitive sentences. The difference with our first task is that now the sentences of a pair are unrelated in a word level, i.e. subjects, objects, and verbs are all different. The results for this second experiment are presented in Table 3.

Model	ρ with cos	ρ with Eucl.
Verbs only	0.449	0.392
Additive	0.581	0.542
Multiplicative	0.287	0.109
Relational	0.334	0.173
Rank-1 approx. of relational	0.333	0.175
Separable	0.332	0.105
Copy-subject	0.427	0.096
Copy-object	0.198	0.144
Frobenius additive	0.428	0.117
Frobenius multiplicative	0.302	0.041
Frobenius tensored	0.332	0.042
Human agreement	0.66	

Table 3: Results for the second dataset (different subjects, objects and verbs)

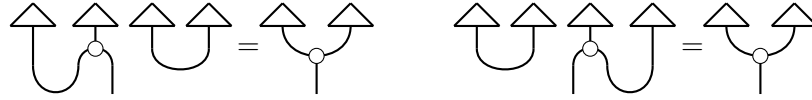
As a general observation about the performance of the various models in the two tasks, we note the high scores achieved by the Frobenius models when one uses the preferred method of measurement, that of cosine similarity. Especially the **Frobenius additive** has been proved to perform better than the Relational model, having the additional advantage that it allows comparison between sentences of different structures (since every sentence vector lives in W).

8 Discussion

The experiments of Sect. 7 revealed an unwelcome property of a method our colleagues and we have used in the past for creating verb tensors in the context of compositional models [7, 13, 12]. The fact that the verb matrix is in effect separable introduces a number of simplifications in the models presented in Sect. 6. More specifically, the Relational model of [7] is reduced to the following:

$$\begin{array}{c} \text{Diagram: A sequence of four upward-pointing triangles. The first and second triangles are connected by a curved line. The second and third triangles are connected by a curved line. The third and fourth triangles are connected by a curved line. The first triangle has a vertical line extending downwards from its base. The second triangle has a vertical line extending downwards from its base. The third triangle has a vertical line extending downwards from its base. The fourth triangle has a vertical line extending downwards from its base. } \end{array} = \begin{array}{c} \text{Diagram: A sequence of two upward-pointing triangles. The first triangle has a vertical line extending downwards from its base. The second triangle has a vertical line extending downwards from its base. } \end{array} \quad \overrightarrow{subj} \overrightarrow{verb} \overrightarrow{obj} = (\overrightarrow{subj} \odot \overrightarrow{verb}^{(l)}) \otimes (\overrightarrow{verb}^{(r)} \odot \overrightarrow{obj})$$

Furthermore, the Frobenius models of [13] get these forms:



which means, for example, that the actual equation behind the successful Frobenius additive model is

$$\overrightarrow{subj\ verb\ obj} = (\overrightarrow{subj} \odot \overrightarrow{verb}^{(l)}) + (\overrightarrow{verb}^{(r)} \odot \overrightarrow{obj}) \quad (18)$$

Despite the simplifications presented above, note that none of these models degenerates to the level of producing “constant” vectors or matrices, as argued for in Sect. 5. Indeed, especially in the first task (Table 2) the Frobenius models present top performance, and the relational models follow closely. The reason behind this lies in the use of Frobenius Δ operators for copying the original dimensions of the verb matrix, a computation that equipped the fragmented system with flow, although not in the originally intended sense. The compositional structure is still fragmented into two parts, but at least now the copied dimensions provide a means to deliver the results of the two individual computations that take place, one for the left-hand part of the sentence and one for the right-hand part. Let us see what happens when we use cosine distance in order to compare the matrices of two transitive sentences created with the **Relational** model (the separable version of a verb matrix \overrightarrow{verb} is denoted by $\overrightarrow{verb}^{(l)} \otimes \overrightarrow{verb}^{(r)}$):

$$\begin{aligned} \langle \overrightarrow{subj_1\ verb_1\ obj_1} | \overrightarrow{subj_2\ verb_2\ obj_2} \rangle &= \\ \langle (\overrightarrow{subj_1} \odot \overrightarrow{verb_1}^{(l)}) \otimes (\overrightarrow{verb_1}^{(r)} \odot \overrightarrow{obj_1}) | (\overrightarrow{subj_2} \odot \overrightarrow{verb_2}^{(l)}) \otimes (\overrightarrow{verb_2}^{(r)} \odot \overrightarrow{obj_2}) \rangle &= \\ \langle \overrightarrow{subj_1} \odot \overrightarrow{verb_1}^{(l)} | \overrightarrow{subj_2} \odot \overrightarrow{verb_2}^{(l)} \rangle \langle \overrightarrow{verb_1}^{(r)} \odot \overrightarrow{obj_1} | \overrightarrow{verb_2}^{(r)} \odot \overrightarrow{obj_2} \rangle \end{aligned}$$

As also computed and pointed out in [6], the two sentences are broken up to a left-hand part and a right-hand part, and two distinct comparisons take place. As long as we compare sentences of the same structure, as we did here, this method is viable. On the other hand, the **Frobenius** models and their simplifications such as the one in (18) do not have this restriction; in principle, all sentences are represented by vectors living in the same space, so any kind of comparison is possible. In case, however, we do compare sentences of the same structure, these models have the additional advantage that also allow comparisons between *different* sentence parts; this can be seen in the dot product of two sentences created by Eq. 18, which gets the following form:

$$\begin{aligned} \langle \overrightarrow{subj_1} \odot \overrightarrow{verb_1}^{(l)} | \overrightarrow{subj_2} \odot \overrightarrow{verb_2}^{(l)} \rangle + \langle \overrightarrow{subj_1} \odot \overrightarrow{verb_1}^{(l)} | \overrightarrow{verb_2}^{(r)} \odot \overrightarrow{obj_2} \rangle + \\ \langle \overrightarrow{verb_1}^{(r)} \odot \overrightarrow{obj_1} | \overrightarrow{subj_2} \odot \overrightarrow{verb_2}^{(l)} \rangle + \langle \overrightarrow{verb_1}^{(r)} \odot \overrightarrow{obj_1} | \overrightarrow{verb_2}^{(r)} \odot \overrightarrow{obj_2} \rangle \end{aligned}$$

9 Using linear regression for entanglement

Corpus-based methods for creating tensors of relational words, such as the models presented so far in this paper, are intuitive and easy to implement. As our experimental work shows, however, this convenience comes with a price. In practice, one would expect that more robust machine learning techniques would produce more reliable tensor representations for composition.

In this section we apply linear regression (following [2]) in order to train verb matrices for a variation of our second experiment, in which we compare elementary verb phrases of the form *verb-object* [18] (so the subjects are dropped). In order to create a matrix for, say, the verb ‘play’, we first collect all instances of the verb occurring with some object in the training corpus, and then we create non-compositional holistic vectors for these elementary verb phrases following exactly the same methodology as if they were words. We now have a dataset with instances of the form $\langle \overrightarrow{obj_i}, \overrightarrow{play\ obj_i} \rangle$ (e.g. the vector of ‘flute’

Model	ρ with cos	ρ with Eucl.
Verbs only	0.331	0.267
Holistic verb-pharse vectors	0.403	0.214
Additive	0.379	0.385
Multiplicative	0.301	0.229
Linear regression	0.349	0.144
Rank-1 approximation of LR matrices	0.119	0.082
Human agreement	0.55	

Table 4: Results for the verb-pharse similarity task

paired with the holistic vector of ‘play flute’, and so on), that can be used to train a linear regression model in order to produce an appropriate matrix for verb ‘play’. The premise of a model like this is that the multiplication of the verb matrix with the vector of a new object will produce a result that approximates the distributional behaviour of all these elementary two-word exemplars used in training. For a given verb, this is achieved by using *gradient descent* in order to minimize the total error between the observed vectors and the vectors predicted by the model, expressed by the following quantity:

$$\frac{1}{2m} \left(\sum_i \overrightarrow{verb} \times \overrightarrow{object_i} - \overrightarrow{verb object_i} \right)^2 \quad (19)$$

where m is the number of training instances. The average cosine similarity between the matrices we got from this method and their rank-1 approximation was only 0.48, showing that in general the level of entanglement produced by this method is reasonably high. This is also confirmed by the results in Table 4; the rank-1 approximation model presents the worst performance, since, as you might recall from the discussion in Sect. 5, separability here means that every verb-object composition is reduced to the left component of the verb matrix, completely ignoring the interaction with the object.

10 Conclusion

The current study takes a closer look to an aspect of tensor-based compositional models of meaning that so far had escaped the attention of researchers. The discovery that a number of concrete instantiations of the categorical framework proposed in [4] produce relational tensors that are in effect separable stresses the necessity of similar tests for any linear model that follows the same philosophy. Another contribution of this work was that it showed this is not necessarily a bad thing. The involvement of Frobenius operators in the creation of verb tensors equips the compositional structure with the necessary flow, so that a comparison between two sentence vectors can be still carried out between individual parts of each sentence. Therefore, approaches such as the Frobenius additive model proposed in this paper can be still considered as viable and “easy” alternatives to more robust machine learning techniques, such as the gradient optimization technique discussed in Sect. 9.

References

- [1] Samson Abramsky & Bob Coecke (2004): *A Categorical Semantics of Quantum Protocols*. In: *19th Annual IEEE Symposium on Logic in Computer Science*, pp. 415–425.
- [2] M. Baroni & R. Zamparelli (2010): *Nouns are Vectors, Adjectives are Matrices*. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- [3] B. Coecke, E. Grefenstette & M. Sadrzadeh (2013): *Lambek vs. Lambek: Functorial Vector Space Semantics and String Diagrams for Lambek Calculus*. *Annals of Pure and Applied Logic*. Available at <http://arxiv.org/abs/1302.0393>.

- [4] B. Coecke, M. Sadrzadeh & S. Clark (2010): *Mathematical Foundations for Distributed Compositional Model of Meaning*. *Lambek Festschrift*. *Linguistic Analysis* 36, pp. 345–384.
- [5] Adriano Ferraresi, Eros Zanchetta, Marco Baroni & Silvia Bernardini (2008): *Introducing and evaluating ukWaC, a very large web-derived corpus of English*. In: *Proceedings of the 4th Web as Corpus Workshop (WAC-4) Can we beat Google*, pp. 47–54.
- [6] E. Grefenstette & M. Sadrzadeh: *Concrete Models and Empirical Evaluations for the Categorical Compositional Distributional Model of Meaning*. *Computational Linguistics*. To appear.
- [7] E. Grefenstette & M. Sadrzadeh (2011): *Experimental Support for a Categorical Compositional Distributional Model of Meaning*. In: *Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- [8] E. Grefenstette & M. Sadrzadeh (2011): *Experimenting with Transitive Verbs in a DisCoCat*. In: *Proceedings of the 2011 EMNLP Workshop on Geometric Models of Natural Language Semantics*.
- [9] Edward Grefenstette, Georgiana Dinu, Yao-Zhong Zhang, Mehrnoosh Sadrzadeh & Marco Baroni (2013): *Multi-Step Regression Learning for Compositional Distributional Semantics*. In: *Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)*. Available at <http://arxiv.org/abs/1301.6939>.
- [10] D. Kartsaklis, M. Sadrzadeh, S. Pulman & B. Coecke (2014): *Reasoning about Meaning in Natural Language with Compact Closed Categories and Frobenius Algebras*. In J. Chubb, A. Eskandarian & V. Harizanov, editors: *Logic and Algebraic Structures in Quantum Computing and Information*, Association for Symbolic Logic Lecture Notes in Logic, Cambridge University Press.
- [11] Dimitri Kartsaklis, Nal Kalchbrenner & Mehrnoosh Sadrzadeh (2014): *Resolving Lexical Ambiguity in Tensor Regression Models of Meaning*. In: *Proceedings of the 52th Annual Meeting of the Association for Computational Linguistics (ACL): Short Papers*, Baltimore, USA. To appear.
- [12] Dimitri Kartsaklis & Mehrnoosh Sadrzadeh (2013): *Prior Disambiguation of Word Tensors for Constructing Sentence Vectors*. In: *Proceedings of the 2013 Conference on Empirical Methods in Natural Language Processing (EMNL)*, Seattle, USA.
- [13] Dimitri Kartsaklis, Mehrnoosh Sadrzadeh & Stephen Pulman (2012): *A Unified Sentence Space for Categorical Distributional-Compositional Semantics: Theory and Experiments*. In: *Proceedings of 24th International Conference on Computational Linguistics (COLING 2012): Posters*, The COLING 2012 Organizing Committee, Mumbai, India, pp. 549–558.
- [14] G Maxwell Kelly (1972): *Many-Variable Functorial Calculus (I)*. In G.M. Kelly, M. Laplaza, G. Lewis & S. MacLane, editors: *Coherence in categories*, Springer, pp. 66–105.
- [15] J. Lambek (2008): *From Word to Sentence*. Polimetrica, Milan.
- [16] T. Landauer & S. Dumais (1997): *A Solution to Plato’s Problem: The Latent Semantic Analysis Theory of Acquisition, Induction, and Representation of Knowledge*. *Psychological Review*.
- [17] C.D. Manning, P. Raghavan & H. Schütze (2008): *Introduction to Information Retrieval*. Cambridge University Press.
- [18] Jeff Mitchell & Mirella Lapata (2010): *Composition in Distributional Models of Semantics*. *Cognitive Science* 34(8), pp. 1388–1439.
- [19] A. Preller & M. Sadrzadeh (2010): *Bell States and Negative Sentences in the Distributed Model of Meaning*. In P. Selinger B. Coecke, P. Panangaden, editor: *Electronic Notes in Theoretical Computer Science, Proceedings of the 6th QPL Workshop on Quantum Physics and Logic*, University of Oxford.
- [20] Mehrnoosh Sadrzadeh, Stephen Clark & Bob Coecke (2013): *The Frobenius Anatomy of Word Meanings I: Subject and Object Relative Pronouns*. *Journal of Logic and Computation* 23(6), pp. 1293–1317.
- [21] Mehrnoosh Sadrzadeh, Stephen Clark & Bob Coecke (2014): *The Frobenius Anatomy of Word Meanings II: Possessive Relative Pronouns*. *Journal of Logic and Computation*. To appear.
- [22] H. Schütze (1998): *Automatic Word Sense Discrimination*. *Computational Linguistics* 24, pp. 97–123.
- [23] Peter Selinger (2011): *A Survey of Graphical Languages for Monoidal Categories*. In Bob Coecke, editor: *New structures for physics*, Springer, pp. 289–355.

The ZX-calculus is incomplete for quantum mechanics

Christian Schröder de Witt
Scherenbergstr. 22, 10439 Berlin
caschroeder@outlook.com

Vladimir Zamdzhiev
Department of Computer Science
University of Oxford
Oxford, United Kingdom
vladimir.zamdzhiev@cs.ox.ac.uk

We prove that the ZX-calculus is incomplete for quantum mechanics. We suggest the addition of a new 'color-swap' rule, of which currently no analytical formulation is known and which we suspect may be necessary, but not sufficient to make the ZX-calculus complete.

1 Introduction

Coecke and Abramsky pioneered the field of categorical quantum mechanics in [1]. Later, from this study, an intuitive graphical calculus (dubbed the ZX-calculus) was developed by Coecke and Duncan [4][3], which can be used as an alternative to Dirac notation in a wide number of applications [5][11][6][8].

Backens recently proved that the ZX-calculus is complete for an important subset of quantum mechanics, namely stabilizer quantum mechanics, i.e. that for stabilizer quantum mechanics, any equation that can be shown to hold in the Dirac formalism can also be shown to hold within the ZX-calculus[2]. For her proof, she relied on operations on a special class of quantum states, namely graph states. This paper addresses the question of whether the ZX-calculus is complete for the whole of quantum mechanics, and the answer is found to be negative.

1.1 Syntax and Semantics of the ZX-calculus

The syntax and semantics of the ZX-calculus are presented below. The semantics are given in Hilbert space. We begin with atomic diagrams. The inputs to the diagrams are located at the bottom and the outputs are located at the top.

$$\left[\begin{array}{c} | \\ | \end{array} \right] = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = I \quad \left[\begin{array}{c} \diagup \quad \diagdown \\ \diagdown \quad \diagup \end{array} \right] = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} = \sigma$$

$$\begin{aligned}
\left[\begin{array}{c} \text{cup} \end{array} \right] &= \langle 00| + \langle 11| & \left[\begin{array}{c} \text{cap} \end{array} \right] &= |00\rangle + |11\rangle & \left[\begin{array}{c} \text{H} \end{array} \right] &= \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = H \\
\left[\begin{array}{c} \text{green dot} \end{array} \right] &= \begin{cases} |0^m\rangle & \mapsto |0^n\rangle \\ |1^m\rangle & \mapsto e^{i\alpha} |1^n\rangle \\ \text{rest} & \mapsto 0 \end{cases} & \left[\begin{array}{c} \text{red dot} \end{array} \right] &= \begin{cases} |+\rangle^m & \mapsto |+\rangle^n \\ |-\rangle^m & \mapsto e^{i\alpha} |-\rangle^n \\ \text{rest} & \mapsto 0 \end{cases}
\end{aligned}$$

where in the last two diagrams m is the number of inputs and n is the number of outputs. The labels of the red and green dots form the circle group under addition. So, admissible values are $\alpha \in [0, 2\pi)$. We also make the convention that we will not write a label for the points when $\alpha = 0$.

We can create compound diagrams from smaller diagrams in two ways - either placing two diagrams next to each horizontally, or plugging the outputs of one diagram to the inputs of another. If

$$\left[\begin{array}{c} \Psi_1 \end{array} \right] = D_1 \quad \text{and} \quad \left[\begin{array}{c} \Psi_2 \end{array} \right] = D_2$$

then

$$\left[\begin{array}{cc} \Psi_1 & \Psi_2 \end{array} \right] = D_1 \otimes D_2$$

and

$$\left[\begin{array}{c} \Psi_1 \\ \Psi_2 \end{array} \right] = D_1 \circ D_2$$

In the latter diagram, the number of outputs of Ψ_2 has to be the same as the number of inputs of Ψ_1 .

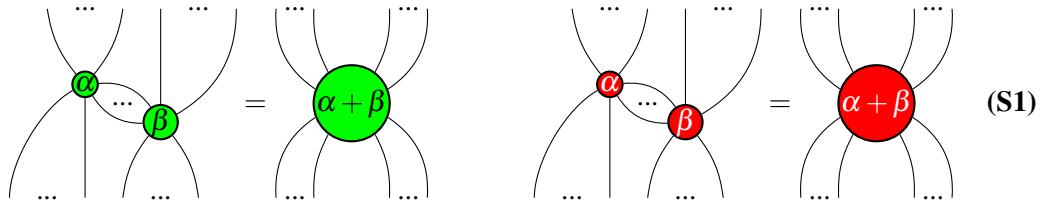
By following the above rules we can represent any pure state map on qubits as a diagram in the ZX-calculus [4].

1.2 ZX Equational Rules

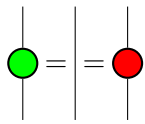
The rules of the ZX-calculus are given by:

"Only the topology matters"

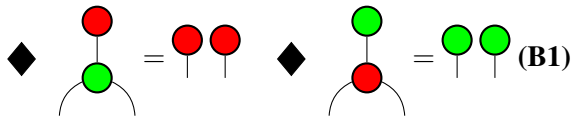
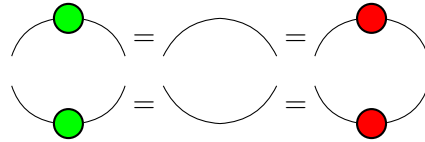
(T)



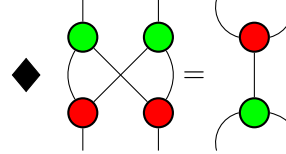
(S1)



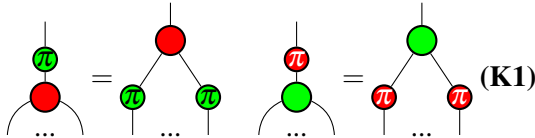
(S2)



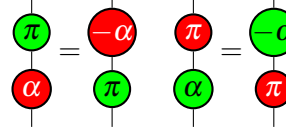
(B1)



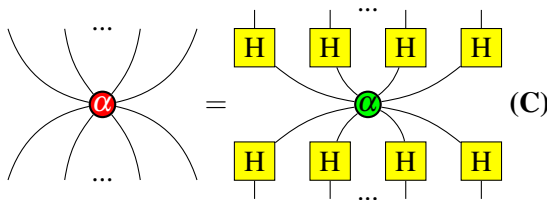
(B2)



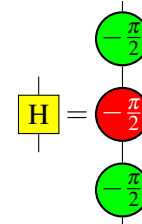
(K1)



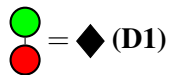
(K2)



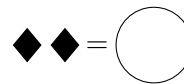
(C)



(EU)



(D1)



(D2)

1.3 Completeness, Soundness and Universality

The original rules of the ZX-calculus, as put forward by Coecke and Duncan [4] did not contain the Euler decomposition of the Hadamard gate, **EU**. Subsequently, Duncan and Perdrix proved [7] that the rule **EU** was not derivable from within the original ZX-calculus. The original ZX-calculus was therefore *incomplete*. Informally, incompleteness signifies that there are equations that can be proven to hold in Dirac-von Neumann notation that cannot be proven in the ZX-calculus. This would reduce the power of the graphical calculus and possibly limit its applications in automated reasoning.

Backens showed in [2] that the current ZX-calculus, which is simply the original ZX-calculus extended by **EU**, is complete for an important fragment of quantum mechanics, namely *stabilizer quantum mechanics* (SQM). Her proof relies on the fact that each SQM state is, under local Clifford operations [9], equivalent to a special entangled state, namely a *graph state* [10]. This allows one to abstract away from matrix representations and instead decide equivalence between different SQM states by performing *local complementations*, a class of graph manipulations, between graph states. In this way, Backens showed that SQM states may be represented by so-called rGS-LC diagrams, which are only equivalent iff they are graphically identical. In this way, equivalence can be decided in the ZX-calculus.

However, ideally, one would wish the ZX-calculus to be as physically expressive as the complete Dirac-von Neumann formalism. To this goal, three important properties of the calculus need to be established: universality, soundness and completeness.

The ZX-calculus is *sound* [4]. That is, if $ZX \vdash D_1 = D_2$ then $\llbracket D_1 \rrbracket = e^{i\phi} \llbracket D_2 \rrbracket$. In other words, if two diagrams are equal under the axioms of the ZX-calculus, then their Hilbert space interpretations are equal up to a global phase.

Secondly, the ZX-calculus is *universal*, meaning that it can express any quantum state and gate. This is easily proven by showing that the ZX-calculus can express any of the set of universal quantum gates [4].

Finally, *completeness* is the converse of soundness. That is, if $\llbracket D_1 \rrbracket = \llbracket D_2 \rrbracket$ then $ZX \vdash D_1 = D_2$. In the next section, we will show that the ZX-calculus does not have this property.

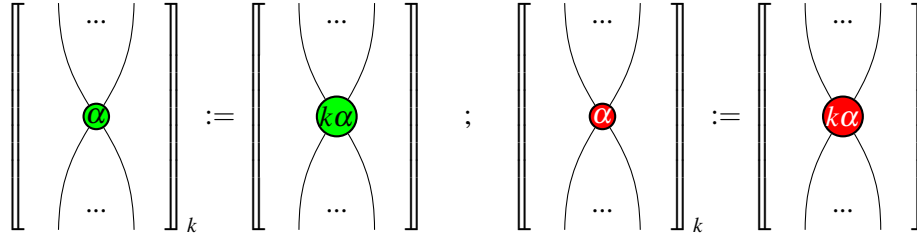
2 Incompleteness

Before we present the proof, let us recall a standard result from quantum mechanics, namely the *Euler decomposition* of single-qubit gates [12]. By this result, any single-qubit unitary gate can be expressed (up to a global phase) through just three consecutive rotations in appropriate bases. For the ZX-calculus, this means that there always exist real angles $\alpha_i, \beta_i, \gamma_i, \phi_i$ such that for any ZX diagram D with one input and one output, we have:

$$\left\| \begin{array}{c} | \\ \boxed{D} \\ | \end{array} \right\| = e^{i\phi_1} \left\| \begin{array}{c} \alpha_1 \\ \beta_1 \\ \gamma_1 \end{array} \right\| = e^{i\phi_2} \left\| \begin{array}{c} \alpha_2 \\ \beta_2 \\ \gamma_2 \end{array} \right\|$$

We prove the incompleteness of the ZX-calculus by using a similar argument to that of Duncan and Perdrix in [7], where they show that the Euler decomposition of the Hadamard gate is not derivable within the ZX-calculus.

In particular, we can define alternative models for the ZX-calculus by setting:

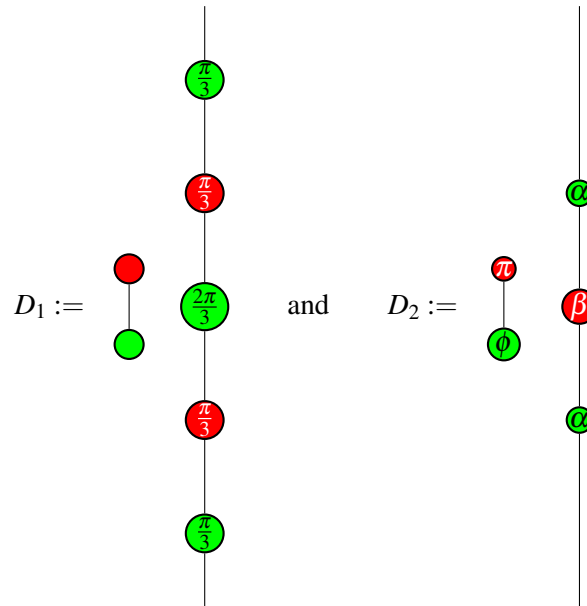


$$[\![\cdot]\!]_k := [\![\cdot]\!] , \text{ otherwise}$$

where $k \in \mathbb{Z}$ and $[\![\cdot]\!]$ is the standard interpretation functor for ZX diagrams in Hilbert space. In other words, we multiply all angles in our diagrams by an integer k and consider the corresponding interpretation.

These models are sound when $k = 4p + 1$ for $p \in \mathbb{Z}$. This can be easily verified by checking that each of the equational rules remains valid under this interpretation.

Consider the following two ZX diagrams:



where

$$\begin{aligned}\alpha &:= -\arccos\left(\frac{5}{2\sqrt{13}}\right) \approx 0.2561\pi \\ \beta &:= -2\arcsin\left(\frac{\sqrt{3}}{4}\right) \approx -0.2851\pi \\ \phi &:= \arcsin\left(\frac{\sqrt{3}}{4}\right) - \alpha \approx 0.3987\pi\end{aligned}$$

Then, we have

$$\llbracket D_1 \rrbracket = \llbracket D_2 \rrbracket$$

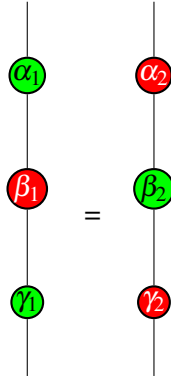
The two scalar factors are introduced so that the equality is exact, otherwise it would be true up to the global phase $e^{i\phi}$.

Let us assume for contradiction that D_1 and D_2 are equal under the axioms of the ZX-calculus, i.e. $ZX \vdash D_1 = D_2$. Since $\llbracket \cdot \rrbracket_{-3}$ provides a sound model of the calculus, it must also be the case that $\llbracket D_1 \rrbracket_{-3} = \lambda \llbracket D_2 \rrbracket_{-3}$, for some $\lambda \in \mathbb{C}$.

However, it is easy to check that this is not true ($\llbracket D_1 \rrbracket_{-3}$ is equal to a scalar times the identity, whereas $\llbracket D_2 \rrbracket_{-3}$ isn't). Therefore the two diagrams D_1 and D_2 are not equal under the axioms of the ZX-calculus, even though they have equal Hilbert space interpretations. This means the ZX-calculus is incomplete.

3 Conclusion and Future Work

The primary contribution of this work is showing that the ZX-calculus is incomplete for quantum mechanics. A natural question to ask is what additional rules can be added to the calculus in order to increase its proving power. The proof that we have used doesn't use any special properties of the presented diagrams – it will be straightforward to apply the same proof to another pair of single-qubit unitary gates where one of them is the Euler decomposition of the other. To eliminate this class of counter-examples, we believe that a "color-swap" rule of the form:

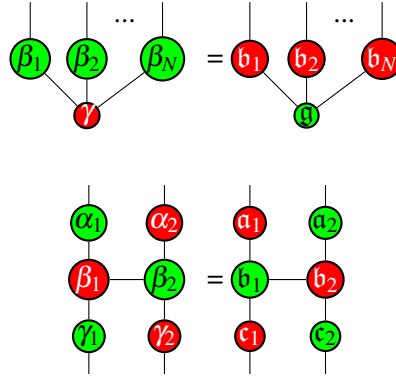


might be needed. This would require identifying functions f_1, f_2, f_3 , s.t. the above rule is valid and

$$\begin{aligned}\alpha_1 &= f_1(\alpha_2, \beta_2, \gamma_2) \\ \beta_1 &= f_2(\alpha_2, \beta_2, \gamma_2) \\ \gamma_1 &= f_3(\alpha_2, \beta_2, \gamma_2)\end{aligned}$$

In other words, an analytic solution for converting from ZXZ to XZX Euler decompositions of single-qubit unitary gates is required.

Whether the addition of such a 'color-swap' would be sufficient to render the ZX-calculus complete is currently unknown. Some simple candidates for further possible non-derivable equalities are presented here:



We suggest to conduct numerical investigations into the question of whether such non-derivable equalities between complex ZX-calculus diagrams exist.

Acknowledgements

We would like to thank Aleks Kissinger, Miriam Backens and Bob Coecke for constructive discussions. One of the authors also wishes to gratefully acknowledge support from the EPSRC and the Scatcherd European Scholarship.

References

- [1] S. Abramsky & B. Coecke (2004): *A categorical semantics of quantum protocols*. In: *Logic in Computer Science, 2004. Proceedings of the 19th Annual IEEE Symposium on*, pp. 415–425, doi:10.1109/LICS.2004.1319636.
- [2] Miriam Backens (2012): *The ZX-calculus is complete for stabilizer quantum mechanics*. In: *Proceedings 9th International Workshop on Quantum Physics and Logic*, Brussels, Belgium October 10-12, 2012. Available at <http://arxiv.org/abs/1307.7025>.

- [3] Bob Coecke & Ross Duncan (2008): *Interacting Quantum Observables*. In Luca Aceto, Ivan Damgård, LeslieAnn Goldberg, MagnúsM. Halldórsson, Anna Ingólfssdóttir & Igor Walukiewicz, editors: *Automata, Languages and Programming, Lecture Notes in Computer Science* 5126, Springer Berlin Heidelberg, pp. 298–310, doi:10.1007/978-3-540-70583-3_25. Available at http://dx.doi.org/10.1007/978-3-540-70583-3_25.
- [4] Bob Coecke & Ross Duncan (2011): *Interacting quantum observables: categorical algebra and diagrammatics*. *New Journal of Physics* 13(4), p. 043016. Available at <http://stacks.iop.org/1367-2630/13/i=4/a=043016>.
- [5] Bob Coecke, Ross Duncan, Aleks Kissinger & Quanlong Wang (2012): *Strong Complementarity and Non-locality in Categorical Quantum Mechanics*. In: *Proceedings of the 2012 27th Annual IEEE/ACM Symposium on Logic in Computer Science, LICS '12*, IEEE Computer Society, Washington, DC, USA, pp. 245–254, doi:10.1109/LICS.2012.35. Available at <http://dx.doi.org/10.1109/LICS.2012.35>.
- [6] Ross Duncan & Maxime Lucas (2013): *Verifying the Steane code with Quantomatic*. In: *Proceedings 10th International Workshop on Quantum Physics and Logic*, Barcelona, Spain, July 17-19, 2013. Available at <http://arxiv.org/abs/1306.4532>.
- [7] Ross Duncan & Simon Perdrix (2009): *Graph States and the Necessity of Euler Decomposition*. In Klaus Ambos-Spies, Benedikt Löwe & Wolfgang Merkle, editors: *Mathematical Theory and Computational Practice, Lecture Notes in Computer Science* 5635, Springer Berlin Heidelberg, pp. 167–177, doi:10.1007/978-3-642-03073-4_18. Available at http://dx.doi.org/10.1007/978-3-642-03073-4_18.
- [8] Ross Duncan & Simon Perdrix (2010): *Rewriting Measurement-Based Quantum Computations with Generalised Flow*. In Samson Abramsky, Cyril Gavoille, Claude Kirchner, Friedhelm Meyer auf der Heide & PaulG. Spirakis, editors: *Automata, Languages and Programming, Lecture Notes in Computer Science* 6199, Springer Berlin Heidelberg, pp. 285–296, doi:10.1007/978-3-642-14162-1_24. Available at http://dx.doi.org/10.1007/978-3-642-14162-1_24.
- [9] D. Gottesman (1999): *The Heisenberg Representation of Quantum Computers*. *Group22: Proceedings of the XXII International Colloquium on Group Theoretical Methods in Physics* (eds. S. P. Corney, R. Delbourgo, and P. D. Jarvis), pp. 32–43.
- [10] M. Hein, J. Eisert & H. J. Briegel (2004): *Multiparty entanglement in graph states*. *Physical Review A* 69(6), p. 062311, doi:10.1103/PhysRevA.69.062311. Available at <http://link.aps.org/doi/10.1103/PhysRevA.69.062311>.
- [11] Anne Hillebrand (2011): *Superdense Coding with GHZ and Quantum Key Distribution with W in the ZX-calculus*. In Bart Jacobs, Peter Selinger & Bas Spitters, editors: *Proceedings 8th International Workshop on Quantum Physics and Logic*, Nijmegen, Netherlands, October 27-29, 2011, *Electronic Proceedings in Theoretical Computer Science* 95, Open Publishing Association, pp. 103–121, doi:10.4204/EPTCS.95.10.
- [12] Michael A Nielsen & Isaac L Chuang (2010): *Quantum computation and quantum information*. Cambridge university press.

BELIEF PROPAGATION IN MONOIDAL CATEGORIES (EXTENDED ABSTRACT)

JASON MORTON*

ABSTRACT. We discuss a categorical version of the celebrated belief propagation algorithm. This provides a way to prove that some algorithms which are known or suspected to be analogous, are actually identical when formulated generically. It also highlights the computational point of view in monoidal categories.

1. INTRODUCTION

We discuss a categorical version of the celebrated belief propagation algorithm [26]. This provides a way to prove that some algorithms which are known or suspected to be analogous, are actually identical when formulated generically. It also highlights the computational point of view in monoidal categories. The approach could also make possible software implementations that use a categorical formulation of algorithms to enable generic programming.

The setting for this algorithm, and for many computational questions concerning monoidal categories, is a *diagram* (Def. 2.2). By a diagram we mean an equivalence class of monoidal words over a finite tensor scheme, usually with certain additional properties. An interpretation [27] assigns values to each of the variables in the tensor scheme in a functorial way. Basic questions to ask of such a diagram, once interpreted in a particular category, include

- (1) compute a (possibly partial) contraction,
- (2) solve the word problem (are two diagrams equivalent, i.e. do they have the same interpretation) or compute a normal form for a diagram,
- (3) solve the implementability problem (construct a word equivalent to a target using a library of allowed morphisms), and
- (4) choose morphisms in a diagram to best approximate a more general diagram (possibly allowing the approximating diagram itself to vary).

Many practical questions are instances of one of these problems including computational challenges in probabilistic graphical models, quantum programming

*Department of Mathematics, Pennsylvania State University.

and logic [5, 12], the tensor network state approach to quantum condensed matter and quantum chemistry, parts of computational complexity theory including constraint satisfaction and counting constraint satisfaction problems, and many database operations.

For general monoidal categories, these problems are difficult: at least $\#P$ -hard (1), undecidable (2), undecidable (3), and NP -hard (4) respectively. The news is similarly bad for approximate versions of each. Nevertheless, given their practical significance, many tractable special cases, approximate algorithms, and heuristics exist to solve these problems in restricted cases.

Perhaps most prominent among such algorithms is the belief propagation algorithm [26], and its many extensions and analogs. From the categorical point of view, these extensions and analogs should just be the same abstract algorithm operating in different categories (e.g. probabilistic graphical models vs. sets and relations). These analogies have been drawn explicitly in many areas. For example, the connection between belief propagation and turbo coding theory was described in [20] and the connection to survey propagation for SAT problems is explored in [18].

We outline a general categorical form of the belief propagation algorithm for Problem 1 and look at how it specializes. We categorized another class of algorithms for Problem 1 in previous work [25]. We also touch briefly on Problems 2 and 3. Approaches to Problem 4 often require a solution to Problem 1 as a component.

By describing algorithms in terms of monoidal categories, the common structure of problems can be understood and computational knowledge can be more readily shared across disciplines. Creating general tools that work for any category with suitable properties, and can be specialized automatically once the monoidal category interface of a domain was specified, would be a significant advance. Given the rapidly expanding universe of applied problems given categorical interpretations, such an abstraction has the potential to be as useful as convex programming or numerical linear algebra. We only take a tiny step in this direction in the present work.

2. TENSOR SCHEMES AND WORD PROBLEMS

We assume the reader is familiar with monoidal categories [17]. We mainly consider the strict version here.

Definition 2.1. A (finite) *tensor scheme* (also called a *monoidal signature* or *monoidal alphabet*) \mathcal{T} is a finite set $\text{Ob}(\mathcal{T})$ of object variables (which must include a monoidal identity object I), a finite set $\text{Mor}(\mathcal{T})$ of morphism variables, and functions $\text{dom}, \text{cod} : \text{Mor}(\mathcal{T}) \rightarrow \text{Ob}(\mathcal{T})$.

The monoidal language $\mathcal{T}^{\otimes, \circ}$ generates the free monoidal category over \mathcal{T} . It consists of all valid morphism words that can be formed from $\text{Mor}(\mathcal{T})$, and identity morphisms. Constructively, $\mathcal{T}^{\otimes, \circ}$ is described as follows.

- (1) For all $A \in \text{Ob}(\mathcal{T})$, id_A is a word.
- (2) Each $f \in \text{Mor}(\mathcal{T})$ is a word.
- (3) Given words u, u' , $u \otimes u'$ is a word with domain $\text{dom}(u) \otimes \text{dom}(u')$ and codomain $\text{cod}(u) \otimes \text{cod}(u')$.
- (4) Given words w, w' with $\text{dom}(w') = \text{cod}(w)$, $w \circ w'$ is a word.

Given a word in a monoidal language, we can attach an *interpretation* by considering it as defining a morphism in a monoidal category. By the universal property, we can safely consider the word in the free monoidal category, which already imposes some equivalences such as $\text{id}_A \circ f = f$ and $(f \otimes g) \circ (f' \otimes g') = (f \circ f') \otimes (g \circ g')$. Two words are equivalent if they represent the same morphism in the free monoidal category.

Definition 2.2. An equivalence class of words in the free monoidal category over a tensor scheme is called a *diagram*.

Further notions of equivalence arise if we add additional relations. So far we have no normal form for words; for example $(f \otimes g) \circ (f' \otimes g')$ is not preferred over the equivalent $(f \circ f') \otimes (g \circ g')$.

Because there is a coherent graphical language for the free monoidal category over a tensor scheme, the word problem is not too difficult. Adding adjectives (special types of monoidal categories) and relations, or fixing values, so that the category is no longer free may make it easier or harder.

Proposition 2.3. *The word problem and implementability problem in a monoidal category over a finite tensor scheme are undecidable.*

To prove this proposition one just needs to embed a known undecidable problem in some non-free monoidal category; see [24] for an example of how to do this.

Proposition 2.3 puts us at a distinct disadvantage as compared to computational commutative algebra, where the word problem for polynomials is always at least computable with Gröbner bases [3]. We call a monoidal category *decidable* if its word problem is decidable.

The existence of coherent graphical languages for some types of monoidal categories means that the word problem can be reduced to graph isomorphism [7]. Hence the word problem for the free closed category and free compact closed category over a finite tensor scheme are in **LOGSPACE** and **P** [16] respectively. Normal forms for such graphs were explored in [8, 2], see also [21].

One could produce normal forms for words in X-categories (for some adjective X such as “traced” or “dagger”) with coherent graphical languages indirectly by this method, by transforming a word to normal form graph and then to a word again by some deterministic scheme.

It would be preferable for some purposes to have a confluent terminating rewriting system that attached a direction to the equalities of the X-category. For example $(f \circ f') \otimes (g \circ g') \mapsto (f \otimes g) \circ (f' \otimes g')$. Term rewriting and computing normal forms in monoidal categories is a field in its infancy; see [13, 23] for some of what is known.

A final method for the word problem is to apply a functor to a category which is complete with respect to the category of interest but might have an easier word problem. Finite dimensional vector spaces over a field of characteristic zero are complete for traced symmetric monoidal categories [10] and finite dimensional Hilbert spaces are complete for dagger compact closed categories [28]. Thus we can potentially certify *inequality* of words once we have bounded dimension, for example obtaining numerical methods for the word problem in categories by assigning random morphisms in the category of finite-dimensional vector spaces and linear transformations.

We assume our category is small, so the set of all morphisms from A to B is a *hom set* denoted $\text{Mor}(A, B)$. We can specialize the word problem to particular hom sets such as $\text{Mor}(I, I)$. In this case it is sometimes called a contraction problem (Problem 1). For example, in *semiringed* categories, the word problem for morphisms in $\text{Mor}(I, I)$ generalizes counting constraint satisfaction problems [25].

An important word problem for current purposes is determining equality of I -valued points, i.e. morphisms of type $\text{Mor}(I, A)$ for some object A . The reason for this is as follows.

We want to generalize algorithms such as belief propagation that work over the category of vector spaces and linear transformations (or a probabilistic version thereof). In the generalization, we can no longer assume that objects A are sets with points (such as probability distributions in the classical belief propagation algorithm). However, messages are still morphisms of type $\text{Mor}(I, A)$ for each object A and we still express the belief propagation equations in terms of equality of morphisms in each $\text{Mor}(I, A)$. Deciding if two vectors are equal up to numerical tolerance becomes deciding a word problem in $\text{Mor}(I, A)$. These messages must also be stored somehow.

Thus for our algorithm to run efficiently, we need the word problem for I -valued points to be efficiently decidable and their representation to be efficient. Thus, we will generally assume that I -valued points can be stored and compared efficiently. This can be made precise by introducing a size function.

In the classical setting for belief propagation there is a monoid homomorphism, $\text{size} : \text{Ob}(\mathcal{T})^\otimes \rightarrow \mathbb{N}$, from the free monoid generated by the objects of our tensor scheme to the natural numbers. This sends monoidal product to multiplication (or to the addition of log dimensions if we prefer to do something closer to counting wires). For example the dimension of a vector space fits this description. Then for each object A , words in $\text{Mor}(I, A)$ require $O(\text{size}(A))$ storage and the word problem for I -valued points of A is linear in the size of A .

2.1. Compact closed, spidered, and dungeon categories. A compact closed category is a monoidal category with duals for objects and a compatible symmetric braiding, and string diagrams define the free compact closed category over a tensor scheme [11].

Definition 2.4. A *spidered category* is a strict symmetric monoidal category equipped with a special commutative Frobenius structure [4] $(A, m, u, \delta, \epsilon, \sigma^F)$ on each object A .

Note that the morphisms of a spidered category need not be Frobenius, or even monoid or comonoid homomorphisms (in fact requiring this trivializes the structure).

To such a spidered category we now add duals for objects to obtain a compact closed category with additional structure. We call a compact closed category which is a spidered category in a compatible way a *dungeon category*.

Definition 2.5. A *dungeon category* is a compact closed category $(\mathcal{C}, \sigma^C, i, e)$ such that

- (i) Each object has a special commutative Frobenius structure $(A, m, u, \delta, \epsilon, \sigma^F)$ with $\sigma_{A(*), A(*)}^F = \sigma_{A(*), A(*)}^C$, and
- (ii) Any two morphisms f, g which are constructed from the identity id_A , the symmetric braiding $\sigma_{A,A}$, the Frobenius morphisms, and the dualizing cup and cap morphisms i_A, e_A for A , and have the same domain (tensor product of zero or more copies of A and A^*) and the same codomain (another such tensor product) are equal.

In other words, a “directed spider” morphism as in (ii) depends only on the number of inward and outward directed arrows, which way they point, and their order, with the second condition dropped up to application of $\sigma_{A,A}$.

Dungeon categories are a good setting for generalized belief propagation because we can bend wires and have spiders that play the role of variables in the probabilistic setting for belief propagation.

3. SUM-PRODUCT AND BELIEF PROPAGATION FOR CONTRACTION

The contraction problem in semiringed categories is $\#P$ -hard. There are many cases however where the problem becomes tractable. The main examples include trees (or diagrams where sections are merged to yield a tree) and categories for which a categorical generalization of holographic algorithms can be made to hold [25].

Most abstractly, the sum product algorithm [14] is simply the observation that when a diagram is a tree, one can perform contraction according to the tree. If the given diagram is not a tree, we can group nodes into a tree decomposition [9] to force it to be a tree, and then run sum-product. This is known as the *junction tree algorithm* [15] and has also been extended to the quantum case [19].

For actual computations, we can improve on the abstract sum-product algorithm by using an optimized message-passing version, which among other benefits permits parallelization.

3.1. Belief propagation in factor graphs. First we review BP in probabilistic factor graph models with discrete variables (see e.g. [22, Ch. 14]). The algorithm operates on a *factor graph*, a bipartite graph with one part discrete random variables $v \in V$ and one part factors $u \in U$. Each factor assigns a real number to each combination of states of the variables it is connected to. Multiplying factors and normalizing if needed gives a joint probability distribution.

Belief propagation is a message passing algorithm. Each message is a probability distribution over the states one variable v can take, so a vector in the associated vector space V_v . Each factor f_U at node u is a tensor in $\otimes_{v \in \text{nbhd}(u)} V_v$. Thus a factor defines $|\text{nbhd}(u)|$ linear maps $f_{u,v} : \otimes_{i \in \text{nbhd}(u) \setminus v} V_i \rightarrow V_v$, one for each $v \in \text{nbhd}(u)$. First we describe how to compute messages locally at each node in the factor graph, then how to assemble these into a complete algorithm.

Messages at variables. Suppose we have designated an edge e incident on the variable as output and the rest as input. Compute the pointwise (Hadamard) product of the incoming messages, and output it as the outgoing message along e . Since we are in a probabilistic category, this Hadamard product includes a rescaling so that the message is a probability distribution. If there are no incoming messages, output the uniform message.

Messages at factors. Suppose we have designated one of the edges (u, v) connected to a variable v incident on the factor u as output and the rest as input. Compute the tensor product of the incoming messages, apply $f_{u,v} :$

$\otimes_{i \in \text{nbhd}(u) \setminus v} V_i \rightarrow V_v$, and output the result as the outgoing message along the edge to v .

Resulting algorithm. This defines a system of *BP equations* describing the fixed points of the update rules. The initial messages can be chosen to be uniform distributions. When applied to a factor graph which is a tree, the algorithm converges after iterating a number of times equal to the diameter of the tree. Choosing a root, this can be completed in two “passes,” leaves to root then root to leaves, updating messages only as they change. More generally we consider a convergence threshold for the BP equations. It is a theorem that belief propagation is exact on trees, and it can work surprisingly well even when this is not satisfied.

3.2. Categorical Belief Propagation. In order for our generalized BP algorithm to operate, the category to which it is applied will need a few basic features; in particular it must be *reshapable*.

Suppose we have a morphism $f : \text{dom}(f) \rightarrow \text{cod}(f)$ and decompositions of its domain and codomain into monoidal products of other objects: $\text{dom}(f) = A_1 \otimes \cdots \otimes A_n$ and $\text{cod}(f) = A_{n+1} \otimes \cdots \otimes A_m$. Suppose I and J are ordered disjoint subsets of $[m] := \{1, 2, \dots, m\}$ whose union is $[m]$. These fix an alternate decomposition of $A_1 \otimes \cdots \otimes A_m$ into two objects (with arbitrary reordering), $D = \otimes_{i \in I} A_i$, $C = \otimes_{j \in J} A_j$. We require that there exists a unique morphism $r_{I,J}(f) : D \rightarrow C$ called the *reshaping of f from I to J* determined from this data. The reshaping does not depend on the route, that is for any I', J' we have $r_{I,J} r_{I',J'} = r_{I,J}$. If this works for all morphisms, we say the category is *reshapable*.

We now assume that we are working in a *dungeon category* (which is therefore reshapable) in which I -valued points can be stored and compared efficiently (e.g. space and time complexity linear in some “size” monoid homomorphism). This will guarantee the efficiency of belief propagation by an argument analogous to the classical case.

3.2.1. Bipartite with one part spiders. Assume we have a *dungeon category* (a strict compact closed category equipped with a compatible special commutative Frobenius structure on each object). This allows for the special spider morphisms to generalize and axiomatize the role played by variables in the traditional belief propagation algorithm for factor graphs. Consider a bipartite diagram with one part consisting of spiders (generalized variables) and the other arbitrary morphisms (generalized factors). Both spiders and factors are reshaped before composition. The reshaping of the spiders simply leads to a Hadamard (elementwise) product in the case of vector spaces and linear transformations.

The BP algorithm is expressed in terms of messages. When objects are sets, messages are elements of the set. For example, if objects are vector spaces, the messages will be vectors. In general messages for object A are I -valued points (elements of $\text{Mor}(I, A)$). For example any probability distribution can be expressed this way as a stochastic matrix applied to the unit Frobenius morphism (the unit “creates” a variable with a uniform distribution).

Using the compact closed structure to reshape morphisms, we still have that each “factor” morphism at node u defines $|\text{nbhd}(u)|$ different morphisms $f : \bigotimes_{i \in \text{nbhd}(u) \setminus v} V_i \rightarrow V_v$, one for each $v \in \text{nbhd}(u)$. First we describe how to compute messages locally at each node, then how to assemble these into a complete algorithm.

Messages at spiders. Apply the reshaped spider to incoming messages, and output the result as the outgoing message. If there are no incoming messages, treat the spider as a Frobenius unit.

Messages at “factor” morphisms. Suppose we have designated one of the wires incident on the factor f as output and the rest as input. Compute the monoidal product of the incoming messages, apply the reshaped f , and output the result as the outgoing message.

The system of BP equations are now equalities of I -valued morphisms, describing the fixed points of the update rules. The initial messages can be chosen to be units at the spiders. The nice behavior of the algorithm on diagrams which are trees should be preserved, once suitable definitions are made so that we can talk about convergence.

A spider is just a special kind of morphism. To get the **general bipartite** version, replace the message procedure at spiders with another copy of the factor message procedure.

3.3. Examples. In the category BoolRel each object is a two element set or a monoidal (here Cartesian) product of such. Morphisms are relations. Categorical belief propagation becomes the survey propagation algorithm for solving constraint satisfaction problems.

Augmenting the category FinRel of finite sets and relations with a positive integer multiplying each element of a relation yields the semiring category NFinRel (discussed in the context of counting constraint satisfaction problems such as computing partition functions in [25]). This category also corresponds approximately to sufficient statistics in the analysis of contingency tables [1].

We can add additional flexibility to FinRel to obtain database categories [29]. Then categorical belief propagation becomes a query planning algorithm.

The category in which numerical linear algebra takes place is vector spaces and linear transformations, where the vector spaces are augmented by orthonormal bases which define spiders. Composition is matrix multiplication, and tensor product is Kronecker product. Consider the *dual numbers* over a field \mathbb{F} , $\mathbb{D} = \mathbb{F}[\epsilon]/\langle\epsilon^2\rangle$. Reverse-mode automatic differentiation is categorical belief propagation in the category of vectors and matrices over \mathbb{D} .

We also conjecture [6] that algorithms commonly used in quantum condensed matter physics such as DMRG [30] and its many extensions can be considered as an instance of the categorical belief propagation algorithm.

REFERENCES

- [1] A. Agresti. *Categorical data analysis*. Wiley, 2013.
- [2] L. Babai and E. M. Luks. Canonical labeling of graphs. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 171–183. ACM, 1983.
- [3] O. Bachmann, G.-M. Greuel, C. Lossen, G. Pfister, and H. Schönemann. *A Singular introduction to commutative algebra*. Springer, 2007.
- [4] B. Coecke, D. Pavlovic, and J. Vicary. A new description of orthogonal bases. *Mathematical Structures in Computer Science*, 13(1), 2008.
- [5] B. Coecke and R. Duncan. Interacting quantum observables: categorical algebra and diagrammatics. *New Journal of Physics*, 13(4):043016, 2011.
- [6] A. Critch and J. Morton. Algebraic geometry of matrix product states. *arXiv preprint arXiv:1210.2812*, 2012.
- [7] L. Dixon and A. Kissinger. Open-graphs and monoidal theories. *Mathematical Structures in Computer Science*, 23(02):308–359, 2013.
- [8] M. Fürer, W. Schnyder, and E. Specker. Normal forms for trivalent graphs and graphs of bounded valence. In *Proceedings of the fifteenth annual ACM symposium on Theory of computing*, pages 161–170. ACM, 1983.
- [9] R. Halin. S-functions for graphs. *Journal of Geometry*, 8(1-2):171–186, 1976.
- [10] M. Hasegawa, M. Hofmann, and G. Plotkin. Finite dimensional vector spaces are complete for traced symmetric monoidal categories. In *Pillars of computer science*, pages 367–385. Springer, 2008.
- [11] G. M. Kelly and M. L. Laplaza. Coherence for compact closed categories. *Journal of Pure and Applied Algebra*, 19:193–213, 1980.
- [12] A. Kissinger, A. Merry, B. Frot, B. Coecke, D. Quick, L. Dixon, M. Soloviev, R. Duncan, and V. Zamdzhiev. Quantomatic. *Software available on-line at <http://sites.google.com/site/quantomatic/>*, 2014.
- [13] A. Kissinger. Pictures of processes: Automated graph rewriting for monoidal categories and applications to quantum computing. *arXiv preprint arXiv:1203.0202*, 2012.
- [14] F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *Information Theory, IEEE Transactions on*, 47(2):498–519, 2001.
- [15] S. L. Lauritzen and D. J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 157–224, 1988.
- [16] E. M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *Journal of Computer and System Sciences*, 25(1):42–65, 1982.

- [17] S. Mac Lane. *Categories for the working mathematician*, volume 5. Springer verlag, 1998.
- [18] E. Maneva, E. Mossel, and M. J. Wainwright. A new look at survey propagation and its generalizations. *Journal of the ACM (JACM)*, 54(4):17, 2007.
- [19] I. L. Markov and Y. Shi. Simulating quantum computation by contracting tensor networks. *SIAM Journal on Computing*, 38(3):963–981, 2008.
- [20] R. J. McEliece, D. J. C. MacKay, and J.-F. Cheng. Turbo decoding as an instance of Pearl’s belief propagation algorithm. *Selected Areas in Communications, IEEE Journal on*, 16(2):140–152, 1998.
- [21] A. A. Mena. Trivalent graph isomorphism in polynomial time. *arXiv preprint arXiv:1209.1040*, 2012.
- [22] M. Mezard and A. Montanari. *Information, physics, and computation*. Oxford University Press, 2009.
- [23] S. Mimram. Towards 3-dimensional rewriting theory. http://www.pps.univ-paris-diderot.fr/~smimram/docs/mimram_3drt.pdf, 2014.
- [24] J. Morton and J. Biamonte. Undecidability in tensor network states. *Physical Review A*, 86(3):030301, 2012.
- [25] J. Morton and J. Turner. Generalized counting constraint satisfaction problems with determinantal circuits. *arXiv preprint arXiv:1302.1932*, to appear in *Linear Algebra and its Applications*, 2013.
- [26] J. Pearl. Reverend bayes on inference engines: A distributed hierarchical approach. In *AAAI*, pages 133–136, 1982.
- [27] P. Selinger. A survey of graphical languages for monoidal categories. *New Structures for Physics*, pages 275–337, 2009.
- [28] P. Selinger. Finite dimensional hilbert spaces are complete for dagger compact closed categories. *Electronic Notes in Theoretical Computer Science*, 270(1):113–119, 2011.
- [29] D. I. Spivak. Functorial data migration. *Information and Computation*, 217:31–51, 2012.
- [30] S. R. White. Density matrix formulation for quantum renormalization groups. *Phys. Rev. Lett.*, 69:2863–2866, Nov 1992.

Translating measurement-based quantum computations with gflow into quantum circuits

Jisho Miyazaki¹, Michal Hajdušek^{1,2} and Mio Murao^{1,3}

¹ Department of Physics, Graduate School of Science, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan

² Singapore University of Technology and Design, 20 Dover Drive, Singapore

³ Institute for Nano Quantum Information Electronics, The University of Tokyo,
4-6-1 Komaba, Meguro-ku, Tokyo, Japan

Abstract

Causal flow (flow) and generalized flow (gflow) are ordering relations on a graph state that guarantee deterministic implementations of a unitary by measurement-based quantum computation (MBQC). There is a method called star pattern transformation to write a circuit decomposition of a unitary described by MBQC on a graph with flow. The straight extension of this method onto graphs with gflow, however, leads to a circuit including acausal gates, which is not well-defined in the quantum circuit model. In this work, we present a generalized method to translate any unitary implemented by MBQC with gflow into a quantum circuit without dealing with acausal gates. We also present another translation method that first maps to a quantum circuit including acausal gates and then erasing them to obtain an ordinary quantum circuit. By comparing these two methods, we study how the depth of quantum computation can be compressed in MBQC compared to the quantum circuit model. This QPL2014 submission is based on the work [1].

1 Introduction

Measurement-based quantum computation (MBQC) originally proposed in [2] is a model of quantum computation in which unitary operations are implemented by measuring qubits of a multipartite entangled state. The depth complexity of this model has been intensively investigated [3, 4, 5]. For several algorithms including the quantum Fourier transformation, it has been shown that MBQC requires smaller quantum computational depth than that of a variation of quantum circuit model where classical control of operations is not included [3]. This advantage originates from constant-time implementability of Clifford gates [6].

Causal flow (or just flow) [7] and generalized flow (or just gflow) [8] are ordering relations on a graph that guarantee deterministic MBQC irrespective of the choices of measurement angles. In order to implement unitary operations deterministically, we have to change the measurement angles of qubits according to the results of former measurements. The number of steps required to perform this adaptive measurements is added to the quantum computational depth of MBQC. Flow and gflow determine the ordering of measurements, and thus give an upper bound for the quantum computational depth complexity. The depth for MBQC on a graph given by gflow is lower than that by flow in general, since gflow is a generalization of flow.

There is a method called the *star pattern transformation* (SPT) to translate a representation of a unitary by MBQC on a graph with flow to a representation by compact circuit without ancilla qubit [7], and vice versa [4]. This correspondence implies that flow can be

considered as an ordering relation similar to that of the quantum circuit model, namely, a partial order constructed by elementary gates. However MBQC allows more general ordering relations, and one of such examples is gflow. If we use SPT to convert MBQC on a graph with gflow but without flow into a quantum circuit, we cannot avoid obtaining an ill-defined quantum circuit decomposition including an acausal loop structure [8] (for example, see the left-bottom circuit in Fig. 1). This impossibility is expected to originate from the constant-time implementability of Clifford gates since the quantum computational depth calculated by gflow is usually less than that of quantum circuit model. To understand how the acausal loop structure can be interpreted to be cancelled in MBQC and the depth of quantum computation can be compressed in MBQC compared to the quantum circuit model, we analyze graphs with gflow but no flow and the circuit decompositions of the unitaries implemented by MBQC on these graphs in this work.

2 Results

We have proven that a set of paths connecting the input vertices to the output vertices always exists on a graph with gflow. Such paths exist for a graph with flow and they are used as wires of the quantum circuit when SPT is applied on the graph with flow. However, for a graph with gflow but without flow, we cannot regard these paths as wires for directly applying SPT, otherwise, acausal CZ -gates appear (see the left-bottom circuit in Fig. 1). To solve this problem, we have constructed two methods of translation.

In the first method, we transform the graph state with no flow into a graph state with flow followed by extra $CNOT$ -gates before applying SPT (see the right-top circuit in Fig. 1). By applying SPT on the graph state with flow, we obtain a well-defined quantum circuit.

In the second method, we first apply SPT on the graph with no flow (represented by the left-bottom circuit in Fig. 1). We show that the effect of acausal CZ -gates can be interpreted by adding ancilla qubits and performing post-selection. In this sense, an acausal CZ -gate is equivalent to a circuit proposed by Bennett and Schumacher [9] and by Svetlichny [10] to simulate closed timelike curve by postselection (for correspondence, see Fig. 2). All acausal CZ -gates can be cancelled by taking appropriate circuit transformations that lead to an ordinary circuit without acausal gates (see the right-bottom circuit in Fig. 1). We show that the two translation methods result in the same quantum circuit as represented in the right-bottom circuit in Fig. 1).

The obtained quantum circuit can be further transformed to the form where single-qubit elementary gates corresponding to the vertices in the same measurement step of MBQC can be performed in parallel (see fig. 3). Using this transformation, we show that if a unitary performed by MBQC requires d measurement steps according to gflow, the resulting quantum circuit has d layers each of them consisting of a set of parallelized single-qubit elementary gates followed by Clifford gates. Since Clifford gates can be performed in a constant step in MBQC, a unitary represented by this quantum circuit can be implemented by $c * d$ steps where c denote a constant. On the other hand, performing Clifford gates may require steps

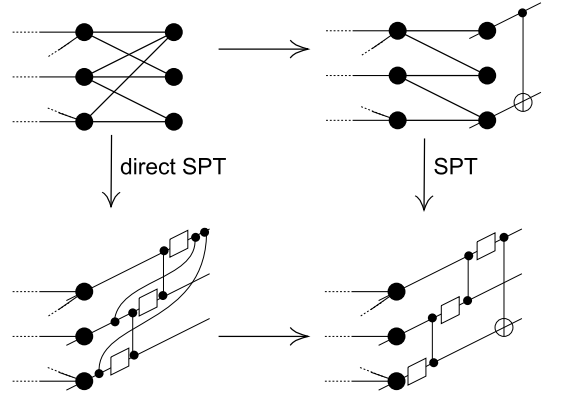


Figure 1: Two methods to translate MBQC (left-top) into circuit (right-bottom).

depending on the system size in the quantum circuit model.

In Ref. [11], a quantum circuit representing a unitary implemented by MBQC on a graph with flow is transformed so that the single-qubit elementary gates are parallelized. Our method is a generalization of this method for graphs with gflow.

In Ref. [12], a category theoretical method to translate MBQC into compact circuits is proposed and shown to be applicable for at least any graph with gflow. It might be possible to describe our translation method as changes of diagrams appearing in the categorical translation method, since both methods are based on local transformation of the graph state with no flow into another graph state with causal flow. Although the category theoretical method seems to be applicable for more general class of graphs, our method provides an explicit characterization of the entanglement between layers of the graph state, and shows a clear relationship between the depth calculated by gflow on a graph and the depth of the quantum circuit obtained from MBQC on the graph.

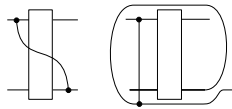


Figure 2: (Left) An acausal CZ -gate appearing in the quantum circuit obtained by directly applying SPT. (Right) The circuit proposed in Refs. [9] and [10] are equivalent to the acausal CZ -gate.

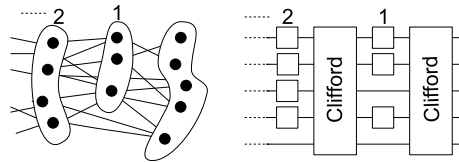


Figure 3: (Left) Gflow of MBQC. Sets of vertices numbered 1 and 2 are groups of vertices measured in a same step. (Right) The corresponding quantum circuit. The number of a set of parallelized single-qubit elementary gates corresponds to the number appearing in the gflow figure

Acknowledgments

We thank D. Markham, K. Nakago and E. Pius for helpful discussions. This work was supported by Project for Developing Innovation Systems of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. M. M. and M. H. acknowledge support from JSPS by KAKENHI (Grant No. 23540463, No. 23240001 and No.23-01770). J. M. is supported by Leading Graduate Course for Frontiers of Mathematical Sciences and Physics. The authors also gratefully acknowledge the ELC project (Grant-in-Aid for Scientific Research on Innovative Areas MEXT KAKENHI (Grant No. 24106009)) for encouraging the research presented in this paper.

References

- [1] J. Miyazaki, M. Hajdušek, and M. Murao, *eprint arXiv* : 1310.4043, (2013).
- [2] R. Raussendorf and H. J. Briegel, *Phys. Rev. Lett.*, **86**, 5188, (2001).
- [3] D. E. Browne, E. Kashefi and S. Perdrix, *In: Proc. of TQC 2010*.
- [4] A. Broadbent and E. Kashefi, *Theoretical Computer Science*, **410**, 26, (2009).
- [5] M. Mhalla and S. Perdrix, *eprint arXiv* : quant-ph/0412071, (2004).
- [6] R. Raussendorf, D. E. Browne, and H. J. Briegel, *Phys. Rev. A*, **68**, 022312 (2003).
- [7] V. Danos and E. Kashefi, *Phys. Rev. A*, **74**, 052310, (2006).
- [8] D. E. Browne, E. Kashefi, M. Mhalla, and S. Perdrix, *New Journal of Physics*, **9**, 250, (2007).
- [9] C. H. Bennett and B. Schumacher (unpublished). See also [<http://web.archive.org/web/20070206131550/http://www.research.ibm.com/people/b/bennetc/QUPONBshort.pdf>]
- [10] G. Svetlichny, *eprint arXiv* : 0902.4898, (2009).
- [11] R. Dias da Silva, E. Pius, and E. Kashfi, *eprint arXiv* : 1301.0351, (2013).
- [12] R. Duncan and S. Perdrix, *In: Proc. of 37th ICALP*, 285-296 (2010).

On monogamy of non-locality and macroscopic averages: examples and preliminary results

Rui Soares Barbosa

Quantum Group
Department of Computer Science
University of Oxford
`rui.soares.barbosa@cs.ox.ac.uk`

We explore a connection between monogamy of non-locality and a weak macroscopic locality condition: the locality of the average behaviour. These are revealed by our analysis as being two sides of the same coin.

Moreover, we exhibit a structural reason for both in the case of Bell-type multipartite scenarios, shedding light on but also generalising the results in the literature [15, 14]. More specifically, we show that, provided the number of particles in each site is large enough compared to the number of allowed measurement settings, and whatever the microscopic state of the system, the macroscopic average behaviour is local realistic, or equivalently, general multipartite monogamy relations hold.

This result relies on a classical mathematical theorem by Vorob'ev [17] about extending compatible families of probability distributions defined on the faces of a simplicial complex – in the language of the sheaf-theoretic framework of Abramsky and Brandenburger [2], such families correspond to no-signalling empirical models, and the existence of an extension corresponds to locality or non-contextuality. Since Vorob'ev's theorem depends solely on the structure of the simplicial complex, which encodes the compatibility of the measurements, and not on the specific probability distributions (i.e. the empirical models), our result about monogamy relations and locality of macroscopic averages holds not just for quantum theory, but for any empirical model satisfying the no-signalling condition.

In this extended abstract, we illustrate our approach by working out a couple of examples, which convey the intuition behind our analysis while keeping the discussion at an elementary level.

Keywords: monogamy of non-locality, macroscopic averages, Bell inequalities, no-signalling models, simplicial complexes, Vorob'ev's theorem.

1 Introduction

Bell's theorem [8] showed that the quantum world is non-local: the correlations between the outcomes of measurements on two entangled (space-like separated) particles are too strong to be explainable by a common 'local' cause. The usual monogamy of non-locality relations impose a limit on the amount of non-locality shared by one party with multiple other parties. For example, in a tripartite (A , B and C) system where each experimenter has two measurement settings available, there is a trade-off between the strengths of violation of a Bell inequality by the subsystem composed of A and B and the subsystem composed of A and C . More explicitly, for a bipartite Bell inequality $\mathcal{B}(-, -) \leq R$, the added inequality $\mathcal{B}(A, B) + \mathcal{B}(A, C) \leq R + R$ holds, even if each of $\mathcal{B}(A, B) \leq R$ and $\mathcal{B}(A, C) \leq R$ might be violated.

In [15], the authors consider multipartite macroscopic systems, consisting of a large number of particles at each site, which are described by quantum mechanics. At each site, only 'macroscopic'

measurements are available: e.g. magnetisation along some direction, which arises as a sort of average of the individual spin measurements in that direction for each particle in the site. The authors are concerned only with the average behaviour over all the microscopic particles – this can be obtained from the mean values of intensities measured macroscopically (see section 3.1 for a more detailed explanation). They show that, whatever the quantum state of the system (so regardless of the form and strength of the entanglement between the particles), and provided the number of particles at each site is large enough when compared to the number of different measurement settings available, there is a local realistic explanation for these macroscopic average correlations. The reason for such classicality is that non-local effects are diluted by averaging due to the restrictions imposed by monogamy.

However, monogamy holds more generally than just for quantum mechanics. In reference [14], it is shown that all no-signalling theories satisfy monogamy relations for the violation of any bipartite Bell-type inequality. More specifically, given a general bipartite Bell inequality $\mathcal{B}(A, B) \leq R$, they consider a scenario with one Alice, A , and k independent copies of Bob, $B^{(1)}, \dots, B^{(k)}$, with k equal to the number of measurement settings available to Bob. They show that a monogamy relation for the bipartite inequality, $\sum_{m=1}^k \mathcal{B}(A, B^{(m)}) \leq kR$, is satisfied by any no-signalling theory.

The methods used in the two above-cited papers are manifestly similar. This observation, also made in [15], leads one to conjecture that the results about local macroscopic averages also hold in general for any no-signalling theory. We show that this turns out to be the case: our investigation establishes a clear structural connection between the two papers leading to a generalisation of the results of both. Indeed, our main result for multipartite scenarios (proposition 5.1) can be read in two ways: on the one hand, it generalises the result of [14], concerning deriving monogamy relations from the no-signalling condition, from bipartite to multipartite Bell inequalities with an arbitrary number of sites; on the other hand, it generalises the result of [15], about the classicality of macroscopic average behaviour in multipartite models, from quantum models to all no-signalling models. Let us spell out the consequences of proposition 5.1 from each of these perspectives.

- Let $\mathcal{B}(A, B, C, \dots) \leq R$ be a general Bell-type inequality over n sites A, B, C, \dots with respectively k_A, k_B, k_C, \dots measurement settings available. Then, consider a scenario with a single copy of one of the sites, say A , and with r_B copies of site B , $B^{(1)}, \dots, B^{(r_B)}$, r_C copies of site C , $C^{(1)}, \dots, C^{(r_C)}$, etc. The monogamy relation for the satisfaction of the Bell inequality,

$$\sum_{m_B=1}^{r_B} \sum_{m_C=1}^{r_C} \dots \mathcal{B}(A, B^{(m_B)}, C^{(m_C)}, \dots) \leq r_B r_C \dots R, \quad (1)$$

is satisfied by any no-signalling model if and only if the number of copies of each site is at least the number of measurement settings at that site; i.e. $r_B \geq k_B$, $r_C \geq k_C$, etc. Reference [14] addressed the particular case $n = 2$ for which it proved the ‘if’ side of this result when $\vec{r} = \vec{k}$.

- From the other perspective, suppose that we have a scenario with n ‘macroscopic’ sites A, B, C, \dots with respectively k_A, k_B, k_C, \dots measurement settings available, and suppose that each of these macroscopic sites is constituted by a number r_i ($i \in \{A, B, C, \dots\}$) of microscopic sites. We assume that each of the k_i measurement settings available at site i corresponds to performing a similar measurement on all the microscopic sites (say, particles) that constitute it: the expected value of the macroscopic measurement then tells us the average behaviour among all the microscopic sites (see section 3.1 for more details). We show that if the number of microscopic sites forming each macroscopic site is at least the number of measurement settings at that site, i.e. if $r_i \geq k_i$ for all sites $i \in \{A, B, C, \dots\}$, then any no-signalling empirical model on a microscopic scenario has local average macroscopic behaviour. Reference [15] proved this result, but restricted to the case of quantum mechanical correlations. We show that having local

macroscopic averages is not a particular property of quantum mechanics distinguishing it from super-quantum correlations (this is not to say that we cannot distinguish them by other, more refined notions of macroscopic locality, cf. e.g. [13], and see section 3.1 for further details).

Moreover, it becomes apparent that the two items above are essentially two ways of looking at the same thing.

More important perhaps than these generalisations is that our analysis highlights the **structural reason** why these results hold. This is related to a characterisation due to Vorob'ev of the measurement scenarios that are inherently local or non-contextual. The idea is that quotienting a large scenario by the identification (of sites that are 'copies' or instances of the same site, or of microscopic sites forming a single macroscopic site)

$$A^{(1)} \sim \dots \sim A^{(r_A)} \quad B^{(1)} \sim \dots \sim B^{(r_B)} \quad C^{(1)} \sim \dots \sim C^{(r_C)} \quad \dots \quad ,$$

along which one considers the monogamy relation or takes the average, yields such an inherently local scenario.

Another important aspect of this work is its potential for further generalisation: as indicated in section 6, the same ideas can potentially be applied to yield monogamy relations for violation of contextuality inequalities, or to study macroscopic averages in more general scenarios as well.

The central aim of this extended abstract is to convey the intuition behind this structural proof. We mainly focus on showing a couple of simple examples whose geometric realisations can be easily visualised. We try to keep the presentation at an elementary level, ignoring some of the more involved technical details to focus on the central ideas and intuitions. A longer version of this work, containing all the technical details and full proofs, as well as presenting things in greater generality, is under preparation.

2 Measurement scenarios and empirical models

We quickly summarise some of the basic ideas of the sheaf-theoretic framework of Abramsky and Brandenburger [2], which provides a unified treatment of non-locality and contextuality in the general setting of no-signalling probabilistic models. For the purpose of the present document, we shall be mainly concerned with non-locality. Still, the geometric structures of this framework provide an appropriate setting in which to understand – and visualise – monogamy and macroscopic averages. For some other interesting results stemming from this sheaf-theoretic approach to non-locality and contextuality, the reader is referred to [4, 5, 3, 11, 1, 10].

2.1 Measurement scenarios

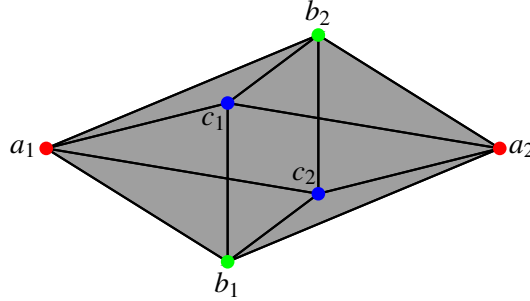
A **measurement scenario** is given by an abstract simplicial complex Σ on the (finite) set X of allowed measurements¹ (or equivalently, by a cover \mathcal{U} of X : the corresponding simplicial complex is obtained by down closure; and conversely, the maximal faces of a simplicial complex form a cover of X). Each face of the complex is called a measurement context. The intuition is that measurements in the same context can be performed together. Examples include multipartite Bell-type scenarios, Kochen-Specker configurations, and more.

¹An abstract simplicial complex on a set (of vertices) X is a family of subsets of X , called faces, that is downwards-closed and contains all the singletons $\{x\}$, $x \in X$. This is interpreted as a combinatorial description of a geometrical object given as a collage of points (the singletons), line segments (sets of two elements), triangles (sets of three elements), and their higher-dimensional counterparts.

Let us take as an example the simplest scenario in which monogamy relations arise, in their most familiar form. We consider a Bell-type scenario with three sites (A , B and C) and two possible measurement settings available to the experimenter at each site (a_1 and a_2 for A , b_1 and b_2 for B , and c_1 and c_2 for C). As usual, the choice of measurement at each site can be done independently of the other sites. Formally, the set of available measurements is $X = \{a_1, a_2, b_1, b_2, c_1, c_2\}$ and the cover of maximal contexts is

$$\mathcal{U} = \{\{a_i, b_j, c_k\} \mid i, j, k \in \{1, 2\}\}$$

The corresponding simplicial complex is a *hollow* octahedron, depicted below:



Note that this complex can be described in a more compositional way as

$$\mathcal{D}_2^{*3} = \mathcal{D}_2 * \mathcal{D}_2 * \mathcal{D}_2,$$

where \mathcal{D}_2 is the discrete simplicial complex on two vertices², corresponding to the scenario available to each experimenter, and $*$ stands for the simplicial join operation³, which captures parallel composition of scenarios. For more details on this, see e.g. [16].

In the explicit examples of empirical models in the rest of this text, we shall take all measurements to have two possible outcomes: 0 and 1. This is irrelevant as none of the results we consider is sensitive to the sets of outcomes, as long as there are at least two outcomes per measurement. With this extra assumption, the scenario \mathcal{D}_2^{*3} above is also customarily known as the $(3, 2, 2)$ scenario: the numbers stand for 3 sites, 2 measurement settings at each site, and 2 outcomes for each measurement.

2.2 Empirical models and extendability

While a measurement scenario is an abstract description of a set of possible experiments, empirical models represent particular (real or hypothetical) probabilistic results of these experiments (one can think of frequencies tabulated from runs of the experiments on ensembles of identically prepared systems).

Given a measurement scenario \mathcal{U} , an **empirical model** is a compatible family of probability distributions $(\mu_C)_{C \in \mathcal{U}}$, where each μ_C is a distribution on joint outcomes of the measurements in context C . Compatibility here means that μ_C and $\mu_{C'}$ marginalise to the same distribution on outcomes of measurements in $C \cap C'$. In the case of multipartite scenarios, this corresponds to the usual **no-signalling** condition.

For such an empirical model, we are concerned with the existence of a global probability distribution μ_X on the joint outcomes of all the measurements that marginalises to all the distributions μ_C . It is

²The discrete complex on n vertices, \mathcal{D}_n , is the minimal simplicial complex on n vertices, containing no faces of dimension higher than 0 (lines, triangles, etc.). Formally, $\mathcal{D}_n := \{\emptyset, \{1\}, \dots, \{n\}\}$.

³Given simplicial complexes Σ_1 and Σ_2 on vertex sets X_1 and X_2 respectively, their simplicial join is defined on the vertex set $X_1 \sqcup X_2$ as $\Sigma_1 * \Sigma_2 := \{\sigma_1 \sqcup \sigma_2 \mid \sigma_1 \in \Sigma_1, \sigma_2 \in \Sigma_2\} = \{\sigma \subseteq X_1 \sqcup X_2 \mid \sigma \cap X_1 \in \Sigma_1 \wedge \sigma \cap X_2 \in \Sigma_2\}$.

shown in [2] that such a global extension exists iff the model admits a non-contextual (or local, in the particular case of multipartite scenarios) hidden variable explanation. So, the set of joint outcomes of all measurements ($\{0, 1\}^X$ in the case of dichotomic measurements) can be seen as a canonical hidden variable space. Obstructions to such extensions are witnessed by violations of Bell-type inequalities by the probability distributions μ_C (cf. [4] for a general scheme, based on logical consistency conditions, for deriving complete sets of Bell-type inequalities on any measurement scenario).

Let us consider some examples. Take the tripartite scenario from section 2.1. An empirical model for this scenario is a collection of no-signalling probabilities of the form $p(a_i, b_j, c_k = x, y, z)$, where x, y, z range over the possible outcomes of the respective measurements. An example of a valid empirical model is represented in the following table - this is the model obtained by preparing a 3-qubit system in the W state and allowing Z and X measurements at each site (on each qubit).

	000	001	010	011	100	101	110	111
$a_1 b_1 c_1$	$3/8$	$1/24$	$1/24$	$1/24$	$1/24$	$1/24$	$1/24$	$3/8$
$a_1 b_1 c_2$	$1/3$	$1/12$	0	$1/12$	0	$1/12$	$1/3$	$1/12$
$a_1 b_2 c_1$	$1/3$	0	$1/12$	$1/12$	0	$1/3$	$1/12$	$1/12$
$a_1 b_2 c_2$	$1/6$	$1/6$	$1/6$	0	$1/6$	$1/6$	$1/6$	0
$a_2 b_1 c_1$	$1/3$	0	0	$1/3$	$1/12$	$1/12$	$1/12$	$1/12$
$a_2 b_1 c_2$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	0	$1/6$	0
$a_2 b_2 c_1$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	0	0
$a_2 b_2 c_2$	0	$1/3$	$1/3$	0	$1/3$	0	0	0

(2)

Another example is the super-quantum tripartite box known as Svetlichny box [7]:

	000	001	010	011	100	101	110	111
$a_1 b_1 c_1$	$1/4$	0	0	$1/4$	0	$1/4$	$1/4$	0
$a_1 b_1 c_2$	$1/4$	0	0	$1/4$	0	$1/4$	$1/4$	0
$a_1 b_2 c_1$	$1/4$	0	0	$1/4$	0	$1/4$	$1/4$	0
$a_1 b_2 c_2$	0	$1/4$	$1/4$	0	$1/4$	0	0	$1/4$
$a_2 b_1 c_1$	$1/4$	0	0	$1/4$	0	$1/4$	$1/4$	0
$a_2 b_1 c_2$	0	$1/4$	$1/4$	0	$1/4$	0	0	$1/4$
$a_2 b_2 c_1$	0	$1/4$	$1/4$	0	$1/4$	0	0	$1/4$
$a_2 b_2 c_2$	0	$1/4$	$1/4$	0	$1/4$	0	0	$1/4$

(3)

Finally, let us also consider a non-symmetric example (with respect to B and C):

	000	001	010	011	100	101	110	111
$a_1 b_1 c_1$	$1/4$	$1/4$	0	0	0	0	$1/4$	$1/4$
$a_1 b_1 c_2$	$1/4$	$1/4$	0	0	0	0	$1/4$	$1/4$
$a_1 b_2 c_1$	$1/4$	$1/4$	0	0	0	0	$1/4$	$1/4$
$a_1 b_2 c_2$	$1/4$	$1/4$	0	0	0	0	$1/4$	$1/4$
$a_2 b_1 c_1$	$1/4$	$1/4$	0	0	0	0	$1/4$	$1/4$
$a_2 b_1 c_2$	$1/4$	$1/4$	0	0	0	0	$1/4$	$1/4$
$a_2 b_2 c_1$	0	0	$1/4$	$1/4$	$1/4$	$1/4$	0	0
$a_2 b_2 c_2$	0	0	$1/4$	$1/4$	$1/4$	$1/4$	0	0

(4)

All three examples above are non-local.

3 Relating monogamy and macroscopic averages: a first example

3.1 Macroscopic average behaviour

Let us start by clarifying what we mean by average macroscopic behaviour. This is the same as the macroscopic correlations considered in [15], and we also contrast it with the notion of macroscopic locality suggested in [13] and [6].

In order to understand the averaging process, we start by considering single-site measurements. Let us first take a (microscopic) measurement with l possible outcomes. One can imagine that, in such a measurement, a *single* particle is subjected to an interaction a , a measurement process of some sort, resulting in the particle colliding with one of the l detectors corresponding to the measurement outcomes. The nature of this interaction might be probabilistic; repeating the experiment many times on identically prepared systems allows us to collect statistical data $p(x|a)$ or $p(a = x)$ (with $x \in \{1, \dots, l\}$) corresponding to the probability of the detector x being clicked given that one has decided to measure a .

Now we introduce a change to this setup. In a macroscopic experiment, the experimenter receives a beam of N particles instead of a single particle. The same interaction is applied (simultaneously) to all the particles in the beam, dividing it into smaller beams that collide with each of the detectors $1, \dots, l$. The information one can obtain from such an experiment is the number of particles that collide with each detector, i.e. the intensity of each of the smaller beams. Note that instead of beams of photons, we could also think of regions of a magnetic material where measuring the magnetisation in a certain direction corresponds to making a spin measurement on all the N particles in the region. The details are not essential to the discussion, so we shall keep talking about ‘beams’.

In order to simplify the discussion, but with no crucial loss, we assume that the microscopic measurements are dichotomic, i.e. $l = 2$, and take the possible outcomes to be 0 and 1. Then, the result of the macroscopic measurement (an intensity) can be represented by a single number I_1 proportional to the number of particles that hit the detector corresponding to outcome 1. (Note that this number can be normalised to yield a number in $[0, 1]$ representing the proportion of particles that hit detector 1.)

Of course, given the probabilistic nature of the microscopic measurements, every time the whole experiment is run, with the same preparation of the initial state of the beam, the number of particles hitting detector 1 differs slightly. But if N is large enough, a realistic detector won’t be able to discern, and so count, individual particles. For the purpose of this paper, we are concerned only with the mean, or expected, value of these intensities. This can easily be obtained from such a macroscopic experiment, and it is what is usually taken to be the *value* of the macroscopic observable (e.g. total magnetisation). This mean intensity can also be interpreted as giving the average behaviour among the particles in the beam or region: if we would randomly select one of the N particles and subject it to the microscopic measurement, one would get the outcome 1 with probability I_1 (assuming the intensities are normalised as mentioned above); i.e. $I_1 = \sum_{i=1}^N p_i(a = 1)$. Observe that the situation is analogous to statistical mechanics, where a macrostate arises as an averaging over an extremely large number of microstates, and hence several different microstates can correspond to the same macrostate.

We, like Ramanathan et al. [15], are interested in multipartite correlations arising from this sort of macroscopic measurements done across different sites. We can think of a beam of photons being sent to each of a number of experimenters in spatially separated locations who can make several different measurements, or we can think (as suggested in [15]) of magnetisation measurements along several different directions done in a number of regions of a many-spin system. In any case, we have a sort of average macroscopic Bell experiment: the (mean) values of the macroscopic intensities (the

intensity of an outcome $\langle x_1, \dots, x_n \rangle$ of a multipartite macroscopic measurement is the product of the intensities of the outcome x_i for each site i) indicate the behaviour of a randomly chosen tuple of particles: one from each of the beams, or sites. We shall show (as a consequence of proposition 5.1) that, as long as there are enough particles (microscopic sites) in each macroscopic site when compared to the number of possible measurement settings the experimenters can choose from, such average macroscopic behaviour is always local no matter which no-signalling theory accounts for the underlying microscopic correlations.

We should mention how this relates to discussions about macroscopic locality in the literature. In [6], a bipartite setup similar to the one described above is considered. One difference is that the authors assume that the beams are composed of independently and identically prepared pairs of particles. This also seems to be the case, implicitly, in [13] (at least in the final paragraph of section I). A run of this macroscopic experiment can be seen as running the same microscopic bipartite Bell experiment multiple times and recording only how many times one obtains a certain outcome, disregarding the information of which particles from each of the beams were originally paired. The mean values of the intensities (or, equivalently, the behaviour of a random pair) that we described above are rather boring in this situation with identically prepared pairs: what we get is a diluted version of the probabilities for one of the identical microscopic empirical models⁴. However, these authors are not simply interested in the mean values of these intensities (each of these is, after normalisation, a value in the interval $[0, 1]$), but rather on the more fundamental probability distributions over $[0, 1]$ from which these means are calculated – recall that each time the macroscopic experiment is run, one measures slightly different intensities, so the observed intensities fall in a distribution around the mean value. Their aim is to witness non-locality on the fluctuations. More specifically, they are concerned with the question of whether these distributions of intensities can be explained by local hidden variables.

Even though some information is inevitably lost in such an experiment (particularly regarding the original pairings) Bancal et al. [6] show that one can witness non-locality at this level, if the detectors are perfect, in the sense that they can measure the intensity of beams with maximal precision, to a sensitivity of one particle. This clearly becomes impractical as N grows. At the opposite end, in the idealised limit where the resolution of the detectors is very bad, one would always observe the same intensities: namely, our mean value of intensities with no fluctuations around it. Navascués and

⁴ Suppose that N pairs $\{\langle a^{(m)}, b^{(m)} \rangle\}_{m=1, \dots, N}$ are prepared in the same empirical model described by probability distributions $p(a, b = x, y)$ where a and b range over the possible measurement settings on each of the sites. Then the average behaviour of a (arbitrary) pair is given by:

$$\begin{aligned}
 \tilde{p}(a, b = x, y) &= \frac{1}{N^2} \sum_{m_A, m_B=1}^N p(a^{(m_A)}, b^{(m_B)} = x, y) \\
 &= \frac{1}{N^2} \sum_{m=1}^N p(a^{(m)}, b^{(m)} = x, y) + \frac{1}{N^2} \sum_{m_A \neq m_B=1}^N p(a^{(m_A)}, b^{(m_B)} = x, y) \\
 &= \frac{N}{N^2} p(a, b = x, y) + \frac{1}{N^2} \sum_{m_A \neq m_B=1}^N p(a^{(m_A)} = x) p(b^{(m_B)} = y) \\
 &= \frac{1}{N} p(a, b = x, y) + \frac{N^2 - N}{N^2} p(a = x) p(b = y) \\
 &= \frac{1}{N} p(a, b = x, y) + \left(1 - \frac{1}{N}\right) p(a = x) p(b = y)
 \end{aligned}$$

which is the initial microscopic model very ‘diluted’ by a local model (corresponding to the pairs that were not prepared originally as a pair).

Wunderlich [13] suggest that it is physically reasonable that, when N is large enough, one could detect changes on intensity values of the order of \sqrt{N} . These authors propose the notion of macroscopic locality to mean that the distributions of observed intensities with a resolution of order \sqrt{N} admit a local hidden variable explanation. They show that this principle of macroscopic locality is satisfied by quantum mechanics, but is not valid in general for all no-signalling theories: more accurately, the set of correlations satisfying it is Q^1 , the first level of the hierarchy of semidefinite programs approximating the quantum set [12].

In this sense, the kind of macroscopic correlations we (and Ramanathan et al. [15]) consider seems more restricted, since we show that these are local no matter which no-signalling theory accounts for the underlying microscopic correlations. However, there are some important differences, which we now summarise:

- Firstly, we do not consider the beams to consist of identically prepared pairs (or tuples) of particles. In the bipartite setting of [6] above, pairs of particles were identically prepared and then a particle of each pair was sent to Alice and the corresponding one to Bob (although the pairing is lost as the particles are lumped together in a beam). In our setting, the particles may be in different states and there is no restrictions on which groups of particles of Alice and of Bob (and possibly of others, as we allow for an arbitrary number of sites) are entangled – the ‘microstate’ of the system can be very highly non-local. The only restriction we impose is that of no-signalling.
- Secondly, our aim is not to explain the distribution of the intensities, with their fluctuations around the mean value, by a local model, as in [13, 6]. Rather, the (products of the) mean intensities themselves, which are taken as the value of the macroscopic observable, give us a description of the behaviour of the average pair or tuple of particles in the beams. It is this average behaviour that we aim to explain by a local model. We prove this is indeed always possible for any no-signalling microscopic theory provided there are enough particles compared to measurement settings available at each site. The reason, again, has to do with monogamy, which ‘dilutes’ the non-locality.

Despite the latter difference, note that since such average behaviour corresponds to the mean, or expected, values of the macroscopic measurements of intensities considered as in [13, 6], our result also implies that macroscopic CHSH-type inequalities (i.e. inequalities involving only the expected values of macroscopic experiments) can never be violated by no-signalling microscopic theories. It is only looking at higher order moments (which correspond to other characteristics of the distribution, such as variance, skewness, kurtosis, etc.) that one may get a difference between quantum mechanics and general no-signalling theories.

3.2 Macroscopic average behaviour: examples

Let us see how this averaging works for our tripartite example. We regard sites B and C as forming one macroscopic site, M , and site A as forming another⁵. The idea is that we will average over the behaviour of the ‘microscopic’ particles B and C . In order to be lumped together, B and C must be symmetric, i.e. of the same ‘type’. In particular, we need to know which measurements on the site

⁵In this example, ‘macroscopic’ means one or two particles only, allowing us to keep the example small enough to be visualised. This is sufficient to get local averages given that we are only considering two measurement settings per site: recall from section 1 that the condition is that the number of ‘microscopic’ particles in a site, or copies of a site, should be at least the number of measurement settings available at that site (except possibly for one of the sites, A in this example, where we can consider a single particle or copy).

B correspond to which measurements on the site C . Here, we consider a symmetry of the system which makes the identifications $b_1 \sim c_1$ and $b_2 \sim c_2$. We will name m_1 and m_2 the ‘macroscopic’ measurements resulting from these identifications.

Given an empirical model on the tripartite scenario, one can consider the partial model on the subsystem composed of sites A and B only, whose probabilities are given by marginalisation (in quantum mechanics, this corresponds to partial trace):

$$p(a_i, b_j = x, y) := \sum_z p(a_i, b_j, c_k = x, y, z) .$$

Note that this expression is independent of c_k due to no-signalling. Similarly, one can consider the partial model on the subsystem composed of A and C only.

The average behaviour under the identification of B with C is then a bipartite model with two ‘macroscopic’ sites A and M , given as an average of probability distributions of the partial models:

$$p(a_i, m_j = x, y) := \frac{p(a_i, b_j = x, y) + p(a_i, c_j = x, y)}{2} . \quad (5)$$

Let us see how such average models look like for the particular empirical models 2-4 from section 2.2. The first two examples are symmetric with respect to B and C , meaning that the restriction of the model to sites A and B is the same as the restriction of the model to sites A and C . Consequently, it is also equal to the macroscopic averaged model, since the latter arises as an average of the two partial models. The table for the macroscopic model emerging from example 2 is:

	00	01	10	11
$a_1 m_1$	5/24	1/24	1/24	5/24
$a_1 m_2$	1/6	1/12	1/6	1/12
$a_2 m_1$	1/6	1/6	1/12	1/12
$a_2 m_2$	1/6	1/6	1/6	0

and for example 3 we have:

	00	01	10	11
$a_1 m_1$	1/4	1/4	1/4	1/4
$a_1 m_2$	1/4	1/4	1/4	1/4
$a_2 m_1$	1/4	1/4	1/4	1/4
$a_2 m_2$	1/4	1/4	1/4	1/4

Note that both these (macroscopic average) bipartite models are local. Now let us consider example 4. The partial models on sites A and B and on sites A and C are represented in the following tables:

	00	01	10	11
$a_1 b_1$	1/2	0	0	1/2
$a_1 b_2$	1/2	0	0	1/2
$a_2 b_1$	1/2	0	0	1/2
$a_2 b_2$	0	1/2	1/2	0

	00	01	10	11
$a_1 c_1$	1/4	1/4	1/4	1/4
$a_1 c_2$	1/4	1/4	1/4	1/4
$a_2 c_1$	1/4	1/4	1/4	1/4
$a_2 c_2$	1/4	1/4	1/4	1/4

Note that the model in the left is non-local (even maximally violating a Bell inequality), while the one in the right is local. The ‘macroscopic’ average model is obtained as an average of these two:

	00	01	10	11
$a_1 m_1$	3/8	1/8	1/8	3/8
$a_1 m_1$	3/8	1/8	1/8	3/8
$a_1 m_1$	3/8	1/8	1/8	3/8
$a_1 m_1$	1/8	3/8	3/8	1/8

This model is also local, like the other average models above: a global probability distribution for this model is

$$\begin{aligned} & \frac{1}{8}[a_1 a_2 b_1 b_2 = 0000] + \frac{1}{8}[a_1 a_2 b_1 b_2 = 0001] \\ & + \frac{1}{8}[a_1 a_2 b_1 b_2 = 0100] + \frac{1}{8}[a_1 a_2 b_1 b_2 = 0110] \\ & + \frac{1}{8}[a_1 a_2 b_1 b_2 = 1001] + \frac{1}{8}[a_1 a_2 b_1 b_2 = 1011] \\ & + \frac{1}{8}[a_1 a_2 b_1 b_2 = 1110] + \frac{1}{8}[a_1 a_2 b_1 b_2 = 1111] . \end{aligned}$$

We shall see that these three examples are in no way special. Indeed, our analysis will clarify that the macroscopic average behaviour is local no matter which no-signalling tripartite empirical model we start from.

3.3 Monogamy and macroscopic averages: Bell inequalities

Now, we make a very simple observation that establishes the connection between monogamy of non-locality and locality of these macroscopic averages. Consider any Bell inequality $\mathcal{B}(-, -) \leq R$ for a scenario with two parties, each with two available measurements. Such an inequality is determined by a set of coefficients $\alpha(i, j, x, y)$ and a bound R . We have that:

$$\begin{aligned} & \mathcal{B}(A, M) \leq R \\ \Leftrightarrow & \sum_{i, j, x, y} \alpha(i, j, x, y) p(a_i, m_j = x, y) \leq R \\ \Leftrightarrow & \{ \text{definition of the average probabilities, eq. 5} \} \\ & \sum_{i, j, x, y} \alpha(i, j, x, y) \frac{p(a_i, b_j = x, y) + p(a_i, c_j = x, y)}{2} \leq R \\ \Leftrightarrow & \{ \text{re-arranging terms} \} \\ & \sum_{i, j, x, y} \alpha(i, j, x, y) p(a_i, b_j = x, y) + \sum_{i, j, x, y} \alpha(i, j, x, y) p(a_i, c_j = x, y) \leq 2R \\ \Leftrightarrow & \mathcal{B}(A, B) + \mathcal{B}(A, C) \leq 2R \end{aligned}$$

That is, the ‘macroscopic’ average model, $p(a_i, m_j = \dots)$ on sites A and M , satisfies the Bell inequality, $\mathcal{B}(A, M) \leq R$, if and only if the ‘microscopic model’ (on sites A , B and C) is monogamous with respect to violating it; i.e. the bipartite partial models $p(a_i, b_j = \dots)$ and $p(a_i, c_j = \dots)$ satisfy the monogamy relation $\mathcal{B}(A, B) + \mathcal{B}(A, C) \leq 2R$, and so cannot both violate the Bell inequality. This is an instance of a more general equivalence between Bell inequalities on ‘macroscopic’ averages and the monogamy of violation of the same inequality at the ‘microscopic’ level. As a consequence, a macroscopic average model satisfies all Bell inequalities (i.e. it is local) if and only if the microscopic model is monogamous with respect to violating all those inequalities. This is the case, in particular, of all the models in the tripartite scenario we are analysing, such as the examples considered in section 3.2. In the next section, we give a reason why this is so based on the structure of the scenario.

4 A structural explanation

4.1 Vorob'ev's theorem

A classical mathematical result due to Vorob'ev [17], and motivated by a problem in game theory, deals with the following question, here rephrased in our terms: for which measurement scenarios \mathcal{U} (or Σ) is it so that any no-signalling empirical model $(\mu_C)_{C \in \mathcal{U}}$ defined on it admits a global extension, i.e. is local or non-contextual? Vorob'ev derived a necessary and sufficient condition on the simplicial complex Σ for this to be the case. We present a simplified yet equivalent version of Vorob'ev's condition, which happens to be known in relational database theory as acyclicity, an important property of database schemata (cf. [1] for more on the connection between relational database theory and the study of locality and non-contextuality). The idea is that such a scenario can be constructed by adding one measurement at a time in such a way that the new measurement is added to only one maximal context. Equivalently, it can be de-constructed by removing at each step a measurement belonging to a single maximal context.

Definition 4.1. Let Σ be a simplicial complex. Given a maximal face C , let π_C denote the vertices of Σ which belong to C and not to any other maximal faces. If $\pi_C \neq \emptyset$ for some C , then we say that there is a **Graham-reduction** step from Σ to the subcomplex

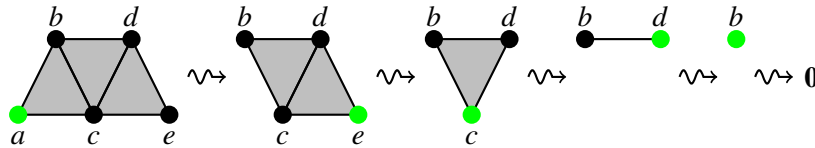
$$\Sigma' := \{\sigma \in \Sigma \mid \sigma \cap \pi_C = \emptyset\} = \{\sigma \setminus \pi_C \mid \sigma \in \Sigma\}$$

and write $\Sigma \rightsquigarrow \Sigma'$.

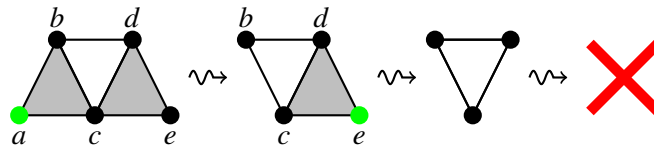
The complex Σ is said to be **acyclic** if it is Graham-reducible to the empty complex⁶, i.e. if there exists a series of Graham-reduction steps from Σ to the empty complex:

$$\Sigma = \Sigma_0 \rightsquigarrow \Sigma_1 \rightsquigarrow \dots \rightsquigarrow \Sigma_r = \mathbf{0}$$

The following is an example of a successful Graham reduction to $\mathbf{0}$, witnessing the acyclicity of the simplicial complex on the left.



On the contrary, the following complex is not acyclic: Graham reduction always fails, hitting a 'cycle'.



Theorem 4.2 ([17], rephrased and with simplified condition). *Let Σ be a simplicial complex. Then any empirical model defined on Σ is extendable if and only if Σ is acyclic.*

⁶The empty complex, $\mathbf{0}$ is the only simplicial complex on \emptyset , that is, with no vertices.

4.2 Structural reason: tripartite example

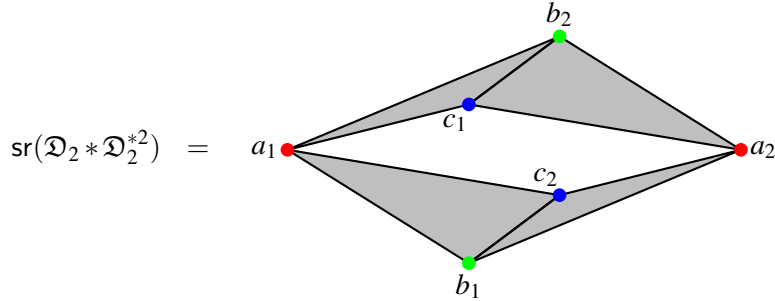
We mentioned above that, for the scenario we are considering, any empirical model gives rise to local average behaviour correlations. The structural reason for this is the fact that the quotient of the scenario by the identification of sites B and C is acyclic. Let us look at this in more detail.

Our scenario is represented by the simplicial complex $\mathfrak{D}_2 * \mathfrak{D}_2 * \mathfrak{D}_2$, where the factors corresponds to sites A , B and C . This is the hollow octahedron we depicted before, in section 2.2. Given that we want to identify B and C , we regard this complex as⁷

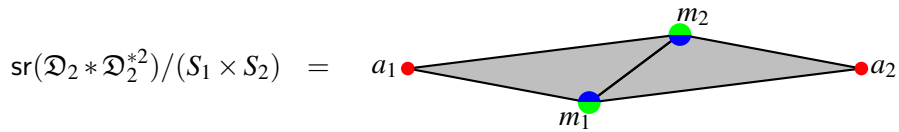
$$\Sigma_{n=2, k_1=2, k_2=2, r_1=1, r_2=2} := \mathfrak{D}_2 * \mathfrak{D}_2^{*2} = \mathfrak{D}_2 * (\mathfrak{D}_2 * \mathfrak{D}_2),$$

with sites B and C ‘grouped’ together in the second factor on which the identification $b_i \sim c_i$ acts.

We shall explicitly see what the quotient is. The first step is so-called semiregularisation, where we remove edges between vertices that are being identified as such edges are unnecessary. So, in this case, we must remove the edges $\{b_1, c_1\}$ and $\{b_2, c_2\}$, obtaining the following simplicial complex:



Now, taking the quotient, we will identify the measurements b_1 and c_1 as m_1 , and b_2 and c_2 as m_2 . We obtain the following simplicial complex:



Observe that the set of maximal faces (i.e. the cover of maximal contexts) is

$$\{\{a_1, m_1, m_2\}, \{a_2, m_1, m_2\}\}.$$

So, more things are compatible than in the usual bipartite scenario, which has cover

$$\{\{a_1, m_1\}, \{a_1, m_2\}, \{a_2, m_1\}, \{a_2, m_2\}\}.$$

As it happens, any empirical model defined on the original complex $\mathfrak{D}_2 * \mathfrak{D}_2^2$ will give rise to another model defined on the quotient scenario, by taking averages along the faces being identified. Therefore, not only are the probabilities $p(a_i, m_j = \dots)$ defined via an average, giving a model on

⁷The $\Sigma_{n, \vec{k}, \vec{r}}$ notation on the left-hand side will be introduced in section 5; it is provided here just for reference: n stands for the number of ‘macroscopic’ sites, k_i for the number of measurement settings available at site i , and r_i for the number of ‘microscopic’ sites in, or copies of, site i . The reader is referred to section 2.1 for the notation on the right-hand side.

the usual bipartite scenario above, so are the probabilities $p(a_i, m_1, m_2 = \dots)$, yielding a model on the more compatible bipartite scenario that arises as a quotient. The probability distribution on the triangle $\{a_i, m_1, m_2\}$ is obtained as an average of the probability distributions on the top and bottom triangles that gave rise to it, namely of $p(a_i, b_1, c_2 = \dots)$ and $p(a_i, c_1, b_2 = \dots)$.

The quotient complex we have obtained does satisfy the Vorob'ev condition of acyclicity. This is easy to see: one can remove the vertices, for example, in the order a_1, a_2, m_1, m_2 . Therefore, no matter which empirical model $p(a_i, b_j, c_k = \dots)$ we start from, the model of average macroscopic behaviour, $p(a_i, m_j = \dots)$, is local. In particular, it satisfies any Bell inequality. Hence, by the equivalence discussed in section 3.3, the original tripartite model also satisfies a monogamy relation for any of these bipartite Bell inequalities.

4.3 A non-acyclic example

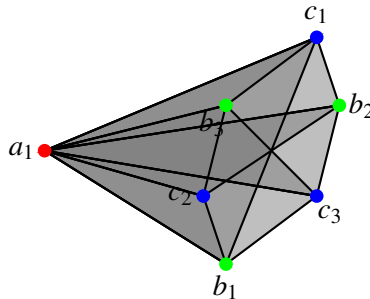
Let us now consider an example where one does not get monogamy relations, or equivalently, where one does not necessarily get local macroscopic averages. Suppose that we again have a tripartite (A, B, C) scenario, but that this time B and C have 3 available measurement settings each. In a compositional notation (as explained in footnote 7 at the start of section 4.2) and since we are again interested in identifying the sites B and C , this scenario is represented by the simplicial complex

$$\Sigma_{n=2, k_1=2, k_2=3, r_1=1, r_2=2} := \mathfrak{D}_2 * \mathfrak{D}_3^{*2} = \mathfrak{D}_2 * (\mathfrak{D}_3 * \mathfrak{D}_3).$$

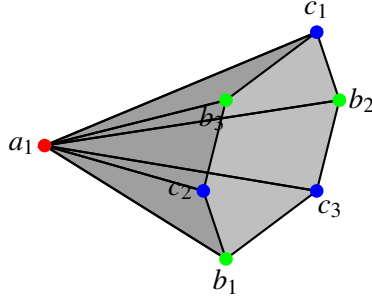
The maximal contexts are

$$\begin{aligned} \mathcal{U} = \{ & \{a_1, b_1, c_1\}, \{a_1, b_1, c_2\}, \{a_1, b_1, c_3\}, \{a_1, b_2, c_1\}, \{a_1, b_2, c_2\}, \{a_1, b_2, c_3\}, \\ & \{a_1, b_3, c_1\}, \{a_1, b_3, c_2\}, \{a_1, b_3, c_3\}, \{a_2, b_1, c_1\}, \{a_2, b_1, c_2\}, \{a_2, b_1, c_3\}, \\ & \{a_2, b_2, c_1\}, \{a_2, b_2, c_2\}, \{a_2, b_2, c_3\}, \{a_2, b_3, c_1\}, \{a_2, b_3, c_2\}, \{a_2, b_3, c_3\} \} \end{aligned}$$

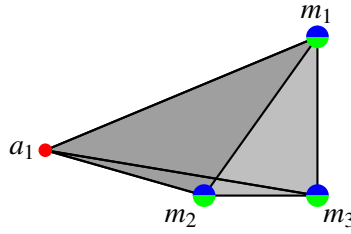
and half the simplicial complex is depicted below (one should imagine the other half, a mirror image of this consisting of the faces that include a_2 instead of a_1 , collated to it; we choose to omit that part as it would make the picture more confusing and hard to visualise):



We consider the identifications $b_i \sim c_i$ ($i = 1, 2, 3$). Again, we first discard the edges between identified measurements, namely $\{b_1, c_1\}$, $\{b_2, c_2\}$, and $\{b_3, c_3\}$. The resulting complex, $\text{sr}(\mathfrak{D}_2 * \mathfrak{D}_3^{*2})$, is depicted below (as above, we just depict half of it):



The quotient then identifies b_i with c_i , yielding the following simplicial complex (half of it, as before):



Collating the missing half of the picture, this is a hollow triangular bipyramid, a complex with 6 two-dimensional maximal faces:

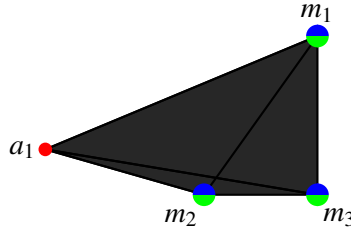
$$\begin{aligned} & \{ \{a_1, m_1, m_2\}, \{a_1, m_2, m_3\}, \{a_1, m_3, m_1\}, \\ & \{a_2, m_1, m_2\}, \{a_2, m_2, m_3\}, \{a_2, m_3, m_1\} \} \end{aligned}$$

which clearly does not satisfy the acyclicity condition of Vorob'ev's theorem. Indeed, one can find empirical models for the original measurement scenario whose 'quotient' average macroscopic behaviour is non-local. The reason for this is that we have too many measurement settings available and not enough microscopic sites (or independent copies of Bob) to dilute the information these measurements can obtain.

The situation becomes different if there is another site D (with measurements d_1, d_2, d_3) and the sites B, C and D are identified as forming the same macroscopic site (or as being three copies of Bob). The complex in this case is

$$\Sigma_{n=2, k_1=2, k_2=3, r_1=1, r_2=3} := \mathfrak{D}_2 * \mathfrak{D}_3^{*3} = \mathfrak{D}_2 * (\mathfrak{D}_3 * \mathfrak{D}_3 * \mathfrak{D}_3),$$

and its quotient is a solid, rather than hollow, triangular bipyramid (two filled tetrahedrons collated together). In keeping with the previous examples, we depict only half of the simplicial complex:



The set of maximal faces of this complex (i.e. the cover of maximal contexts of this scenario) is

$$\{ \{a_1, m_1, m_2, m_3\}, \{a_2, m_1, m_2, m_3\} \},$$

from which it is clear that the complex is acyclic, ensuring that any no-signalling model satisfies all monogamy relations, and that all average macroscopic models are local. The point we are hinting at is that, in order to guarantee monogamy and local averages, there must be at least as many particles (microscopic sites) in each macroscopic site as there are measurement settings available at that site.

5 General multipartite scenarios

We now look at multipartite scenarios in general. We consider the situation already mentioned in the item list in section 1: we have n (macroscopic) sites $1, \dots, n$ (also denoted by A, B, C, \dots); each site i has k_i measurement settings; and we have r_i copies of site i , or microscopic sites constituting the macroscopic site i . If we write A for a (macroscopic) site, then $A^{(1)}, \dots, A^{(r_A)}$ denote the several copies of it or microscopic sites constituting it, and $a_1^{(m)}, \dots, a_{k_A}^{(m)}$ are the measurements for the m -th copy or microscopic site $A^{(m)}$, where $m \in \{1, \dots, r_A\}$.

Such a situation is therefore determined by the positive integers $n, k_1, \dots, k_n, r_1, \dots, r_n$. The simplicial complex representing this scenario is

$$\Sigma_{n, \vec{k}, \vec{r}} := \mathcal{D}_{k_1}^{*r_1} * \dots * \mathcal{D}_{k_n}^{*r_n}$$

For example, as already mentioned, our main example measurement scenario, the tripartite simplicial complex $\mathcal{D}_2 * \mathcal{D}_2 * \mathcal{D}_2$ where we want to lump together the second and third sites, would be written as

$$\Sigma_{n=2, k_1=2, k_2=2, r_1=1, r_2=2} = \mathcal{D}_2 * \mathcal{D}_2^{*2}$$

Other examples were also provided in the previous section.

Now, we have a symmetry of the scenario which identifies the appropriate copies or microscopic sites that are to be lumped together. It identifies the measurements:

$$\begin{aligned} a_j^{(1)} &\sim \dots \sim a_j^{(r_A)} \quad (\forall j \in \{1, \dots, k_A\}), \\ b_j^{(1)} &\sim \dots \sim a_j^{(r_A)} \quad (\forall j \in \{1, \dots, k_A\}), \\ &\text{etc.} \end{aligned}$$

Formally, this symmetry is given by an action of the group $S_{r_1} \times \dots \times S_{r_n}$, where S_l is the symmetric group in l elements. We are interested in knowing under which conditions the quotient of $\Sigma_{n, \vec{k}, \vec{r}}$ (or rather, of its semiregularisation) by this symmetry is acyclic.

Proposition 5.1. *The quotient of the measurement scenario $\text{sr}(\Sigma_{n, \vec{k}, \vec{r}})$ by the symmetry above is acyclic iff one of the following holds:*

- (i) *each site has at least as many microscopic sites or copies as it has measurement settings, i.e.*
 $\forall_{i \in \{1, \dots, n\}}. k_i \leq r_i$;
- (ii) *one of the sites has a single copy and the condition above is satisfied by all the other sites, i.e.*
 $\exists_{i_0}. \left(r_{i_0} = 1 \wedge \forall_{i \in \{1, \dots, \widehat{i_0}, \dots, n\}}. k_i \leq r_i \right)$.

Proof. See [16] for the proof of this result. The examples of sections 4.2 and 4.3 provide some intuition. \square

The way in which this proposition splits into two cases might strike one as strange at first sight. The first case is better suited for a reading of the result in terms of macroscopic averages, whereas the second case resembles more closely the usual monogamy relations, where one deals with the correlations shared by a single party with several others. As mentioned in section 1, we can read the result of proposition 5.1 as a generalisation of the results in [15] and [14].

From the former’s perspective, suppose that we have several (a large number of) particles distributed over n sites, with r_i particles at site i . The group $S_{r_1} \times \cdots \times S_{r_n}$ captures the symmetry of the system: we can interchange any of the r_i particles within the same site i . Now assume there are k_i measurement settings available at each site i . Microscopically, we need to consider k_i possible measurements for each particle. But we consider that, macroscopically, only the averaged behaviour is accessible, with the corresponding measurements being lumped together as k_i averaged measurements. The fact that the quotient is acyclic as long as there are enough particles in each site means the following: no matter what the statistics for all the original microscopic measurements are (as long as they satisfy no-signalling), the average behaviour is classical, in the sense that it admits a local hidden variable description. This generalises the paper’s result because it holds for any no-signalling theory and not just for quantum mechanics.

Note, however, that by augmenting the number of (macroscopic) measurements that one performs, it would in principle be possible to detect non-locality on the average macroscopic correlations. However, this soon becomes impractical if one has a large (say $\approx 10^{23}$) number of particles in each site. So, it seems that the limitation on our experimental capability of performing enough measurements makes the average behaviour appear local.

From the point of view of [14], we start with an n -partite scenario with k_i measurement settings for each site i . Then the question is: fixing the first site (or any other for that matter), how many copies of the other sites do we need to consider so that the monogamy relation for the violation of any n -partite Bell-type inequality holds? (see equation 1 for the general form of such a monogamy relation.) That is, with how many copies of the other sites can Alice violate the same Bell-type inequality? The authors of the paper consider only the case $n = 2$ and show that one can take k_2 copies of the second site in order to get the monogamy relations. Our proposition above generalises this for any n .

Moreover, our proposition is a complete characterisation. Not only does it say that it suffices to take k_i copies of each site i , it also says that taking less than that is not enough. That is, if one takes less copies of some site, there is a no-signalling empirical model that violates the monogamy relations. Similarly, the interpretation in terms of locality of macroscopic averages is also an ‘if and only if’. This is another way in which our result generalises both papers.

6 Conclusions and outlook

This work explores a connection between monogamy of non-locality and the locality of average macroscopic behaviour in multipartite scenarios. We show that both can be explained by a structural property of the simplicial complex representing the compatibility of measurements in the scenario: after taking a quotient by an appropriate symmetry along which one takes the average or considers the monogamy relation, the resulting simplex should be acyclic, hence inherently local or non-contextual according to Vorob’ev’s theorem. This means, in particular, that the proof is independent of quantum mechanics and works more generally for any no-signalling theory. In the present document, we have motivated and illustrated the main ideas behind this analysis via some simple example measurement scenarios.

The language of simplicial complexes, as used in the sheaf-theoretic framework [2], allows one

to describe not only the Bell-type multipartite scenarios familiar from discussions of non-locality that we have been considering, but also more general contextuality scenarios, such as Kochen-Specker configurations [9]. In upcoming work, we develop a scheme formalising our analysis in this more general setting. The result for Bell-type scenarios stated in proposition 5.1, whose full proof will also appear there, can be seen as a first instance or application of that scheme. Future work includes applying this scheme in different kinds of scenarios to yield monogamy relations for contextuality inequalities and to study non-contextuality of macroscopic averages.

Acknowledgements

I thank Samson Abramsky, Adam Brandenburger, and Shane Mansfield for valuable guidance, discussions, and comments on several versions of this work. I also thank Miguel Navascués for some very important clarifications. Finally, I thank audiences of the seminars at Paris Diderot and ParisTech for their helpful feedback.

I gratefully acknowledge support from the Marie Curie Initial Training Network MALOA – From MAThematical LOGic to Applications, PITN-GA-2009-238381, and from FCT – Fundação para a Ciência e Tecnologia (the Portuguese Foundation for Science and Technology), PhD grant SFRH/BD/94945/2013.

References

- [1] Samson Abramsky (2013): **Relational databases and Bell’s theorem**. In Val Tannen, Limsoon Wong, Leonid Libkin, Wenfei Fan, Wang-Chiew Tan & Michael Fourman, editors: *In search of elegance in the theory and practice of computation - Essays dedicated to Peter Buneman, Lecture Notes in Computer Science* 8000, Springer Berlin Heidelberg, pp. 13–35, doi:10.1007/978-3-642-41660-6_2. Eprint available at arXiv:1208.6416 [cs.LO].
- [2] Samson Abramsky & Adam Brandenburger (2011): **The sheaf-theoretic structure of non-locality and contextuality**. *New Journal of Physics* 13(11), p. 113036, doi:10.1088/1367-2630/13/11/113036. Eprint available at arXiv:1102.0264 [quant-ph].
- [3] Samson Abramsky & Carmen Constantin (2013): **A classification of multipartite states by degree of non-locality**. In: *(Pre)Proceedings of 10th Workshop on Quantum Physics and Logic (QPL X), ICFO Barcelona*.
- [4] Samson Abramsky & Lucien Hardy (2012): **Logical Bell inequalities**. *Physical Review A* 85(6), p. 062114, doi:10.1103/PhysRevA.85.062114. Eprint available at arXiv:1203.1352 [quant-ph].
- [5] Samson Abramsky, Shane Mansfield & Rui Soares Barbosa (2012): **The Cohomology of Non-Locality and Contextuality**. *Electronic Proceedings in Theoretical Computer Science* 95, pp. 1–15, doi:10.4204/EPTCS.95.1. Eprint available at arXiv:1111.3620 [quant-ph].
- [6] Jean-Daniel Bancal, Cyril Branciard, Nicolas Brunner, Nicolas Gisin, Sandu Popescu & Christoph Simon (2008): **Testing a Bell inequality in multipair scenarios**. *Physical Review A* 78(6), p. 062110, doi:10.1103/PhysRevA.78.062110. Eprint available at arXiv:0810.0942 [quant-ph].
- [7] Jonathan Barrett, Noah Linden, Serge Massar, Stefano Pironio, Sandu Popescu & David Roberts (2005): **Nonlocal correlations as an information-theoretic resource**. *Physical Review A* 71(2), p. 022101, doi:10.1103/PhysRevA.71.022101. Eprint available at arXiv:quant-ph/0404097.
- [8] John S. Bell (1964): **On the Einstein-Podolsky-Rosen paradox**. *Physics* 1(3), pp. 195–200.
- [9] Simon Kochen & Ernst P. Specker (1967): **The problem of hidden variables in quantum mechanics**. *Journal of Mathematics and Mechanics* 17(1), pp. 59–87.

- [10] Shane Mansfield (2013): **The mathematical structure of non-locality and contextuality**. DPhil thesis, University of Oxford.
- [11] Shane Mansfield & Rui Soares Barbosa (2013): **Extendability in the sheaf-theoretic approach: Construction of Bell models from Kochen-Specker models**. In: *(Pre)Proceedings of 10th Workshop on Quantum Physics and Logic (QPL X), ICFo Barcelona*. Eprint available at arXiv:1402.4827 [quant-ph].
- [12] Miguel Navascués, Stefano Pironio & Antonio Acín (2008): **A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations**. *New Journal of Physics* 10(7), p. 073013, doi:10.1088/1367-2630/10/7/073013. Eprint available at arXiv:0803.4290 [quant-ph].
- [13] Miguel Navascués & Harald Wunderlich (2009): **A glance beyond the quantum model**. *Proceedings of the Royal Society A* 466(2115), pp. 881–890, doi:10.1098/rspa.2009.0453. Eprint available at arXiv:0907.0372 [quant-ph].
- [14] Marcin Pawłowski & Časlav Brukner (2009): **Monogamy of Bell’s inequality violations in nonsignaling theories**. *Physical Review Letters* 102(3), p. 030403, doi:10.1103/PhysRevLett.102.030403. Eprint available at arXiv:0810.1175 [quant-ph].
- [15] Ravishankar Ramanathan, Tomasz Paterek, Alastair Kay, Paweł Kurzyński & Dagomir Kaszlikowski (2011): **Local realism of macroscopic correlations**. *Physical Review Letters* 107(6), p. 060405, doi:10.1103/PhysRevLett.107.060405. Eprint available at arXiv:1010.2016 [quant-ph].
- [16] Rui Soares Barbosa: DPhil thesis, University of Oxford. Forthcoming.
- [17] Nikolai N. Vorob’ev (1962): **Consistent families of measures and their extensions**. *Theory of Probability and its Applications (Teoriya Veroyatnostei i ee Primeneniya)* 7(2), pp. 147–163 (English: N. Greenleaf, trans.), 153–159 (Russian).

Stochastic Relational Presheaves and Dynamic Logic for Contextuality

Kohei Kishida*

Department of Computer Science
University of Oxford
Oxford, United Kingdom
kohei.kishida@cs.ox.ac.uk

Presheaf models [7, 26, etc.] provide a formulation of labelled transition systems that is useful for, among other things, modelling concurrent computation. This paper aims to extend such models further to represent stochastic dynamics such as shown in quantum systems. After reviewing what presheaf models represent and what certain operations on them mean in terms of notions such as internal and external choices, composition of systems, and so on, I will show how to extend those models and ideas by combining them with ideas from other category-theoretic approaches to relational models [15] and to stochastic processes [11, 3, 17, etc.]. It turns out that my extension yields a transitional formulation of sheaf-theoretic structures that Abramsky and Brandenburger [1] proposed to characterize non-locality and contextuality. An alternative characterization of contextuality will then be given in terms of a dynamic modal logic of the models I put forward.

1 Introduction

The goal of this paper is to devise a formalism of semantic structure for dynamic logic that is suitable for expressing stochastic dynamics such as shown in quantum systems. Essential features of stochastic dynamics I aim to capture include

- the distinction and interaction between *internal* and *external choices*, that is, non-deterministic branchings that are made within a system and that are made by external agents or experimenters;
- the distinction and interaction between what is *globally* the case in an entire system and what is *locally* the case in a subsystem.

In particular, the resulting semantics and logic shall be general enough to accommodate both the presence and the absence of (typically quantum) *non-locality* and *contextuality*, but at the same time expressive enough to provide logical characterization for non-locality and contextuality.

I achieve my goal by integrating three frameworks of categorical approaches that have been proposed to modelling non-deterministic and stochastic processes. Firstly, my formalism will be based on

- (i) Presheaves as labelled transition models for concurrency (Winskel *et al.* [7, 26], etc.). I show how the presheaf structure can be used to capture notions that are essential to my goal, such as internal and external choices, composition of multipartite systems, and so on.

Then I extend this setting in two aspects, by admitting non-trees and by adding probabilities. I attain these extensions by integrating the following ideas into my formalism.

*Kishida's research has been supported by the grant FA9550-12-1-0136 of the U.S. AFOSR. Grateful acknowledgment goes to the anonymous referees for insightful comments, which improved this paper.

- (ii) Kripke relational semantics in terms of Kripke frames as functors from labels to the category **Rel** of sets and relations (Hermida [15]). Integrating this idea with the presheaf framework admits presheaf-like models as transition systems of non-tree forms. I will also lay out motivation for admitting non-trees. (One mode of this integration has already been given in Sobociński [24]; yet the mode of integration I propose in this paper is different and not equivalent.)
- (iii) The category of stochastic maps, or equivalently the Kleisli category of the distribution monad (the idea goes back at least to Lawvere [18]; it is also studied recently by Fritz [11], Baez *et al.* [3], Fong [9], etc., in the former formulation, by Jacobs [17], etc., in the latter formulation). How to add probabilities to presheaf models is a question posed in the concluding part of Varacca [25]; I answer this question by using structures closely related, though not equivalent, to stochastic maps.

These extensions give semantic structures on which I define a dynamic and probabilistic logic.

To demonstrate that the resulting semantics and logic achieve the goal mentioned above, I will show how they capture non-locality and contextuality. In particular, the semantics gives an alternative, transitional formulation to a sheaf-theoretic approach to non-locality and contextuality (Abramsky and Brandenburger [1], etc.). This approach provides a sheaf-theoretic expression for, among other things, measurement scenarios in quantum mechanics, and characterizes non-locality and contextuality found in such scenarios in terms of non-existence of global sections. The transitional formulation I give to this approach leads to an alternative, dynamic-logical characterization of non-locality and contextuality.

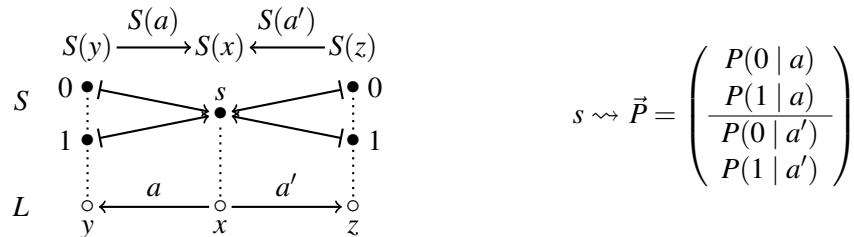
2 Presheaf Models for Measurements

This section reviews presheaves over trees as labelled transition systems (see [7, 26]). Rather than giving new definitions or theorems, this section is concerned with conceptually laying out how to use the familiar notions of presheaf and fibration to represent features of non-deterministic processes that are essential to the goal of this paper.

2.1 Trees and Presheaves of Non-Deterministic Choices

Here I lay out the key idea of how to use a presheaf over a tree as a labelled transition system, or LTS for short, in a manner suitable for representing different kinds of non-determinacy in stochastic processes.

As in the standard terminology, by a “measurement scenario of (n, k, ℓ) -type” let us mean a Bell-type scenario of (typically quantum) measurements that involves n parts (or experimenters), each of which (or whom) chooses one from k measurements, each of which has ℓ outcomes. For instance, in a $(1, 2, 2)$ scenario, Alice chooses one from two measurements, a and a' , each of which has two outcomes, 0 and 1. This simple scenario can be represented by the following tree L and presheaf S over L .



The binary branching in L represents the choice Alice makes outside the system, choosing from two measurements a and a' . Then regard S as a transition system, reading “ \mapsto ” backward as transition “ \leftarrow ”;

each such edge of transition in S is labelled with an edge in L —for instance, those in $S(a)$ above are labelled with a , representing possible outcomes the system has for Alice’s choice of a . So the binary branching in $S(a)$ represents the system having two outcomes for measurement a . One of our objectives is to assign probabilities to such branchings, so that, in the picture above for instance, the state s can be (at least partially) specified by the vector of probabilities \vec{P} to the right of the picture above.

Note that the representation just given involves two kinds of choice. Put in general terms, when we describe a system and agents external to the system,

- The agents may be able to choose from different ways to interfere or interact with the system. We call these choices *external choices*, and represent them with branching in the base tree.
- The system may behave by itself non-deterministically—sometimes in response to external choices, but sometimes simply as time passes—with several possible outcomes. We call these choices *internal choices*, and represent them with branching in function components of the presheaf.

In short, external choice resides in the base tree L ; internal choice resides in (function components of) the presheaf S . This is the slogan for our use of presheaves S over trees L as L -labelled transition systems.

In fact, not just the distinction between internal and external choices, the presheaf structure also gives us several useful ways to control descriptions of these choices—for instance, to shift the boundary between the internal and external. We will see this in subsection 2.2. Before doing so, it is useful to observe that the presheaves over a tree are equivalent to the *fibrations* over the tree (which should be quite obvious from the picture above). Let us recall

Definition 1. A bundle (i.e., monotone map) $\pi : S \rightarrow L$ of posets is called a *fibration* (over L) if, whenever $x \leq_L \pi(t)$, there is a unique $s \in \pi^{-1}(x)$ such that $s \leq_S t$. Write **Fib** for the category of posets and fibrations.

We should note that if $\pi : S \rightarrow L$ is a fibration and L is a tree then S is also a tree. Then it is easy to show the following (we provide a proof rather as a review of notation).

Fact 1. $\mathbf{Sets}^{L^{\text{op}}} \simeq \mathbf{Fib}/L$ for any poset L .

Proof. A presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$ yields a fibration with the projection $\pi : S \rightarrow L$ from the dependent sum $S := \sum_{x \in L} S(x) = \{ (x, s) \mid x \in L \text{ and } s \in S(x) \}$ to L and the order \leq_S on S such that $(x, s) \leq_S (y, t)$ iff $x \leq_L y$ and $s = S(x, y)(t)$. A fibration $\pi : S \rightarrow L$ yields a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$ by letting $S(x) = \pi^{-1}(x)$ for $x \in L$ and, whenever $x \leq_L y$, defining $S(x, y) : \pi^{-1}(y) \rightarrow \pi^{-1}(x)$ so that $S(x, y)(t)$ for $t \in \pi^{-1}(y)$ is the unique $s \in \pi^{-1}(x)$ such that $s \leq_S t$.

Given presheaves $S, T : L^{\text{op}} \rightarrow \mathbf{Sets}$ and corresponding fibrations $\pi_S : S \rightarrow L, \pi_T : T \rightarrow L$, the natural transformations from S to T are just the monotone maps $f : S \rightarrow T$ over L (meaning $\pi_T \circ f = \pi_S$), but any such monotone map f can easily be shown to be a fibration. \square

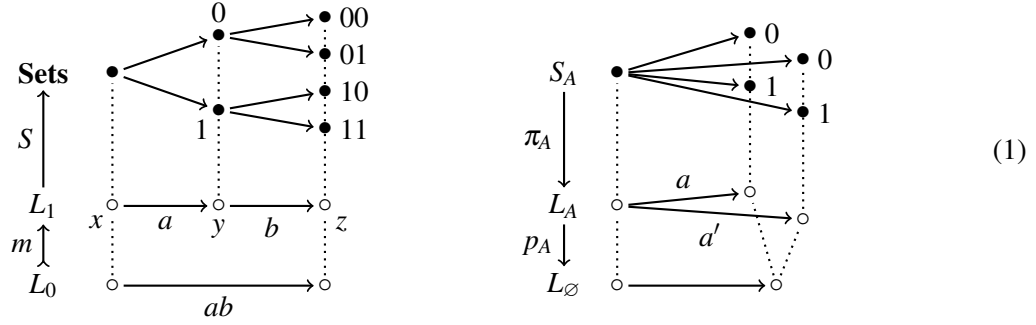
We will invoke this presheaf-fibration equivalence extensively in the rest of this paper.

2.2 Controlling System Descriptions

Given presheaf-fibration descriptions of non-deterministic processes with internal and external choices, we can take further advantage of operations on the presheaf-fibration structure to control the descriptions.

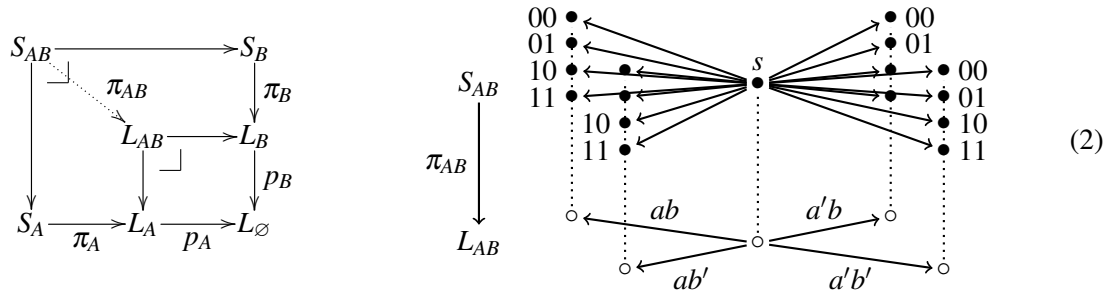
A family of operations that will later prove useful is done by change of base. One such operation is to precompose a given presheaf $S : L_1^{\text{op}} \rightarrow \mathbf{Sets}$ with an embedding $m : L_0 \hookrightarrow L_1$, obtaining a new presheaf $S \circ m^{\text{op}} : L_0^{\text{op}} \rightarrow \mathbf{Sets}$. Since some points, or “stages”, of L_1 are “omitted” in L_0 , the precomposition makes the model “forget” what takes place at these omitted stages. For instance, take $m : L_0 \hookrightarrow L_1$ as on the left of (1) below, and let a and b represent measurements by Alice and by Bob. Then a presheaf

$S : L_1^{\text{op}} \rightarrow \mathbf{Sets}$ carries information as to the original states (in $S(x)$), the possible outcomes of a (in $S(y)$), and then the possible further outcomes of b (in $S(z)$). In contrast, the presheaf $S \circ m^{\text{op}} : L_0^{\text{op}} \rightarrow \mathbf{Sets}$ carries the same information as to the original states (in $S(x)$) and the outcomes of both measurements (in $S(z)$), but it has no information as to the process in between (or, indeed, even as to whether a is performed before, after, or at the same time as b).



Another is to compose fibrations $\pi : S \rightarrow L_0$ and $p : L_0 \rightarrow L_1$, obtaining a new fibration $p \circ \pi : S \rightarrow L_1$. In π , branchings in L_0 represent external choices, but some of them are internal choices in p ; so the composition “internalizes” these external choices. Take π_A and p_A as on the right of (1) above. π_A describes Alice as an agent external to a system who externally chooses from measurements a and a' . On the other hand, $p_A \circ \pi_A$ describes a bigger system encompassing Alice—so that we simply watch the bigger system internally choose from the four outcomes, “Alice performs a and gets outcome 0”, etc.

In fact, such composition of fibrations can be used to compose descriptions of several systems into a description of a multipartite system. The fibration π_A in the picture above describes a $(1, 2, 2)$ -scenario for Alice. Take an isomorphic $\pi_B : S_B \rightarrow L_B$ to describe a $(1, 2, 2)$ -scenario for Bob. Then a fibration $\pi_{AB} : S_{AB} \rightarrow L_{AB}$ for the composed $(2, 2, 2)$ -scenario is obtained as follows:



That is, $\pi_{AB} = \pi_A \times_{L_0} \pi_B : S_A \times_{L_0} S_B \rightarrow L_A \times_{L_0} L_B$. Put more conceptually, we use L_0 as a clock for synchronizing events in Alice’s scenario and ones in Bob’s, and then take simultaneous pairs of events from Alice’s and Bob’s scenarios. We should note that the pair of projections from S_{AB} and L_{AB} to S_A and L_A represents the restriction of a description of what is globally the case in the bipartite system to a description of what is locally the case in Alice’s subsystem—this is a tool crucial for the purpose of this paper, of capturing non-locality and contextuality. We will see, for instance, that this projection has a role in characterizing the no-signalling property in fibrational terms in section 3.

It may need stressing that S_{AB} described above is just a cartesian product (taken fiberwise over L_0)—rather than anything similar to a tensor product—of S_A and S_B ; hence it does not by itself express any correlation between Alice’s and Bob’s measurement outcomes. It is rather a transition-system expression for the 4×4 entries in a probability table describing a $(2, 2, 2)$ -scenario. Any correlation will be expressed by assigning probabilities to transitions in S_{AB} ; we will see how in section 3.

3 Adding Probabilities to Presheaves

This short section lays out how to add probabilities to the presheaf representation of non-deterministic processes given in section 2. The definitions provided here will later be generalized in subsection 4.2, after a generalization of the presheaf representation is proposed in subsection 4.1.

3.1 Stochastic Presheaves

Recall that in a description of a non-deterministic process with a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$, for any edge $e = (x, y)$ of L and state $s \in S(x)$, the inverse image $S(e)^{-1}(s) \subseteq S(y)$ is the set of states to which the system may internally choose to transition from s when e is externally chosen. Now we want to give probability to such an internal choice; so let us achieve just that, with the following series of definitions. They use the notion of R -distribution for a commutative semiring R ; see [1, §2.3] for its definition. In particular, throughout this paper all distributions are assumed to be normalized and with finite support.

Definition 2. Fix a commutative semiring R . Then we define an R -map as any surjection $f : Y \twoheadrightarrow X$ equipped with, for each $s \in X$, an R -distribution d_s^f on $f^{-1}(s) \subseteq Y$. (We say that an R -map is *on* its underlying surjection.)

Obviously, we can achieve what we wanted above with an $\mathbb{R}_{\geq 0}$ -map f on $S(e) : S(y) \rightarrow S(x)$ (assuming $S(e)$ is surjective): The distribution d_s^f assigns to each $t \in S(e)^{-1}(s)$ the probability $d_s^f(t)$ with which the system transitions from s to t (when e is chosen). To do this for the entire presheaf, we give

Definition 3. Given two R -maps $f : Z \twoheadrightarrow Y$ and $g : Y \twoheadrightarrow X$, let their composition $g \circ f : Z \twoheadrightarrow X$ have, for each $s \in X$, an R -distribution $d_s^{g \circ f}$ on $f^{-1}(g^{-1}(s)) \subseteq Z$ such that

$$d_s^{g \circ f}(u) = d_s^g(f(u)) \cdot d_{f(u)}^f(u). \quad (3)$$

Write $R\text{-}\mathbf{Map}$ for the category of sets and R -maps. (Clearly, the unique R -map on the identity map $1_X : X \rightarrow X$ is the identity on X in $R\text{-}\mathbf{Map}$.)

The point of (3) should be clear: When $s = g(t)$ and $t = f(u)$, the system transitions from s to t with probability $r = d_s^g(t)$ and from t to u with probability $r' = d_t^f(u)$; so it transitions from s to u with probability $r \cdot r' = d_s^g(t) \cdot d_t^f(u) = d_s^{g \circ f}(u)$. (Note that the system can go from s to u through at most one t , since f is a function.) Then, finally,

Definition 4. An R -presheaf over a category \mathbf{C} is a contravariant functor from \mathbf{C} to $R\text{-}\mathbf{Map}$. (We say that an R -presheaf is *on* its underlying presheaf.)

So, given a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$ over a tree L as an L -LTS, we assign probabilities to the internal choices in S by simply taking an R -presheaf on S .

The presheaf-fibration equivalence (Fact 1) partially extends to R -presheaves: We can define “ R -fibrations” and prove that the equivalence extends to an essentially surjective and full functor from the category of rooted R -presheaves over a rooted tree L to that of rooted R -fibrations over L (we however omit the definitions and proof in this abstract). This extended version is limited and no longer an equivalence, but good enough for practical purposes. The core idea is that, given an R -presheaf $S : L^{\text{op}} \rightarrow \mathbf{Rel}$ that has a root $s \in S(x)$, the “horizontal” assignment of probabilities $d_s^{S(x,y)}(t)$ to all states $t \in S$ can be turned into a “vertical” assignment of probabilities $d_y^\pi(t)$ on the fibration $\pi : S \rightarrow L$ that corresponds to the underlying presheaf of S .

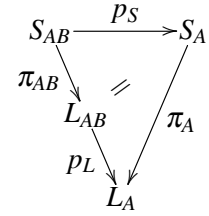
Lastly, note that, though it may be proper to reserve the term “probability” to values of $\mathbb{R}_{\geq 0}$ -distributions, in this paper I apply the term broadly to values of R -distributions in general. Other interesting cases of R include \mathbb{B} , the booleans, and \mathbb{R} , all the reals, both of which are discussed in [1].

3.2 Example: No-Signalling

Let us say that a commutative semiring R is “normalizable” if, for every family $\{r_i\}_{i \in I}$ of elements of R such that $c := \sum_{i \in I} r_i \neq 0$, there is a family $\{r'_i\}_{i \in I}$ of elements of R such that $r'_j \cdot c = r_j$ for each $j \in I$ and $\sum_{i \in I} r'_i = 1$. For instance, $\mathbb{R}_{\geq 0}$ is normalizable. Now, in $R\text{-}\mathbf{Map}$ for normalizable R , we have the following fact (we omit a proof in this abstract).

Fact 2. Suppose R is normalizable. Then a factorization $h = g \circ f : Z \twoheadrightarrow Y \twoheadrightarrow X$ of surjections yields a surjection ϕ from the R -maps on h to the R -maps on g such that, for any R -map h^R on h , the R -map $\phi(h^R)$ on g has $d_s^{\phi(h^R)}(t) = \sum_{u \in f^{-1}(t)} d_s^{h^R}(u)$ for all $s \in X$ and $t \in Y$; or, in other words, $\phi(h^R)$ is the *marginal* of h^R along the identification of states $u \in Z$ via the quotient map $f : Z \twoheadrightarrow Y$.

Let us apply this fact to the diagram in (2), writing $p_S : S_{AB} \twoheadrightarrow S_A$ and $p_L : L_{AB} \twoheadrightarrow L_A$ for the pair of projections. Take $h = p_L \circ \pi_{AB}$ and $g = \pi_A$, with $f = p_S$. Then $\phi(h^R)$ (on π_A) is the marginal of h^R (on $p_L \circ \pi_{AB}$) along the restriction of description from the bipartite to Alice’s system. Note that, however, this involves probabilities on p_L , that is, with which Bob chooses from measurements b and b' . Different probabilities on p_L may lead to different $\phi(h^R)$ —or maybe not, if the probabilities on π_{AB} satisfy the no-signalling property. Thus we have



Theorem 1. An R -presheaf π_{AB}^R on the presheaf π_{AB} for a multipartite system satisfies no-signalling iff, for each pair of projections S_L and p_L , $\phi(p_L^R \circ \pi_{AB}^R)$ is the same regardless of the choice of R -map p_L^R on p_L .

4 Stochastic Relational Presheaves

In section 2 we saw how presheaves over trees—which are themselves trees—can be used as LTSs; and in section 3 we saw how to add probabilities to such systems. Generalizing this, this section obtains similarly labelled transition systems with probabilities that are however not trees.

4.1 Relational Presheaves

We first show how to implement LTSs of a non-tree shape using a presheaf-like structure. The core idea in using presheaves as LTSs was the following, functorial one: Let a tree L represent a series of external choices; assign to each stage in L the set of possible states at that stage; and connect states from different stages with internally chosen transitions. This idea involves no intrinsic reason why this connection of transitions should be (reverse) functional, i.e., why the functor we take should be a presheaf.

In fact, here is a reason the functor we take should *not* always be a presheaf. Consider the following two objectives, each of which may, conceivably, be well motivated.

- (i) For our functor S from the tree L , we may like to take, as values $S(x)$ for stages $x \in L$, the sets of states in Hilbert spaces instead of just any sets, to express quantum processes straightforwardly.
- (ii) We may consider a non-deterministic process that involves both branching and colliding (so cannot be a tree, forward or backward). In fact, when we do a quantum measurement a in one basis and then another a' in another basis, the system may transition from a state s to t_0 (after a) to u (after a'), but may also transition from s to $t_1 \neq t_0$ (after a) to the same u (after a').

The use of a presheaf, and in particular of functions $S(e)$ for edges e of L —which forces the transition system to be a tree—cannot accommodate both (i) and (ii). To accommodate a non-tree as in (ii) in a

tree formalism, it is a standard technique to “unfold” or “unravel” the non-tree into a tree, duplicating the single state u to u_0 following t_0 and u_1 following t_1 . This, however, does not go well with (i), since the set $S(x)$ encompassing u_0, u_1 , and all the required duplicates may have to be much more complicated than just the set of states of a Hilbert space. This is why we should at least sometimes let $S(e)$, for edges e of L , be relations in general rather than functions. Then, in (ii), the state s can be connected to both t_0 and t_1 while both t_0 and t_1 connected to u .

So, instead of the category **Sets** of sets and functions, we take the category **Rel** of sets and relations as the codomain of our functors (see [6] and [10, esp. Ch. II] for categorical characterizations of **Rel** and its generalizations). For the sake of notation, let us enter

Definition 5. **Rel** is the category of sets and relations. Its objects are sets, and its arrows from a set X to another Y are relations $f \subseteq X \times Y$, written $f : X \rightrightarrows Y$ as well. We write $s \xrightarrow{f} t$ instead of $(s, t) \in f$, and, identifying $f : X \rightrightarrows Y$ with $f : X \rightarrow \mathcal{P}(Y)$, sometimes write $f(s) = \{t \in Y \mid s \xrightarrow{f} t\}$. The composition $g \circ f : X \rightrightarrows Z$ of $f : X \rightrightarrows Y$ and $g : Y \rightrightarrows Z$ is defined so that $s \xrightarrow{g \circ f} u$ iff $s \xrightarrow{f} t \xrightarrow{g} u$ for some $t \in Y$.

Rel is a dagger compact category. Firstly, it has a \dagger structure: Any $f : X \rightrightarrows Y$ has a unique opposite relation $f^\dagger : Y \rightrightarrows X$, so that $s \xrightarrow{f^\dagger} t$ iff $t \xrightarrow{f} s$. Also, even though the cartesian product is no longer the product in the categorical sense in **Rel**, it is still a monoidal product \otimes . In addition, the identification of $f : X \rightrightarrows Y$ with $f : X \rightarrow \mathcal{P}(Y)$ is just one aspect of the fact that **Rel** is the Kleisli category $\mathbf{Kl}(\mathcal{P})$ of the powerset monad \mathcal{P} on **Sets**. Now, let us finally provide

Definition 6. A *relational presheaf* over a category **C** is a covariant functor from **C** to **Rel**.¹

So we generalize presheaves with relational presheaves as our LTSs. We must note that relational presheaves are covariant and not contravariant. Thus, given an edge $e = (x, y)$ of a tree L , the system’s transition from states at stage x to ones at stage y is represented by a relation $S(e) : S(x) \rightrightarrows S(y)$ in a relational presheaf $S : L \rightarrow \mathbf{Rel}$, whereas by a function $S(e) : S(y) \rightarrow S(x)$ in a presheaf $S : L^{\text{op}} \rightarrow \mathbf{Sets}$.

It may be worth noting that, although relational presheaves over a tree of labels are themselves LTSs, they are also a generalization of the ordinary kind of LTSs in the following sense. As Hermida [15] observes, given a set L of labels, the (ordinary) transition systems labelled by L are, in our terminology, the relational presheaves over the free monoid L^* generated by L . Our notion of relational presheaf as a LTS generalizes this by replacing L^* —a tree in which every (type of) edge is followed by every other (type of) edge—with a general tree, and permitting different stages to have different sets of states.

It is also worth noting that a small part of the presheaf-fibration equivalence (Fact 1) applies to relational presheaves, as relational presheaves over a tree L can be regarded as “open” bundles over L : The equivalence extends to an essentially surjective and faithful functor from the category of rooted and open bundles over a rooted tree L to that of rooted relational presheaves over L . (Again, we omit the definitions and proof in this abstract.)

4.2 Adding Probabilities to Relational Presheaves

We added probabilities to presheaves as LTSs in section 3. In this subsection, we add probabilities to relational presheaves, which we introduced in subsection 4.1. This can be done by simply replacing the functional elements of the definitions in section 3 with relational elements. (We should recall that, in the generalization given in subsection 4.1, a relation $f : X \rightrightarrows Y$ generalizes a function $f : Y \rightarrow X$ of the opposite direction.)

¹Rosenthal [23] defines a relational presheaf as a “lax” functor; Sobociński [24] follows this “lax” definition in his account of relational presheaves as LTSs. In contrast, I define a relational presheaf “strongly”.

Definition 7. We define an *R-relation* as an “entire” relation $f : X \rightarrowtail Y$ (i.e., such that each $s \in X$ has some $t \in Y$ with $s \xrightarrow{f} t$) equipped with, for each $s \in X$, an *R-distribution* d_s^f on Y with support

$$\text{supp}(d_s^f) \subseteq f(s). \quad (4)$$

(We say that an *R-relation* is *on* its underlying relation.) Given two *R-relations* $f : X \rightarrowtail Y$ and $g : Y \rightarrowtail Z$, let their composition $g \circ f : X \rightarrowtail Z$ have, for each $s \in X$, an *R-distribution* $d_s^{g \circ f}$ on Z such that

$$d_s^{g \circ f}(u) = \sum_{t \in Y} d_s^f(t) \cdot d_t^g(u). \quad (5)$$

Write **R-Rel** for the category of sets and *R-relations*. (It should be clear that the unique *R-relation* on the identity relation $1_X : X \rightarrowtail X$ is the identity on X in **R-Rel**.)

This notion of *R-relation* is closely related to that of *stochastic map*. We discuss this relationship in subsection 4.3; it will be significant to the discussion that (4) has “ \subseteq ” as opposed to “ $=$ ”.

Let us compare the equation (5) with the one (3) for *R-maps*. For *R-maps* $f : Z \rightarrow Y$ and $g : Y \rightarrow X$, there is at most one state $t \in Y$ through which the system may transition from a given $s \in X$ to a given $u \in Z$; so the probability of the transition from s to u is just the probability of this particular path, given by the product of the two transitions, from s to t and from t to u . In contrast, for *R-relations* $f : X \rightarrowtail Y$ and $g : Y \rightarrowtail Z$, there can be many paths through which the system may transition from $s \in X$ to $u \in Z$; yet, since these paths are mutually exclusive, we can just sum their probabilities up to obtain the probability of the transition from s to u . Lastly, enter

Definition 8. An *R-relational presheaf* over a category **C** is a covariant functor from **C** to **R-Rel**. (We say that an *R-relational presheaf* is *on* its underlying relational presheaf.)

This definition provides a structure that integrates the three frameworks (i)–(iii) mentioned in Introduction: An *R-relational presheaf* $S : L \rightarrow \mathbf{R-Rel}$ over a tree L forms an L -LTS in which internal choices take place with probabilities and possibly in a non-tree fashion.

Example 1. Let a tree L represent a branching family of series of quantum measurements, gates, and other operations that can be performed. Then, for stages $x \in L$, let $S(x)$ be sets of states in (possibly, though not necessarily, identical) Hilbert spaces, and, for each edge $e = (x, y)$ of L , let $S(e) : S(x) \rightarrowtail S(y)$ be the $\mathbb{R}_{\geq 0}$ -relation that models the operation e in Hilbert-space terms, such as projections (branching with probabilities) to the suitable measurement basis. If L is moreover a free monoid and $S(x)$ are all identical (as in Hermida’s [15] formulation of transition systems mentioned in subsection 4.1), models amount essentially to ones given in Baltag and Smets [4].

This example gives a straightforward representation of quantum protocols. So it is not surprising at all that we can find non-local or contextual behaviors in such representations. Yet, using more general values than Hilbert spaces, *R-relational presheaves* can model not only the presence but also the absence of non-locality and contextuality, and indeed characterize contextuality, as we will see in section 5.

4.3 Relation to Other Work and Other Formulations

The notion of *R-relation* is closely related to that of *stochastic map*, or equivalently to Kleisli maps of the *distribution monad*.² A stochastic map from a set X to another Y is an X -indexed family of $\mathbb{R}_{\geq 0}$ -distributions on Y , with the composition defined exactly by (5). This can also be rewritten using

²I thank an anonymous referee for his/her comments regarding the relation between **R-Rel** and $\mathbf{Kl}(\mathcal{D}_{\mathbb{R}_{\geq 0}})$, which prompted me to write this subsection as a reply.

Definition 9. Given any set X , write $\mathcal{D}_R(X)$ for the set of R -distributions on X . This gives rise to the R -distribution functor $\mathcal{D}_R : \mathbf{Sets} \rightarrow \mathbf{Sets}$ (see [16] as well as [1, §2.3]), which is in fact a monad on \mathbf{Sets} (see [16]).

Then the stochastic maps f from a set X to another Y are exactly the functions $f : X \rightarrow \mathcal{D}_{\mathbb{R}_{\geq 0}}(Y)$, the Kleisli maps of $\mathcal{D}_{\mathbb{R}_{\geq 0}}$. Moreover the Kleisli composition amounts to (5), and so the category **Stoch** of sets and stochastic maps is the Kleisli category $\mathbf{Kl}(\mathcal{D}_{\mathbb{R}_{\geq 0}})$ of $\mathcal{D}_{\mathbb{R}_{\geq 0}}$ (see [17, §2]).

This is closely related to **R -Rel**, but not exactly the same (aside from R generalizing $\mathbb{R}_{\geq 0}$): In short, an $\mathbb{R}_{\geq 0}$ -relation on a relation f is a stochastic map with an extra piece of information, namely, the underlying relation f . To express this formally, consider the following subfunctor of $\times \langle \mathcal{P}, \mathcal{D}_R \rangle : \mathbf{Sets} \rightarrow \mathbf{Sets}$.

$$\mathcal{P}\mathcal{D}_R : X \mapsto \sum_{S \in \mathcal{P}(X)} \mathcal{D}_R(S) = \{ (S, d) \in \mathcal{P}(X) \times \mathcal{D}_R(X) \mid d \in \mathcal{D}_R(S) \}.$$

(We identify $d \in \mathcal{D}_R(X)$ and $d \in \mathcal{D}_R(S)$, as long as $\text{supp}(d) \subseteq S, X$.) Then the R -relations f from a set X to another Y are exactly the functions $f : X \rightarrow \mathcal{P}\mathcal{D}_R(Y)$, with a $\mathcal{P}(Y)$ component. The two sets $\mathcal{P}\mathcal{D}_R(Y)$ and $\mathcal{D}_R(Y)$ are related by the projection $p : (S, d) \mapsto d$ and a section $s : d \mapsto (\text{supp}(d), d)$, but $s \circ p \neq 1$ since we have “ \subseteq ” as opposed to “ $=$ ” in (4). Thus an R -relation $f : X \rightarrow \mathcal{P}\mathcal{D}_R(Y)$ carries properly more information, of the underlying relation, than a stochastic map $f : X \rightarrow \mathcal{D}_R(Y)$. More categorically put, postcomposing p and s with Kleisli maps gives a retraction and a section of categories so that

Fact 3. **Stoch** = $\mathbf{Kl}(\mathcal{D}_{\mathbb{R}_{\geq 0}})$ is a retract of **$\mathbb{R}_{\geq 0}$ -Rel**, but the retraction is not faithful.

The extra piece of information may appear redundant, as long as we are concerned with probabilities of transitions; yet that piece of information sometimes proves useful. In such a model as in (1) or (2), the underlying relational presheaf S describes the “logical” constraint of which states can be “logically” connected to which states; for instance, on the left of (1), state 00 can follow 0 but cannot 1. When we add the “physical” information of probabilities to S by taking an R -relational presheaf on S , the “logical” information is sometimes entailed by supports, but not always so: If the edge connecting states 0 and 00 in (1) has probability 0, then the support cannot tell us whether state 00 can “logically” follow state 0 or 1. It is useful to retain the “logical” constraint so as to consider a family of physical models satisfying it, as opposed to just one model—it is as useful as having a table of 4×4 entries that accommodates a family of probability assignments to outcomes in a $(2, 2, 2)$ -scenario. And for this purpose we need to retain the underlying relations, hence using **R -Rel** as opposed to **Stoch**.

Lastly, it may be useful to note that, as one may have expected,

Fact 4. $\mathcal{P}\mathcal{D}_R$ is a monad on **Sets**, and **R -Rel** is its Kleisli category $\mathbf{Kl}(\mathcal{P}\mathcal{D}_R)$.

(We omit a proof.) This puts **R -Rel** in the tradition [18, 12, 20, 21, 8, 17, etc.] of using algebras for monads to represent stochastic relations.

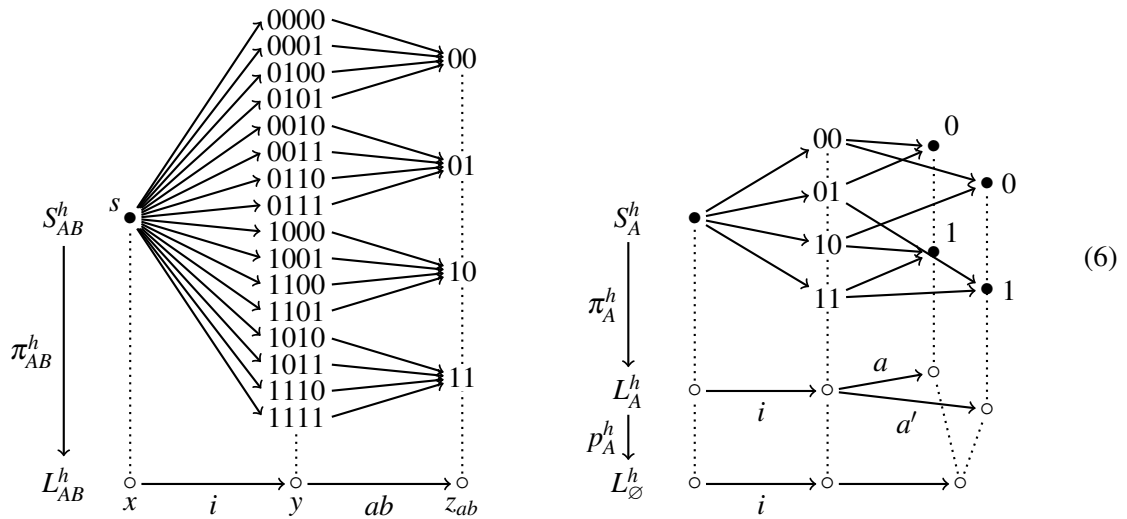
5 Dynamic Logic for Contextuality

So far we have laid out R -presheaves and R -relational presheaves as labelled and stochastic transition systems. Now we demonstrate that these models are good enough for representing essential features of stochastic dynamics such as shown in quantum systems, by showing that they can characterize non-locality and contextuality; in fact, the dynamic logic of those transition systems is expressive enough to express this characterization in logical terms.

5.1 Deterministic Hidden-Variable Models

In their sheaf-theoretic approach to non-locality and contextuality, Abramsky and Brandenburger [1] provided a characterization of non-locality and contextuality in terms of “global sections” of certain presheaves; see [1, esp. §3 and §8]. We can “translate” this characterization into our setting of stochastic relational presheaves as follows.

Suppose we have an $\mathbb{R}_{\geq 0}$ -presheaf representing an “empirical model” for a (n, k, ℓ) -scenario that satisfies no-signalling in the sense of subsection 3.2. (The characterization given in [1] is more general than just about (n, k, ℓ) -scenarios, though I only take (n, k, ℓ) -scenarios here. We can translate the characterization in full generality, but omit it in this abstract.) As an example, let us take an $\mathbb{R}_{\geq 0}$ -presheaf E on the presheaf S_{AB} in (2) (and assume no-signalling). Then E is realized by a (factorizable) hidden-variable model if and only if it has a “global section” (Theorem 8.1 of [1])—meaning, in our terms, that there exists an $\mathbb{R}_{\geq 0}$ -relational presheaf H on the relational presheaf S_{AB}^h in



(complete the picture by adding edges ab' , $a'b$, and $a'b'$ to L_{AB}^h) from which E is obtained by forgetting the middle stage y with the change-of-base operation as on the left of (1), that is, $E = H \circ m^{\text{op}}$ for the embedding $m : L_{AB} \hookrightarrow L_{AB}^h$ that omits y . Here $E = H \circ m^{\text{op}}$ means that $E(ab) = H(ab \circ i) = H(ab) \circ H(i)$, and hence that, by (5),

$$d_s^{E(ab)}(u) = \sum_{t \in H(y)} d_s^{H(i)}(t) \cdot d_t^{H(ab)}(u). \quad (7)$$

This is exactly to “reproduce the empirically observed probabilities $[d_s^{E(ab)}]$ by averaging over the hidden variables $[t \in H(y)]$ with respect to the distribution $[d_s^{H(i)}]$ ” ([1], p. 11).

From this characterization, the following features of H should be obvious: The set $H(y)$, which is forgotten in E , is a set of *hidden variables*; moreover, they are *deterministic* hidden variables, as $H(ab)$ is an $\mathbb{R}_{\geq 0}$ -relation on a *function*, as opposed to just any relation, from $H(y)$ to $H(z_{ab}) = E(z_{ab})$. Thus, the contextuality in a labelled and stochastic transition system E amounts to the failure of E to have such a deterministic hidden-variable model H . A little more formally,

Theorem 2. *For an empirical model E (in the sense of [1]), the following are equivalent.*

- (i) *E has a realization by a factorizable hidden-variable model.*

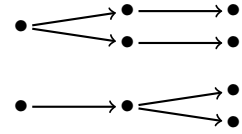
(ii) E has a global section.

(iii) The $\mathbb{R}_{\geq 0}$ -relational presheaf for E is obtained by forgetting the middle stage of a deterministic hidden-variable model.

Proof. “(i) iff (ii)” is Theorem 8.1 of [1]. “(ii) iff (iii)” is essentially due to the fact that the equation for “averaging over” in [1] (p. 11) is identical to (7). \square

Note that the underlying relational presheaf S_{AB}^h of H (or any general ones for (n, k, ℓ) -scenarios) is not provided *ad hoc*, but canonically obtained, in the manner of (2), as the fibered cartesian product $S_A^h \otimes_{L_\phi^h} S_B^h$ of the obvious hidden-variable models $S_A^h : L_A^h \rightarrow \mathbf{Rel}$ for Alice (as in (6) above) and for Bob.

To extract an essential idea from the discussion so far, contextuality means in transitional terms that a model is inconsistent with the first shape of branching to the right (in which the system internally chooses from hidden variables before external choices are made), but has to have the second shape (in which the system internally chooses outcomes when external choices are made). And the distinction between these two shapes is one of the things modal logic is good at. Thus we carry on to consider the modal logic of our labelled and stochastic transition systems.



5.2 Dynamic Logic of Stochastic Relational Presheaf Models

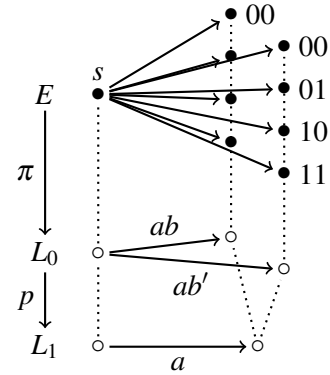
We lay out here how to use R -relational presheaves as a semantic structure for modal, dynamic logic. It turns out that the logic it gives rise to is expressive enough to capture in logical terms the characterization of contextuality we saw in subsection 5.1. (See [14] for general exposition of dynamic logic. A modal logic of stochastic relations expressed by algebras for a monad is also found in [8].)

Let us fix some (propositional) language; for our purposes it needs to have \wedge and \neg . Then we fix a set of labels of transition; for instance, we use labels a, b, ab, ab' , etc., for a $(2, 2, 2)$ -scenario. For each such label e , we add “dynamic modalities” $[e]$ and $\langle e \rangle$ to the language; we may also like to use probability modalities $P(- | e) \geq r$ for reals r . Since we take the base logic to be classical, $\langle e \rangle$ can be defined as $\neg[e] \neg$, and \top, \vee , biconditional \leftrightarrow and exclusive disjunction \oplus can be defined as usual. So we put

$$\varphi ::= p \mid \varphi \wedge \varphi \mid \neg \varphi \mid [e] \varphi \mid P(\varphi | e) > r \mid P(\varphi | e) = r \mid P(\varphi | e) < r$$

for (a fixed set of) propositional letters p and the labels e .

For this modal language, $\mathbb{R}_{\geq 0}$ -relational presheaves provide models. Firstly, the labels need interpreting in trees of labels. We take a family of trees—but not necessarily a single tree—and fibrations among them with an initial tree L_0 , so that each label e is an edge of one of the trees. For instance, the $(2, 2, 2)$ -scenario of Alice and Bob described in (2) has four trees L_- of labels and fibrations among them, with L_{AB} initial; the picture to the right describes $L_0 = L_{AB}$, $L_1 = L_A$, and the fibration $p : L_{AB} \rightarrow L_A$. Labels ab and ab' lie in L_0 , which takes both Alice and Bob as external to the system; a lies in L_1 , which takes Alice as external but Bob as internal.



Then we take an R -relational presheaf $\pi : E \rightarrow L_0$ over the initial tree L_0 ; to keep describing the $(2, 2, 2)$ -scenario, let us take E on S_{AB} in (2). Now, finally, we can provide interpretations $\llbracket \varphi \rrbracket$ for sentences φ of the language above by first assigning subsets $\llbracket p \rrbracket$ to propositional letters p and by then extending $\llbracket - \rrbracket$ recursively. We use the classical clauses for the Boolean connectives.

The new clauses of recursion for $\llbracket - \rrbracket$ concern $[e] \varphi$ and $P(\varphi | e) \gtrless r$. Since e may not lie in the initial tree L_0 , let $p : L_0 \rightarrow L_1$ be the fibration to the tree L_1 in which e lies. Then we express the ideas

- $[a] \varphi$ means that φ will be the case when Alice chooses a , regardless of which of b and b' Bob may choose;
- $P(\varphi | a) \gtrless r$ means that the probability with which φ will be the case when Alice chooses a is greater than (equal to, or less than) r , regardless of which of b and b' Bob may choose;

with the following clauses:

- $s \in \llbracket [e] \varphi \rrbracket$ iff $\text{supp}(d_s^{E(e')}) \subseteq \llbracket \varphi \rrbracket$ for all $e' = (\pi(s), x) \in p^{-1}(e)$;
- $s \in \llbracket P(\varphi | e) \gtrless r \rrbracket$ iff $\sum_{t \in \llbracket \varphi \rrbracket \cap S(x)} d_s^{E(e')}(t) \gtrless r$ for all $e' = (\pi(s), x) \in p^{-1}(e)$.

It is worth noting that s may fail to be in any of $\llbracket P(\varphi | e) > r \rrbracket$, $\llbracket \dots = r \rrbracket$ and $\llbracket \dots < r \rrbracket$, when no-signalling fails (this is why the three sentences cannot define each other). On the other hand, $s \in \llbracket P(\varphi | a) = r \rrbracket$ implies that the model satisfies no-signalling regarding a .

Given this semantics, the following axioms and rules are sound (we omit ones regarding probability modalities; a complete axiomatization is an open problem):

- Classical propositional logic.
- Standard axioms and rules for every $[e]$:

$$\frac{\varphi \vdash \psi}{[e] \varphi \vdash [e] \psi}, \quad \vdash [e] \top, \quad [e] \varphi \wedge [e] \psi \vdash [e] (\varphi \wedge \psi).$$

- Moreover, whenever e_0, e_1 are such that $p(e_1) = e_0$ for one of the fibrations p ,

$$[e_0] \varphi \vdash [e_1] \varphi. \quad (8)$$

Now recall that the characterization of contextuality in Theorem 2. That is, a model E fails to be contextual iff obtained from a deterministic hidden-variable model H by forgetting the middle stage—that is, iff consistent with the possibility that, at the middle stage, i.e., after i and before ab as in (6), the states are deterministic hidden variables. So, let us introduce the sentence Det expressing determinacy:

$$\text{Det}(a) := [a](a = 0) \vee [a](a = 1), \quad \text{Det} := \text{Det}(a) \wedge \dots \wedge \text{Det}(b').$$

Then the description of the Popescu-Rohrlich box [22] of $(2, 2, 2)$ -type,

$$\Delta_{\text{PR}} := \langle ab \rangle \top \wedge \dots \wedge \langle a'b' \rangle \top \wedge [ab](a = 0 \leftrightarrow b = 0) \wedge [ab'](a = 0 \leftrightarrow b' = 0) \\ \wedge [a'b](a' = 0 \leftrightarrow b = 0) \wedge [a'b'](a' = 0 \oplus b' = 0),$$

entails $\neg \text{Det}$, using the axioms and rules mentioned above, including the suitable ones of the form (8) such as $[a] \varphi \vdash [ab] \varphi$. Also, a (partial) description Δ_{Hardy} of the Hardy model [13],

$$\Delta_{\text{Hardy}} := \langle i \rangle \langle ab \rangle \top \wedge \dots \wedge \langle i \rangle \langle a'b' \rangle \top \wedge \langle i \rangle \langle ab \rangle (a = 0 \wedge b = 0) \wedge [i] [ab'] (a = 1 \vee b' = 1) \\ \wedge [i] [a'b] (a' = 1 \vee b = 1) \wedge [i] [a'b'] (a' = 0 \vee b' = 0)$$

(note that this description involves label i), entails $\neg [i] \text{Det}$ using the same axioms.

Generalizing these examples, we have the following (we omit a proof).

Theorem 3. *Given a set Δ of sentences (in the dynamic logic just given) of the form either $[i] \varphi$, $\langle i \rangle \varphi$ or $P(\varphi | e \circ i) \gtrless r$, take a set Λ of suitable axioms of the form (8) for the labels in Δ . Then Δ is contextual iff every model that validates Λ validates $\Delta \vdash \neg [i] \text{Det}$. Moreover, Δ is strongly contextual (see [1, §6] for definition) iff every model that validates Λ validates $\Delta \vdash [i] \neg \text{Det}$.*

6 Conclusion

In this paper, we have integrated the three frameworks mentioned in the Introduction for capturing non-deterministic processes, (i)–(iii), by introducing the category ***R-Rel*** of *R*-relations and taking *R*-relational presheaves—functors from trees to ***R-Rel***. The resulting structure captures stochastic dynamics with a good enough expressive power, as demonstrated by the fact that it provides a labelled transitional formulation for the sheaf-theoretic approach of Abramsky and Brandenburger [1] to non-locality and contextuality, and moreover yielding dynamic logic with a modal-logical characterization of contextuality. (In fact, our formalism is partially equivalent to the sheaf-theoretic approach, extending the equivalence between presheaves and fibrations.) Whereas the sheaf-theoretic approach can take advantage of methods of cohomology to calculate conditions for contextuality (see [2]), our approach on the other hand has a certain flexibility in the base trees of measurement labels, so that it can readily express contextuality in not just one round of measurements but within a sequence or protocol of measurements. Thus our approach is expected to complement the sheaf-theoretic approach and extend it to various applications. Of course, applications to other kinds of stochastic dynamics can be expected as well.

References

- [1] Samson Abramsky & Adam Brandenburger (2011): *The Sheaf-Theoretic Structure of Non-Locality and Contextuality*. *New Journal of Physics* 13:113036, doi:10.1088/1367-2630/13/11/113036.
- [2] Samson Abramsky, Shane Mansfield & Rui Soares Barbosa (2012): *Presheaf Models for Concurrency*. In Bart Jacobs, Peter Selinger & Bas Spitters, editors: *Proceedings 8th International Workshop on Quantum Physics and Logic (QPL 2011)*, *Electronic Proceedings in Theoretical Computer Science* 95, pp. 1–14, doi:10.4204/EPTCS.95.1.
- [3] John Baez, Tobias Fritz & Tom Leinster (2011): *A Characterization of Entropy in Terms of Information Loss*. Available at <http://arxiv.org/abs/1106.1791>.
- [4] Alexandru Baltag & Sonja Smets (2006): *LQP: The Dynamic Logic of Quantum Information*. *Mathematical Structures in Computer Science* 16, pp. 491–525, doi:10.1017/S0960129506005299.
- [5] Jonathan Barrett (2007): *Information Processing in Generalized Probabilistic Theories*. *Physical Review A* 75:032304, doi:10.1103/PhysRevA.75.032304.
- [6] Aurelio Carboni & Robert F. C. Walters (1987): *Cartesian Bicategories I*. *Journal of Pure and Applied Algebra* 49, pp. 11–32, doi:10.1016/0022-4049(87)90121-6.
- [7] Gian Luca Cattani & Glynn Winskel (1997): *Presheaf Models for Concurrency*. In Dirk van Dalen & Marc Bezem, editors: *Computer Science Logic: 10th International Workshop, CSL'96, Lecture Notes in Computer Science* 1258, Springer, pp. 58–75, doi:10.1007/3-540-63172-0_32.
- [8] Ernst-Erich Doberkat (2007): *Stochastic Relations: Foundations for Markov Transition Systems*. Chapman & Hall/CRC.
- [9] Brendan Fong (2012): *Causal Theories: A Categorical Perspective on Bayesian Networks*. Master's thesis, University of Oxford.
- [10] Peter J. Freyd & Andre Scedrov (1990): *Categories, Allegories*. North-Holland.
- [11] Tobias Fritz (2009): *A Presentation of the Category of Stochastic Matrices*. Available at <http://arxiv.org/abs/0902.2554>.
- [12] Michèle Giry (1982): *A Categorical Approach to Probability Theory*. In B. Banaschewski, editor: *Categorical Aspects of Topology and Analysis, Lecture Notes in Mathematics* 915, Springer, pp. 68–85, doi:10.1007/BFb0092872.

- [13] Lucien Hardy (1993): *Nonlocality for Two Particles without Inequalities for Almost All Entangled States*. *Physical Review Letters* 71, pp. 1665–1668, doi:10.1103/PhysRevLett.71.1665.
- [14] David Harel, Dexter Kozen & Jerzy Tiuryn (2000): *Dynamic Logic*. MIT Press.
- [15] Claudio Hermida (2011): *A Categorical Outlook on Relational Modalities and Simulations*. *Information and Computation* 209, pp. 1505–1517, doi:10.1016/j.ic.2010.09.009.
- [16] Bart Jacobs (2010): *Convexity, Duality and Effects*. In Christian S. Calude & Vladimiro Sassone, editors: *Theoretical Computer Science: 6th IFIP TC 1/WG 2.2 International Conference, TCS 2010*, Springer, pp. 1–19, doi:10.1007/978-3-642-15240-5_1.
- [17] Bart Jacobs (2011): *Probabilities, Distribution Monads, and Convex Categories*. *Theoretical Computer Science* 412, pp. 3323–3336, doi:10.1016/j.tcs.2011.04.005.
- [18] F. William Lawvere (1962): *The Category of Probabilistic Mappings*. Unpublished manuscript.
- [19] N. David Mermin (1981): *Quantum Mysteries for Anyone*. *Journal of Philosophy* 78, pp. 397–408, doi:10.2307/2026482.
- [20] Prakash Panangaden (1998): *Probabilistic Relations*. In C. Baier, M. Huth, M. Kwiatkowska & M. Ryan, editors: *Preliminary Proceedings of PROBMIV’98*, pp. 59–74.
- [21] Prakash Panangaden (1999): *The Category of Markov Kernels*. *Electronic Notes in Theoretical Computer Science* 22, pp. 171–187, doi:10.1016/S1571-0661(05)80602-4.
- [22] Sandu Popescu & Daniel Rohrlich (1994): *Quantum Nonlocality as an Axiom*. *Foundations of Physics* 24, pp. 397–385, doi:10.1007/BF02058098.
- [23] Kimmo I. Rosenthal (1996): *The Theory of Quantaloids*. Addison Wesley.
- [24] Paweł Sobociński (2012): *Relational Presheaves as Labelled Transition Systems*. In Dirk Pattinson & Lutz Schröder, editors: *Coalgebraic Methods in Computer Science, Lecture Notes in Computer Science 7399*, Springer, pp. 40–50, doi:10.1007/978-3-642-32784-1_3.
- [25] Daniele Varacca (2003): *Probability, Nondeterminism and Concurrency: Two Denotational Models for Probabilistic Computation*. Ph.D. thesis, University of Aarhus.
- [26] Glynn Winskel & Mogens Nielsen (1997): *Presheaves as Transition Systems*. In Doron Peled, Vaughan R. Pratt & Gerard J. Holzmann, editors: *Partial Order Methods in Verification: DIMACS Workshop July 24–26, 1996*, American Mathematical Society, pp. 129–140.

Circuit model implementation of controllization functional on unitary with and without fractional query

Akihito Soeda Shojun Nakayama

Department of Physics
Graduate School of Science
University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan
soeda@phys.s.u-tokyo.ac.jp nakayama@eve.phys.s.u-tokyo.ac.jp

Mio Murao

Department of Physics
Graduate School of Science
University of Tokyo
7-3-1, Hongo, Bunkyo-ku, Tokyo, Japan
Institute for Nano Quantum Information Electronics
University of Tokyo
4-6-1, Komaba, Meguro-ku, Tokyo, Japan
murao@phys.s.u-tokyo.ac.jp

High-level programming languages aim to provide an intuitive interface to programmers, typically, via uses of functionals (or a function of a function), a trend already seen in quantum programming. Since a unitary itself is a function on quantum states, a function that maps a given unitary to its controlled counterpart (namely, the controlled-unitary) is a functional in quantum computation. We prove that such a *controllization* functional is *not* implementable as a quantum circuit, even if we consider implementations using multiple queries to the input unitary, with non-deterministic success probability, and an extra degree of freedom in the relative phase of the controlled unitary. We show, however, with fractional unitary queries and their inverse, probabilistic controllization is possible with multiple calls and the degree of freedom in the relative phase. In addition, it is proven that the inverse query is unnecessary for single-qubit unitaries, by presenting a circuit that implements the inverse query exactly but probabilistically with a single query of the input unitary. Our result shows that the circuit model becomes provably more powerful with fractional queries in the circuit implementation of quantum functionals.

1 Introduction

Despite its success, the circuit model of quantum computation is a low-level programming paradigm, lacking the intuitive features realized in classical computing through high-level programming. The recent development in quantum programming languages has begun to incorporate high-level features in the form of functional programming to offer programmers a more intuitive interface between quantum computers [1, 2]. It is reasonable to expect that such an intuitive interface will accelerate programmers in designing future quantum algorithms.

Perhaps the most apparent difference between a high-level programming and a low-level one is that the former allows manipulation of functions, while the latter can only manipulate computation data. Thus, in a high-level programming, we encounter functions of a function, i.e., functionals. There are three steps between the actual coding and the execution of a program in high-level programming, namely, the coding process itself, compilation of the code into a sequence of low-level instructions, and the

execution of the instructions. Any program coded with a high-level language must result in a set of executable low-level instructions. However, the programming paradigm affects the very efficiency of code compilation and that of execution time, as already seen in classical programming in precompiled headers [3, 4] and dynamic link libraries [5].

Besides its practical implications, functional calculus has already appeared in more conceptual studies of quantum algorithms [6, 7, 8, 9, 10]. This is referred to as *higher-order quantum computation* in the literature [10], where the input of an algorithm is not a quantum state but unitary transformations. Since unitary transformations (or, simply *unitaries*) are a function on quantum states, such an algorithm on unitaries necessarily involves quantum functionals. These quantum functionals are known to provide information processing advantage [9, 10].

Mathematically, these quantum functionals are a map from a unitary to a unitary. The existence and non-existence of some quantum functionals cannot be simply deduced from its consistency with well-known physical principles such as the unitarity of closed-system evolution [9]. The implementability of a quantum functional should not be assumed based solely on physical consistencies.

Certain versatile quantum algorithms, such as DQC1 [11] and Kitaev's phase estimation algorithm (KPEA) [12], are designed to accept unitaries as input. These algorithms are used as a subroutine of quantum factoring [13, 14], quantum simulation [15], and coupled linear equation solver [16]. In both DQC1 and KPEA, the input unitary U is first converted to its controlled form $\Lambda U = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes U$. This conversion to a controlled-unitary, which we name *controllization*, is another example of quantum functional.

The known implementations of this controllization functional require some prior knowledge of the parameters that describe the input unitary. In the most extreme case, where all the parameters are known, Barenco has identified in Ref. [17] a way to implement the functional by decomposing the input unitary into elementary unitary gates of single-qubit unitaries and controlled-NOT. In an optical setting where the input unitary is guaranteed to be constructed from passive linear elements, the controllization functional is implementable without knowing some of the parameters [18]; its proof-of-principle experiment has been conducted [18]. The only known method to implement the controllization functional on a completely unknown unitary is to perform process tomography to identify all the parameters [19], but this method requires an infinite number of copies of the unitary. In fact, Araújo et al. has proven that a *single* copy of unitary is not enough to implement the controllization functional even if the implementation allows an extra degree of freedom in the relative phase of the controlled-unitary, namely, $\Lambda_\eta U = |0\rangle\langle 0| \otimes \mathbb{I} + |1\rangle\langle 1| \otimes \exp(i\eta_U)U$ [20].

In this work, we further relax the requirement on the implementation of the controllization functional to allow non-deterministic success probability and multiple copies of the input unitary, and prove that such controllization functional still does not exist. We then introduce *fractional queries* of the input unitary, i.e., $U^{1/m}$, and their inverse $(U^{1/m})^\dagger$ and find that the relaxed controllization functional becomes implementable. Such a fractional query appears frequently when a problem is given in terms of a Hamiltonian H , e.g., quantum simulation, where the input unitary U is then the time-evolution operator $\exp(-iHt/\hbar)$ for some evolution duration t . In this work, we understand that $U^{1/m} = \exp[-iH(t/m)/\hbar]$ to avoid any ambiguity arising from the m -th root function being multi-valued.

2 Without fractional query

Let us introduce the following parametrization of unitary operations on a N -dimensional Hilbert space,

$$U(N) \ni u(\delta, x_1, x_2, \dots, x_{N^2-1}) = e^{i\delta} \hat{u}(x_1, x_2, \dots, x_{N^2-1}), \quad (1)$$

where $\hat{u}(x_1, x_2, \dots, x_{N^2-1}) \in SU(N)$. The controllization functional (which we denote by f) should satisfy the following condition:

$$f(\text{with } N \text{ uses of } u(\delta, \mathbb{x}_{N^2-1})) = c(\delta, \mathbb{x}_{N^2-1}) \begin{pmatrix} \mathbb{I} \\ e^{i\eta(\mathbb{x}_{N^2-1})} \hat{u} \end{pmatrix} \quad (2)$$

for some complex-valued function $c(\delta, \mathbb{x}_{N^2-1})$ and some real-valued function $\eta(\mathbb{x}_{N^2-1})$. The phase factor $e^{i\eta(\mathbb{x}_{N^2-1})}$ represents the degree of freedom in the relative phase of the controlled-unitary. In certain cases, we are only interested in the difference in the phase of some of the eigenvalues. The extra phase factor preserves the difference. Parameter δ cannot appear in $\eta(\mathbb{x}_{N^2-1})$, otherwise it implies that the difference in the global phase becomes detectable by implementing the controllization functional. The extra coefficient $c(\delta, \mathbb{x}_{N^2-1})$ is a complex-valued function, used to express probabilistic implementation. When $|c(\delta, \mathbb{x}_{N^2-1})| < 1$, it implies that there is a nonzero probability that the controllization functional fails, but we also assume that it succeeds at least with some nonzero probability for all input unitary, i.e.,

$$c(\delta, \mathbb{x}_{N^2-1}) > 0 \text{ for all } \delta \text{ and } \mathbb{x}_{N^2-1}. \quad (3)$$

If we set $N = 1$ and require $c(\delta, \mathbb{x}_{N^2-1}) = 1$ for all the parameters, the functional f reduces to the one considered in Ref. [20]. Notice that Eq. (2) is equivalent to impose

$$f(\text{with } N \text{ uses of } u(\delta, \mathbb{x}_{N^2-1})) = c(\delta, \mathbb{x}_{N^2-1}) \begin{pmatrix} e^{i\eta(\mathbb{x}_{N^2-1})} \mathbb{I} \\ \hat{u} \end{pmatrix} \quad (4)$$

since $c(\delta, \mathbb{x}_{N^2-1})$ and $\eta(\mathbb{x}_{N^2-1})$ are arbitrary. In the remainder of the section, we shall assume this definition of controllization functional and prove that its existence leads to a contradiction.

It suffices to prove the non-existence when the input unitary is a subset of $SU(2)$, parametrized as

$$\hat{u}(t, \theta) = \cos t \mathbb{I} + i \sin t (\sin \theta \sigma_x + \cos \theta \sigma_z). \quad (5)$$

Suppose f exists, then without loss of generality, we may assume that f takes the form of *quantum comb* [6, 8], where

$$f(\text{with } N \text{ uses of } u) = \left(\mathbb{I}_{in} \otimes \langle 00 \dots 0 |_{anc} \right) W_N u W_{N-1} u \dots u W_0 \cdot \left(\mathbb{I}_{in} \otimes |00 \dots 0 \rangle_{anc} \right) \quad (6)$$

for some $N + 1$ unitaries W_0, W_1, \dots, W_N and qubit ancillas in state $|0\rangle$. The operator \mathbb{I}_{in} is the identity operator on the input system. The state $|00 \dots 0 \rangle_{anc}$ represents the extension of the input system by appending ancillas. The bra vector $\langle 00 \dots 0 |_{anc}$ represents the possibility of probabilistic implementation; in other words, we are allowed to postselect the desired outcome. Note that W_k must be independent of u .

According to the linearity of quantum circuits, the left hand side of Eq. (4) can be expanded as

$$\begin{aligned} & f(\text{with } N \text{ uses of } \hat{u}(t, \theta)) \\ &= \sum_{(n_0, n_1, n_2) \in S_N} (\cos t)^{n_0} (\sin t \sin \theta)^{n_1} (\sin t \cos \theta)^{n_2} \\ & \quad \times [\text{sum of all } f(\dots) \text{ such that there are } n_0 \text{ of } \mathbb{I}, n_1 \text{ of } i\sigma_x, \text{ and } n_2 \text{ of } i\sigma_z], \end{aligned} \quad (7)$$

where S_N denotes the set of all non-negative integer triplets (n_0, n_1, n_2) such that their sum satisfies $n_0 + n_1 + n_2 = N$. Comparing Eqs. (4) and (7), we see that there must be three functions f_0 , f_1 , and f_2

such that

$$f_0 = \sum_{k=0}^N \left[\sum_{l=0}^k \alpha_{kl} (\cos \theta)^{k-l} (\sin \theta)^l \right] (\cos t)^{N-k} (\sin t)^k \quad (8)$$

$$f_1 = \sum_{k=0}^N \left[\sum_{l=0}^k \beta_{kl} (\cos \theta)^{k-l} (\sin \theta)^l \right] (\cos t)^{N-k} (\sin t)^k \quad (9)$$

$$f_2 = \sum_{k=0}^N \left[\sum_{l=0}^k \gamma_{kl} (\cos \theta)^{k-l} (\sin \theta)^l \right] (\cos t)^{N-k} (\sin t)^k \quad (10)$$

$$c(0, t, \theta) \dot{u} = f_0 \mathbb{I} + f_1 i\sigma_x + f_2 i\sigma_z. \quad (11)$$

By Eqs. (5) and (11) and the linear independence of \mathbb{I} , $i\sigma_x$, and $i\sigma_z$, we have that

$$c(0, t, \theta) \cos t = f_0 \quad (12)$$

$$c(0, t, \theta) (\sin t) (\cos \theta) = f_1 \quad (13)$$

$$c(0, t, \theta) (\sin t) (\sin \theta) = f_2. \quad (14)$$

From the first two equations, we obtain

$$f_1 = f_0 (\tan t) (\sin \theta), \quad (15)$$

which is equivalent to

$$\begin{aligned} & \sum_{k=0}^N \left[\sum_{l=0}^k \beta_{kl} (\cos \theta)^{k-l} (\sin \theta)^l \right] (\cos t)^{N-k} (\sin t)^k \\ &= \sum_{k'=1}^{N+1} \left[\sum_{l=0}^{k'-1} \alpha_{(k'-1)l} (\cos \theta)^{(k'-1)-l} (\sin \theta)^{l+1} \right] (\cos t)^{N-k'} (\sin t)^{k'}. \end{aligned} \quad (16)$$

Equation (16) holds for all values of t and θ , hence

$$\beta_{N0} = 0. \quad (17)$$

From Eq. (13), however, when $t = \pi/2$ and $\theta = 0$,

$$c(0, \pi/2, 0) = \beta_{N0} = 0, \quad (18)$$

which contradicts with Eq. (3). Therefore, the controllization functional cannot be implemented as a quantum circuit.

3 With fractional query

In this section, we show that if fractional queries $U^{1/m}$ of the input unitary and their inverse $(U^{1/m})^\dagger$ are available, we can implement ΛU with the additional freedom in the relative phase, with an arbitrarily high probability. In general, there is no quantum circuit for the functional on U that converts to U^\dagger , which is forbidden by the linearity of quantum comb. Nevertheless, we will prove that if the input unitary is of a single-qubit, then its exact inverse can be implemented probabilistically, hence the inverse query is unnecessary in this case.

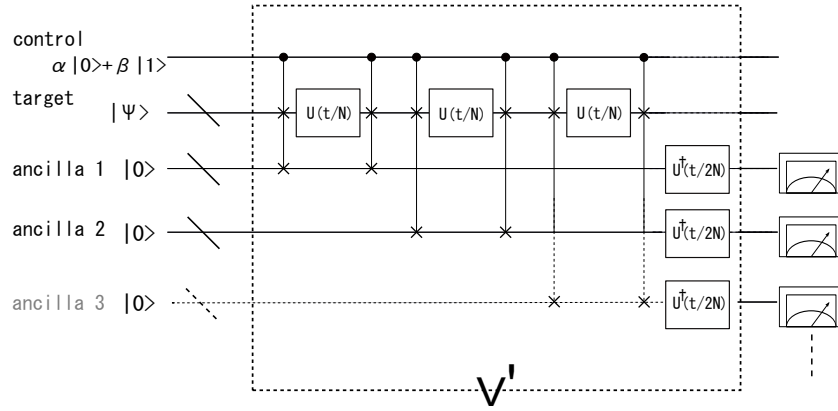


Figure 1: A quantum circuit probabilistically implementing the controllization functional of Eq. (2). V' represents the unitary operation given by the sequence of unitary gates, the Fredkin gates (the controlled-swap gate), the black boxes implementing $U(t/N)$ and $U^\dagger(t/2N)$. For each ancilla state, the measurement in the quantum circuit is given by the projective measurements in the computational basis for each qubit $\{|0\rangle\langle 0|, |1\rangle\langle 1|\}$.

The quantum circuit implementing our controllization functional is presented in Fig. 1. The time-inverted unitary gate $U^\dagger(t/2N)$ immediately before the measurement on each ancilla qubit changes the state before the measurement of each ancilla to $U(t/2N)|0\rangle$, when the initial state of the controlled qubit is $|1\rangle$, and, otherwise, to $U^\dagger(t/2N)|0\rangle$. For any matrix, the absolute value of every diagonal element of a matrix A and its adjoint matrix A^\dagger are the same. If the measurements return the successful outcome, i.e., projection on to $|0\rangle^{\otimes N}$, we obtain

$$\begin{aligned} & \langle 0|_{anc}^{\otimes N} V' (\alpha|0\rangle_{cnt} + \beta|1\rangle_{cnt}) |\psi\rangle_{tgt} |0\rangle_{anc}^{\otimes N} = \\ & |\langle 0|_{anc} U(t/2N) |0\rangle_{anc}|^N \left(e^{-i\varphi'(t)} |0\rangle_{cnt} |\psi\rangle_{tgt} + |1\rangle_{cnt} U(t) |\psi\rangle_{tgt} \right) |0\rangle_{anc}^{\otimes N}. \end{aligned} \quad (19)$$

The subscripts *cnt*, *tgt*, and *anc* denote that the marked state belongs to the control qubit, the target system, or the ancillas, respectively. The phase φ is given by

$$e^{-i\varphi'(t)} = \left(\frac{\langle 0|_{anc} U(t/2N) |0\rangle_{anc}}{\langle 0|_{anc} U(-t/2N) |0\rangle_{anc}} \right)^N \quad (20)$$

and it satisfies

$$\varphi'(t) = \langle H \rangle_0 t + O(1/N), \quad (21)$$

where $\langle H \rangle_0 = \langle 0|H|0\rangle$. The success probability of the measurement in this case is given by

$$|\langle 0|_{anc} U(t/2N) |0\rangle_{anc}|^{2N}, \quad (22)$$

which scales as $1 - O(\Delta H_0^2/2N)$, where $\Delta H_0^2 = \langle 0|(H - \langle H \rangle_0)^2|0\rangle$. In case of success, the effect of ignoring the term $O(1/N)$ appears only in the global phase factor. Therefore, we conclude that the quantum circuit given by Fig. 1 implements $\Lambda U(t)$ with an arbitrarily high probability by choosing N sufficiently large.

Finally for $SU(2)$, recall that

$$(u \otimes \mathbb{I}_2) |\Phi^+\rangle_{12} = (\mathbb{I}_1 \otimes {}^t u) |\Phi^+\rangle_{12} \quad (23)$$

$$u^\dagger = \sigma_y {}^t u \sigma_y, \quad (24)$$

where $|\Phi^+\rangle_{12} = (|00\rangle_{12} + |11\rangle_{12})/\sqrt{2}$. It is then easy to see that the following equation holds

$$\frac{1}{2} u^\dagger |\psi\rangle_3 = \left(\langle \Phi^+ |_{12} \otimes \mathbb{I}_3 \right) \cdot \left(\mathbb{I}_1 \otimes \sigma_y u \sigma_y \otimes \mathbb{I}_3 \right) \cdot |\psi\rangle_1 \otimes |\Phi^+\rangle_{23}. \quad (25)$$

By swapping the first and the third qubit, we probabilistically implement u^\dagger applied on the input state $|\psi\rangle$ by a single query to the input unitary. Each implementation of the inverse query succeeds with probability $1/4$, which implies that a simple substitution of the inverse queries in Fig. 1 results in an exponential reduction of the success probability.

4 Conclusion

In this work, we investigated the circuit implementation of the controllization functional with non-unit success probability, multiple queries of the input unitary, and an extra relative phase. It was proven that the controllization functional is not implementable without fractional queries of the input unitary. On the other hand, if the fractional queries and their inverse are available, we gave an explicit construction of the controllization functional whose success probability becomes arbitrarily high with a suitable choice of parameters. The inverse queries are not necessary if the input unitary is a single-qubit unitary, for which a probabilistic implementation of the inverse query is presented. It is an interesting question whether a such inverse functional exists for unitaries of higher dimensions, although we conjecture it unlikely. Finally, whether fractional queries offer implementation advantage for other quantum functionals is an interesting open question.

References

- [1] S. J. Gay, “Quantum programming languages: survey and bibliography”, *Mathematical Structures in Computer Science* **16**, pp. 581-600 (2006).
- [2] A. S. Green, P. L. Lumsdaine, N. J. Ross, P. Selinger, and B. Valiron, “Quipper: A Scalable Quantum Programming Language”, in G. W. Dueck and D. M. Miller (eds.), *Reversible Computation, Lecture Notes in Computer Science* 7948, Springer, New York, pp. 333-342 (2013).
- [3] D. Vandevorde and N. M. Josuttis, *C++ Templates: The Complete Guide*, 1st edition, Chapter 6, Pearson Education, Boston (2003).
- [4] K. Sung, P. Shirley, and S. Baer, *Essentials of Interactive Computer Graphics : Concepts and Implementation*, Har/Cdr edition, Chapter 2, A K Peters/CRC Press, Boca Raton (2009).
- [5] J. R. Levine, *Linkers and loaders*, 1st edition, Chapter 10, Academic Press, London (2000).
- [6] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Optimal Cloning of Unitary Transformation”, *Phys. Rev. Lett.* **101**, 060401 (2008).
- [7] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Transforming quantum operations: Quantum supermaps”, *Europhys. Lett.* **83**, 30004 (2008).
- [8] G. Chiribella, G. M. D’Ariano, and P. Perinotti, “Theoretical framework for quantum networks”, *Phys. Rev. A* **80**, 022339 (2009).

- [9] T. Colnaghi, G. M. D’Ariano, S. Facchini, and P. Perinotti, “Quantum computation with programmable connections between gates”, *Phys. Lett. A* **376**, pp. 2940-2943 (2012).
- [10] G. Chiribella, G. M. D’Ariano, P. Perinotti, and B. Valiron, “Quantum computations without definite causal structure”, *Phys. Rev. A* **88**, 022318 (2013).
- [11] E. Knill and R. Laflamme, “Power of One Bit of Quantum Information”, *Phys. Rev. Lett.* **81**, 5672 (1998).
- [12] A.Y. Kitaev, “Quantum measurements and the Abelian stabilizer problem”, *Electr. Coll. Comput. Complex* **3**, 3 (1996).
- [13] P. W. Shor, “Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer”, *arXiv:9508027* (1996).
- [14] S. Parker and M. Plenio, “Efficient Factorization with a Single Pure Qubit and logN Mixed Qubits”, *Phys. Rev. Lett.* **85**, 3049 (2000).
- [15] (for reviews see e.g.) K. L. Brown, W. J. Munro, and V. M. Kendon, “Using Quantum Computers for Quantum Simulation”, *Entropy* **12**, 2268-2307 (2010); B. Sanders, “Efficient Algorithms for Universal Quantum Simulation”, in G. W. Dueck and D. M. Miller (eds.), *Reversible Computation, Lecture Notes in Computer Science* 7948, Springer, New York, pp. 1-10 (2013).
- [16] A. W. Harrow, A. Hassidim, and S. Lloyd, “Quantum Algorithm for Linear Systems of Equations”, *Phys. Rev. Lett.* **103**, 150502 (2009).
- [17] A. Barenco, C. H. Bennett, R. Cleve, D. P. DiVincenzo, N. Margolus, P. Shor, T. Sleator, J. A. Smolin, and H. Weinfurter, “Elementary gates for quantum computation”, *Phys. Rev. A* **52**, 3457 (1995).
- [18] X.-Q. Zhou, T. C. Ralph, P. Kalasuwan, M. Zhang, A. Peruzzo, B. P. Lanyon, and J. L. O’Brien, “Adding control to arbitrary unknown quantum operations”, *Nat. Comm.* **2**, 413 (2011).
- [19] I. L. Chuang and M. A. Nielsen, “Prescription for experimental determination of the dynamics of a quantum black box”, *J. Mod. Phys.* **44**, 2455, (1997).
- [20] M. Araújo, A. Feix, F. Costa, and Č. Brukner, “Quantum circuits cannot control unknown operations”, *arXiv:1309.7976* (2013).

On Gács' quantum algorithmic entropy *

Toru Takisaka

Research Institute for Mathematical Sciences
Kyoto University
Kyoto, Japan
takisaka@kurims.kyoto-u.ac.jp

We define an infinite dimensional modification of lower-semicomputability of density operators by Gács with an attempt to fix some problem in the paper. Our attempt is partly achieved by showing the existence of universal operator under some additional assumption. It is left as a future task to eliminate this assumption. We also see some properties and examples which stimulate further research. In particular, we show that universal operator has certain nontrivial form if it exists.

1 Preliminaries

Kolmogorov complexity is the notion of actual information content of finite string in computational point of view. This notion has been proposed by Kolmogorov - Solomonoff - Chaitin in 1960s and used in various areas as a basic tool to represent descriptive complexity. On the other hand, Since Shor's algorithm [1] has been discovered, the research on quantum information has made a great progress and produced various proposals on application to quantum information technology.

Quantum Kolmogorov complexity is one of these branches appeared in early 2000s. Several different definitions are proposed so far [2-4], and some applications to quantum information are recently emerging [5-6]. However, it seems that there is very little progress in this area despite a decade has passed since these suggestions have been made, and a number of elementary facts are still not investigated.

In particular, relationships between them are not clarified. In classical domain, there are several definitions of descriptive complexity and some of them are known as equivalent notions (Levin's coding theorem). This theorem, in some sense, guarantees that these notions are reliable.

It naturally leads us to the following question: can we find any good relationship between these quantum complexities? In particular, if it turns out that some of them are equivalent, it would be helpful to make these notions more reliable and more applicable to other research subject such like quantum information theory.

We particularly have interest on those by Berthiaume et al. [2] and Gács [3] since they are the quantum extension of plain Kolmogorov complexity and universal semimeasure, respectively. Levin's coding theorem claims that prefix Kolmogorov complexity and universal semimeasure are equivalent, so they are expected to be nearly equivalent.

For Berthiaume's definition, there are several results about fundamental facts such like its invariance and relation between classical complexity [7-9]. As compared to this, there are not so much subsequent research of Gács' approach, so we mainly treat his definition.

*Most part of this research was carried out without knowing about Tadaki's work [15]. Quite recently, in March 2014, Prof. Tadaki draw our attention to his work and we noticed that there are substantial overlaps between Tadaki's work and ours. As far as we know, his work seems to be the first one providing an extension of Gács' work to the infinite-dimensional setting. In the present paper, we tried to reflect Tadaki's results as possible as we can. Interested readers should consult Tadaki's article [15]. We are very grateful to Prof. Tadaki for his advices.

In [3], the quantum analogue of lower-semicomputable semimeasure which is named *lower-semicomputable semi-density matrix* is introduced. In the paper, though, proofs of two crucial theorems have some flaw.

Conjecture 1.1. *There is a lower-semicomputable semi-density matrix μ dominating all other such matrices in the sense that for every other such matrix ρ there is a constant $c > 0$ with $\rho \leq c\mu$.*

Conjecture 1.2. *Let $|1\rangle, |2\rangle, \dots$ be a computable orthogonal sequence of states. Also Let \overline{H} and \underline{H} be real-valued functions defined as*

$$\overline{H}(|\psi\rangle) = -\langle\psi|(\log\mu)\psi\rangle, \quad \underline{H}(|\psi\rangle) = -\log\langle\psi|\mu\psi\rangle.$$

Then for $H = \overline{H}$ or $H = \underline{H}$ we have

$$H(|i\rangle) = K(i) + O(1).$$

Here, $K(i)$ is the prefix Kolmogorov complexity of i .

The former is indispensable to define quantum algorithmic entropy, and the latter is expected to be true when we wish to compare Gács' quantum algorithmic entropy and the qubit complexity defined by Berthiaume et al [2].

In this paper, we introduce an infinite dimensional modification of Gács' definition to fix these problems. Our attempt is partly achieved by showing the existence of universal operator under some additional assumption. This is an analogous approach to the one of Tadaki [15], in which the notion of *lower-computable semi-POVM* is introduced, and it is shown that a universal semi-POVM does exist. Still, it seems that this assumption should be derived from our definition itself, so checking whether it is possible or not is our future task. It turns out that, in our modification, if we assume the existence of universal operator then Conjecture 1.2 is also true. We also see some properties and examples which stimulate further research.

Contents of this paper are as follows: in section 2, we recall some classical notions of descriptive complexity for preparation. In section 3, we propose an infinite dimensional modification of lower-semicomputable semi-density matrix, which is defined by Gács to define his quantum algorithmic entropy. We prove some of their properties, and consider the problem about the existence of universal operator.

We assume the readers are familiar with the basic ideas and technics of quantum information theory. The most famous textbook of this area would be Nielsen and Chuang [17], but we also suggest Heinosaari and Ziman [18] as an introduction, which is fairly readable and includes knowledge for infinite dimensional cases. For more exhaustive learning of functional analysis, see Conway [19].

2 Classical notions of descriptive complexity

In this section, we review two classical notions about descriptive complexity which are equivalent in some sense. Proof of any theorem in this section can be found in [12].

2.1 Kolmogorov complexity

(Plain) Kolmogorov complexity $C_M(w)$ of finite binary string w with respect to a Turing machine M is the length of a shortest program which makes M output w :

$$C_M(w) = \min \{ l(v) \mid M(v) = w \}.$$

M is called a *reference machine*. In many cases, some optimal universal Turing machine M_0 is employed as a fixed reference machine and $C(w) := C_{M_0}(w)$ is just called Kolmogorov complexity of w . Here, we say M_0 is *optimal* if for any Turing machine M there exists $c_M > 0$ such that

$$C_{M_0}(w) \leq C_M(w) + c_M.$$

$A \subset \{0, 1\}^*$ is a *prefix set* if for any two disjoint elements $w, v \in A$, w is not a prefix of v , and vice versa: that is, $w \neq vu$ and $v \neq wu$ for any $u \in \{0, 1\}^*$. We call a Turing machine T *Prefix Turing machine* if $\text{dom} T$ is a prefix set. We can enumerate all prefix Turing machines effectively, and there exists an optimal universal prefix Turing machine. For detail, see [12]. We fix some optimal universal prefix Turing machine M_1 and call $K(w) := C_{M_1}(w)$ *prefix Kolmogorov complexity* of w .

2.2 Lower-semicomputable semimeasure

A nonnegative real function $f(w)$ on strings is called a *semimeasure* if $\sum_w f(w) \leq 1$, and a *measure* if the sum is 1. f is *lower-semicomputable* if there is a computable function $\tilde{f} : \{0, 1\}^* \times \mathbb{N} \rightarrow \mathbb{Q}$ such that $\tilde{f}(w, k) \leq \tilde{f}(w, k+1)$ for every $w \in \{0, 1\}^*$, $k \in \mathbb{N}$, and $\tilde{f}(w, k) \xrightarrow{k \rightarrow \infty} f(w)$ for every w . We call \tilde{f} a *lower-approximation* of f (we use this notation for convenience, but probably this function does not have any widely accepted name).

Theorem 2.1. *We can enumerate all lower-semicomputable semimeasures effectively. Namely, there exists $\tilde{m} : \{0, 1\}^* \times \mathbb{N}^2 \rightarrow \mathbb{Q}$ which satisfies following two conditions:*

1. *for any $n \in \mathbb{N}$, $\tilde{m}(-, -, n)$ is a lower-approximation of some lower-semicomputable semimeasure;*
2. *for given lower-semicomputable semimeasure m' , there is $n \in \mathbb{N}$ such that $\tilde{m}(-, -, n)$ is a lower-approximation of m' .*

It is well known that there exists a universal semimeasure in the following sense.

Theorem 2.2. *There is a semicomputable semimeasure \mathbf{m} with the property that for any other semicomputable semimeasure m' there is a constant $c > 0$ such that for all w we have $cm'(w) \leq \mathbf{m}(w)$.*

Proof. We can easily show that

$$\mathbf{m}(w) := \sum_{n=1}^{\infty} 2^{-n} m_n(w)$$

is a universal semimeasure, where $\{m_n\}_{n=1}^{\infty}$ is an effective enumeration of all lower-semicomputable semimeasures. \square

We conclude this section with a theorem due to Levin. It indicates that the notion of universal semimeasure is somewhat equivalent to that of Kolmogorov complexity.

Theorem 2.3 (Levin's coding theorem). $K(w) = -\log \mathbf{m}(w) + O(1)$.

3 Quantization of lower-semicomputable semimeasure

In this section, we define an infinite dimensional modification of lower-semicomputable semi-density matrix defined by Gács [3], and see some properties, examples, and problems.

3.1 Definition and some properties

As a quantum analogue of the set of all binary strings, we introduce the space of indeterminate-length qubit strings, $\mathcal{H} := \bigoplus_{n=0}^{\infty} (\mathbb{C}^2)^{\otimes n}$. We assume an orthonormal basis $\{|0\rangle, |1\rangle\}$ is given for each qubit space \mathbb{C}^2 , so \mathcal{H} has an orthonormal basis $\{|w\rangle\}_{w \in \{0,1\}^*}$, where $|w\rangle = |a_1\rangle \otimes \cdots \otimes |a_n\rangle$ for $w = a_1 \cdots a_n$. We call it the *computational basis* of \mathcal{H} . Notice that the computational basis is indispensable to consider descriptive complexity of qubit strings, just as in classical domain we need to work on $\{0,1\}^*$, not ω .

Let $\mathcal{B}(\mathcal{H})$ be the set of all bounded operator on \mathcal{H} , and $\mathcal{L}(\mathcal{H})$ be the set of all bounded hermitian operator on \mathcal{H} . We also write $\mathbb{C}_q := \{x + yi \mid x, y \in \mathbb{Q}\}$.

Definition 3.1. $\rho \in \mathcal{L}(\mathcal{H})$ is called a semi-density operator if $\rho \geq 0$ and $\text{Tr} \rho \leq 1$. Let $\tilde{\mathcal{S}}(\mathcal{H})$ be the set of all semi-density operators on \mathcal{H} .

$\rho \in \mathcal{L}(\mathcal{H})$ is lower-semicomputable (upper-semicomputable) if there is a computable function $\psi : \mathbb{N} \times \{0,1\}^* \times \{0,1\}^* \rightarrow \mathbb{C}_q$ such that the sequence $\{\rho_n\}_{n=1}^{\infty} \subset \mathcal{L}(\mathcal{H})$ defined by

$$\langle w | \rho_n v \rangle := \psi(n, w, v)$$

satisfies $\rho_n \leq \rho_{n+1}$ ($\rho_n \geq \rho_{n+1}$) and $\rho_n \xrightarrow{n \rightarrow \infty} \rho$ in WOT (i.e. $\langle \psi | \rho_n \psi \rangle \rightarrow \langle \psi | \rho \psi \rangle$ for any $|\psi\rangle \in \mathcal{H}$). WOT is an abbreviation of weak operator topology). We call ψ a lower- (upper-) approximation of ρ .

$\rho \in \mathcal{L}(\mathcal{H})$ is computable if there is a computable function $\psi : \mathbb{N} \times \{0,1\}^* \times \{0,1\}^* \rightarrow \mathbb{C}_q$ which defines $\{\rho_n\}_{n=1}^{\infty} \subset \mathcal{L}(\mathcal{H})$ such that $\|\rho - \rho_n\| < 2^{-n}$, in the same manner as above. We call ψ an approximation of ρ .

Dimension of the string space \mathcal{H} is almost the only difference between the definition by Gács and us. A mode of convergence of $\{\rho_n\}_{n=1}^{\infty}$ needs to be specified when we work on an infinite dimensional space, so we choose WOT, which is equivalent to the pointwise convergence of each matrix coefficient. Gács allows a lower-approximation function to take an algebraic number, but we do not feel it necessary, so we only allow a complex-rational value.

Remark. Perhaps lower-semicomputability of operator can be defined for unbounded hermitian operator, but we will content ourselves with this definition in this paper. We mainly treat lower-semicomputable semi-density operators, which are automatically bounded.

It is equivalent to define the correspondence between ψ and $\{\rho_n\}_{n=1}^{\infty}$ as

$$\psi(n, w, v) = \begin{cases} \langle w + v | \rho(w + v) \rangle & (w \leq v) \\ \langle w + iv | \rho(w + iv) \rangle & (w > v). \end{cases}$$

In this definition, a lower approximation ψ of any lower-semicomputable semi-density operator satisfies $\psi(n, w, v) \leq \psi(n+1, w, v)$. Also notice that the converse is not true; there exists $\{\rho_n\}_{n=1}^{\infty}$ which is not increasing but corresponding ψ is increasing with respect to n . In fact, the matrix

$$\rho := \begin{pmatrix} 3 & 2 & 0 & \cdots \\ 2 & 1 & 0 & \cdots \\ 0 & 0 & 0 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is not positive, but $\langle w + v | \rho(w + v) \rangle \geq 0$ and $\langle w + iv | \rho(w + iv) \rangle \geq 0$ hold for every $w, v \in \{0,1\}^*$.

In contrast, our computability of operator is equivalent to that of its approximation function.

Proposition 3.2. ρ is computable if and only if $\psi(w, v) = \langle w | \rho v \rangle$ is computable (in the classical sense).

Proof. Suppose $\rho \in \mathcal{L}(\mathcal{H})$ is computable and let $\{\rho_n\}_{n=1}^\infty$ be an approximation of ρ . Then Schwarz inequality tells us

$$|\langle w | (\rho - \rho_n) v \rangle| \leq \|\rho - \rho_n\| < 2^{-n},$$

which shows that $\langle w | \rho_n v \rangle$ is an n -digit approximation of $\langle w | \rho v \rangle$.

Conversely, suppose $\psi(w, v) = \langle w | \rho v \rangle$ is computable, i.e. there is a computable function $\tilde{\psi} : \{0, 1\}^* \times \mathbb{N}^2 \rightarrow \mathbb{C}_q$ such that $|\psi(w, v) - \tilde{\psi}(w, v, n)| < 2^{-n}$ for any w, v, n . Then $\tilde{\varphi}(w, v, n) := \tilde{\psi}(w, v, \lceil \frac{w+v+n}{2} + 1 \rceil)$ is an approximation of ρ . In fact, let $\{\sigma_n\}_{n=1}^\infty$ be the sequence of operators induced by $\tilde{\varphi}$. then

$$\|\rho - \sigma_n\| \leq \|\rho - \sigma_n\|_{HS} \leq \sum_{w,v} |\psi(w, v) - \tilde{\varphi}(w, v, n)|^2 < 2^{-n}.$$

Here, $\|\cdot\|_{HS}$ is the Hilbert-Schmidt norm

$$\|\rho\|_{HS} = \sum_{w,v} |\langle w | \rho v \rangle|^2.$$

□

In the classical domain, a function is computable if and only if it is lower- and upper-semicomputable. The same thing can be said in our quantum modification.

Proposition 3.3. ρ is computable if and only if it is lower-semicomputable and upper-semicomputable.

Proof. Let ρ be lower- and upper-semicomputable. Also let $\{\rho_n\}_{n=1}^\infty$ and $\{\bar{\rho}_n\}_{n=1}^\infty$ be a lower- and upper approximation of ρ , respectively. Then $\psi(w, v) = \langle w | \rho v \rangle$ is computable since

$$\begin{aligned} |\langle w | (\rho - \bar{\rho}_n) v \rangle| &\leq \frac{1}{4} \sum_{k=0}^3 |\langle w + i^k v | (\rho - \bar{\rho}_n) (w + i^k v) \rangle| \\ &\leq \frac{1}{4} \sum_{k=0}^3 |\langle w + i^k v | (\rho_n - \bar{\rho}_n) (w + i^k v) \rangle| \\ &\xrightarrow{n \rightarrow \infty} 0 \end{aligned}$$

holds, and we can compute the right side of inequality successively for all n . This means we can construct a computable function $f : \mathbb{N} \rightarrow \mathbb{N}$ which makes $\tilde{\psi}(n, w, v) := \langle w | \rho_{f(n)} v \rangle$ an approximation of ρ . Notice that this proof is slightly different from classical one since lower- and upper-approximation of ρ itself is not a one of ψ .

Conversely, let ρ be computable. Then we can obtain a lower-approximation $\{\tilde{\rho}_n\}_{n=1}^\infty$ of ρ defining

$$\tilde{\rho}_n := \rho_n - 2^{-n+2} I.$$

In fact, $\|\rho - \tilde{\rho}_n\| \rightarrow 0$ so $\tilde{\rho}_n \rightarrow \rho$ in WOT. Using the inequality $\rho \leq \|\rho\| I$ it can be shown

$$\rho_n - \rho_{n+1} < 2^{-n+1} I.$$

Hence

$$\tilde{\rho}_n - \tilde{\rho}_{n+1} = \rho_n - \rho_{n+1} - 2^{-n+1} I \leq 0.$$

Obviously $\{\tilde{\rho}_n\}_{n=1}^\infty$ is induced by a computable function: let $\tilde{\psi}(w, v, n) := \psi(w, v, n) - 2^{-n+2} \delta_{ij}$. Upper-semicomputability of ρ can be shown in the same manner. □

We say a sequence $\{|\psi_n\rangle\}_{n=1}^\infty$ of states is *uniformly computable* if there is a recursive function $\tilde{\psi} : \mathbb{N}^2 \times \{0, 1\}^* \rightarrow \mathbb{C}_q$ such that

$$|\langle w|\psi_n\rangle - \tilde{\psi}(k, n, w)| < 2^{-k}$$

for every $k, n \in \mathbb{N}$ and $w \in \{0, 1\}^*$.

Let m be a lower-semicomputable semimeasure, and $\{|\psi_n\rangle\}_{n=1}^\infty$ be a uniformly computable sequence of states. If it holds $\langle w|\psi_n\rangle\langle\psi_n|v\rangle \in \mathbb{C}_q$ for every $w, v \in \{0, 1\}^*$ and $n \in \mathbb{N}$, then obviously an operator $\sum_n m(n)|\psi_n\rangle\langle\psi_n|$ is lower-semicomputable: in fact, $\{\sum_n \tilde{m}(k, n)|\psi_n\rangle\langle\psi_n|\}_{k=1}^\infty$ is its lower-approximation. It turns out that we can discard the last assumption.

Proposition 3.4. *Let $\{|\psi_n\rangle\}_{n=1}^\infty$ be a uniformly computable sequence of states, and m be a lower-semicomputable semimeasure. Then $\rho := \sum_n m(n)|\psi_n\rangle\langle\psi_n|$ is a lower-semicomputable semi-density operator.*

Proof. For every $k, n \in \mathbb{N}$ and $w \in \{0, 1\}^*$, let $|\psi_{k,n}\rangle$ be a vector (not necessarily a state) which is identified by an equation

$$\langle w|\psi_{k,n}\rangle = \tilde{\psi}(k + w, n, w).$$

Then it is routine to show that $\tilde{\rho}'_k := \sum_n \tilde{m}(n)|\psi_{k,n}\rangle\langle\psi_{k,n}|$ converges to ρ in WOT (actually it converges in norm). We can also show that $\tilde{\rho}_k := \tilde{\rho}'_k - 2^{-(k+1)}I$ forms a lower-approximation of ρ in the same manner as proposition 3.3. \square

It is still open whether the converse is also true or not. Formally, can we find a uniformly computable sequence $\{|\psi_n\rangle\}_{n=1}^\infty$ of states and a lower-semicomputable semimeasure m such that $\rho = \sum_n m(n)|\psi_n\rangle\langle\psi_n|$ for any lower-semicomputable semi-density operator ρ ? But at least, we expect that taking $\{|\psi_n\rangle\}_{n=1}^\infty$ as an orthonormal basis is *not* always possible, since otherwise there is no universal operator, as we see in proposition 3.12 and corollary 3.13.

We conclude this subsection with some examples which shows that some obvious property in classical domain fails to hold in our quantum version. In classical case, it is always possible to take a sequence of *positive* functions as a lower-approximation of semimeasure, since if ψ is a lower-approximation of m then so is $\varphi(x, k) := \max\{\psi(x, k), 0\}$. This is not always true in our quantum modification.

Examples 3.5 ([15]). There is a lower-semicomputable semi-density operator which cannot be approximated by any sequence of positive operators from below. In fact, let ρ be a rank-one projection of which nonzero eigenvector is $\frac{1}{2}|\lambda\rangle + \frac{\sqrt{3}}{2}|0\rangle$. Matrix representation of ρ is

$$\frac{1}{4} \begin{pmatrix} 1 & \sqrt{3} & 0 & \dots \\ \sqrt{3} & 3 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

Obviously ρ is computable, so it is lower-semicomputable. On the other hand, since ρ is rank-one projection, if there is σ such that $0 \leq \sigma \leq \rho$ then $\sigma = c\rho$ ($0 \leq c \leq 1$). But it holds that $\langle \lambda|\rho|\lambda\rangle \notin \mathbb{C}_q$ or $\langle 0|\rho|\lambda\rangle \notin \mathbb{C}_q$ for any $c \in \mathbb{R} \setminus \{0\}$.

The same thing happens even if we allow a lower-approximation function to take an algebraic number, as Gács proposed in [3]. The operator

$$\frac{1}{1+\pi^2} \begin{pmatrix} 1 & \pi & 0 & \dots \\ \pi & \pi^2 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

cannot be approximated by any sequence of positive operators from below. \square

3.2 Problem: the existence of a universal operator

Just like the classical case, we expect that there is a universal semi-density operator in the following sense.

Definition 3.6. A lower-semicomputable semi-density operator μ is universal if for any lower-semicomputable semi-density operator ν there is a real number $c_\nu > 0$ such that $c_\nu \nu \leq \mu$.

Unfortunately, our proof of the existence of a universal operator has somewhat weak form: namely, we need to assume some additional properties for each lower-approximation. We expect that these properties is derived from our definition.

Before stating the assumption and the proof, let us see the reason why we need such an additional assumption. In Gács [3], the following question is said to be solved positively in the same manner as the classical case, but it is not true.

Problem 3.7. Can we enumerate all lower-semicomputable semi-density operators effectively?

To see the difficulty of this problem, let us review a proof of theorem 2.1.

Proof of theorem 2.1. Let $\{\varphi_n\}_{n=1}^\infty$ be an effective enumeration of all partial recursive function. Consider the following algorithm:

Input $n \in \mathbb{N}$.

1. Let $\alpha_w := 0$ for every $w \in \{0, 1\}^*$.
2. Dovetail φ_n , regarding φ_n as a function from $\{0, 1\}^* \times \mathbb{N}$ to \mathbb{Q} . Whenever φ_n halts for an input $\langle w, k \rangle$, go to step 3.
3. Check whether the conditions $\varphi_n(w, k) \geq \alpha_w$ and $(\sum_{v \neq w} \alpha_v) + \varphi_n(w, k) \leq 1$ hold. If so, then let $\alpha_w := \varphi_n(w, k)$. Otherwise, do nothing. go back to step 2.

Let $\tilde{\psi}(w, t, n)$ be the value of α_w after the t -steps computation of the algorithm above for an input n . Obviously $\tilde{\psi}(-, -, n)$ is an lower-approximation of some lower-semicomputable semimeasure. $\tilde{\psi}$ can approximate any lower-semicomputable semimeasure from below, since any lower-approximation of a semimeasure is equal to some φ_n , and $\tilde{\psi}(-, -, n)$ approximates the same semimeasure from below. \square

When we naively interpret this proof into the quantum setting, the corresponding algorithm would be like this:

Input $n \in \mathbb{N}$.

1. Let $\alpha_{w,v} := 0$ for every $w, v \in \{0, 1\}^*$, and let ρ be an operator defined by $\langle w | \rho | v \rangle := \alpha_{w,v}$.
2. Dovetail φ_n , regarding φ_n as a function from $\{0, 1\}^* \times \{0, 1\}^* \times \mathbb{N}$ to \mathbb{C}_q . Whenever φ_n halts for an input $\langle w', v', k \rangle$, go to step 3.
3. Let ρ' be an operator defined by

$$\langle w | \rho' | v \rangle := \begin{cases} \varphi_n(w, v, k) & ((w, v) = (w', v')) \\ \overline{\varphi_n(w, v, k)} & ((w, v) = (v', w')) \\ \alpha_{w,v} & (otherwise). \end{cases}$$

Check whether the condition $\rho' \geq \rho$ and $\text{Tr} \rho' \leq 1$ holds. If so, then let $\alpha_{w',v'} := \varphi_n(w', v', k)$ and $\alpha_{v',w'} := \overline{\varphi_n(w', v', k)}$. Otherwise, do nothing. go back to step 2.

For $\rho \in \mathbb{M}^n(\mathbb{C}_q)$ it is always possible to decide whether $\rho \geq 0$ or not (see [16]), so step 3 always ends in finite time. Let $\tilde{\psi}(w, v, t, n)$ be the value of $\alpha_{w,v}$ after the t -steps computation of the algorithm above for an input n . The problem is that $\tilde{\psi}(-, -, -, n)$ generally does not approximate the same semi-density operator as which is approximated by φ_n from below.

There are at least two main difficulties to construct the algorithm. First, updating process easily fails to maintain the monotonicity of sequence of operators, as long as we try to change the coefficients of the matrix pointwisely. For example, let $\varphi_{n_0} : \{0, 1\}^* \times \{0, 1\}^* \times \mathbb{N} \rightarrow \mathbb{C}_q$ be a recursive function such that

$$\varphi_{n_0}(v, w, k) = \begin{cases} \frac{1}{2} & (w, v \in \{\lambda, 0\}) \\ 0 & (\text{otherwise}), \end{cases}$$

and $t(\lambda, \lambda, 0) \ll t(1, 1, 0) \ll t(\lambda, 1, 0) \ll t(\text{any other input})$, where $t(v, w, k)$ is the time needed to compute $\varphi_{n_0}(v, w, k)$. We would expect that ρ is updated as follows when we run the algorithm above for an input n_0 , but it is not true:

$$0 \rightarrow \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \rightarrow \frac{1}{2} \begin{pmatrix} 1 & 0 & 0 & \dots \\ 0 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \rightarrow \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 & \dots \\ 1 & 1 & 0 & \dots \\ 0 & 0 & 0 & \dots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

Actually ρ is never updated from the third step. Moreover, it turns out that for any $n \in \mathbb{N}$ the operator corresponds to $\lim_{k \rightarrow \infty} \tilde{\psi}(-, -, k, n)$ is diagonal. Hence, if we use the algorithm above, the expected-to-be-universal operator constructed in the same manner as the classical case is also diagonal, which cannot be universal (see proposition 3.13).

Second, as long as we initially set $\alpha_{w,v} := 0$ for every $w, v \in \{0, 1\}^*$, $\tilde{\psi}$ cannot be a lower-approximation of the operator described in the Example 3.5, since any $\tilde{\psi}(-, -, -, n)$ corresponds to a sequence of positive operators.

To avoid these problems, we assume some additional properties for each lower-approximation. This is an analogous approach to the one of Tadaki [15], in which the notion of *lower-computable semi-POVM* is introduced, and it is shown that a universal semi-POVM does exist. The properties are as follows:

1. For a lower-approximation $\{\rho_n\}_{n=1}^\infty$ of any lower-semicomputable operator, each ρ_n has a “finite matrix representation with respect to the computational basis”: that is, there is a recursive function $f : \mathbb{N} \rightarrow \{0, 1\}^*$ such that $P_{f(n)} \rho_n P_{f(n)} = \rho_n$, where $P_w := \sum_{v=\lambda}^w |v\rangle\langle v|$. this property enables us to encode each ρ_n to some natural number, and hence to avoid the difficulty to update the coefficients of the matrix pointwisely.
2. $\{\rho_n\}_{n=1}^\infty$ is a positive but “almost increasing” sequence: that is, there exists a computable density operator σ such that for every $n \in \mathbb{N}$ it satisfies the conditions $\rho_n \geq 0$ and $\rho_{n+1} - \rho_n \geq -\rho^{-(n+1)}$. This is more restrictive than our definition since a sequence $\{\rho_n - \sigma^{-n}\}$ is always increasing and approximates the same element.

It turns out that an operator $\frac{1}{2}(\rho + \sigma)$, which multiplicatively dominates ρ , is also lower-semicomputable semi-density and approximated by a sequence of positive operators. This property enables us to overcome an inability to find $n \in \mathbb{N}$ which makes $\tilde{\psi}$ a lower-approximation of certain operator.

Here we restate our assumption more formally. We would like to call it conjecture since these properties are expected to be derived from our definition.

Conjecture 3.8. For given lower-semicomputable semi-density operator ρ , there exists a sequence $\{\rho_n\}_{n=1}^\infty$ of operator which satisfies the following conditions:

1. $\rho_n \geq 0$, $\rho_n \xrightarrow{n \rightarrow \infty} \rho$ (WOT), and there is a density operator σ such that $\rho_{n+1} - \rho_n \geq -2^{-(n+1)}\sigma$.
2. There is a recursive function ψ and φ such that $\psi(w, v, n) = \langle w | \rho_n v \rangle$ and $\varphi(w, v) = \langle w | \sigma v \rangle$.
3. There is a recursive function $f : \mathbb{N} \rightarrow \{0, 1\}^*$ such that $P_{f(n)} \rho_n P_{f(n)} = \rho_n$.
4. For $\sigma_n := P_{f(n)} \sigma P_{f(n)}$, it holds that $\sigma_{n+1} \geq \sigma_n$.

Proposition 3.9. Assume the conjecture above is true. Then there exists a universal operator.

Proof. First, we show an easy, but crucial fact.

Claim. Let $\rho \in \tilde{S}(\mathcal{H})$ be lower-semicomputable, and $\{\rho_n\}_{n=1}^\infty$, σ , and f be operators and a function described in conjecture 3.8, respectively. Then an operator $\rho' := \frac{1}{2}(\rho + \sigma)$ is lower-semicomputable semi-density, and there exists a lower-approximation $\{\rho'_n\}_{n=1}^\infty$ of ρ' which satisfies the conditions $\rho'_n \geq 0$ and $P_{f(n)} \rho'_n P_{f(n)} = \rho'_n$.

In fact, let $\rho'_n := \frac{1}{2}(\rho_n + (1 - 2^{-n})\sigma_n)$. Then the conditions $\rho'_n \geq 0$ and $P_{f(n)} \rho'_n P_{f(n)} = \rho'_n$ obviously hold, and showing $\rho'_n \xrightarrow{n \rightarrow \infty} \rho'$ is also straightforward. $\{\rho'_n\}_{n=1}^\infty$ is increasing since from the condition 1 and 3 of the conjecture we get

$$\begin{aligned} \rho_{n+1} + (1 - 2^{-(n+1)})\sigma_{n+1} &= P_{f(n+1)}(\rho_{n+1} + (1 - 2^{-(n+1)})\sigma)P_{f(n+1)} \\ &\geq P_{f(n+1)}(\rho_n + (1 - 2^{-n})\sigma)P_{f(n+1)} \\ &= \rho_n + (1 - 2^{-n})\sigma_{n+1}, \end{aligned}$$

and using the condition 4 of the conjecture we get $\rho'_{n+1} \geq \rho'_n$.

Now consider the following algorithm. Here, we let $\mathcal{L}_q(\mathbb{C}^m)$ be the set of all $m \times m$ hermitian matrices of which each coefficient is in \mathbb{C}_q , and often identify an operator in $\mathcal{L}_q(\mathbb{C}^m)$ with that on \mathcal{H} in a canonical way.

Input $n \in \mathbb{N}$.

1. Let $v := 0$ ($v \in \mathcal{B}(\mathcal{H})$).
2. Dovetail φ_n , regarding φ_n as a function from \mathbb{N} to $\bigcup_{m \in \mathbb{N}} \mathcal{L}_q(\mathbb{C}^m)$. Whenever φ_n halts for an input k , go to step 3.
3. Check whether the conditions $\varphi_n(k) \geq v$ and $\text{Tr} \varphi_n(k) \leq 1$ hold. If so, then let $v := \varphi_n(k)$. Otherwise, do nothing. go back to step 2.

Let $\tilde{\psi}(n, t)$ be the value of v after the t -steps computation of the algorithm above for an input n . It can be shown that for every $n \in \mathbb{N}$ there exists $v_n \in \mathcal{B}(\mathcal{H})$ such that $\tilde{\psi}(n, t) \xrightarrow{t \rightarrow \infty} v_n$ in WOT. Obviously v_n is lower-semicomputable semi-density.

Now let $\rho \in \tilde{S}(\mathcal{H})$ be lower-semicomputable, and $\{\rho_t\}$, σ , and f be operators and a function described in conjecture 3.8, respectively. Then there is $n \in \mathbb{N}$ such that $v_n = \frac{1}{2}(\rho + \sigma)$. In fact, there exists $n \in \mathbb{N}$ such that $\varphi_n(t) = \rho'_t$, where ρ'_t is described in the claim above, and $\{\tilde{\psi}(n, t)\}_{t=1}^\infty$ is also a lower-approximation of ρ' . This can be shown using the fact that $\{\tilde{\psi}(n, t)\}_{t=1}^\infty = \{\rho'_{g(n)}(t)\}_{t=1}^\infty$, where $g : \mathbb{N} \rightarrow \mathbb{N}$ is an appropriate nondecreasing, unbounded function (We assume $\rho'_1 = 0$ without loss of generality).

Finally, we can show $\mu := \sum_{n=1}^\infty 2^{-n} v_n$ is universal in the following way:

- Since $\{\sum_{k=1}^n 2^{-k} v_k\}_{n=1}^\infty$ is a Cauchy sequence, μ is well-defined semi-density operator.

- μ dominates any lower-semicomputable semi-density operator ρ , since there is $n \in \mathbb{N}$ such that $2v_n = \rho + \sigma$, so $\rho \leq 2v_n \leq 2^{(n+1)}\mu$.
- μ is also lower-semicomputable since $\varphi(n, w, v) := \sum_{k=1}^n 2^{-k} \psi_k(n, w, v)$ is its lower-approximation, where ψ_k is a lower-approximation of v_k . In fact, for given $\epsilon > 0$ and a unit vector $|\psi\rangle \in \mathcal{H}$, there is an integer k_0 such that $\|\sum_{k=k_0+1}^{\infty} 2^{-k} v_k\| < \epsilon$, and there is an integer $k_1 \geq k_0$ such that $\langle \psi | (v_n - v_{nk_1}) \psi \rangle < \frac{2^n}{k_0} \epsilon$ for every $n \leq k_0$. Hence

$$\begin{aligned}
\langle \psi | (\mu - \mu_{k_1}) \psi \rangle &\leq \langle \psi | \mu \psi \rangle - \sum_{n=1}^{k_0} \langle \psi | v_{nk_1} \psi \rangle \\
&= \sum_{n=1}^{k_0} 2^{-n} \langle \psi | (v_n - v_{nk_1}) \psi \rangle + \langle \psi | (\sum_{n=k_0+1}^{\infty} 2^{-n} v_n) \psi \rangle \\
&< 2\epsilon,
\end{aligned}$$

so $\mu_n \rightarrow \mu$ in WOT. Obviously μ_n is increasing, and $\varphi(n, w, v) \in \mathbb{C}_q$ for every $n \in \mathbb{N}$ and $w, v \in \{0, 1\}^*$. \square

Once we prove the existence of universal semi-density operator, we can define quantum algorithmic entropy \overline{H} and \underline{H} in the same manner as Gács [3]:

$$\overline{H}(|\psi\rangle) = -\langle \psi | (\log \mu) \psi \rangle, \quad \underline{H}(|\psi\rangle) = -\log \langle \psi | \mu \psi \rangle.$$

The following proposition claims that \overline{H} and \underline{H} are the extensions of classical descriptive complexity.

Proposition 3.10. *Assume a universal operator μ exists. Then for any uniformly computable orthonormal system $\{|\psi_n\rangle\}_{n=1}^{\infty}$ (not necessarily a basis),*

$$K(n) = H(|\psi_n\rangle) + O(1),$$

where $H = \overline{H}$ or $H = \underline{H}$. In particular, for any $w \in \{0, 1\}^*$,

$$K(w) = H(|w\rangle) + O(1).$$

We strongly expect this equation holds, since there is an analogous consequence about qubit complexity defined by Berthiaume et al [2]. Our eventual goal is to examine the equivalence of qubit complexity and Gács' quantum algorithmic entropy, so this is a very minimum requirement for us.

Proposition 3.11 ([8]). *For any $w \in \{0, 1\}^*$,*

$$C(w) = QC(|w\rangle) + O(1).$$

Here, $QC(|\psi\rangle)$ is the qubit complexity of $|\psi\rangle$ (see [2]).

Proof of proposition 3.10. The proof is completely the same as the one in [3], but it is valid in our definition. The function $f(n) = \langle \psi_n | \mu \psi_n \rangle$ is lower-semicomputable with $\sum_n f(n) = \text{Tr } \mu \leq 1$, hence $K(n) \leq \underline{H}(n) + O(1)$.

On the other hand, the semi-density operator $\rho = \sum_n \mathbf{m}(n) |\psi_n\rangle \langle \psi_n|$ is lower-semicomputable (proposition 3.4), so

$$K(n) = \langle \psi_n | (-\log \rho) \psi_n \rangle \geq \langle \psi_n | (-\log \mu) \psi_n \rangle + O(1) = \overline{H}(|\psi_n\rangle) + O(1).$$

Notice that the inequality above holds since $g(x) = \log x$ is an operator monotone function. Finally, for any state $|\psi\rangle$ we have an inequality $\overline{H}(|\psi_n\rangle) \geq \underline{H}(|\psi_n\rangle)$, which completes the proof. \square

Remark. The statement which makes the problem in the definition by Gács is “ $\sum_n f(n) = \text{Tr } \mu \leq 1$ ”. His universal operator is actually the sequence $\{\mu_n\}_{n=1}^\infty$ of matrices, and μ in the definition of f is actually some appropriate μ_{k_n} . The value of k_n cannot be the same for all $n \in \mathbb{N}$ since $\{|\psi_n\rangle\}_{n=1}^\infty$ is an infinite sequence of orthogonal states. So we do not know how to show $\sum_n f(n) \leq 1$. Also we do not know what the statement “ $\rho = \sum_n \mathbf{m}(n)|\psi_n\rangle\langle\psi_n|$ is lower-semicomputable” means in his finite dimensional formulation. In short, the proof is stated as if we work on an infinite dimensional setting, and it is one of the main reasons we try to modify his definition into an infinite dimensional version.

We conclude this subsection with an easy corollary which evokes an analogous fact in classical domain: for a universal semimeasure \mathbf{m} and an infinite recursive set $\{w_n\} \subset \{0, 1\}^*$, a function $\mathbf{m}'(n) := \mathbf{m}(w_n)$ is again universal. The following seems to be the quantum version of this fact.

Corollary 3.12. *Assume a universal operator μ exists. Let $\{|\psi_n\rangle\}$ be a uniformly computable orthonormal system of \mathcal{H} . Then a function $\mathbf{m}_\psi(n) := \langle\psi_n|\mu\psi_n\rangle$ is a universal semimeasure.*

3.3 μ is not diagonal

At first glance, one might expect that an operator $\mu_1 := \sum_i \mathbf{m}(i)|i\rangle\langle i|$ is universal. In fact, for corollary 3.12, diagonal entries of universal operator should form a universal semimeasure, so it would be natural to question whether the simplest operator with this property, i.e. a diagonal one, is universal.

It is rather disappointing if the answer is yes, since in this case H is very simple combination of classical complexity:

$$\overline{H}(\sum_w \alpha_w |w\rangle) = - \sum_w |\alpha_w|^2 \log \mathbf{m}(w), \quad \underline{H}(\sum_w \alpha_w |w\rangle) = - \log \sum_w |\alpha_w|^2 \mathbf{m}(w).$$

For good or bad, it turns out μ_1 is not universal.

Proposition 3.13. *There is a lower-semicomputable semi-density operator which cannot be multiplicatively dominated by μ_1 .*

Proof. Assume μ_1 is universal, and let $|\psi_n\rangle := 2^{-\frac{n}{2}} \sum_{l(w)=n} |w\rangle$. Then for corollary 3.12, the function $\overline{\mathbf{m}}(n) := \langle\psi_n|\mu_1\psi_n\rangle = 2^{-n} \sum_{l(w)=n} \mathbf{m}(w)$ must be a universal semimeasure, which is not true. In fact, The function $2^n \overline{\mathbf{m}}$ is also a lower-semicomputable semimeasure which cannot be dominated by $\overline{\mathbf{m}}$. \square

We can derive a more general fact which tells us the set of eigenspaces and eigenvalues of μ should have certain “incomputability”.

Corollary 3.14. *There is no uniformly computable orthonormal basis $\{|\psi_n\rangle\}_{n=1}^\infty$ of \mathcal{H} and lower-semicomputable semimeasure m which makes an operator $\sum_n m(n)|\psi_n\rangle\langle\psi_n|$ universal.*

Proof. Let $|\varphi_n\rangle := 2^{-\frac{n}{2}} \sum_{l=2^{n-1}}^{2^n-1} |\psi_l\rangle$ and consider the same argument as the previous proof. \square

4 Discussion and perspective

We defined an infinite dimensional modification of lower-semicomputability of density operators by Gács, and examined their properties, especially the well-definedness of his quantum algorithmic entropy. We needed some additional assumption to establish well-defined notion, and checking whether this assumption can be eliminated or not is left as a future task.

In particular, the condition 1 of conjecture 3.8 could be relaxed or eliminated in some way. As we saw in the proof of proposition 3.9, for given ρ , we only needed to find v_n which multiplicatively dominates

ρ , not which is equal to ρ itself. The necessity of the condition 1 has arisen from example 3.5, but this operator is actually dominated by $|\lambda\rangle\langle\lambda| + |0\rangle\langle 0|$, which is apparently positively-approximated lower-semicomputable operator. It is likely that there is some nice algorithm to find a dominating, positively-approximated operator, for given ρ .

Acknowledgment

The author would like to thank his supervisor Prof. Masahito Hasegawa for constant support and encouragement, Prof. Kohtaro Tadaki, Prof. Peter Gács and Prof. Takayuki Miyadera for helpful discussions, and two anonymous referees for the valuable comments.

References

- [1] Shor, P. W. Algorithms for quantum computation: Discrete logarithms and factoring, Proceedings of the 35th Annual Symposium on Foundations of Computer Science (Los Alamitos, CA) (G. Goldwasser, ed.), IEEE Computer Society Press, 1994, 124-134.
- [2] Berthiaume, A.; van Dam, W.; Laplante, S. Quantum Kolmogorov complexity. *J. Comput. System Sci.* 2001, 63, 201-221.
- [3] Gács, P. Quantum algorithmic entropy. *J. Phys. A: Math. Gen.* 2001, 34, 6859-6880.
- [4] Vitányi, P.M.B. Quantum Kolmogorov complexity based on classical descriptions. *IEEE Trans. Inform. Theory* 2001, 47, 2464-2479.
- [5] Miyadera, T. Quantum Kolmogorov complexity and information-disturbance theorem. *Entropy* 2011, 13, 778-789.
- [6] Benatti, F.; Krüger, T.; Müller, M.; Siegmund-Schultze, R.; Szkoła, A. Entropy and quantum Kolmogorov complexity: A quantum Brudno's theorem. *Commun. Math. Phys.* 2006, 265, 437-461.
- [7] Müller, M. Strongly universal quantum Turing machines and invariance of Kolmogorov complexity. *IEEE Trans. Inform. Theory* 2008, 54, 763-780.
- [8] Müller, M. On the quantum Kolmogorov complexity of classical strings. *Int. J. Quant. Inf.* 2009, 7, 701-712.
- [9] Müller, M. Quantum Kolmogorov Complexity and the Quantum Turing Machine. Ph.D. thesis 2007.
- [10] Müller, M.; Rogers, C. Quantum Bit Strings and Prefix-Free Hilbert Spaces. Proceedings of the 2008 International Conference on Information Theory and Statistical Learning, ITSL 2008 (Las Vegas, Nevada, USA) CSREA Press 2008, 106-111.
- [11] Neumann, J. von. *Mathematische Grundlagen der Quantenmechanik*. Springer, Berlin, 1932.
- [12] Bernstein, E.; Vazirani, U. Quantum complexity theory. *SIAM J. Comput.* 1997, 26, 1411-1473.
- [13] Li, M.; Vitányi, P. *An Introduction to Kolmogorov Complexity and Its Applications*. Springer 2008.
- [14] Downey, R.; Hirschfeldt, D. *Algorithmic Randomness and Complexity*. Springer 2010.
- [15] Tadaki, K. An extension of Chaitin's halting probability Ω to a measurement operator in an infinite dimensional quantum system. *Math. Log. Quart.* 2006, 52, 419-438.
- [16] Korte, B.; Vygen, J. *Combinatorial Optimization*. Springer 2012.
- [17] Nielsen, M.; Chuang, I. *Quantum Computation and Quantum Information*. Cambridge University Press 2010.
- [18] Heinosaari, T.; Ziman, M. *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press 2011.
- [19] Conway, J. *A Course in Functional Analysis*. Springer 1990.

Parallelized adiabatic gate teleportation

Kosuke Nakago, Michal Hajdušek, Shojun Nakayama, and Mio Murao

Department of Physics Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

To investigate how a causally ordered gate sequence can be parallelized in adiabatic implementations of quantum computation, we extend adiabatic gate teleportation (AGT), a model of quantum computation recently proposed by D. Bacon and S. T. Flammia, Phys. Rev. Lett. **103**, 120504 (2009), to a form deterministically simulating parallelized gate teleportation, which is achievable only by postselection. We introduce a *twisted Hamiltonian*, which is a Hamiltonian of the Heisenberg-type interaction where the coordinates of the second qubit is twisted according to a unitary gate. We develop *parallelized adiabatic gate teleportation* (PAGT) where a sequence of unitary gates is performed in a single step of the adiabatic process. In PAGT, the necessary time for the adiabatic evolution implementing a sequence of gates increases quadratically, however it maps causal order of gates to spatial order of interactions specified by the final Hamiltonian. Using this property, we present a controlled-PAGT scheme to manipulate the order of gates by a control-qubit. In the controlled-PAGT scheme, two differently ordered unitary operations $U^{(1)}U^{(2)}$ and $U^{(2)}U^{(1)}$ are coherently performed depending on the state of a control-qubit by simultaneously applying the twisted Hamiltonians of $U^{(1)}$ and $U^{(2)}$. We show that the twisted Hamiltonian has an ability to perform the transposed unitary gate in addition to the unitary gate and time reverse transformations represented by the transposed unitary gate enable deterministic simulation of an postselected event of parallelized gate teleportation. The details are presented in arXiv:1310.4061.

1 Introduction

The quantum circuit model is a standard model of quantum computation describing the relationship between input and output by a sequence of elementary gates. This model is widely used since it is universal and quantum circuits have a good correspondence to logic circuits used in classical computation. However at the same time, there is a restriction that elementary gates have to be performed without creating any loops in the circuit and the ordering of the gates can be determined in a partial order. The partial order determines the causal structure of a gate sequence. Thus only operations with definite causal order can be performed, whereas such a restriction may not be necessary in quantum mechanics as pointed out in [1].

In [1], an operation beyond causally ordered quantum computation called quantum switch has been investigated. Quantum switch is a super-map of which input is two different single-qubit unitary gates $U^{(1)}$ and $U^{(2)}$, and output is a two-qubit controlled-unitary that coherently performs two differently ordered unitary operations $U^{(1)}U^{(2)}$ and $U^{(2)}U^{(1)}$ depending on the state of a control-qubit. It was proven that quantum switch cannot be implemented within the quantum circuit model with a fixed causal order, if each of $U^{(1)}$ and $U^{(2)}$ is allowed to be used only once.

In this paper, we show that by implementing the quantum computation adiabatically, one can map the causal order of gates to the spatial order in the construction of a Hamiltonian, and that the causal order can be manipulated by arranging the spatial order of the Hamiltonian. To achieve this task, we extend adiabatic gate teleportation (AGT), which is a model of quantum computation proposed by Bacon and Flammia [2].

To implement a unitary gate operation U in AGT, the Hamiltonian $H(s)$ of the system consisting of three qubits, an input qubit, a mediating qubit and an output qubit, is slowly changed from the initial Hamiltonian H_{ini} to the final Hamiltonian H_{fin} as given by $H(s(t)) = (1 - s(t))H_{ini} + s(t)H_{fin}$, where $H_{ini} = I \otimes H_U^{AGT}$ and $H_{fin} = H_I^{AGT} \otimes I$ and the time dependent parameter $s(t)$ varies from 0 to 1, using an interaction Hamiltonian $H_U^{AGT} = -\omega(I \otimes U)(X \otimes X + Z \otimes Z)(I \otimes U^\dagger)$ with a coupling constant ω . In AGT, computation is performed in the degenerate subspace of ground states (thus it is also considered as open loop holonomic computation [3]). It starts by preparing an initial state which is a ground state of the initial Hamiltonian, and ends with a ground state of the final Hamiltonian, which is considered to be the solution of a given problem.

In spite of its naming, AGT is not a scheme based on quantum teleportation [4] that requires to be achieved by entanglement assisted local operations and classical communications (LOCC), since AGT requires direct interactions between the qubits. However, AGT deterministically implements a map that is achieved by postselected gate teleportation of an arbitrary unitary gate [5] by using adiabatic dynamics of an interacting multi-qubit ground state. This association of deterministically “simulating” a postselection event of a measurement using adiabatic dynamics is useful for analyzing the role of causal order and parallelizability in quantum computation.

A sequence of unitary gates can be probabilistically implemented in a *parallel* manner by combining multiple gate teleportation [6]. Each unitary gate operation can be simultaneously performed for different resource states and all the measurements can be simultaneously performed in this case. However, the original AGT scheme proposed by [2] cannot simulate this parallelized property of postselected gate teleportation. A sequence of unitary gate operations has to be performed by *sequentially* applying the AGT scheme in time for each gate operation following the causal structure of the gate sequence.

2 Main results

Parallelized AGT (PAGT)

To simulate parallelized gate teleportation, we modify the Hamiltonian of the original AGT scheme to use the Hamiltonians of the *Heisenberg-type* spin interactions. We introduce *twisted Hamiltonian* H_U given by $H_U := -\omega(I \otimes U)(I \otimes I + X \otimes X - Y \otimes Y + Z \otimes Z)(I \otimes U^\dagger) = -4\omega(I \otimes U)|\Phi\rangle\langle\Phi|(I \otimes U^\dagger)$ where $|\Phi\rangle := (|00\rangle + |11\rangle)/\sqrt{2}$ is a maximally entangled state, as a resource to implement a unitary gate denoted by U in our parallelized AGT (PAGT). Comparing to the interaction Hamiltonian used in AGT, H_U contains the terms $I \otimes I$ and $-Y \otimes Y$. The first term only gives an energy shift of a ground state, but the second term $-Y \otimes Y$ is necessary for parallelization. Similarly to AGT, we consider a three-qubit system consisting of an arbitrary input qubit $|\phi\rangle$ and a maximally entangled state $|\Phi\rangle$. The Hamiltonians for the adiabatic evolution to implement U on the third qubit for an input state $|\phi\rangle$ are given by $H_{ini} = I \otimes H_U$ and $H_{fin} = H_I \otimes I$. The initial state $|\phi\rangle \otimes |\Phi\rangle$ is a ground state of H_{ini} and the adiabatic evolution $H(s(t))$ leads to a ground state of H_{fin} given by $|\Phi\rangle \otimes U|\phi\rangle$. The initial and final Hamiltonians of PAGT given by H_U provides the same effects as AGT for simply implementing a single gate operation U .

We construct a PAGT scheme where consecutive gates $U^{(L)} \dots U^{(2)}U^{(1)}$ are implemented in a single adiabatic shift by using the twisted Hamiltonians. For simplicity we show a case of $L = 2$. The main idea originates from the quantum circuit representation of parallelized multiple gate teleportation using a five-qubit system achieved by postselection shown in Fig. 1 (a). The PAGT scheme mimics this circuit by dragging the whole system adiabatically using the PAGT Hamiltonians. To implement $U^{(2)}U^{(1)}$, we simply add two twisted Hamiltonians, $H_{U^{(1)}}$ for the 2nd and 3rd qubits and $H_{U^{(2)}}$ for the 4th and 5th qubits for H_{ini} . H_{ini} is represented by $H_{ini} = H_{U^{(1)}}^{23} + H_{U^{(2)}}^{45}$ where H_U^{ij} represents a product of a twisted

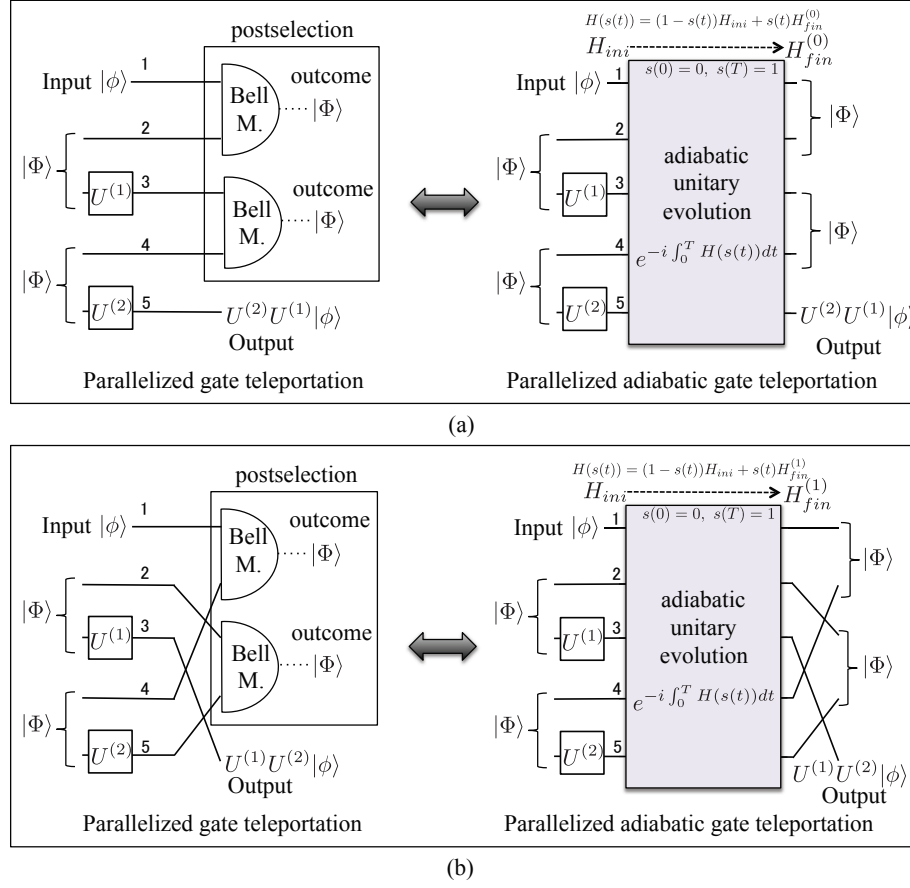


Figure 1: Circuit representations of parallelized gate teleportation for $L = 2$ and corresponding circuits representation of PAGT. (a) A circuit to implement $U^{(2)}U^{(1)}$. (b) A circuit to implement $U^{(1)}U^{(2)}$.

Hamiltonian H_U of the i th and j th qubits and I for other qubits. For H_{fin} , we add H_I for the 1st and 2nd qubits, and also for the 3rd and 4th qubits. We denote this final Hamiltonian by $H_{fin}^{(0)}$ corresponding to Figure 1 (a). $H_{fin}^{(0)}$ is represented by $H_{fin}^{(0)} = H_I^{12} + H_I^{34}$.

Thus the causally ordered operation $U^{(2)}U^{(1)}$ is implemented by *simultaneously* applying multiple twisted Hamiltonians in a single adiabatic shift, while the time parameter $s(t)$ in $H(s(t))$ does not affect the causal order of the desired operation. Note that H_{fin} is independent of the form of the unitary gates, and it only specifies the order of gate operations. If we choose the final Hamiltonian to be $H_{fin} = H_{fin}^{(1)} = H_I^{14} + H_I^{25}$ but using the same H_{ini} , the resulting PAGT scheme implements a differently ordered operation $U^{(1)}U^{(2)}$ corresponding to the case of Figure 1 (b).

In PAGT, the causal order of gate operations is mapped to the spatial order of interactions in the final Hamiltonian, whereas information of which gates to be applied is encoded in twisted angles of in the initial twisted Hamiltonian. All twisted Hamiltonians are simultaneously applied, although the speed of adiabatic evolution should be slowed down due to decreasing energy gaps. Thus PAGT does not contribute to speed up performing a sequence of gate operations by parallelizing, but to eliminate the control of causal order of gate operations in the time domain. Using this property, we present the controlled-

PAGT scheme that performs the controlled-unitary operations implemented by quantum switch.

Controlled-PAGT (C-PAGT)

We consider a six-qubit system consisting of a control-qubit represented by a subscript C and a five-qubit system introduced in the PAGT scheme. The initial Hamiltonian of controlled-PAGT H_{ini}^{C-PAGT} for two gate operations $U^{(1)}$ and $U^{(2)}$ is a sum of two twisted Hamiltonians of the six-qubit system given by $H_{ini}^{C-PAGT} = I_C \otimes (H_{U^{(1)}}^{23} + H_{U^{(2)}}^{45})$. The final Hamiltonian is given by $H_{fin}^{C-PAGT} := |0\rangle\langle 0|_C \otimes H_{fin}^{(0)} + |1\rangle\langle 1|_C \otimes H_{fin}^{(1)}$, where $H_{fin}^{(0)}$ and $H_{fin}^{(1)}$ are the final Hamiltonians used in the PAGT scheme. Using this final Hamiltonian, the control qubit manipulates the causal order of gate operations determined by $H_{fin}^{(0)}$ or $H_{fin}^{(1)}$. Note that this final Hamiltonian is the resource to determine just the order of gate operations but independent of applied gates themselves. After the adiabatic evolution by $H(s(t))$ from H_{ini}^{C-PAGT} to H_{fin}^{C-PAGT} , we perform a CSWAP (controlled-SWAP) operation between the 2nd and 4th qubits controlled by the control-qubit followed by another CSWAP operation between the 1st and the 3rd qubits. Then the controlled-PAGT scheme implements a unitary gate $|0\rangle\langle 0|_C \otimes U^{(2)}U^{(1)} + |1\rangle\langle 1|_C \otimes U^{(1)}U^{(2)}$ introduced in [1], by simultaneously applying the twisted Hamiltonians $H_{U^{(1)}}$ and $H_{U^{(2)}}$. In this case, the input state $|0\rangle_C \otimes |\phi\rangle + |1\rangle_C \otimes |\phi\rangle$ evolves to $|0\rangle_C \otimes U^{(2)}U^{(1)}|\phi\rangle + |1\rangle_C \otimes U^{(1)}U^{(2)}|\phi\rangle$. It is shown in [1] that implementing this operation in a standard quantum circuit with definite causal structure requires to call one of single-qubit unitary gates $U^{(1)}$ or $U^{(2)}$ at least twice.

Analysis of the twisted Hamiltonian

We also investigate why the twisted Hamiltonian enables parallelized adiabatic gate teleportation. The twisted Hamiltonian H_U allows to simulate another gate teleportation property implementing a transposed unitary gate operation U^T , which is not possible by the original AGT scheme using a non-Heisenberg type interaction Hamiltonian. We analyze this ability to perform U^T and the relationship to the ability to deterministically simulating acausal operations represented by postselected gate teleportation. It turns out that we can interpret that time reverse transformations represented by the transposed unitary gate enable deterministic simulation of a postselected event of parallelized gate teleportation.

References

- [1] G. Chiribella, G.M. D'Ariano, P. Perinotti, and B. Valiron, Phys. Rev. A **88**, 022318 (2013).
- [2] D. Bacon and S.T. Flammia, Phys. Rev. Lett. **103**, 120504 (2009).
- [3] P. Zanardi and M. Rasetti, Phys. Lett. A **264**, 94 (1999).
- [4] C. H. Bennett, G. Brassard, C. Crépeau, R. Jozsa, A. Peres, W. K. Wootters, Phys. Rev. Lett. **70**, 1895 (1993).
- [5] G. Svetlichny, International Journal of Theoretical Physics, **50** 3903 (2011); C. H. Bennett, in Proceedings of QUPON, Wien, 2005, <http://www.research.ibm.com/people/b/bennetc/>; S. Lloyd *et al.*, Phys. Rev. Lett. **106**, 040403 (2011).
- [6] D. Gottesman and I. L. Chuang, Nature **402**, 390 (1999).
- [7] M. Hayashi, *Quantum information an introduction* (Springer, Berlin, 2006).
- [8] G. Chiribella, G. M. D'Ariano, and P. Perinotti, Phys. Rev. A **80**, 022339 (2009).

Globalness of separable maps in terms of time and space resources

Seiseki Akibue

Department of Physics Graduate School of Science
The University of Tokyo, Tokyo 113-0033, Japan
akibue@eve.phys.s.u-tokyo.ac.jp

Masaki Owari

NTT Communication Science Laboratories
NTT corporation, Kanagawa, 243-0198, JAPAN

Go Kato

Mio Murao

Department of Physics Graduate School of Science, The University of Tokyo, Tokyo 113-0033, Japan

We propose a new approach to analyze globalness of separable maps and to distinguish them from local operations and classical communications (LOCC) and non-separable maps. In this approach, all quantum operations are restricted in the local spaces, but classical communication connecting different times (super communication) described by fictitious *acausal classical correlations* without globally causal structure can be used as additional resources relaxing LOCC constraints on time. We define a map described by local operations and super communication (LOSC) and prove that the intersection of the set of LOSC and the set of completely positive trace preserving (CPTP) maps is equivalent to the set of separable maps. We show that LOSC is closely related to the framework of the process matrix introduced by Oreshkov et. al. for analyzing quantum correlations with no causal order. We also investigate the gap between separable maps and LOCC in a conventional approach, by analyzing the amount of entanglement, which relaxes LOCC constraints on space, required in entanglement assisted LOCC implementations of separable maps for the orthogonal basis state discrimination problems. Combining these two approaches, the globalness of separable maps can be interpreted as a correspondence between the time and space resources, namely, acausal classical correlations and entanglement.

1 Introduction and summary

When we perform a quantum information processing task between two spatially separated parties, transmission of quantum states (quantum communication) between the parties are necessary in general. However, reliable quantum communications is much harder than reliable classical communication due to decoherence. Therefore we want to reduce quantum communications as much as possible by using a class of quantum maps implementable without quantum communication. *Local operations and classical communication* (LOCC) is such a class of maps, which is implementable by a sequence of conditional local quantum measurements and classical communications of measurement outcomes.

In spite of its clear operational meaning, analysis of quantum information processing tasks under the restriction of LOCC is hard in general since the mathematical structure of LOCC is highly complicated [1]. A slightly larger class of quantum maps with a simpler mathematical structure called *separable maps* is often used instead of LOCC. For some quantum information processing tasks, e.g. convertibility of pure bipartite states [8], separable maps achieving the tasks are also LOCC. On the other hands, the gap between these two classes has been observed in quantum information tasks such as state discrimination [9, 10, 11] and entanglement distillation [12].

This gap between LOCC and separable maps can be considered as “*globalness*” of separable maps not included in LOCC and can be investigated in terms of additional resources required to perform

separable maps on top of LOCC. Entanglement, which is defined as non-local quantum correlation non-increasing under LOCC [2, 3, 4], has been frequently analyzed as the resource (*space resource*) to achieve global quantum maps together with LOCC [5, 6, 7]. For the state discrimination tasks, the globalness of separable maps discriminating a set of orthogonal product states can be characterized by the amount of entanglement required in an entanglement assisted LOCC implementation of the separable map.

This conventional approach is useful to find the existence of a gap between LOCC and separable maps. However, this approach does not fully characterize separable maps, since the gap between separable maps and non-separable maps cannot be characterized by the amount of entanglement. For example, just one-ebit is enough to implement a class of two-qubit entangling gates called controlled-unitary gates. Thus, it is meaningful to consider another resource to characterize separable maps. In this paper we propose a totally new approach by considering (fictitious) *acausal classical correlations* as a candidate of such resources to discriminate separable maps from both LOCC and non-separable maps.

In each step of a LOCC protocol, parties communicate each other's measurement outcome. The effects of these communications can be considered as a subclass of correlations having globally causal structure. We consider a generalization of communications to correlations without globally causal structure, which is referred to as acausal classical correlations in this paper. We interpret acausal classical correlations as "*time resource*" on top of LOCC in contrast to space resource. We define a new class of quantum maps called *local operation and super communication* (LOSC) and show that LOSC is closely related to a class of separable maps and also to the framework of the *process matrix* introduced by [13] for analyzing quantum correlations with no causal order. We note that more general acausal correlations not restricted in classical correlations have been considered as resources for information processing in the context of closed time-like curves (CTCs) predicted by general relativity. For example, post-selected closed time like curves (pCTCs) [14, 15, 16] are shown to remarkably enhance the performance of quantum computation [16].

In this paper, we investigate globalness of separable maps in terms of acausal classical correlations, and also entanglement necessary to discrimination of product basis, and derive the following results:

1. We prove that the intersection of the class of LOSC and that of completely positive trace preserving (CPTP) maps is equivalent to that of separable maps. We also show that LOSC can be simulated by local post-selection without communication with constant probability. This result gives a new characterization of separable maps in terms of acausal classical correlations.
2. In quantum mechanics, correlations without globally causal structure between parties are allowed [13] in contrast to classical mechanics. We analyze when introduced time resources, acausal classical correlations, can be physically realizable in quantum mechanics. For this purpose, we investigate a class of classical correlations, which we call *valid channels*, that are consistent with quantum mechanics. We define valid channels as classical correlations with which all local instruments generate CPTP maps, and prove the one-to-one correspondence between a valid channel and a classical process matrix. We further prove that a valid channel only generates one-way LOCC in a bipartite system. This result indicates that acausal classical correlations required to generate separable maps in bipartite systems without entanglement also generates at least one non-CPTP map.
3. Separable maps are implementable by using local operations and particular types of acausal classical correlations (LOSC) without using shared entanglement (space resources). On the other hand, shared entanglement is necessary if we implement separable maps by using local operations and classical correlations with globally causal structure (LOCC) or by using local operations and one-way classical communications (one-way LOCC). There is a tradeoff between the restrictions

on classical correlations and the requirements on shared entanglement. As an analysis of entanglement resources, we derive an amount of the optimal entanglement resource that is necessary to perfectly discriminate orthonormal basis states by one-way LOCC. We construct a two-way LOCC protocol which discriminates product basis states and consumes less amount of entanglement resources than the best one-way LOCC protocol.

The results 1, 2, and 3 will be derived in Section 2, 3, and 4, respectively.

2 Analysis of time resource

In this section, we introduce a framework describing classical correlations between spatially separated parties under an assumption called *valid probability conditions* and define LOCC as a class of quantum maps described by local operations and classical correlations satisfying the valid probability conditions. We analyze separable maps by investigating properties of LOCC. For simplicity, we mainly consider bipartite cases in this paper, but the extension of LOCC to multipartite cases is straightforward.

We assume that Alice receives a quantum state in Hilbert space \mathcal{H}_1 and classical information denoted by x , performs an instrument $\{A_a^{(x)}\}$ depending on x , and send a measured quantum state in Hilbert space \mathcal{H}_2 and a measurement outcome a . Bob does the same by using an instrument $\{B_b^{(y)}\}$ where y denotes classical information and b denotes a measurement outcome. Note that $\{A_a^{(x)}\}$ and $\{B_b^{(y)}\}$ are Choi-Jamiołkowski (CJ) matrices of quantum instruments, i.e.

$$\{A_a^{(x)}\} \in QI(\mathcal{H}_1 : \mathcal{H}_2), \quad \{B_b^{(y)}\} \in QI(\mathcal{H}_3 : \mathcal{H}_4), \quad (1)$$

where QI is the set of CJ matrices of quantum instruments defined by

$$QI(\mathcal{H}_1 : \mathcal{H}_2) := \left\{ \{M_m^{(c)}\} \mid \left(\forall c, \forall m, M_m^{(c)} \in \text{Pos}(\mathcal{H}_1 \otimes \mathcal{H}_2) \right) \wedge \left(\forall c, \text{tr}_2 \left[\sum_m M_m^{(c)} \right] = \mathbb{I}_1 \right) \right\}. \quad (2)$$

We consider a classical communication channel between Alice and Bob where input classical information is given by a, b and output classical information is given by x, y . In LOCC, a classical communication channel is considered noiseless and the direction of communications is determined by the causal order between the local instruments $\{A_a^{(x)}\}$ and $\{B_b^{(y)}\}$, e.g. Alice can signal to Bob if the event of Alice sending classical information a is in the causal past of the event of Bob receiving classical information y . We extend the notion of classical communication to classical correlation $p(x, y|a, b)$, which includes a usual classical communication channel as a special case. We assume that the classical correlation is given by a valid conditional probability distribution satisfying

$$p(x, y|a, b) \geq 0 \quad (3)$$

$$\forall a, \forall b, \sum_{x, y} p(x, y|a, b) = 1. \quad (4)$$

We call these two conditions *valid probability conditions* and classical correlations that satisfy the valid probability conditions *super communications*. The total map obtained by the classical correlations and local instruments is given by

$$\sum_{x, y, a, b} p(x, y|a, b) A_a^{(x)} \otimes B_b^{(y)}. \quad (5)$$

A super communication is not based on a causal order, but it has an interesting physical meaning presented in subsection 2.2 and it provides a new interpretation of the gap between the class of separable maps and that of LOCC. We give an example of the super communication.

Example: A noiseless classical communication channel from Alice to Bob described by

$$p(x, y|a, b) = 1 \text{ if } x = 0 \wedge y = a, \quad p(x, y|a, b) = 0 \text{ else.} \quad (6)$$

This classical correlation gives a one-way LOCC protocol such that

$$\sum_{a,b} A_a^{(0)} \otimes B_b^{(a)} = \sum_a A_a \otimes B^{(a)}, \quad (7)$$

where $A_a = A_a^{(0)}$ and $B^{(a)} = \sum_b B_b^{(a)}$. This protocol is implemented by a procedure such that Alice first performs an instrument $\{A_a\} \in QI(\mathcal{H}_1 : \mathcal{H}_2)$, sends her output a to Bob and Bob performs a CPTP map $B^{(a)} \in CPTP(\mathcal{H}_3 : \mathcal{H}_4)$ depending on her output, where $CPTP$ is the set of CJ matrices of CPTP maps defined by

$$CPTP(\mathcal{H}_1 : \mathcal{H}_2) := \{M \in \text{Pos}(\mathcal{H}_1 \otimes \mathcal{H}_2) | \text{tr}_2[M] = \mathbb{I}_1\}. \quad (8)$$

2.1 LOSC and separable maps

We first present a formal definition of a class of maps LOSC for bipartite cases. The definition of LOSC for multipartite cases is given later.

Definition 1. Bipartite LOSC is a linear completely positive map $\Gamma : \mathbf{Pos}(\mathcal{H}_1 \otimes \mathcal{H}_3) \rightarrow \mathbf{Pos}(\mathcal{H}_2 \otimes \mathcal{H}_4)$ whose CJ matrix is given by

$$\Gamma = \sum_{a,b} A_a^{(b)} \otimes B_b^{(a)}, \quad (9)$$

where $\{A_a^{(b)}\} \in QI(\mathcal{H}_1 : \mathcal{H}_2)$ and $\{B_b^{(a)}\} \in QI(\mathcal{H}_3 : \mathcal{H}_4)$. We denote the set of LOSC as $LOSC(\mathcal{H}_1, \mathcal{H}_3 : \mathcal{H}_2, \mathcal{H}_4)$.

It is easy to see that the class of LOSC is a subset of the set of the total maps represented by Eq. (5) with super communications defined in Eq. (3,4) since we use only a special super communication in the definition of LOSC. In Theorem 1, we show that the two sets are actually identical. LOSC can be interpreted as the situation where a party performs local operations depending on measurement outcomes of the other parties. However a local operation of Alice can also depend on her own measurement outcome since Bob can send back Alice's measurement outcome through his local operation. We use such a property in the proof.

Theorem 1. LOSC is equivalent to the set of total maps with super communications.

We give the definition of multipartite LOSC in terms of the Kraus operator.

Definition 2. $LOSC(\mathcal{H}_1, \mathcal{H}_3, \dots, \mathcal{H}_{2n-1} : \mathcal{H}_2, \mathcal{H}_4, \dots, \mathcal{H}_{2n})$ is the set linear completely positive maps $\Gamma : \mathbf{Pos}(\mathcal{H}_1 \otimes \mathcal{H}_3 \otimes \dots \otimes \mathcal{H}_{2n-1}) \rightarrow \mathbf{Pos}(\mathcal{H}_2 \otimes \mathcal{H}_4 \otimes \dots \otimes \mathcal{H}_{2n})$ such that

$$\Gamma(\rho) = \sum_{\mathbf{m}} \left(E_{m_1}^{(\mathbf{m} \setminus \{m_1\} | 1)} \otimes \dots \otimes E_{m_n}^{(\mathbf{m} \setminus \{m_n\} | n)} \right) \rho \left(E_{m_1}^{(\mathbf{m} \setminus \{m_1\} | 1)} \otimes \dots \otimes E_{m_n}^{(\mathbf{m} \setminus \{m_n\} | n)} \right)^\dagger, \quad (10)$$

where $\mathbf{m} = \{m_1, \dots, m_n\}$, each m_i takes an element of finite set M_i in the summation, $\mathbf{m} \setminus \{m_i\} = \{m_1, \dots, m_{i-1}, m_{i+1}, \dots, m_n\}$ and

$$\forall i, \forall \mathbf{m} \setminus \{m_i\}, \sum_{m_i \in M_i} E_{m_i}^{(\mathbf{m} \setminus \{m_i\} | i)^\dagger} E_{m_i}^{(\mathbf{m} \setminus \{m_i\} | i)} = \mathbb{I}_{2i-1}, \quad (11)$$

where \mathbb{I}_{2i-1} is an identity operator on \mathcal{H}_{2i-1} .

Table 1: i) The table of the Kraus operators $\{E_a^{(b|A)}\}$ and ii) that of $\{E_b^{(a|B)}\}$. By simple calculation, we can certify $\Gamma_{sep}(\rho) = \sum_{a,b} (E_a^{(b|A)} \otimes E_b^{(a|B)}) \rho (E_a^{(b|A)} \otimes E_b^{(a|B)})^\dagger$ can discriminate the set of states, $\forall b; \sum_a E_a^{(b|A)\dagger} E_a^{(b|A)} = \mathbb{I}_A$ and $\forall a; \sum_b E_b^{(a|B)\dagger} E_b^{(a|B)} = \mathbb{I}_B$, which are the completeness condition in Eq.(11).

i)	<div><div>a</div><div>b</div></div>	1	2	3
	1	$ 1\rangle_{A'}\langle 0 _A$	$ 2\rangle_{A'}\langle 0 _A$	$ 3\rangle_{A'}\langle 0+1 _A$
	2	$ 8\rangle_{A'}\langle 1-2 _A$	$ 9\rangle_{A'}\langle 1 _A$	$ 4\rangle_{A'}\langle 0-1 _A$
	3	$ 7\rangle_{A'}\langle 1+2 _A$	$ 6\rangle_{A'}\langle 2 _A$	$ 5\rangle_{A'}\langle 2 _A$
ii)	<div><div>a</div><div>b</div></div>	1	2	3
	1	$ 1\rangle_{B'}\langle 0+1 _B$	$ 2\rangle_{B'}\langle 0-1 _B$	$ 3\rangle_{B'}\langle 2 _B$
	2	$ 8\rangle_{B'}\langle 0 _B$	$ 9\rangle_{B'}\langle 1 _B$	$ 4\rangle_{B'}\langle 2 _B$
	3	$ 7\rangle_{B'}\langle 0 _B$	$ 6\rangle_{B'}\langle 1-2 _B$	$ 5\rangle_{B'}\langle 1+2 _B$

Even for multipartite cases, we can also show that the set of the total maps with super communications is equivalent to the class of LOCC. Note that an arbitrary finite round LOCC map is a LOCC map since we can regard $\{E_{m_i}^{(\mathbf{m} \setminus \{m_i\} | i)}\}$ in Eq.(10) as a generalized measurement performed by i -th party depending on the measurement outcomes of the other parties $\mathbf{m} \setminus \{m_i\}$ with an outcome m_i . We denote the set of LOCC as $LOCC(\mathcal{H}_1, \dots, \mathcal{H}_{2n-1} : \mathcal{H}_2, \dots, \mathcal{H}_{2n})$. Then we obtain

$$LOCC(\mathcal{H}_1, \dots, \mathcal{H}_{2n-1} : \mathcal{H}_2, \dots, \mathcal{H}_{2n}) \subset LOCC(\mathcal{H}_1, \dots, \mathcal{H}_{2n-1} : \mathcal{H}_2, \dots, \mathcal{H}_{2n}). \quad (12)$$

If $\Gamma \in LOCC(\mathcal{H}_1, \dots, \mathcal{H}_{2n-1} : \mathcal{H}_2, \dots, \mathcal{H}_{2n})$ is LOCC, every $\{E_{m_i}^{(\mathbf{m} \setminus \{m_i\} | i)}\}$ must be decomposed into

$$E_{m_i}^{(\mathbf{m} \setminus \{m_i\} | i)} = F_{m_{i,l}}^{(\mathbf{m}_l | i, l)} F_{m_{i,l-1}}^{(\mathbf{m}_{l-1} | i, l-1)} \dots F_{m_{i,2}}^{(\mathbf{m}_2 | i, 2)} F_{m_{i,1}}^{(i, 1)}, \quad (13)$$

where $m_i = (m_{i,l}, m_{i,l-1}, \dots, m_{i,1})$ and $\{F_{m_{i,l}}^{(\mathbf{m}_l | i, l)}\}$ is the measurement performed by i -th party at l -th step in the LOCC protocol, which depends on the past measurement outcomes $\mathbf{m}_l = \{m_{n,l-1}, \dots, m_{2,1}, m_{1,l-1}, \dots, m_{1,1}\}$. Note that the measurement $\{F_{m_{i,1}}^{(i, 1)}\}$ at the first step does not depend on any measurement outcomes since there is no past measurement. Thus a LOCC map is a LOCC map if and only if every $\{E_{m_i}^{(\mathbf{m} \setminus \{m_i\} | i)}\}$ can be decomposed into the form of Eq.(13), where a definite causal structure exists. We give an example of a LOCC map that is not a LOCC map.

Example: In [9], it was shown a set of states defined by

$$\begin{aligned} |\psi_{1(2)}\rangle_{AB} &= |0\rangle_A |0 \pm 1\rangle_B, & |\psi_{3(4)}\rangle_{AB} &= |0 \pm 1\rangle_A |2\rangle_B \\ |\psi_{5(6)}\rangle_{AB} &= |2\rangle_A |1 \pm 2\rangle_B, & |\psi_{7(8)}\rangle_{AB} &= |1 \pm 2\rangle_A |0\rangle_B, & |\psi_9\rangle_{AB} &= |1\rangle_A |1\rangle_B, \end{aligned} \quad (14)$$

where $|a \pm b\rangle := (|a\rangle \pm |b\rangle)/\sqrt{2}$, is not distinguishable by a LOCC map perfectly. However we can see that this map is an element of LOCC by using the LOCC elements shown in Table 1. This map is implementable if there exists a classical pCTC and probabilistically simulatable by using local post-selection, which will be shown in subsection 2.2.

Next, we present a theorem showing the relationship between LOCC maps and separable maps. A generalization for multipartite cases is straightforward.

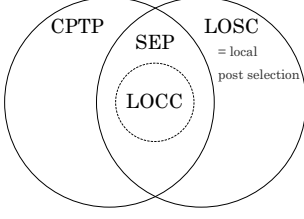


Figure 1: The intersection of LOSC and the set of CPTP maps is equivalent to the set of separable maps (SEP). The set of LOCC is strictly smaller than SEP. Normalized LOSC is equivalent to the set of local post-selections.

Theorem 2. (Relationship between LOSC maps and separable maps)

$$LOSC(\mathcal{H}_1, \mathcal{H}_3 : \mathcal{H}_2, \mathcal{H}_4) \cap CPTP(\mathcal{H}_1 \otimes \mathcal{H}_3 : \mathcal{H}_2 \otimes \mathcal{H}_4) = Sep(\mathcal{H}_1, \mathcal{H}_3 : \mathcal{H}_2, \mathcal{H}_4), \quad (15)$$

where $Sep(\mathcal{H}_1, \mathcal{H}_3 : \mathcal{H}_2, \mathcal{H}_4)$ is the set of separable maps.

2.2 LOSC and local post-selection

First we define a class of maps referred to as a *local post-selection*, and then present a theorem about the equivalence between LOSC the *local post-selection*.

Definition 3. A completely positive and trace preserving map $\hat{\mathcal{E}} : \mathbf{Pos}(\mathcal{H}_1 \otimes \mathcal{H}_3) \rightarrow \mathbf{Pos}(\mathcal{H}_2 \otimes \mathcal{H}_4)$ is a *local post-selection* if there exists a local measurement \mathcal{E} such that

$$\hat{\mathcal{E}}(\rho) = \frac{\mathcal{E}(\rho)}{\text{tr}[\mathcal{E}(\rho)]} \quad (16)$$

for all states $\rho \in \mathbf{D}(\mathcal{H}_1 \otimes \mathcal{H}_3)$, where \mathcal{E} can be decomposed into

$$\mathcal{E}(\rho) = \sum_{\{k_1, k_2\} \in \mathbf{K}} (F_{k_1}^{(1)} \otimes F_{k_2}^{(2)}) \rho (F_{k_1}^{(1)} \otimes F_{k_2}^{(2)})^\dagger, \quad (17)$$

where $\{F_{k_i}^{(i)} | \sum_{k_i} F_{k_i}^{(i)\dagger} F_{k_i}^{(i)} = \mathbb{I}_{2i-1}\}$ is a local measurement operator and \mathbf{K} is a particular subset of the set of all measurement outcomes, i.e.

$$\sum_{\{k_1, k_2\} \in \mathbf{K}} F_{k_1}^{(1)\dagger} F_{k_1}^{(1)} \otimes F_{k_2}^{(2)\dagger} F_{k_2}^{(2)} \leq \mathbb{I}_{1,3}. \quad (18)$$

Theorem 3. A linear completely positive map $\hat{\Gamma}$ is a normalized LOSC map if and only if $\hat{\Gamma}$ is a local post-selection.

In this section, we have defined a set of completely positive maps, LOSC, which is realized by local operations and classical correlations satisfying the valid probability conditions. Some of such classical correlations violate causality, however, LOSC can be simulated by the local post-selection. Thus LOSC gives a new interpretation of the gap between the set of LOCC and that of separable maps, i.e. a separable map not included in LOCC uses a power of acausal super communications.

We summarize the inclusion relations of the set of LOCC, that of separable maps (SEP), that of CPTP maps and LOSC in Fig. 1.

3 VALID CHANNEL CONDITION

In this section, we consider local operations and classical correlations with a more restricted condition called a *valid channel condition*, which all the deterministic classical correlations compatible with quantum mechanics must satisfy. *Valid channel* defined by classical correlations satisfying the valid channel condition has a good mathematical structure.

We consider that Alice and Bob receive quantum states in Hilbert space $\mathcal{H}_1, \mathcal{H}_3$ and classical information x, y , perform local quantum instruments $\{A_a^{(x)}\} \in QI(\mathcal{H}_1 : \mathcal{H}_2)$, $\{B_b^{(y)}\} \in QI(\mathcal{H}_3 : \mathcal{H}_4)$ and broadcast quantum states in Hilbert space $\mathcal{H}_2, \mathcal{H}_4$ and classical information a, b . In the previous section, we consider the case that classical correlations satisfy the valid probability conditions. Here, the classical correlations satisfy a stronger condition, i.e. Alice and Bob can freely choose their local instruments but the total map is required to be completely positive and trace preserving (CPTP) for *any* combinations of local instruments. We give a formal definition as follows.

Definition 4. For fixed numbers of inputs and outputs $a \in A, b \in B, x \in X, y \in Y$, a valid channel is a classical correlation $p(x, y|a, b)$ satisfying the valid channel condition such that

$$\sum_{a,b,x,y} p(x, y|a, b) A_a^{(x)} \otimes B_b^{(y)} \in CPTP(\mathcal{H}_1 \otimes \mathcal{H}_3 : \mathcal{H}_2 \otimes \mathcal{H}_4) \quad (19)$$

for any local instruments $\{A_a^{(x)}\} \in QI(\mathcal{H}_1 : \mathcal{H}_2)$ and $\{B_b^{(y)}\} \in QI(\mathcal{H}_3 : \mathcal{H}_4)$.

If the event of Alice sending her quantum state in \mathcal{H}_2 is in the causal past of the event of Bob receiving his quantum state in \mathcal{H}_3 , Alice can send a quantum state to Bob in principle. Then a quantum map after connecting \mathcal{H}_2 and \mathcal{H}_3 by a quantum channel $CPTP(\mathcal{H}_2 : \mathcal{H}_3)$ must be a CPTP map $CPTP(\mathcal{H}_1 : \mathcal{H}_4)$, i.e. the right hand side of Eq. (19) becomes a subset of $CPTP(\mathcal{H}_1 \otimes \mathcal{H}_3 : \mathcal{H}_2 \otimes \mathcal{H}_4)$. Under such a condition, the classical correlation $p(x, y|a, b)$ always describes a classical communication channel from Alice to Bob. However, quantum mechanics does not contradict more potential of the valid channel if the events of Alice and Bob receiving input states and their sending output states are not specified in space-time. We discuss on this point more in detail in subsection 3.2.

Note that a total map with a valid channel is a separable map, i.e. the CPTP map defined by Eq. (19) is a separable map for all $\{A_a^{(x)}\}$ and $\{B_b^{(y)}\}$ if $p(x, y|a, b)$ is a valid channel, since the valid channel automatically satisfies the valid probability conditions defined by Eq. (3,4) shown as follows. First, Eq. (3) is satisfied, since if we take

$$A_a^{(x)} = |a\rangle\langle a|_1 \otimes |x\rangle\langle x|_2, \quad B_b^{(y)} = |b\rangle\langle b|_3 \otimes |y\rangle\langle y|_4, \quad (20)$$

the whole CJ matrix is a diagonal matrix whose diagonal elements are $p(x, y|a, b)$ and thus positive. Second, Eq. (4) is satisfied by taking $\dim(\mathcal{H}_1) = \dim(\mathcal{H}_2) = \dim(\mathcal{H}_3) = \dim(\mathcal{H}_4) = 1$ and

$$\forall x, A_a^{(x)} = \delta_{a,i} \quad \forall y, B_b^{(y)} = \delta_{b,j}, \quad (21)$$

for fixed i, j .

3.1 The valid channel and the process matrix

The valid channel is closely related to the *process matrix*, which has been developed in [13]. Thus, we can apply the properties of the process matrix derived in [13] to the valid channel. In this subsection, we

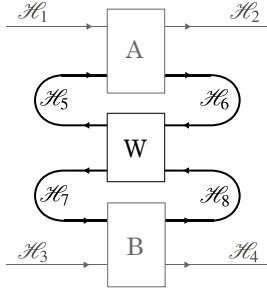


Figure 2: Bipartite process matrix representing a quantum correlation between Alice and Bob. The CJ matrix of a quantum channel from Alice to Bob or from Bob to Alice is a bipartite process matrix. The CJ matrix of the probability mixture of the quantum channels is also a bipartite process matrix. There exists a 'weird' bipartite process matrix that cannot be understood in terms of causal order, i.e. the weird process matrix cannot be decomposed into a probability mixture of the quantum channels. It is shown that any process matrix represents a quantum channel from Alice to Bob if Alice can send a quantum state in \mathcal{H}_2 to Bob through \mathcal{H}_3 [17]. However, quantum mechanics does not contradict the existence of such a weird process matrix if the events of Alice and Bob receiving input states in \mathcal{H}_1 , \mathcal{H}_3 and their sending output states in \mathcal{H}_2 , \mathcal{H}_4 are not specified in space-time.

provide a short review on the process matrix formalism and show the correspondence between the valid channel and the process matrix.

A framework to treat quantum correlations between multiparties without an assumption of a global causal structure has been developed recently [13, 17, 18]. In this framework, a quantum correlation between Alice and Bob is described by a process matrix $W \in \text{Pos}(\mathcal{H}_5 \otimes \mathcal{H}_6 \otimes \mathcal{H}_7 \otimes \mathcal{H}_8)$ such that

$$W * (A \otimes B) \in \text{CPTP}(\mathcal{H}_1 \otimes \mathcal{H}_3 : \mathcal{H}_2 \otimes \mathcal{H}_4) \quad (22)$$

for any CPTP maps $A \in \text{CPTP}(\mathcal{H}_1 \otimes \mathcal{H}_5 : \mathcal{H}_2 \otimes \mathcal{H}_6)$ and $B \in \text{CPTP}(\mathcal{H}_3 \otimes \mathcal{H}_7 : \mathcal{H}_4 \otimes \mathcal{H}_8)$ illustrated in Fig. 2, where we use the *link product* defined in [17]:

Definition 5. The link product of two operators $M \in L(\otimes_{m \in \mathcal{M}} \mathcal{H}_m)$ and $N \in L(\otimes_{n \in \mathcal{N}} \mathcal{H}_n)$ is the operator $M * N \in L(\otimes_{m \in \mathcal{M} \cup \mathcal{N}} \mathcal{H}_m)$ given by

$$M * N = \text{tr}_{\mathcal{M} \cap \mathcal{N}} [(\mathbb{I}_{\mathcal{N} \setminus \mathcal{M}} \otimes M^{T_{\mathcal{M} \cap \mathcal{N}}})(N \otimes \mathbb{I}_{\mathcal{M} \setminus \mathcal{N}})], \quad (23)$$

where the set-subscript \mathcal{A} is a shorthand for $\otimes_{i \in \mathcal{A}} \mathcal{H}_i$, $\mathcal{A} \setminus \mathcal{B} = \{i \in \mathcal{A}, i \notin \mathcal{B}\}$ for two sets \mathcal{A} and \mathcal{B} , $\mathbb{I}_{\mathcal{A}}$ is the identity operator on $\otimes_{i \in \mathcal{A}} \mathcal{H}_i$, and the superscript $T_{\mathcal{A}}$ is the partial transpose on $\otimes_{i \in \mathcal{A}} \mathcal{H}_i$.

We can interpret the valid channel as a special type of a process matrix, a *classical* process matrix, which describes a quantum correlation between Alice and Bob when they perform local classical (diagonal) operations.

Theorem 4. There is a one-to-one correspondence between the valid channels and the classical process matrices, where a classical process matrix is the process matrix defined by

$$W = \sum_{x,y,a,b} p(x,y|a,b) |x\rangle\langle x|_5 \otimes |y\rangle\langle y|_7 \otimes |a\rangle\langle a|_6 \otimes |b\rangle\langle b|_8 \quad (24)$$

for fixed orthonormal bases $\{|x\rangle_5\}$, $\{|y\rangle_7\}$, $\{|a\rangle_6\}$ and $\{|b\rangle_8\}$.

3.2 The valid channel and one-way LOCC

In this subsection, we define a subset of the valid channel, a *classical communication channel*, which describes classical correlations based on causal order between Alice and Bob. We show that the set of one-way LOCC is equivalent to the set of the total maps with the classical communication channel. We also show that the set of one-way LOCC is equivalent to the set of the total maps with valid channels corresponding to *causally separable* process matrices, which is a subset of the process matrix. These results clarify the relationship between the valid channel, LOCC and the process matrix.

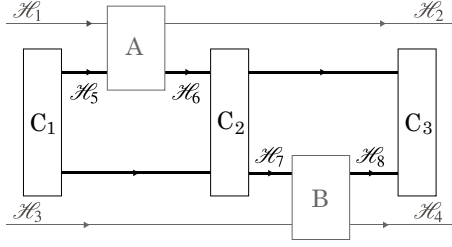


Figure 3: A deterministic quantum channel from Alice to Bob. C_i represents a CPTP map. Note that C_1 is a CPTP map that prepares a state and C_3 is a CPTP map that discards a state. Bob cannot signal to Alice with this channel.

Definition 6. A classical communication (CC) channel from Alice to Bob is a probability distribution $p(x, y|a, b)$ such that

$$p(x, y|a, b) = \sum_{\lambda} q(\lambda) p(x|\lambda) p(y|a, \lambda), \quad (25)$$

where $q(\lambda)$ is a probability distribution.

It is trivial that any one-way LOCC from Alice to Bob can be obtained by a CC channel from Alice to Bob. We can also show that any total map with the CC channel from Alice to Bob is implementable by a one-way LOCC from Alice to Bob as follows. A total map obtained by the CC channel is given by

$$\sum_{\lambda, a, b, x, y} q(\lambda) p(x|\lambda) p(y|a, \lambda) A_a^{(x)} \otimes B_b^{(y)} = \sum_m A_m \otimes B^{(m)}, \quad (26)$$

where

$$m = (a, \lambda), \quad A_{a, \lambda} = \sum_x q(\lambda) p(x|\lambda) A_a^{(x)}, \quad B^{(a, \lambda)} = \sum_{b, y} p(y|a, \lambda) B_b^{(y)}. \quad (27)$$

We can easily verify that $\{A_m\}$ is an instrument and $B^{(m)}$ is a CPTP map. Thus, the total map is implementable by a one-way LOCC from Alice to Bob. Note that it was also shown that the CC channel is a valid channel since the total map is always a CPTP map for any local operations of Alice and Bob.

Theorem 5. There is a one-to-one correspondence between convex mixtures of CC channels and causally separable classical process matrices.

We briefly review a causally separable process matrix. In [13, 17], it was shown that a process matrix $W_{A \rightarrow B}$ is implementable by a deterministic quantum channel from Alice to Bob, which is illustrated in Fig. 3, if and only if $W_{A \rightarrow B}$ satisfies the following conditions:

$$W_{A \rightarrow B} \in \text{Pos}(\mathcal{H}_5 \otimes \mathcal{H}_6 \otimes \mathcal{H}_7 \otimes \mathcal{H}_8) \quad (28)$$

$$W_{A \rightarrow B} = \mathbb{I}_8 \otimes W_{5,6,7}, \quad \text{tr}_7[W_{5,6,7}] = \mathbb{I}_6 \otimes \rho_5, \quad \text{tr}_5[\rho_5] = 1. \quad (29)$$

Definition 7. A process matrix W is called causally separable if W can be decomposed into

$$W = qW_{A \rightarrow B} + (1 - q)W_{B \rightarrow A}. \quad (30)$$

Any classical bipartite process matrix between two parties is known to be causally separable. However, in more than two parties cases, there exists a causally non-separable classical process matrix.

To summarize this section, we have showed that (i) the set of convex mixture of one-way LOCC from Alice to Bob and from Bob to Alice is equivalent to the set of the total maps with valid channels corresponding to causally separable classical process matrices, (ii) in the bipartite case, the set of the total maps with valid channels is equivalent to the set of convex mixture of one-way LOCC since all the classical process matrices are causally separable [13]. Although we can not discuss in detail in this paper, it has been shown that (iii) in the multipartite case (more than two parties), the set of the total maps with valid channels is strictly larger than the set of convex mixture of one-way LOCC. We summarize the inclusion relations in Fig. 4.

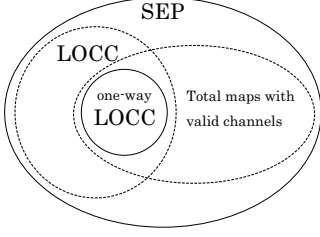


Figure 4: The set of convex mixture of one-way LOCC is included by the set of total maps with valid channels. In a bipartite system, the two sets are identical, however they are different in a multipartite system. In any parties system, the set of convex mixture of one-way LOCC is equivalent to the set of total maps with valid channels corresponding to causally separable classical process matrices.

4 ANALYSIS OF SPACE RESOURCES

The states discrimination has been intensively studied for the purpose of finding the gap between LOCC and separable maps. A LOCC indistinguishable set of product states like the nine states defined in Eq. (14), which was introduced by Bennett et al. [9], can certainly be discriminated by separable maps.

Thus, existence of such a set clearly illustrates the gap. In this section, we focus on the space resources, entanglement, which is necessary to discriminate a LOCC indistinguishable product orthonormal basis. As a result, we characterize the amount of entanglement resource to discriminate an orthonormal basis by one-way LOCC protocols, and a two-way LOCC protocol which consumes entanglement resource much less than the lower bound for one-way LOCC protocols. In this section, we only treat orthonormal basis on a bipartite system $\mathcal{H}_A \otimes \mathcal{H}_B$, and entanglement resource $|\Phi\rangle$ on the ancilla system $\mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$. We assume that Alice and Bob have $\mathcal{H}_{AA'} := \mathcal{H}_A \otimes \mathcal{H}_{A'}$ and $\mathcal{H}_{BB'} := \mathcal{H}_B \otimes \mathcal{H}_{B'}$, respectively.

4.1 Entanglement resource for one-way LOCC discrimination

In this subsection, we focus on characterizing the amount of entanglement resource to discriminate an orthonormal basis by one-way LOCC in terms of the Schmidt rank. Note that we do not assume the orthonormal basis as a product basis; thus, all the results in this subsection are valid for all possibly entangled orthonormal basis. Although we assume that one-way LOCC means one-way LOCC from Alice ($\mathcal{H}_{AA'}$) to Bob ($\mathcal{H}_{BB'}$) in this subsection, all the results can be easily extend to one-way LOCC from Bob to Alice.

For an orthonormal basis $\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}$ on $\mathcal{H}_A \otimes \mathcal{H}_B$, we define $r_{\min}(\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B})$ as the minimum of the Schmidt rank of a state $|\Phi\rangle_{A'B'} \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}$ such that a set $\{|\psi_j\rangle_{AB} \otimes |\Phi\rangle_{A'B'}\}_{j=1}^{d_A d_B} \subset \mathcal{H}_{AA'} \otimes \mathcal{H}_{BB'}$ can be perfectly discriminated by one-way LOCC, namely,

$$r_{\min}(\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}) := \min_{|\Phi\rangle} \left\{ r_{\text{sch}}(|\Phi\rangle) \mid \exists \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}, \text{ s.t. } |\Phi\rangle \in \mathcal{H}_{A'} \otimes \mathcal{H}_{B'}, \right. \\ \left. \text{and } \{|\psi_j\rangle_{AB} \otimes |\Phi\rangle_{A'B'}\}_{j=1}^{d_A d_B} \text{ is one-way LOCC distinguishable} \right\}, \quad (31)$$

where $r_{\text{sch}}(|\Phi\rangle)$ is the Schmidt rank of $|\Phi\rangle$. Thus, $r_{\min}(\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B})$ is the optimal entanglement resource to discriminate $\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}$ by one-way LOCC in terms of the Schmidt rank.

Our main result in this subsection is the following theorem which characterizes $r_{\min}(\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B})$ completely.

Theorem 6. For all orthonormal basis $\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B} \subset \mathcal{H}_A \otimes \mathcal{H}_B$,

$$r_{\min}(\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}) = d_{\min}(\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}),$$

where $d_{\min} \left(\{ |\psi_j\rangle_{AB} \}_{j=1}^{d_A d_B} \right)$ is defined by

$$d_{\min} \left(\{ |\psi_j\rangle_{AB} \}_{j=1}^{d_A d_B} \right) := \min_{\mathcal{H}_A = \bigoplus_k \mathcal{M}_k} \max_k \left\{ \dim \mathcal{M}_k \mid \forall j, \exists k, \text{ s.t. } |\psi_j\rangle_{AB} \in \mathcal{M}_k \otimes \mathcal{H}_B \right\}.$$

4.2 Entanglement resource to distinguish the nine states

In this subsection, we focus on orthogonal product bases and study the amount of entanglement necessary to discriminate the basis states by LOCC. In particular, as a example, we treat the nine states defined by Eq. (14) and their generalization, and study the entanglement resource necessary to discriminate them by LOCC.

For the nine states $\{ |\psi_j\rangle_{AB} \}_{j=1}^9$ defined by Eq. (14), we can easily see that there is no non-trivial subspace $\mathcal{M} \subset \mathcal{H}_A$ satisfying for all j , either $|\psi_j\rangle \in \mathcal{M} \otimes \mathcal{H}_B$ or $|\psi_j\rangle \in \mathcal{M}^\perp \otimes \mathcal{H}_B$, where \mathcal{M}^\perp is the orthogonal complement of \mathcal{M} . Thus, from Theorem 6, the optimal entanglement resource $|\Phi\rangle$ necessary to discriminate the nine states by one-way LOCC satisfies $r_{Sch}(|\Phi\rangle) = 3$. On the other hands, here, we propose a two-way LOCC protocol by which the nine states can be discriminated by consuming just 1-e-bit. The protocol can be described as follows: First, we extend the dimension of \mathcal{H}_B from 3 to d by adding additional states $|i\rangle_B$ for $3 \leq i \leq d$. Then, apply a global unitary V_d given by

$$V_d := |0\rangle\langle 0|_A \otimes (|3\rangle\langle 1| + |1\rangle\langle 3| + |0\rangle\langle 0| + |2\rangle\langle 2| + |4\rangle\langle 4| + \cdots + |d\rangle\langle d|)_B + (|1\rangle\langle 1| + |2\rangle\langle 2|)_A \otimes I_B.$$

As a result, the nine states $\{ |\psi_j\rangle_{AB} \}_{j=1}^9$ are transformed to

$$\begin{aligned} |\psi'_{1(2)}\rangle_{AB} &= |0\rangle_A |0 \pm 3\rangle_B, & |\psi'_{3(4)}\rangle_{AB} &= |0 \pm 1\rangle_A |2\rangle_B, \\ |\psi'_{5(6)}\rangle_{AB} &= |2\rangle_A |1 \pm 2\rangle_B, & |\psi'_{7(8)}\rangle_{AB} &= |1 \pm 2\rangle_A |0\rangle_B, & |\psi'_9\rangle_{AB} &= |1\rangle_A |1\rangle_B. \end{aligned} \quad (32)$$

It is easily see that these states are distinguishable by LOCC. V_d can be implemented with 1-e-bit since the operator Schmidt rank of V_d is 2 [19]. We further define the Schmidt strength $H(U)$ of a unitary operator U on $\mathcal{H}_A \otimes \mathcal{H}_B$ as the Shannon entropy of the operator Schmidt coefficients of U . Then, by the definition, since $U^{\otimes n}$ generates $nH(U)$ e-bits from a product state, $nH(U)$ e-bits is necessary to implement $U^{\otimes n}$, when n is large. Hence, the Schmidt strength $H(V_d)$ is a lower bound of entanglement resource which is necessary to implement V_d [20]. A numerical calculation strongly suggests that $H(V_d)$ decreases to 0 when $d \rightarrow \infty$.

The above discussion can be generalized to an orthogonal product basis $\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}$ defined by

$$\begin{aligned}
|\psi_{1(2)}\rangle &= \frac{1}{\sqrt{2}}|1\rangle \otimes (|1\rangle \pm |2\rangle), \\
|\psi_{3(4)}\rangle &= \frac{1}{\sqrt{2}}(|d_A - 1\rangle \pm |d_A\rangle) \otimes |1\rangle, \\
|\psi_{m_1+5}\rangle &= \sum_{k=2}^{d_A-2} e^{i\frac{2\pi}{d_A-3}m_1 \cdot (k-2)} |k\rangle \otimes |1\rangle, \\
|\psi_{d_A+m_2+2}\rangle &= \sum_{k=2}^{d_A-1} e^{i\frac{2\pi}{d_A-2}m_2 \cdot (k-2)} |k\rangle \otimes |2\rangle, \\
|\psi_{2d_A+m_3}\rangle &= |d_A\rangle \otimes \sum_{k=2}^{d_B} e^{i\frac{2\pi}{d_B-1}m_3 \cdot (k-2)} |k\rangle, \\
|\psi_{2d_A+d_B-1+(d_A-1)m_5+m_4}\rangle &= \sum_{k=1}^{d_A-1} e^{i\frac{2\pi}{d_A-1}m_4 \cdot (k-1)} |k\rangle \otimes |m_5+3\rangle
\end{aligned} \tag{33}$$

where m_i for $i = 1, 2, 3, 4, 5$, satisfies $0 \leq m_1 \leq d_A - 4$, $0 \leq m_2 \leq d_A - 3$, $0 \leq m_3 \leq d_B - 2$, $0 \leq m_4 \leq d_A - 2$, and $0 \leq m_5 \leq d_B - 3$. We consider discrimination of $\{|\psi_j\rangle_{AB}\}_{j=1}^{d_A d_B}$. Theorem 6 guarantees that in order to discriminate this basis by one-way LOCC from \mathcal{H}_A to \mathcal{H}_B , an entanglement resource with the Schmidt rank d_A is necessary. Similarly, to discriminate it by one-way LOCC from \mathcal{H}_B to \mathcal{H}_A , an entanglement resources with the Schmidt rank d_B is necessary. On the other hands, by means of the similar discussion as the nine states, it is enough to use 1-eit as an resource in case of two-way LOCC. Therefore, in the limit of large d_A , there is infinite gap between one-way LOCC and two-way LOCC in terms of entanglement resource to distinguish the orthogonal product basis defined by Eq. (33).

Acknowledgements

This work is supported by Project for Developing Innovation Systems of MEXT, Japan and JSPS by KAKENHI (Grant No. 23540463). We also gratefully acknowledge to the ELC project (Grant-in-Aid for Scientific Research on Innovative Areas MEXT KAKENHI (Grant No. 24106009)) for encouraging this research.

References

- [1] E. Chitambar, D. Leung, L. Mancinska, M. Ozols, A. Winter, Everything You Always Wanted to Know About LOCC (But Were Afraid to Ask), arXiv:1210.4583 (2012)
- [2] V. Vedral, M. B. Plenio, M. A. Rippin, and P. L. Knight, Quantifying Entanglement, Phys. Rev. Lett. 78, 2275, (1997)
- [3] Martin B. Plenio and Shashank Virmani, An introduction to entanglement measures, Quant. Inf. Comp. vol. 7, 1-51, (2007)
- [4] Michal Horodecki, Pawel Horodecki, and Ryszard Horodecki, Limits for Entanglement Measures, Phys. Rev. Lett. 84, 1414, (2000)
- [5] W. Dur, G. Vidal, and J. I. Cirac, Optimal Conversion of Nonlocal Unitary Operations, Phys. Rev. Lett. 89, 057901, (2002)

- [6] A. Soeda, P. S. Turner, and M. Murao, Entanglement cost of implementing controlled-unitary operations, arXiv:1008.1128 (2010)
- [7] Dan Stahlke and Robert B. Griffiths, Entanglement requirements for implementing bipartite unitary operations, Phys. Rev. A 84, 032316 (2011)
- [8] Vlad Gheorghiu and Robert B. Griffiths, Separable operations on pure states, Phys. Rev. A 78, 020304 (2008)
- [9] Charles H. Bennett et.al., Quantum nonlocality without entanglement, Phys. Rev. A 59, 1070-1091 (1999)
- [10] J. Niset and N. J. Cerf, Multipartite nonlocality without entanglement in many dimensions, Phys. Rev. A 74, 52103 (2006)
- [11] David P. DiVincenzo, Tal Mor, Peter W. Shor, John A. Smolin, Barbara M. Terhal, Unextendible Product Bases, Uncompletable Product Bases and Bound Entanglement, Comm. Math. Phys. 238 (2003)
- [12] Eric Chitambar and Runyao Duan, Nonlocal Entanglement Transformations Achievable by Separable Operations, Phys. Rev. Lett. 103, 110502 (2009)
- [13] Ognian Oreshkov, Fabio Costa, Caslav Brukner, Quantum correlations with no causal order, Nature communications 3, 1092 (2012).
- [14] Charles H. Bennett. talk at QUPON, Vienna, Austria, May (2005)
- [15] Seth Lloyd, Lorenzo Maccone, Raul Garcia-Patron, Vittorio Giovannetti, and Yutaka Shikano, Quantum mechanics of time travel through post-selected teleportation, Phys. Rev. D 84, 025007 (2011)
- [16] Seth Lloyd, Lorenzo Maccone, Raul Garcia-Patron, Vittorio Giovannetti, Yutaka Shikano, Stefano Pirandola, Lee A. Rozema, Ardavan Darabi, Yasaman Soudagar, Lynden K. Shalm, and Aephraim M. Steinberg, Closed Timelike Curves via Postselection: Theory and Experimental Test of Consistency, Phys. Rev. Lett. 106, 040403 (2011)
- [17] Giulio Chiribella, Theoretical framework for quantum networks, Phys. Rev. A 80, 022339 (2009).
- [18] Giulio Chiribella, Quantum computations without definite causal structure, Phys. Rev. A 88, 022318 (2012)
- [19] Scott Cohen, All unitaries having operator Schmidt rank 2 are controlled unitaries, Phys. Rev. A 87, 022329 (2013).
- [20] E. Wakakuwa and M. Murao, Resource Compression for Distributed Quantum Computation, arXiv:1310.3991 (2013).

Completeness of Hardy Non-locality: Consequences & Applications

Shane Mansfield

Quantum Group
Department of Computer Science
University of Oxford
`shane.mansfield@cs.ox.ac.uk`

Completeness results due to Mansfield & Fritz show that Hardy’s paradox provides a necessary and sufficient condition for logical non-locality in all $(2, 2, l)$ and $(2, k, 2)$ scenarios. This paper considers a variety of consequences and applications of these results. This includes a proof that Bell states, despite being perhaps the most studied and utilised of entangled states, appear to be anomalous in terms of non-locality, in the sense that they are the only entangled two-qubit states which are not logically non-local. Much of the literature on Hardy’s paradox is concerned with the probability of witnessing a paradox, which has experimental motivations, and is often considered to be a measure of the quality of Hardy non-locality. We demonstrate that it is possible to witness Hardy non-locality with certainty for a simple tripartite quantum system. We also discuss results relating to strong non-locality, and to the complexity of logical non-locality.

1 Introduction

Since the fundamental insight of Bell [10, 11], it is known that quantum mechanics gives rise to stronger-than-classical, non-local correlations. Under some seemingly natural assumptions of locality and realism, it can be shown that any empirical correlations should satisfy certain Bell inequalities, which can be violated quantum-mechanically, from which Bell’s conclusion follows.

A more intuitive, logical approach to non-locality proofs was pioneered by Heywood & Redhead [24], Greenberger, Horne, Shimony & Zeilinger [21, 20] (which was formulated in a simplified form by Mermin [30, 31]) and Hardy [22, 23] (also treated by Mermin [32]). This kind of non-locality proof disregards the exact values of the joint outcome probabilities and only records which of them are non-zero and which are zero. In other words, one distinguishes only between possible outcomes and impossible outcomes, and this turns out to be sufficient for demonstrating non-locality in quantum mechanics. Several other logical non-locality proofs of this kind have subsequently appeared (e.g. [12, 15, 19]). We will refer to such proofs as *logical non-locality proofs*.

We work within a general framework for logical non-locality proofs which was developed in [27]. More specifically, we are interested in logical non-locality in (n, k, l) Bell scenarios, where n is the number of sites, k is the number of allowed measurements at each site, and l is the number of possible outcomes for each measurement. This framework is inspired by the relational hidden variable framework of Abramsky [2]. Though not as general, it might also be considered as a precursor to the unified sheaf-theoretic framework for non-locality and contextuality [3, 7, 28], which it predates, in that it can be understood as a purely possibilistic version of the sheaf-theoretic framework for such scenarios. The advantage of the present framework is that it comes with a particular representation for $n = 2$ and (as we will see in this paper) $n = 3$ scenarios, which can provide a powerful means of reasoning about empirical models.

Hardy’s logical non-locality proof or ‘paradox’ [22, 23] is considered to be the simplest non-locality proof for quantum mechanics. In [27], the author and Fritz proved a number of completeness theorems showing that it is a necessary and sufficient condition for logical non-locality in all $(2, k, 2)$ and $(2, 2, l)$ scenarios, thereby subsuming all other logical non-locality proofs or ‘paradoxes’. However, for the $(2, 3, 3)$ and $(3, 2, 2)$ [26] scenarios logical conditions for locality have been found which can be violated without the occurrence of a Hardy paradox.

In this paper, we will see that these completeness results have many interesting consequences. As minor points, we show that they lead to insights on the complexity of non-locality and to a constructive argument that the Popescu-Rohrlich box is the only *strongly non-local* $(2, 2, 2)$ model, by which we mean, roughly speaking, that even at the level of possibilities the model cannot be factored into a local and a non-local part.

They also lead to a proof that Bell states are not logically non-local. Since all other entangled two-qubit states can be shown to admit a Hardy paradox [23] this proves the surprising result that the Bell states are the only such states that are not logically non-local. Together with recent results indicating that all n -partite entangled qubit states for $n > 3$ are logically non-local [5], this shows that the Bell states are anomalous in the landscape of entangled qubit states, in spite of the fact that they are perhaps the most studied and utilised of these.

Finally, much of the literature on Hardy’s paradox is concerned with the *paradoxical probability*; i.e. the probability of witnessing the joint outcome from which the logical argument follows. This has experimental motivations, and is often considered to be a measure of the quality of Hardy non-locality. For Hardy’s family of non-local quantum models, the maximum paradoxical probability is $(5\sqrt{5} - 11)/2 \approx 0.09$. It has been shown, however, that it is possible to achieve a paradoxical probability of 0.125 for a generalisation of Hardy’s paradox in a tripartite quantum system [19]. It has also been shown that a ‘ladder’ version of Hardy’s paradox, which allows k measurement settings to each party, can give rise to a higher paradoxical probability which approaches 0.5 as $k \rightarrow \infty$. More recently, it has been shown by Chen et al. that another generalisation of Hardy’s paradox can be witnessed with probability ≈ 0.4 for certain high-dimensional bipartite quantum systems [16]. The measurement scenarios for both of these logical non-locality proofs fall within the scope of the completeness results for Hardy non-locality. The ladder paradox was already discussed in [27], but here we will show explicitly that each Chen et al. paradox contains within it many different Hardy paradoxes. Moreover, we will see that their paradoxical probability might more accurately be described as the sum of the paradoxical probabilities for these Hardy paradoxes, all of which occur within the one model.

Using the completeness of Hardy non-locality we will achieve a striking improvement on these results, demonstrating by a much simpler argument that if such a summing of paradoxical probabilities is considered, it is possible to witness Hardy non-locality with certainty for a tripartite quantum system. Interestingly, the argument relies on the same state and measurements as the GHZ experiment [20]. The author has become aware of a similar proposal by Cabello [13]; the result contained in this paper has the advantage of being far simpler, both in terms of the argument and of the empirical model in question. We also show that Hardy non-locality can be achieved with certainty for a particular non-quantum, no-signalling $(2, 2, 2)$ model, which turns out to be the Popescu-Rohrlich no-signalling box [35].

2 Background

The empirical model for the original Hardy non-locality proof, Table 1, concerns the $(2, 2, 2)$ scenario. Each of the two parties can make one of two measurements on their subsystem, giving rise to outcomes

Table 1: An empirical model containing a Hardy paradox. This is a possibilistic table with ‘1’ denoting ‘possible’ and ‘0’ denoting ‘impossible’. The blank entries are not relevant to the argument.

		Bob	
		\uparrow	\downarrow
Alice	\uparrow	1	
	\downarrow		
	G		0
	W	0	

which we label here $\{\uparrow, \downarrow\}$ for the first measurement and $\{G, W\}$ for the second. The precise probabilities of obtaining the various joint outcomes are not required; it is only necessary to specify which are possible and which are not. A ‘1’ in the table signifies that it is possible (with probability greater than zero) to obtain the corresponding joint outcome, and a ‘0’ signifies that it is not possible to obtain the joint outcome.

Definition 2.1. Up to re-labelling of measurements and outcomes, any possibilistic $(2, 2, 2)$ empirical model containing the arrangement of ‘1’s and ‘0’s shown in Table 1 is said to *contain a Hardy paradox* (i.e. it admits Hardy’s logical non-locality proof) and we say that the joint outcome (\uparrow, \uparrow) *witnesses a Hardy paradox*.

Any probabilistic empirical model can be transformed into a possibilistic one in a canonical way via *possibilistic collapse*: the process by which all non-zero probabilities are conflated to ‘1’ [27, 3, 28]. Containing a Hardy paradox is just one way in which a model may be demonstrated to be non-local at the possibilistic level (other examples were mentioned in the introduction). More generally, we have the following definition.

Definition 2.2. A model which is non-local at the possibilistic level is said to be *logically non-local*.

Definition 2.1 defines Hardy non-locality for $(2, 2, 2)$ scenarios. However, it is still possible to consider Hardy paradoxes in $(2, 2, l)$ scenarios, simply by course-graining outcomes; see Table 2. It is also possible, more generally, to define Hardy non-locality in $(2, k, l)$ models as arising whenever some 2×2 subtable (i.e. restricting attention to only two of k measurements at each site) contains a Hardy paradox.

Definition 2.3. Any possibilistic $(2, k, l)$ empirical model containing a 2×2 subtable which is isomorphic (up to re-labelling of measurements and outcomes) to table 2 is said to contain a (coarse-grained) Hardy paradox.

Wang & Markham have described a generalisation of Hardy’s logical non-locality proof to $(n, 2, 2)$ scenarios, which they have used to demonstrate that all symmetric n -partite qubit states for $n > 2$ are logically non-local [36]. This kind of generalisation has been described elsewhere by Ghosh, Kar and Sarkar [19], and is also considered in [15] and [17].

We write $p(o|m) = 1$ if it is possible with probability greater than zero to obtain joint outcome o when joint measurement m is made, and $p(o|m) = 0$ otherwise, and let measurements and outcomes both be labelled by $\{0, 1\}$ at each site.

Table 2: A $(2, 2, l)$ scenario with a coarse-grained Hardy paradox.

	o'_1	\dots	o'_l	$o_1 \dots o_{m_2}$	$o_{m_2+1} \dots o_l$
o'_1	1				0 \dots 0
\vdots					
o'_l					
o_1				0 \dots 0	
\vdots				\vdots \ddots \vdots	
o_{m_1}				0 \dots 0	
o_{m_1+1}	0				
\vdots	\vdots				
o_l	0				

Definition 2.4. For any $(n, 2, 2)$ scenario, a generalised Hardy paradox occurs if (up to re-labelling of measurements and outcomes) the following possibilistic conditions are satisfied.

- $p(0, \dots, 0 \mid 0, \dots, 0) = 1$
- $p(\pi(1, 0, \dots, 0) \mid \pi(1, 0, \dots, 0)) = 0$ for all permutations π
- $p(0, \dots, 0 \mid 1, \dots, 1) = 0$

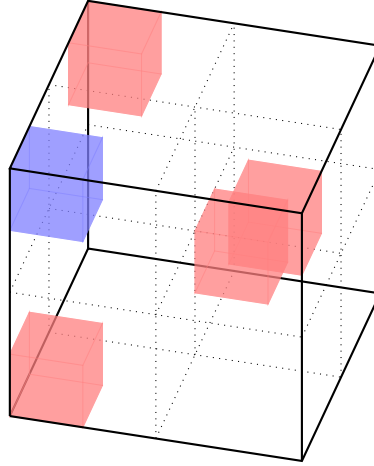
The $n = 3$ generalisation of the Hardy paradox can be represented in a three dimensional version of the tabular representation [26]; see Figure 1. The advantage of the representation is that it provides a powerful visual means of analysing models. The axes correspond to different sites, the cubes to joint measurement choices, and individual entries to outcomes, similarly to the $n = 2$ case. The properties of the tabular representation generalise in the obvious way to the third dimension.

For example, in [27] it was shown that for such scenarios an empirical model is local if and only if every ‘1’ in its table can be completed to a *deterministic grid*; i.e. a choice of ‘1’ for each box of the table, such that these ‘1’s form a rectangular arrangement by aligning in outcome rows/columns where possible. Deterministic grids correspond to global sections of the event sheaf in the sheaf-theoretic approach [26]. The ‘1’ in Table 1 cannot be completed to a deterministic grid, regardless of the unspecified entries. This characterisation generalises in the obvious way to the three dimensional representation for $n = 3$ models, so that we can similarly see that the blue entry in Figure 1 cannot be completed to a deterministic grid, and therefore any $(3, 2, 2)$ model containing this arrangement of ‘1’s and ‘0’s, or red and blue boxes, is logically non-local.

It can be shown that Hardy non-locality completely characterises logical non-locality in a variety of scenarios.

Theorem 2.5 (Mansfield & Fritz [27]). *For any $(2, k, 2)$ or $(2, 2, l)$ scenario, an empirical model is logically non-local if and only if it contains a (coarse-grained) Hardy paradox.*

Figure 1: The $n = 3$ Hardy paradox. The blue entry corresponds to Boolean ‘1’ or ‘possible’, and the red entries to ‘0’ or ‘impossible’. The blank entries are unspecified.



3 Hardy Subsumes Other Paradoxes

An immediate consequence of Theorem 2.5 is that in the relevant scenarios Hardy’s paradox subsumes all other paradoxes, in the sense that any model which can be demonstrated to be logically non-local necessarily contains a Hardy paradox.

The ladder paradox [12] has been proposed as a generalisation of the original Hardy paradox and was used for experimental tests of quantum non-locality [9]. Up to symmetries, there is one ladder paradox for any number of settings k ; i.e. for each $(2, k, l)$ scenario. It was observed in [27] that, by Theorem 2.5, any ladder paradox necessarily contains a Hardy paradox, and, moreover, it was explicitly demonstrated how this comes about.

We turn our attention now to a very recent paper by Chen et al. [16] which attempts to provide a different generalisation of Hardy’s paradox for high-dimensional (qudit) systems. In the present terminology, their argument applies to $(2, 2, l)$ Bell scenarios. This will also be relevant to the discussion in section 7.

Proposition 3.1. *The occurrence of a Chen et al. paradox (Table 3) implies the occurrence of a Hardy paradox.*

Proof. This follows directly from Theorem 2.5, but one can also prove the proposition more directly. Suppose one of the starred entries corresponding to outcomes (o'_i, o_j) of Table 3 is non-zero. We write $p(i, j) > 0$ for short. Then we can see from the table that for the joint measurement represented by the upper-right box, we must have $p(r, j) = 0$ for all $r > (l - j)$. Similarly, for the measurement represented by the lower-left box, $p(i, s) = 0$ for all $s > (l - i)$. In the lower-right box, we have $p(r, s) = 0$ when $r \leq (l - j)$ and $s \leq (l - i)$. This is a $(2, 2, l)$ Hardy paradox. \square

The proof shows that every non-zero starred entry in Table 3 witnesses a (coarse-grained) Hardy paradox.

Table 3: The Chen et al. paradox occurs when at least one of the starred entries is non-zero. The relevant outcomes for each joint measurement are either those above or those below the diagonal.

	*	...	*	0	...	0	
		⋮	⋮		⋮	⋮	
			*			0	
	0	...	0				
		⋮	⋮	0			
			0	⋮	⋮		
				0	...	0	

4 Complexity of Logical Non-locality

It was mentioned in [27] that Theorem 2.5 can tell us something about the computational complexity of recognising logical non-locality, which in the relevant scenarios is equivalent to deciding whether a Hardy paradox occurs. Here, we give an explicit proof.

Proposition 4.1. *Polynomial algorithms can be given for deciding non-locality in $(2,2,l)$ and $(2,k,2)$ models.*

Proof. For $(2,k,2)$ scenarios, deciding whether a model in the tabular form is local or non-local simply amounts to checking all 2×2 sub-tables for such a Hardy paradox, which gives an algorithm that is polynomial in the size of the input table: we check for the 64 possible Hardy configurations in each of $\binom{k}{2}^2$ sub-tables, which is clearly $O(k^4)$. For $(2,2,l)$ scenarios, one has to check each ‘1’ in the table to see whether it can be completed to a deterministic grid. There are $4l^2$ entries in the table, and each check is clearly $O(l^2)$. Again, we have an algorithm that is polynomial in the size of the input. \square

It was conjectured in [27] that the general decidability problem for possibilistic local models with k as the free input is NP-hard when $n > 2, l \geq 2$ or $n \geq 2, l > 2$; as is the case for probabilistic models [34]. It was shown that the problem is NP by Abramsky in [2], and it has since been proved to be NP-complete by Abramsky, Gottlob & Kolaitis [6]. This gives some reason to suspect that it is not possible to obtain a classification of conditions that are necessary and sufficient for logical non-locality in full generality.

Table 4: Stages in the proof of proposition 5.1.

	1		1	0			1	0			
					1						
	1		0	1							
	0	1		1							
(a)											
	1	0	1	0			1	0			
	0		0	1			0	1			
	1	0	0	1							
	0	1	1								
(b)											
	1	0	1	0			1	0			
	0	1	0	1			0	1			
	1	0	0	1							
	0	1	1	0							
(c)											

5 Strong Contextuality & the PR Box

Strong non-locality [3] is a strictly stronger form of non-locality than logical non-locality. Whereas a model is said to be logically non-local if and only if there is some ‘1’ in its table which cannot be completed to a deterministic grid, a model is said to be *strongly non-local* if and only if no ‘1’ in its table can be completed to a deterministic grid. In the sheaf-theoretic language, strong non-locality is the property that no assignment of outcomes that is possible in the model can belong to a global assignment. Another way of putting this is that at the level of possibilities, a strongly contextual model contains no local part. We note that PR boxes represent the maximally non-local $(2,2,2)$ correlations, and that they are not realisable by quantum systems.

Theorem 2.5 can be used to provide the first constructive proof of a result originally due to Lal [3, 25] that the only strongly non-local $(2,2,2)$ models are the Popescu-Rohrlich no-signalling boxes [35].

Proposition 5.1. *The only strongly non-local no-signalling $(2,2,2)$ models are the PR boxes.*

Proof. By Theorem 2.5, strong non-locality is equivalent to the property that every ‘1’ witnesses a Hardy paradox. Simply by using this characterisation of strong contextuality and the requirement that the model must be no-signalling we can prove the required result. In the tabular representation, no-signalling translates to the condition that whenever a ‘1’ occurs in a table, its outcome row and column must contain at least one ‘1’ per measurement setting, for otherwise the possibility of witnessing a particular outcome for one party would depend on the measurement choice of the other (see [27] for a more detailed discussion).

For any choice of measurements there must be some possible outcome. This possible assignment is represented by a ‘1’ in the table, and it must witness a Hardy paradox. After re-labelling as necessary, we can represent the model as in Table 1. For this to be a no-signalling model, it is necessary to fill in ‘1’s as in Table 4 (a). Using the fact that the ‘1’s in the lower-right box must also witness Hardy paradoxes, we must fill in ‘0’s as in Table 4 (b). By no-signalling, the remaining unspecified entry in the upper-left box must be a ‘1’, and by the fact that it must witness a Hardy paradox, the remaining entry in the lower-right box must be a ‘0’. We thus arrive at Table 4 (c), and the only no-signalling empirical model whose possibilistic collapse has this form is the PR box. \square

6 Bell States are Anomalous

Projective measurements can be prescribed for almost all entangled two-qubit states such that the resulting empirical model will contain a Hardy paradox [23]. The prescription breaks down for the maximally entangled states, or the familiar Bell states. This naturally raises the question of whether there are any projective measurements that can be chosen for the maximally entangled states such that the resulting empirical model contains a Hardy paradox. Some failed attempts at finding a logical non-locality proof for the Bell states are described in [14]. The question gains even more importance in light of Theorem 2.5 which shows that this is equivalent to asking whether the maximally entangled states are logically non-local.

Somewhat surprisingly, we answer this question in the negative, and show that no projective measurements can be chosen that lead to a Hardy paradox (and thus logical non-locality) for a maximally entangled state. To the author's knowledge, this is the first full proof of the fact. A related result showing that if the same two measurements are available at each site then it is impossible to realise a Hardy paradox was proved independently by Abramsky & Constantin [4]. The proof we are about to present holds for any number of measurements per qubit, and without the restriction that the same set of measurements should be available for each qubit.

This is surprising since it shows that the Bell states are the only entangled two-qubit states not to be logically non-local. Moreover, a forthcoming work by Abramsky, Constantin & Ying claims that all n -qubit entangled states are logically non-local for $n > 2$ [5]. In [27] and [26] it was shown that the completeness of Hardy's paradox for logical non-locality immediately breaks down beyond the scenarios to which Theorem 2.5 apply. Beyond qubit states, therefore, since there are more ways of being logically non-local, it appears less likely that such an anomaly might arise. Indeed, logical non-locality has already been demonstrated for higher-dimensional maximally entangled states; e.g. the GHZ states [20] (see also section 7), or the states described in [13].

It appears, therefore, that despite being perhaps the most studied and utilised states in the field of quantum information, the Bell states are actually anomalous in the landscape of entangled states.

Theorem 6.1. *Bell states are not logically non-local.*

Proof. We prove the statement for the Bell state

$$|\phi^+\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle).$$

Since all other maximally entangled states are equivalent to this one up to local unitaries, which can easily be incorporated into the local measurements, the proof will extend to all maximally entangled states.

Any quantum mechanical empirical model obtained by making local projective measurements on $|\phi^+\rangle$ will necessarily give rise to a $(2, k, 2)$ model. By Theorem 2.5 we know that Hardy's paradox completely characterises logical non-locality for such scenarios, and that logical non-locality implies the occurrence of a Hardy paradox in some $(2, 2, 2)$ sub-model. It therefore suffices to show that for any observables $\{A_1, A_2\}$ for the first qubit and $\{B_3, B_4\}$ for the second qubit the resulting model does not contain a Hardy paradox.

The $+1$ and -1 eigenvectors for these measurements will be given by

$$\begin{aligned} |0_i\rangle &= \cos \frac{\theta_i}{2} |0\rangle + e^{i\phi_i} \sin \frac{\theta_i}{2} |1\rangle \\ |1_i\rangle &= \sin \frac{\theta_i}{2} |0\rangle + e^{-i\phi_i} \cos \frac{\theta_i}{2} |1\rangle \end{aligned}$$

where $\{(\theta_i, \phi_i)\}_{i \in \{1,2,3,4\}}$ label the coordinates of the $+1$ eigenvector of the respective measurements on the Bloch sphere. The amplitudes of the outcomes of the various joint measurements are calculated to be:

$$\begin{aligned} \langle 0_j 0_k | \psi \rangle &= \frac{1}{\sqrt{2}} \left(\cos \frac{\theta_j}{2} \cos \frac{\theta_k}{2} + e^{-i(\phi_j + \phi_k)} \sin \frac{\theta_j}{2} \sin \frac{\theta_k}{2} \right) \\ \langle 0_j 1_k | \psi \rangle &= \frac{1}{\sqrt{2}} \left(\cos \frac{\theta_j}{2} \sin \frac{\theta_k}{2} + e^{-i(\phi_j - \phi_k)} \sin \frac{\theta_j}{2} \cos \frac{\theta_k}{2} \right) \\ \langle 1_j 0_k | \psi \rangle &= \frac{1}{\sqrt{2}} \left(\sin \frac{\theta_j}{2} \cos \frac{\theta_k}{2} + e^{i(\phi_j - \phi_k)} \cos \frac{\theta_j}{2} \sin \frac{\theta_k}{2} \right) \\ \langle 1_j 1_k | \psi \rangle &= \frac{1}{\sqrt{2}} \left(\sin \frac{\theta_j}{2} \sin \frac{\theta_k}{2} + e^{i(\phi_j + \phi_k)} \cos \frac{\theta_j}{2} \cos \frac{\theta_k}{2} \right) \end{aligned}$$

where $j \in \{1, 2\}$ and $k \in \{3, 4\}$. We see that $\langle 0_j 0_k | \psi \rangle = e^{-i(\phi_j + \phi_k)} \langle 1_j 1_k | \psi \rangle$ and $\langle 0_j 1_k | \psi \rangle = \langle 1_j 0_k | \psi \rangle$ for each choice of measurements. Thus the symmetry of the underlying state manifests itself as a symmetry in the probabilities of the joint outcomes for each choice of measurements:

$$p(01 | AB) = p(10 | AB) \tag{1}$$

$$p(00 | AB) = p(11 | AB). \tag{2}$$

We know from Proposition 5.1 that the only strongly contextual $(2, 2, 2)$ models are the PR boxes, which are not quantum realisable [35]. So while the PR box satisfies these symmetries, it cannot be realised by measurements on $|\phi^+\rangle$. We show that there is a unique possibilistic $(2, 2, 2)$ model (up to re-labelling) which satisfies the symmetries (1) and (2) and is logically but not strongly non-local.

If a model is not strongly non-local then there exists at least one global assignment compatible with the model, or in tabular form at least one deterministic grid. Up to re-labelling this is represented in Table 5 (a). By the symmetry (2) there must exist a second global assignment, as in Table 5 (b). It is clear from the configuration of the table that none of the entries that have already been specified can witness a Hardy paradox. If the model is logically non-local, therefore, at least one of the unspecified entries in Table 5 (b) must witness a Hardy paradox. Up to re-labelling, this can be represented as in Table 5 (c). By the symmetry (1) the table must be completed to Table 5 (d). This (up to re-labelling) is the only possibilistic empirical model that respects the symmetries and is logically non-local without being strongly non-local. The question now is whether it can be realised by measurements on $|\phi^+\rangle$.

Consider the measurement statistics for the joint measurement $A_1 B_3$ required by Table 5. If these are to arise from quantum observables A_1 and B_3 , then $\langle \phi^+ | 0_1 0_3 \rangle = \langle \phi^+ | 1_1 1_3 \rangle = \frac{1}{\sqrt{2}}$ and $\langle \phi^+ | 0_1 1_3 \rangle = \langle \phi^+ | 1_1 0_3 \rangle = 0$. So, either $|0_1\rangle = |0_3\rangle = |0\rangle$ and $|1_1\rangle = |1_3\rangle = |1\rangle$ up to an overall sign or vice versa. The eigenvectors of both observables are $\{|0\rangle, |1\rangle\}$, so they must simply be Pauli X operators (up to a common sign, which would allow for re-labelling the outcomes):

$$A_1 = B_3 = \pm X. \tag{3}$$

Table 5: Stages in the proof of Theorem 6.1.

	1	1
	1	1

(a)

	1	1
	1	1
	1	1

(b)

	1	1
	0	1
	1	1
	1	1
	1	1

(c)

	B_3	B_4
A_1	1 0	1 0
	0 1	0 1
A_2	1 1	1 0
	1 1	0 1

(d)

A similar argument applies for the joint measurements A_1B_4 and A_2B_4 , showing that

$$A_1 = B_4 = \pm X, \quad (4)$$

$$A_2 = B_4 = \pm X. \quad (5)$$

Equations (3–5) imply that

$$A_1 = A_2 = B_3 = B_4 = \pm X;$$

but therefore the measurement statistics for A_2B_3 must be the same as for each of the other joint measurements, and it is not possible to realise Table 5 (d). This completes the proof that no quantum mechanical logically non-local empirical model can be obtained by considering (any number of) local projective measurements on the Bell state. \square

Symmetry is important here: the symmetry of the underlying state manifests itself as a symmetry of the probabilities of outcomes for each joint measurement. By Theorem 2.5, logical non-locality implies a particular relationship between certain probabilities in each of these distributions (a Hardy paradox). However, quantum mechanically there cannot exist local projective measurements that realise these correlations and respect the symmetries at the same time. On the other hand, there exists a whole family of no-signalling empirical models which are logically non-local and respect the symmetries. These are the no-signalling models with support as in Table 5 (d), and the PR box.

These models have some interesting properties in their own right [26]: despite not being realisable quantum mechanically, they may lie within the Tsirelson bound, coming arbitrarily close to the local polytope. They can be seen, however, to violate information causality, which has been proposed as a physical principle that might characterise quantum correlations, by means of the same protocol described in [33]. In fact, similar families of models to this one have already been considered in this context in [8].

We also note that Fritz [18] has considered quantum analogues of Hardy’s paradox. These are not realisable quantum mechanically, but can arise in more general no-signalling empirical models. An interesting point is that table 5 (d) contains two such paradoxes, and so the fact that any model with this support is not quantum realisable also follows more directly from these results.

7 Hardy Non-locality with Certainty

While Hardy’s paradox is considered to be an ‘almost probability free’ non-locality proof, much of the literature on Hardy’s paradox (e.g. [12, 16]) is concerned with the value of the *paradoxical probability*; i.e. the probability of obtaining the particular outcome assignment that witnesses a Hardy paradox (Definition 2.1). This is motivated as being especially relevant for experimental tests. In this section, we will show how Hardy non-locality can be demonstrated in such a way that even this probability becomes irrelevant.

The author has recently become aware of a similar proposal by Cabello [13]. We note that the present approach, which was also presented in [26], has the advantage of being much simpler both in terms of the argument and the empirical model required (in fact, the model has already been experimentally realised).

As previously mentioned, Hardy [23] prescribed measurements for all entangled two-qubit states (excluding the maximally entangled ones) such that the resulting empirical model contains a Hardy paradox. For this family of models the maximum paradoxical probability is

$$p_{\max} = \frac{5\sqrt{5} - 11}{2} \approx 0.09. \quad (6)$$

Table 6: The PR box.

	$\frac{1}{2}$	0	$\frac{1}{2}$	0
	0	$\frac{1}{2}$	0	$\frac{1}{2}$
	$\frac{1}{2}$	0	0	$\frac{1}{2}$
	0	$\frac{1}{2}$	$\frac{1}{2}$	0

A model has also been found for which the tripartite Hardy paradox can be witnessed with probability 0.125 [19], and in [17] it is demonstrated that for a generalised no-signalling theory it is possible to witness a $(2, 2, 2)$ Hardy paradox with probability 0.5. It was shown that the ladder generalisation of Hardy's paradox could achieve a paradoxical probability approaching 0.5 for $(2, k, 2)$ scenarios, as $k \rightarrow \infty$. For the $(2, 2, l)$ scenario, Chen et al. [16] have recently argued that it is possible to achieve a paradoxical probability of ≈ 0.4 in the large d limit for two qudit systems with the paradox presented in Table 3. From our Proposition 3.1, it becomes clear that they are essentially summing the probabilities of witnessing $(l - 1)^2/2$ (coarse-grained) Hardy paradoxes.

In this section, we use completeness of Hardy non-locality to achieve a striking improvement on these results, demonstrating by simple arguments that by considering such a summing of paradoxical probabilities Hardy non-locality can be witnessed with certainty for a tripartite quantum system, which turns out to be the familiar GHZ-Mermin model [21, 20, 30, 31]. We also show that Hardy non-locality can be witnessed with certainty for a particular non-quantum $(2, 2, 2)$ empirical model, which turns out to be the PR box. We begin with the $(2, 2, 2)$ case.

Proposition 7.1. *The PR box witnesses a Hardy paradox with certainty.*

Proof. The probabilistic version of the PR box is given in Table 6. We have already observed in the proof of Proposition 5.1 that every joint outcome that has non-zero probability witnesses a Hardy paradox. Therefore, each non-zero entry in the table represents a joint outcome that witnesses Hardy non-locality with paradoxical probability of 0.5, and it is clear that for each joint measurement the probability of obtaining an outcome that witnesses a Hardy paradox is 1. \square

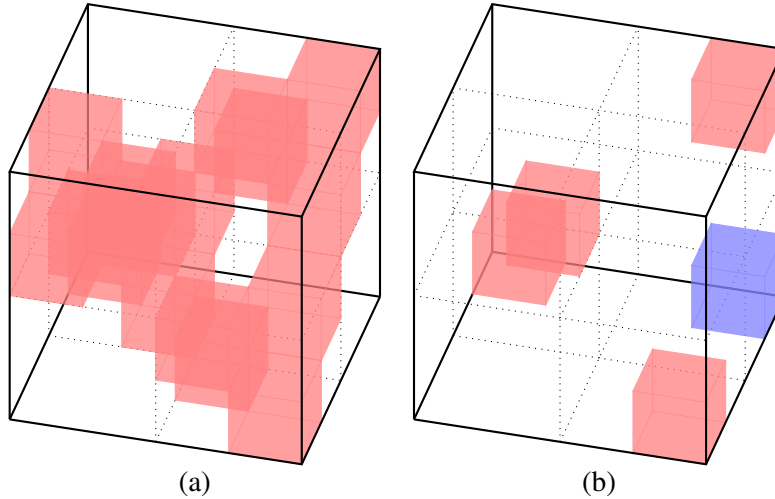
The PR box achieves the maximum individual paradoxical probability of 0.5 for a no-signalling model to witness a Hardy paradox found by Choudhary et al. in [17], but by a much simpler argument and using a familiar and well-studied model. More importantly, we see that since every joint outcome witnesses a Hardy paradox, the more relevant parameter, the probability of witnessing Hardy non-locality, is actually 1 for any choice of measurements. However, it can be shown that for any quantum realisable $(2, k, 2)$ empirical model it is not possible to use this method of summing paradoxical probabilities to witness Hardy non-locality with higher probability.

Proposition 7.2. *For any quantum realisable $(2, k, 2)$ empirical model, the probability of witnessing a Hardy paradox cannot be improved by summing the paradoxical probabilities for different paradoxes occurring within the same model.*

Table 7: The relevant portion of the GHZ-Mermin possibilistic empirical model. The suppressed rows of the table $\{XXY, XYX, YXX, YYY\}$ have full support. See Figure 2 (a) for the three dimensional representation of this model.

	000	001	010	011	100	101	110	111
$X \ X \ X$	1	0	0	1	0	1	1	0
$X \ Y \ Y$	0	1	1	0	1	0	0	1
$Y \ X \ X$	0	1	1	0	1	0	0	1
$Y \ Y \ X$	0	1	1	0	1	0	0	1

Figure 2: (a) The GHZ model. We represent only the red, impossible outcomes; all other entries are possible. (b) Hardy's paradox within the GHZ model; the blue outcome is possible.



Proof. First, we note that it suffices to prove the proposition for $(2,2,2)$ models, since a $(2,k,2)$ model contains a Hardy paradox if and only if some $(2,2,2)$ sub-model contains a Hardy paradox. In order to obtain an improvement in the probability of witnessing Hardy non-locality it would have to be the case that for some joint measurement, more than one Hardy paradox could be witnessed. Working within our formalism, it is clear that any such empirical model must belong to the family of models with support given by Table 5 (d) up to re-labelling of measurements and outcomes. In this family, for any joint measurement the probability of witnessing Hardy non-locality is always '1'. However, it was shown in the proof of Theorem 6.1 that no model in this family is quantum realisable, and the result follows. \square

We now consider the $(3,2,2)$ empirical model used in the Mermin version [30] of the GHZ logical non-locality proof [20]. It should be noted that the GHZ argument is not of the tripartite Hardy form mentioned in section 2. Here, we need only consider a subset of the measurement contexts, shown in Table 7 in orthodox notation, and three dimensional representation in Figure 2 (a).

Proposition 7.3. *The GHZ model witnesses a Hardy paradox with certainty.*

Proof. The three dimensional representation makes it easy to identify a tripartite Hardy paradox, which is shown in Figure 2 (b). It can be expressed algebraically as follows.

- $p(1, 1, 1 \mid Y, Y, Y) > 0$
- $p(1, 1, 0 \mid Y, Y, X) = p(1, 0, 1 \mid Y, X, Y) = p(0, 1, 1 \mid X, Y, Y) = 0$
- $p(0, 0, 0 \mid X, X, X) = 0$

Up to re-labelling, this is the form of the n -partite Hardy paradox defined in section 2. Moreover, it can similarly be demonstrated that any joint outcome for the measurement context YYY witnesses a Hardy paradox (a more careful treatment is given in [26]). The paradoxical probability is $p(1, 1, 1 \mid Y, Y, Y) = 0.125$. However, since every outcome to the measurement YYY witnesses some Hardy paradox, then it is clear that the probability of witnessing Hardy non-locality is actually 1. \square

This provides a much simpler tripartite Hardy argument than that of Ghosh, Kar and Sarkar [19], using a simpler empirical model (theirs also assumed the GHZ state, but with different measurements). We obtain the same maximum of 0.125 for the individual paradoxical probabilities, but in fact we do better than that, since every possible outcome to the joint measurement YYY witnesses some Hardy paradox, and therefore we witness Hardy non-locality with certainty. The model here is exactly the GHZ-Mermin model, since the observables available at each subsystem are simply the X and Y operators. So we have proved that the GHZ experiment [20] witnesses Hardy non-locality with certainty.

Corollary 7.4. *The GHZ experiment [20] witnesses Hardy non-locality with certainty.*

Mermin gave logical non-locality proofs for n -partite generalisations of the GHZ state [29] for all $n > 2$. Again, his arguments were not of the Hardy form, but we can also show how to generalise Proposition 7.3 to some of the $\text{GHZ}(n)$ models.

Proposition 7.5. *All $\text{GHZ}(n)$ models for $n = 3 \bmod 4$ witness an n -partite Hardy paradox with certainty.*

Proof. See [26]. \square

It should be pointed out that even though we can say with certainty that some Hardy paradox will be witnessed in these models, the individual paradoxical probabilities are $\frac{1}{2^n}$, and so the maximum for these individual paradoxical probabilities is obtained for the tripartite GHZ model.

This kind of result does not hold for $\text{GHZ}(n)$ models for which $n \neq 3 \bmod 4$, as it can be shown that these models do not contain n -partite Hardy paradoxes. This is because any $(n, 2, 2)$ Hardy paradox must take the form of one of the paradoxes in the proof of Proposition 7.5, but it can easily be verified that the counting arguments for identifying such paradoxes in $\text{GHZ}(n)$ models work if and only if $n = 3 \bmod 4$.

8 Conclusions

The completeness theorems proved by the author and Fritz in [27] showed that the occurrence of a Hardy paradox is a necessary and sufficient condition for logical non-locality in $(2, 2, l)$ and $(2, k, 2)$ Bell scenarios. In this paper we have seen that these lead to a variety of interesting consequences and applications.

One direct consequence is that the Hardy paradox must subsume all other non-locality arguments for $(2, 2, l)$ and $(2, k, 2)$ scenarios, and we have demonstrated this for the ladder paradoxes and the Chen

et al. paradox. We have also seen that for $(2,2,l)$ and $(2,k,2)$ scenarios, polynomial algorithms can be given for deciding non-locality, which is significant since the general problem is known to be NP-complete [6]. Furthermore, the theorems have been used to provide the first constructive proof that the PR boxes are the only strongly contextual $(2,2,2)$ models, as well as the first full proof that the Bell states, despite being maximally entangled, are the only entangled two-qubit states that are not logically non-local.

Together with forthcoming work by Abramsky, Constantin & Ying [5] which will show that all entangled n -partite qubit states are logically non-local for $n > 2$, as well as the fact that higher-dimensional maximally entangled states have been shown to be logically non-local (e.g. [20, 13]) this singles out the Bell states as being anomalous in the landscape of entangled states. This is quite surprising in light of the fact that they are perhaps the most studied and utilised of entangled states. We mention, however, that it remains to be seen whether the result still holds when we allow for POVM's.

In section 7, we have taken advantage of the perspective gained within our framework to demonstrate a striking improvement on the probability of witnessing a Hardy paradox, which is often used in the literature as a measure of the quality of Hardy non-locality. With much simpler arguments, it has been demonstrated that a tripartite quantum system can in fact witness Hardy non-locality with certainty. Interestingly, the empirical model used for this proof was exactly that of the GHZ-Mermin non-locality proof, and has already been experimentally realised [20]. Though it is not quantum realisable, we have also shown that the PR box [35] has this property.

With regard to open questions, we note that Abramsky has proved a correspondence between possibilistic empirical models and relational database theory [1]. It remains to be explored whether the completeness theorems of this chapter might find applications in database theory, or indeed whether similar results already exist in the field that might lead to further insights on non-locality.

Acknowledgements

The author thanks Samson Abramsky, Tobias Fritz, Ray Lal and Rui Soares Barbosa for comments and discussions.

References

- [1] Samson Abramsky (2012): *Relational databases and Bell's theorem*. *arXiv:1208.6416* ArXiv preprint.
- [2] Samson Abramsky (2013): *Relational hidden variables and non-locality*. *Studia Logica* 101(2), pp. 411–452, doi:10.1007/s11225-013-9477-4.
- [3] Samson Abramsky & Adam Brandenburger (2011): *The sheaf-theoretic structure of non-locality and contextuality*. *New Journal of Physics* 13(11), p. 113036.
- [4] Samson Abramsky & Carmen Constantin (2013): *A classification of multipartite states by degree of non-locality*. In: *Proceedings of Quantum Physics & Logic*.
- [5] Samson Abramsky, Carmen Constantin & Shenggang Ying (2013): Forthcoming.
- [6] Samson Abramsky, Georg Gottlob & Phokion G Kolaitis (2013): *Robust constraint satisfaction and local hidden variables in quantum mechanics*. In: *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, AAAI Press, pp. 440–446.
- [7] Samson Abramsky & Lucien Hardy (2012): *Logical Bell inequalities*. *Physical Review A* 85(6), p. 062114.
- [8] Jonathan Allcock, Nicolas Brunner, Marcin Pawłowski & Valerio Scarani (2009): *Recovering part of the quantum boundary from information causality*. *arXiv:0906.3464* ArXiv preprint.

- [9] M. Barbieri, C. Cinelli, F. De Martini & P. Mataloni (2005): *Test of quantum nonlocality by full collection of polarization entangled photon pairs*. *The European Physical Journal D-Atomic, Molecular, Optical and Plasma Physics* 32(2), pp. 261–267. Available at <http://www.springerlink.com/index/k5y7k27ttjga27un.pdf>.
- [10] John S. Bell (1964): *On the Einstein-Podolsky-Rosen paradox*. *Physics* 1(3), pp. 195–200.
- [11] John S. Bell (1987): *Speakable and unspeakable in quantum mechanics: Collected papers on quantum philosophy*. Cambridge University Press.
- [12] D. Boschi, S. Branca, F. De Martini & L. Hardy (1997): *Ladder proof of nonlocality without inequalities: Theoretical and experimental results*. *Physical Review Letters* 79(15), pp. 2755–2758. Available at <http://link.aps.org/doi/10.1103/PhysRevLett.79.2755>.
- [13] Adán Cabello (1998): *Ladder proof of nonlocality without inequalities and without probabilities*. *Physical Review A* 58(3), p. 1687.
- [14] Adán Cabello (2000): *Nonlocality without inequalities has not been proved for maximally entangled states*. *Physical Review A* 61(2), p. 022119.
- [15] José L. Cereceda (2004): *Hardy's nonlocality for generalized n-particle GHZ states*. *Physics Letters A* 327(5-6), pp. 433 – 437.
- [16] Jing-Ling Chen, Adán Cabello, Zhen-Peng Xu, Hong-Yi Su, Chunfeng Wu & LC Kwek (2013): *Hardy's paradox for high-dimensional systems*. *Physical Review A* 88(6), p. 062116.
- [17] Sujit K Choudhary, Sibasish Ghosh, Guruprasad Kar, Samir Kunkri, Ramij Rahaman & Anirban Roy (2008): *Hardy's non-locality and generalized non-local theory*. *arXiv:0807.4414 ArXiv preprint*.
- [18] Tobias Fritz (2011): *Quantum analogues of Hardy's nonlocality paradox*. *Foundations of Physics* 41(9), pp. 1493–1501.
- [19] Sibasish Ghosh, G. Kar & Debasis Sarkar (1998): *Hardy's nonlocality for entangled states of three spin-1/2 particles*. *Physics Letters A* 243(5), pp. 249–255.
- [20] Daniel M. Greenberger, Michael A. Horne, Abner Shimony & Anton Zeilinger (1990): *Bells theorem without inequalities*. *American Journal of Physics* 58(12), pp. 1131–1143.
- [21] Daniel M. Greenberger, Michael A. Horne & Anton Zeilinger (1989): *Going Beyond Bell's Theorem*. In M. Kafatos, editor: *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, Kluwer, pp. 69–72.
- [22] Lucien Hardy (1992): *Quantum mechanics, local realistic theories, and Lorentz-invariant realistic theories*. *Physical Review Letters* 68(20), pp. 2981–2984.
- [23] Lucien Hardy (1993): *Nonlocality for two particles without inequalities for almost all entangled states*. *Physical Review Letters* 71(11), pp. 1665–1668.
- [24] Peter Heywood & Michael Redhead (1983): *Nonlocality and the Kochen-Specker paradox*. *Foundations of Physics* 13, pp. 481–499. Available at <http://dx.doi.org/10.1007/BF00729511>.
- [25] Raymond Lal (2013): *A sheaf-theoretic approach to the contextuality of cluster states*. Forthcoming.
- [26] Shane Mansfield (2013): *The mathematical structure of non-locality & contextuality*. D.Phil. thesis, Oxford University. Available at <http://ora.ox.ac.uk/objects/uuid:394bb375-db3f-4a12-bdd8-cd1ab5809573>.
- [27] Shane Mansfield & Tobias Fritz (2012): *Hardy's non-locality paradox and possibilistic conditions for non-locality*. *Foundations of Physics* 42, pp. 709–719. Available at <http://dx.doi.org/10.1007/s10701-012-9640-1>.
- [28] Shane Mansfield & Rui Soares Barbosa (2013): *Extendability in the Sheaf-theoretic Approach: Construction of Bell Models from Kochen-Specker Models*. In: *Proceedings of Quantum Physics & Logic X*.
- [29] N David Mermin (1990): *Extreme quantum entanglement in a superposition of macroscopically distinct states*. *Physical Review Letters* 65(15), pp. 1838–1840.

- [30] N. David Mermin (1990): *Quantum mysteries revisited*. *American Journal of Physics* 58(8), pp. 731–734.
- [31] N. David Mermin (1990): *Simple unified form for the major no-hidden-variables theorems*. *Physical Review Letters* 65(27), pp. 3373–3376.
- [32] N.D. Mermin (1994): *Quantum mysteries refined*. *American Journal of Physics* 62, p. 880.
- [33] Marcin Pawłowski, Tomasz Paterek, Dagomir Kaszlikowski, Valerio Scarani, Andreas Winter & Marek Żukowski (2009): *Information causality as a physical principle*. *Nature* 461(7267), pp. 1101–1104.
- [34] I. Pitowsky (1991): *Correlation polytopes: their geometry and complexity*. *Mathematical Programming* 50(1), pp. 395–414. Available at <http://www.springerlink.com/index/njt1m90jtgrw6607.pdf>.
- [35] Sandu Popescu & Daniel Rohrlich (1994): *Quantum nonlocality as an axiom*. *Foundations of Physics* 24(3), pp. 379–385.
- [36] Zizhu Wang & Damian Markham (2012): *Nonlocality of symmetric states*. *Physical Review Letters* 108(21), p. 210407.

QPEL: Quantum Program and Effect Language

Robin Adams

Radboud University Nijmegen

`r.adams@cs.ru.nl`

We present the syntax and rules of deduction of QPEL (Quantum Program and Effect Language), a language for describing both quantum programs, and properties of quantum programs — *effects* on the appropriate Hilbert space. We show how semantics may be given in terms of *state-and-effect triangles*, a categorical setting that allows semantics in terms of Hilbert spaces, C^* -algebras, and other categories. We prove soundness and completeness results that show the derivable judgements are exactly those provable in all state-and-effect triangles.

1 Introduction

There is a growing number of quantum programming languages, and there is a need for a syntactic method of reasoning about these quantum programs: both in the hope of making automated tools for proving the correctness of programs, and because experience in other fields shows that many problems that are difficult when treated semantically

We present QPEL, a syntax for both describing quantum programs, and properties of quantum programs (quantum predicates, or effects). This system should be useful for reasoning about quantum programs and proving their correctness, as well as showing more generally how a language for quantum effects may be added on top of any quantum programming language. The system is loosely based on Selinger’s Quantum Programming Language (QPL) [14].

The part of the system that describes quantum programs is a *linear* type theory (see [6, 7]). This is a linear type theory: we are not able to duplicate or erase data. Duplication of quantum data would violate the no-cloning theorem. We do allow deletion of data (which corresponds to partial measurement), but this must be done explicitly in the program.

The part of the system that describes quantum predicates is based on the fact that the effects on a Hilbert state or C^* -algebra form an *effect algebra* - in fact, an *effect module* over the appropriate effect monoid [13].

There is a categorical structure called the *state-and-effect triangle* that has been shown to generalise several different ways of giving semantics to quantum computing, including Hilbert spaces and C^* -algebras. The first version of QPEL we present captures all and only the structure of a state-and-effect triangle. We show how to give semantics in an arbitrary triangle, and prove a Soundness and Completeness Theorem. We proceed to discuss what would need to be added to the system to represent other features of a quantum programming language, particularly qubits.

2 Preliminaries

2.1 Notation

If E and F are expressions involving partial functions, we write:

- $E = F$ to denote: E and F are both defined, and their values are equal;
- $E \simeq F$ to denote: E is defined if and only if F is defined, in which case their values are equal;
- $E \xrightarrow{\sim} F$ to denote: if E is defined, then F is defined and their values are equal.

2.2 Effect Algebras and Effect Monoids

We represent the effects on a quantum system by the elements of an *effect module* over an *effect monoid* M , whose elements we call *scalars*. The canonical example is the effects on a Hilbert space form an effect module over $[0, 1]$, with the scalars being probabilities.

Definition 1 (Partial Commutative Monoid). A *partial commutative monoid* consists of a set M ; an element $0 \in M$, the *zero*; and a partial binary operation $\odot : M^2 \rightarrow M$, the (*partial*) *sum*; such that, writing $x \perp y$ (' x is *orthogonal* to y ') iff $x \odot y$ is defined:

- $x \odot y \simeq y \odot x$
- $x \odot (y \odot z) \simeq (x \odot y) \odot z$
- $x \odot 0 = x$

for all $x, y, z \in M$.

Definition 2 (Effect Algebra). An *effect algebra* is a partial commutative monoid E with a (total) function $(-)^{\perp} : E \rightarrow E$, the *orthosupplement*, such that

- $x \odot y = 0^{\perp}$ iff $y = x^{\perp}$.
- If $x \perp 0^{\perp}$ then $x = 0$.

We write 1 for 0^{\perp} .

Definition 3 (Effect Monoid). An effect monoid is an effect algebra E with a binary operation $\cdot : E^2 \rightarrow E$, the *multiplication*, such that

- $(x \odot y) \cdot z \xrightarrow{\sim} (x \cdot z) \odot (y \cdot z)$
- $x \cdot (y \odot z) \xrightarrow{\sim} (x \cdot y) \odot (x \cdot z)$
- $1 \cdot x = x \cdot 1 = x$
- $x \cdot (y \cdot z) = (x \cdot y) \cdot z$

Effect monoids were introduced in [10]. An effect monoid is a monoid in the category of effect algebras.

Definition 4 (Effect Module). An *effect module* over an effect monoid E is an effect algebra A with a binary operation $\cdot : E \times A \rightarrow A$ called *scalar multiplication* such that:

- $r \cdot (x \odot y) \xrightarrow{\sim} (r \cdot x) \odot (r \cdot y)$
- $(r \odot s) \cdot x \xrightarrow{\sim} (r \cdot x) \odot (s \cdot x)$
- $(r \cdot s) \cdot x = r \cdot (s \cdot x)$
- $1 \cdot x = x$

2.2.1 Examples

1. For any Hilbert space H , the set of effects over H forms an effect module over the effect monoid $[0, 1]$, with $F \otimes G = F + G$ iff $F + G$ is an effect [13].
2. Given a C^* -algebra A , the set of *effects* in A (positive elements below the unit) form an effect module over the real numbers $[0, 1]$.

2.3 Convex Sets

Definition 5. Given an effect monoid E , the *distribution monad* $\mathcal{D}_E : \mathbf{Set} \rightarrow \mathbf{Set}$ is defined as follows.

$$\mathcal{D}_E X = \{ \phi : X \rightarrow E : \text{supp } \phi \text{ is finite, } \sum_{x \in X} \phi(x) \text{ exists and is equal to } 1 \}$$

where $\text{supp } \phi = \{x \in X : \phi(x) \neq 0\}$.

The category \mathbf{Conv}_E of *convex sets* and *affine functions* over E is the Eilenberg-Moore category of \mathcal{D}_E . A convex set may thus be thought of as a set X together with a function mapping any finite tuple $\langle r_1, \dots, r_n \rangle$ of elements of E that sum to 1, and any tuple $\langle x_1, \dots, x_n \rangle$ of elements of X , to an element $r_1 x_1 + \dots + r_n x_n$ of X .

Theorem 6. *The distribution monad is a strong monad. It is a commutative monad iff E is commutative.*

Corollary 6.1. *If E is commutative, then \mathbf{Conv}_E is a symmetric monoidal category.*

The convex set $A \otimes B$ consists of all sums $r_1(a_1, b_1) + \dots + r_n(a_n, b_n)$ ($r_1 \otimes \dots \otimes r_n = 1, a_i \in A, b_i \in B$), quotiented by the appropriate equivalence relation. An affine function $f : A \otimes B \rightarrow C$ in \mathbf{Conv}_M is determined by the values $f(a, b)$ for $a \in A$ and $b \in B$.

Theorem 7. *The hom-functors $\mathbf{EMod}_E[-, E] \dashv \mathbf{Conv}_E[-, E] : \mathbf{EMod}_E^{\text{op}} \rightleftarrows \mathbf{Conv}_E$ form an adjunction.*

Proof. To appear in [12]. The special case $E = [0, 1]$ was proved in [9]. \square

3 Syntax and Rules of Deduction

We begin with a system that represents a symmetric monoidal closed category with distributive coproducts, with an effect module of predicates over each object.

The syntax of the system is given by the following grammar:

Type	$A ::= A \otimes A \mid I \mid A + B$
Term	$M ::= x \mid \langle M, M \rangle \mid \text{let } \langle x, x \rangle = M \text{ in } M \mid \langle \rangle \mid \text{let } \langle \rangle = M \text{ in } N \mid \text{inl}(M) \mid \text{inr}(M) \mid$ $(\text{case } M \text{ of } \text{inl}(x) \mapsto M \mid \text{inr}(x) \mapsto M) \mid \text{measure } \phi \mapsto M \mid \dots \mid \phi \mapsto M$
Proposition	$\phi ::= 0 \mid \phi \otimes \phi \mid \phi^\perp \mid \phi \cdot \psi$
Context	$\Gamma ::= \langle \rangle \mid \Gamma, x : A$
Judgement	$\mathcal{J} ::= \Gamma \vdash \phi \text{ prop} \mid \Gamma \vdash \phi \leq \psi$

We write $\Gamma \vdash \phi \perp \psi$ for $\Gamma \vdash \phi \leq \psi^\perp$ and $\Gamma \vdash \phi \equiv \psi$ for the two judgements $\Gamma \vdash \phi \leq \psi$ and $\Gamma \vdash \psi \leq \phi$. We write 1 for 0^\perp .

The intended interpretation is that a term $x_1 : A_1, \dots, x_n : A_n \vdash M : B$ represents a quantum program that takes n inputs of type A_1, \dots, A_n , and returns an output of type B . The type $A \otimes B$ represents a pair of values of A and B (possibly entangled). A program of type $A + B$ represents a value that is either of type A or of type B , with the 'or' understood classically. Thus, we may represent the type of classical bits as $I + I$.

We discuss how to represent qubits in the system in Section 5.

Structural Rule

$$\frac{\Gamma, x : A, y : B, \Delta \vdash \mathcal{J}}{\Gamma, y : B, x : A, \Delta \vdash \mathcal{J}}$$

Term Formation

$$\begin{array}{c} \frac{\Gamma \vdash M : A \quad \Delta \vdash N : B}{\Gamma, \Delta \vdash \langle M, N \rangle : A \otimes B} \quad \frac{\overline{x : A \vdash x : A} \quad \Gamma \vdash M : A \otimes B \quad \Delta, x : A, y : B \vdash N : C}{\Gamma, \Delta \vdash \text{let } \langle x, y \rangle = M \text{ in } N : C} \\[10pt] \frac{}{\vdash \langle \rangle : I} \quad \frac{\Gamma \vdash M : I \quad \Delta \vdash N : A}{\Gamma, \Delta \vdash \text{let } \langle \rangle = M \text{ in } N : A} \\[10pt] \frac{\Gamma \vdash M : A}{\Gamma \vdash \text{inl}(M) : A + B} \quad \frac{\Gamma \vdash M : B}{\Gamma \vdash \text{inr}(M) : A + B} \\[10pt] \frac{\Gamma \vdash M : A + B \quad \Delta, x : A \vdash N : C \quad \Delta, y : B \vdash P : C}{\Gamma, \Delta \vdash \text{case } M \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P : C} \\[10pt] \frac{\Gamma \vdash \phi_1 \otimes \dots \otimes \phi_n \text{ prop} \quad \Delta \vdash M_1 : A \quad \dots \quad \Delta \vdash M_n : A}{\Gamma, \Delta \vdash \text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n : A} \end{array}$$

Figure 1: Rules of Deduction of QPEL

3.1 Rules of Deduction

The rules of deduction are given in Figures 1–4. The rules called ‘Commuting Conversions’ are based on the rules in [6]. Note the rule for the formation of $\phi \otimes \psi$. In order to be able to form the proposition $\phi \otimes \psi$, we must have a derivation that ϕ and ψ are orthogonal (i.e. $\phi \leq \psi^\perp$).

3.2 Metatheorems

We can prove the following properties, which show that the typing system is well behaved.

Lemma 8.

1. If $\Gamma \vdash M = N : A$ then $\Gamma \vdash M : A$ and $\Gamma \vdash N : A$.
2. If $\Gamma \vdash \phi \leq \psi$ then $\Gamma \vdash \phi \text{ prop}$ and $\Gamma \vdash \psi \text{ prop}$.
3. If $\Gamma \vdash \phi \otimes \psi \text{ prop}$ then $\Gamma \vdash \phi \perp \psi$.
4. **Substitution** If $\Gamma \vdash M : A$ and $\Delta, x : A, \Delta' \vdash \mathcal{J}$ then $\Delta, \Gamma, \Delta' \vdash [M/x] \mathcal{J}$.
5. **Functionality** If $\Gamma \vdash M = N : A$ and $\Delta, x : A \vdash P : B$ then $\Gamma, \Delta \vdash [M/x]P = [N/x]P : B$.
6. **Functionality** If $\Gamma \vdash M = N : A$ and $\Delta, x : A \vdash \phi \text{ prop}$ then $\Gamma, \Delta \vdash [M/x]\phi \equiv [N/x]\phi$.
7. If $\Gamma, x : A, y : B \vdash M : C$, $\Delta \vdash N : A \otimes B$ and $\Theta, z : C \vdash P : D$, then

$$\Gamma, \Delta, \Theta \vdash [\text{let } \langle x, y \rangle = N \text{ in } M/z]P = (\text{let } \langle x, y \rangle = N \text{ in } [M/z]P) : D$$

Equality of Terms

$$\frac{\Gamma \vdash M : A}{\Gamma \vdash M = M : A} \quad \frac{\Gamma \vdash M = N : A}{\Gamma \vdash N = M : A} \quad \frac{\Gamma \vdash M = N : A \quad \Gamma \vdash N = P : A}{\Gamma \vdash M = P : A}$$

Congruences

$$\frac{\Gamma \vdash M = M' : A \quad \Delta \vdash N = N' : B}{\Gamma, \Delta \vdash \langle M, N \rangle = \langle M', N' \rangle : A \otimes B}$$

$$\frac{\Gamma \vdash M = M' : A \otimes B \quad \Delta, x : A, y : B \vdash N = N' : C}{\Gamma, \Delta \vdash (\text{let } \langle x, y \rangle = M \text{ in } N) = (\text{let } \langle x, y \rangle = M' \text{ in } N') : C}$$

$$\frac{\Gamma \vdash M = M' : I \quad \Delta \vdash N = N' : A}{\Gamma, \Delta \vdash (\text{let } \langle \rangle = M \text{ in } N) = (\text{let } \langle \rangle = M' \text{ in } N') : A}$$

$$\frac{\Gamma \vdash M = N : A}{\Gamma \vdash \text{inl}(M) = \text{inl}(N) : A + B} \quad \frac{\Gamma \vdash M = N : B}{\Gamma \vdash \text{inr}(M) = \text{inr}(N) : A + B}$$

$$\frac{\Gamma \vdash M = M' : A + B \quad \Delta, x : A \vdash N = N' : C \quad \Delta, y : B \vdash P = P' : C}{\Gamma, \Delta \vdash (\text{case } M \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P) = (\text{case } M' \text{ of } \text{inl}(x) \mapsto N' \mid \text{inr}(y) \mapsto P') : C}$$

$$\frac{\begin{array}{c} \Gamma \vdash \phi_1 \otimes \dots \otimes \phi_n \text{ prop} \\ \Gamma \vdash \phi_1 \equiv \psi_1 \dots \Gamma \vdash \phi_n \equiv \psi_n \\ \Delta \vdash M_1 = N_1 : A \dots \Delta \vdash M_n = N_n : A \end{array}}{\Gamma, \Delta \vdash (\text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n) = (\text{measure } \psi_1 \mapsto M_1 \mid \dots \mid \psi_n \mapsto M_n) : A}$$

 β -conversions

$$\frac{\Gamma \vdash M : A \quad \Delta \vdash N : B \quad \Theta, x : A, y : B \vdash P : C}{\Gamma, \Delta, \Theta \vdash (\text{let } \langle x, y \rangle = \langle M, N \rangle \text{ in } P) = [M/x, N/y]P : C}$$

$$\frac{\Gamma \vdash M : A}{\Gamma \vdash (\text{let } \langle \rangle = \langle \rangle \text{ in } M) = M : A}$$

$$\frac{\Gamma \vdash M : A \quad \Gamma, x : A \vdash N : C \quad \Gamma, y : B \vdash P : C}{\Gamma \vdash (\text{case } \text{inl}(M) \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P) = [M/x]N : C}$$

$$\frac{\Gamma \vdash M : B \quad \Gamma, x : A \vdash N : C \quad \Gamma, y : B \vdash P : C}{\Gamma \vdash (\text{case } \text{inr}(M) \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P) = [M/y]P : C}$$

 η -conversions

$$\frac{\Gamma \vdash M : A \otimes B}{\Gamma \vdash M = \text{let } \langle x, y \rangle = M \text{ in } \langle x, y \rangle : A \otimes B} \quad \frac{\Gamma \vdash M : I}{\Gamma \vdash M = (\text{let } \langle \rangle = M \text{ in } \langle \rangle) : I}$$

$$\frac{\Gamma \vdash M : A + B}{\Gamma \vdash M = (\text{case } M \text{ of } \text{inl}(x) \mapsto \text{inl}(x) \mid \text{inr}(y) \mapsto \text{inr}(y)) : A + B}$$

Figure 2: Rules of Deduction of QPEL

Commuting Conversions

$$\begin{array}{c}
\frac{\Gamma \vdash M : A \otimes B \quad \Delta, x : A, y : B \vdash N : C \otimes D \quad \Theta, t : C, u : D \vdash P : E}{\Gamma, \Delta, \Theta \vdash (\text{let } \langle x, y \rangle = M \text{ in let } \langle t, u \rangle = N \text{ in } P) = (\text{let } \langle t, u \rangle = (\text{let } \langle x, y \rangle = M \text{ in } N) \text{ in } P) : E} \\
\\
\frac{\Gamma \vdash M : A \otimes B \quad \Delta, x : A, y : B \vdash N : I \quad \Theta \vdash P : C}{\Gamma, \Delta, \Theta \vdash (\text{let } \langle x, y \rangle = M \text{ in let } \langle \rangle = N \text{ in } P) = (\text{let } \langle \rangle = (\text{let } \langle x, y \rangle = M \text{ in } N) \text{ in } P) : C} \\
\\
\frac{\Gamma \vdash M : A + B \quad \Delta, x : A \vdash N : C \otimes D \quad \Delta, y : B \vdash P : C \otimes D \quad \Theta, z : C, t : D \vdash Q : E}{\Gamma, \Delta, \Theta \vdash (\text{let } \langle z, t \rangle = (\text{case } M \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P) \text{ in } Q) \\ = (\text{case } M \text{ of } \text{inl}(x) \mapsto (\text{let } \langle z, t \rangle = N \text{ in } Q) \mid \text{inr}(y) \mapsto (\text{let } \langle z, t \rangle = P \text{ in } Q)) : E} \\
\\
\frac{\Gamma \vdash M : A + B \quad \Delta, x : A \vdash N : I \quad \Delta, y : B \vdash P : I \quad \Theta \vdash Q : C}{\Gamma, \Delta, \Theta \vdash (\text{let } \langle \rangle = (\text{case } M \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P) \text{ in } Q) \\ = (\text{case } M \text{ of } \text{inl}(x) \mapsto \text{let } \langle \rangle = N \text{ in } Q \mid \text{inr}(y) \mapsto \text{let } \langle \rangle = P \text{ in } Q) : C} \\
\\
\frac{\Gamma \vdash M : A + B \quad \Delta, x : A \vdash N : C + D \quad \Delta, y : B \vdash P : C + D \quad \Theta, z : C \vdash Q : E \quad \Theta, t : D \vdash R : E}{\Gamma, \Delta, \Theta \vdash \text{case } M \text{ of } \text{inl}(x) \mapsto (\text{case } N \text{ of } \text{inl}(z) \mapsto Q \mid \text{inr}(t) \mapsto R) \mid \\ \text{inr}(y) \mapsto (\text{case } P \text{ of } \text{inl}(z) \mapsto Q \mid \text{inr}(t) \mapsto R) \\ = \text{case } (\text{case } M \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P) \text{ of } \text{inl}(z) \mapsto Q \mid \text{inr}(t) \mapsto R : E}
\end{array}$$

Rules for Measurement

$$\begin{array}{c}
\frac{\Gamma \vdash 1 \leq \phi_1 \otimes \dots \otimes \phi_n \quad \Delta \vdash M_1 : A \dots \Delta \vdash M_n : A}{\Gamma, \Delta \vdash (\text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n) = (\text{measure } \phi_{p(1)} \mapsto M_{p(1)} \mid \dots \mid \phi_{p(n)} \mapsto M_{p(n)})} \quad p \text{ a permutation of } \{1, \dots, n\} \\
\\
\frac{\Gamma \vdash 1 \leq \phi_1 \otimes \dots \otimes \phi_n \quad \Delta \vdash M_1 : A \dots \Delta \vdash M_{n+1} : A}{\Gamma \vdash (\text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n \mid 0 \mapsto M_{n+1}) = \text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n : A} \\
\\
\frac{\Gamma \vdash M : A}{\Gamma \vdash (\text{measure } 1 \mapsto M) = M : A} \\
\\
\frac{\Gamma \vdash 1 \leq \phi \otimes \psi \otimes \chi_1 \otimes \dots \otimes \chi_n \quad \Delta \vdash M : A \quad \Delta \vdash P_1 : A \dots \Delta \vdash P_n : A}{\Gamma, \Delta \vdash (\text{measure } \phi \otimes \psi \mapsto M \mid \chi_1 \mapsto P_1 \mid \dots \mid \chi_n \mapsto P_n) \\ = (\text{measure } \phi \mapsto M \mid \psi \mapsto M \mid \chi_1 \mapsto P_1 \mid \dots \mid \chi_n \mapsto P_n)}
\end{array}$$

Proposition Formation

$$\frac{}{\Gamma \vdash 0 \text{ prop}} \quad \frac{\Gamma \vdash \phi \text{ prop}}{\Gamma \vdash \phi^\perp \text{ prop}} \quad \frac{\Gamma \vdash \phi \perp \psi}{\Gamma \vdash \phi \otimes \psi \text{ prop}} \quad \frac{\vdash \phi \text{ prop} \quad \Gamma \vdash \psi \text{ prop}}{\Gamma \vdash \phi \cdot \psi \text{ prop}}$$

Figure 3: Rules of Deduction of QPEL

Derivability

$$\begin{array}{c}
\frac{\Gamma \vdash \phi \text{ prop}}{\Gamma \vdash \phi \leq \phi} \quad \frac{\Gamma \vdash \phi \leq \psi \quad \Gamma \vdash \psi \leq \chi}{\Gamma \vdash \phi \leq \chi} \quad \frac{\Gamma \vdash \phi \text{ prop}}{\Gamma \vdash 0 \leq \phi} \\
\\
\frac{\Gamma \vdash \phi \leq \psi}{\Gamma \vdash \psi^\perp \leq \phi^\perp} \quad \frac{\Gamma \vdash \phi \text{ prop}}{\Gamma \vdash \phi \leq \phi^{\perp\perp}} \quad \frac{\Gamma \vdash \phi \text{ prop}}{\Gamma \vdash \phi^{\perp\perp} \leq \phi} \\
\\
\frac{\Gamma \vdash \phi \perp \psi}{\Gamma \vdash \phi \leq \phi \otimes \psi} \quad \frac{\Gamma \vdash \phi \leq \psi \quad \Gamma \vdash \psi \leq \chi^\perp}{\Gamma \vdash \phi \otimes \chi \leq \psi \otimes \chi} \\
\\
\frac{\Gamma \vdash \phi \perp \psi}{\Gamma \vdash \phi \otimes \psi \leq \psi \otimes \phi} \quad \frac{\Gamma \vdash \phi \otimes \psi \perp \chi}{\Gamma \vdash \psi \otimes \chi \perp \phi} \\
\\
\frac{\vdash \phi \perp \psi \quad \Gamma \vdash \chi \text{ prop}}{\Gamma \vdash \phi \cdot \chi \perp \psi \cdot \chi} \quad \frac{\vdash \phi \text{ prop} \quad \Gamma \vdash \psi \perp \chi}{\Gamma \vdash \phi \cdot \psi \perp \phi \cdot \chi} \\
\Gamma \vdash (\phi \otimes \psi) \cdot \chi \equiv \phi \cdot \chi \otimes \psi \cdot \chi \quad \Gamma \vdash \phi \cdot (\psi \otimes \chi) \equiv \phi \cdot \psi \otimes \phi \cdot \chi \\
\\
\frac{\Gamma \vdash \phi \text{ prop}}{\Gamma \vdash 1 \cdot \phi \equiv \phi} \quad \frac{\vdash \phi \text{ prop}}{\Gamma \vdash \phi \cdot 1 \equiv \phi} \quad \frac{\vdash \phi \text{ prop} \quad \vdash \psi \text{ prop} \quad \Gamma \vdash \chi \text{ prop}}{\Gamma \vdash \phi \cdot (\psi \cdot \chi) \equiv (\phi \cdot \psi) \cdot \chi} \\
\\
\frac{\vdash \phi \text{ prop} \quad \vdash \psi \text{ prop}}{\vdash \phi \cdot \psi \equiv \psi \cdot \phi}
\end{array}$$

Figure 4: Rules of Deduction of QPEL

8. If $\Gamma \vdash M : A + B$, $\Delta, x : A \vdash N : C$, $\Delta, y : B \vdash P : C$ and $\Theta, z : C \vdash Q : D$, then

$$\begin{aligned}
& \Gamma, \Delta, \Theta \vdash [\text{case } M \text{ of } \text{inl}(x) \mapsto N \mid \text{inr}(y) \mapsto P/z] Q \\
& = \text{case } M \text{ of } \text{inl}(x) \mapsto [N/z] Q \mid \text{inr}(y) \mapsto [P/z] Q : D
\end{aligned}$$

Proof. The proofs are mostly straightforward. We note that the proof of 7 and 8 involve the observation that the rules for tensor product and tensor unit allow us to define local definition. This is a result about linear type theory which, to the best of my knowledge, is new. We define let $x = M$ in N for let $\langle x, y \rangle = \langle M, \langle \rangle \rangle$ in let $\langle \rangle = y$ in N . \square

4 Semantics

4.1 State and Effect Triangles

Let E be a commutative effect monoid. Recall the adjunction $\mathbf{EMod}_E[-, E] \dashv \mathbf{Conv}_E[-, E] : \mathbf{EMod}_E^{\text{op}} \rightleftarrows \mathbf{Conv}_E$.

Definition 9 (State-and-Effect Triangle). A *state-and-effect triangle* consists of:

- a symmetric monoidal category \mathcal{V} with binary coproducts that distribute over the tensor, such that the tensor unit is terminal;
- an effect monoid E ;
- a functor $P : \mathcal{V} \rightarrow \mathbf{EMod}_E^{\text{op}}$ that preserves finite coproducts and the terminal object;

- a symmetric monoidal functor $S : \mathcal{V} \rightarrow \mathbf{Conv}_E$;
- given a finite set $r_1, \dots, r_n \in PA$ such that $r_1 \odot \dots \odot r_n = 1$, an arrow $\text{meas}_A(r_1, \dots, r_n) : A \rightarrow n \cdot I$ in \mathcal{V} ;
- a natural transformation $\alpha : P \rightarrow \mathbf{Conv}_E[S-, E]$;
- a natural transformation $\beta : S \rightarrow \mathbf{EMod}_E[P-, E]$;

such that

1. given a permutation p on $\{1, \dots, n\}$, we have

$$\text{meas}_A(r_{p(1)}, \dots, r_{p(n)}) = \pi_p \circ \text{meas}_A(p_1, \dots, p_n)$$

where $\pi_p : n \cdot I \rightarrow n \cdot I$ satisfies

$$\pi_p \circ \kappa_i = \kappa_{p(i)}$$

2. $\text{meas}_A(p_1, \dots, p_n, 0) = \kappa_1 \circ \text{meas}_A(p_1, \dots, p_n) : A \rightarrow n \cdot I \rightarrow (n+1) \cdot I$
3. $\text{meas}_A(p \odot q, r_1, \dots, r_n) = [\kappa_1, \kappa_1, \kappa_2, \dots, \kappa_{n+1}] \circ \text{meas}_A(p, q, r_1, \dots, r_n)$
4. meas_A is natural in A ; i.e. given $f : A \rightarrow B$,

$$\text{meas}_A(r_1, \dots, r_n) \circ f = \text{meas}_B(Pf(r_1), \dots, Pf(r_n))$$

5. $\rho \circ \mu = 1$, where $\mu_{AB} : SA \otimes SB \rightarrow S(A \otimes B)$ is given by the symmetric monoidal functor structure of S .
6. $\alpha_A(p)(x) = \beta_A(x)(p)$ for all A, x, p .

We think of the arrows in \mathcal{V} as *computations*, the arrows $SA \rightarrow SB$ as *state transformers*, and the arrows $PA \rightarrow PB$ as *predicate transformers*.

Examples The following are all examples of state-and-effect triangles:

- Take \mathcal{V} to be the category $\mathbf{FdHilb}_{\text{Un}}$ of finite-dimensional Hilbert spaces with unitary maps, PH to be the set of *effects* on H (positive operators less than I), and SH to be the set of *density matrices* on H . See [13].
- Take \mathcal{V} to be $\mathbf{Kl}(\mathcal{D}_E)$, the Kleisli category of the distribution monad \mathcal{D}_E . S is the canonical functor from the Kleisli category to the Eilenberg-Moore category. For $X \in \text{Set}$, PX is the set of all functions $X \rightarrow \mathcal{D}_E(2)$, equivalently the set of functions $X \rightarrow E$.
- Take \mathcal{V} to be $\mathbf{CStar}_{\text{PU}}^{\text{op}}$, the category of C^* -algebras and positive unital maps. PA is the set of all effects on a , $[0, 1]_A = \{a \in A : 0 \leq a \leq 1\}$. SA is the set of all positive unital maps $A \rightarrow \mathbb{C}$.

See [11] for more details on all of these.

Remarks

1. We do not want S always to be a strong monoidal functor. Intuitively, $S(A) \otimes S(B)$ gives the mixtures of pure states of $A \otimes B$, while $S(A \otimes B)$ also includes entangled states, and these will not be isomorphic in general.
2. The condition $\alpha_A(p)(x) = \beta_A(x)(p)$ can also be written as $\alpha = G\beta \circ \eta P$ or as $\beta = F\alpha \circ \varepsilon S$, where $F = \mathbf{EMod}_E[-, E] : \mathbf{EMod}_E^{\text{op}} \rightarrow \mathbf{Conv}_E$ and $G = \mathbf{Conv}_E[-, E] : \mathbf{Conv}_E \rightarrow \mathbf{EMod}_E^{\text{op}}$.

4.2 Semantics

Definition 10. Given any state-and-effect triangle, we interpret the syntax as follows.

- We associate with every type A an object $\llbracket A \rrbracket$ of \mathcal{V} .
- We associate with every context Γ an object $\llbracket \Gamma \rrbracket$ of \mathcal{V} .
- We associate with every term $\Gamma \vdash M : A$ an arrow $\llbracket M \rrbracket = \llbracket \Gamma \vdash M : A \rrbracket : \llbracket \Gamma \rrbracket \rightarrow \llbracket A \rrbracket$ in \mathcal{V} .
- We associate with every proposition ϕ such that $\Gamma \vdash \phi$ prop, an element $\llbracket \phi \rrbracket \in P[\llbracket \Gamma \rrbracket]$.
- We associate with every proposition ϕ such that $\vdash \phi$ prop, an element $\langle \phi \rangle \in E$.

If $\Gamma \vdash M : A$ and $\Delta \vdash N : B$, then

$$\llbracket \Gamma, \Delta \vdash \langle M, N \rangle : A \otimes B \rrbracket = \llbracket M \rrbracket \otimes \llbracket N \rrbracket .$$

If $\Gamma \vdash M : A \otimes B$ and $\Delta, x : A, y : B \vdash N : C$, then $\llbracket \Gamma, \Delta \vdash \text{let } \langle x, y \rangle = M \text{ in } N : C \rrbracket$ is the arrow

$$\llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{\llbracket M \rrbracket \otimes 1} \llbracket A \rrbracket \otimes \llbracket B \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{\llbracket N \rrbracket} \llbracket C \rrbracket$$

If $\Gamma \vdash \phi_i$ prop and $\Delta \vdash M_i : A$, then $\llbracket \Gamma, \Delta \vdash \text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n : A \rrbracket$ is

$$\llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{\text{meas}_A(\llbracket \phi_1 \rrbracket, \dots, \llbracket \phi_n \rrbracket) \otimes 1} (n \cdot I) \otimes \llbracket \Delta \rrbracket \xrightarrow{\cong} n \cdot \llbracket \Delta \rrbracket \xrightarrow{\llbracket M_1 \rrbracket, \dots, \llbracket M_n \rrbracket} \llbracket A \rrbracket$$

Lemma 11. 1. If $\Gamma, x : A \vdash M : B$ and $\Delta \vdash N : A$, then $\llbracket \Gamma, \Delta \vdash [N/x]M : B \rrbracket$ is the arrow

$$\llbracket \Gamma \rrbracket \otimes \llbracket \Delta \rrbracket \xrightarrow{1 \otimes \llbracket N \rrbracket} \llbracket \Gamma \rrbracket \otimes \llbracket A \rrbracket \xrightarrow{\llbracket M \rrbracket} \llbracket B \rrbracket$$

2. If $\Gamma, x : A \vdash \phi$ prop and $\Delta \vdash M : A$ then

$$\llbracket [M/x]\phi \rrbracket = P(1_{\llbracket \Gamma \rrbracket} \otimes \llbracket M \rrbracket)(\llbracket \phi \rrbracket)$$

3. If $x : A \vdash \phi$ prop and $\vdash M : A$ then

$$\langle [M/x]\phi \rangle = \beta_A(\llbracket M \rrbracket(*))(\langle \phi \rangle)$$

where $SI = \{*\}$.

Definition 12. In a state-and-effect triangle, a judgement $\Gamma \vdash M = N : A$ is *true* iff $\llbracket \Gamma \vdash M : A \rrbracket = \llbracket \Gamma \vdash N : A \rrbracket$. A judgement $\Gamma \vdash \phi \leq \psi$ is *true* iff $\llbracket \Gamma \vdash \phi \text{ prop} \rrbracket \leq \llbracket \Gamma \vdash \psi \text{ prop} \rrbracket$, in the order in the effect module $P[\llbracket \Gamma \rrbracket]$.

Theorem 13 (Soundness). *Any derivable judgement is true in any state-and-effect triangle.*

Theorem 14 (Completeness). *Any judgement that is true in every state-and-effect triangle is derivable.*

Proof. Define a state-and-effect triangle as follows.

The category \mathcal{V} is the category with objects the types, and arrows $A \rightarrow B$ the terms M such that $x : A \vdash M : B$, quotiented by: $M = N$ iff $x : A \vdash M = N : B$. For types A and B , the tensor product is $A \otimes B$ and the coproduct is $A + B$.

The effect monoid M is the set of all propositions ϕ such that $\vdash \phi$ prop, quotiented by: $\phi = \psi$ iff $\vdash \phi \leq \psi$ and $\vdash \psi \leq \phi$. We have that $\phi \perp \psi$ is defined iff $\vdash \phi \odot \psi$ prop (equivalently, iff $\vdash \phi \leq \psi$, in which case the partial sum is $\phi \odot \psi$). The zero element is 0, and the orthocomplement of ϕ is ϕ^\perp . The product of ϕ and ψ is $\phi \cdot \psi$.

The functor P is defined by: PA is the set of all propositions ϕ such that $x : A \vdash \phi$ prop, quotiented by: $\phi = \psi$ iff $x : A \vdash \phi \equiv \psi$.

The functor S is defined by: SA is the set of all terms M such that $\vdash M : A$, quotiented by: $M = N$ iff $\vdash M = N : A$. The convex structure is given by: if $\phi_1 \otimes \dots \otimes \phi_n = 1$, then $\phi_1 M_1 + \dots + \phi_n M_n = \text{measure } \phi_1 \mapsto M_1 \mid \dots \mid \phi_n \mapsto M_n$.

We have $\text{meas}_A(\phi_1, \dots, \phi_n) = \text{measure } \phi_1 \mapsto \text{in}_1(\langle \rangle) \mid \dots \mid \phi_n \mapsto \text{in}_n(\langle \rangle)$, where the terms $\text{in}_i(M)$ are the n canonical terms such that $x : A \vdash \text{in}_i(x) : \overbrace{A + \dots + A}^n$.

The transformation α is given by $\alpha_A(x : A \vdash \phi \text{ prop})(\vdash M : A) \equiv (\vdash [M/x]\phi \text{ prop})$, and so β is given by $\beta_A(\vdash M : A)(x : A \vdash \phi \text{ prop}) \equiv (\vdash [M/x]\phi \text{ prop})$.

If a judgement is true in every state-and-effect triangle, it is true in this triangle, and therefore is derivable. \square

5 Qubits

In order to represent qubits, we extend the system with a new type **qbit**; terms new 1 and $U[M]$ where U is any n -ary quantum gate (i.e. unitary operation on \mathbb{C}^{2^n}); and a new proposition $M = 1$. We write $M = 0$ for $(M = 1)^\perp$.

We extend the system with the following rules of deduction.

$$\frac{}{\vdash \text{new } 1 : \mathbf{qbit}} \quad \frac{\Gamma \vdash M : \mathbf{qbit}^{\otimes n}}{\Gamma \vdash U[M] : \mathbf{qbit}^{\otimes n}} \quad U \text{ is an } n\text{-ary quantum gate} \quad \frac{\Gamma \vdash M : \mathbf{qbit}}{\Gamma \vdash M = 1 \text{ prop}}$$

We add rules of deduction such that the following equations hold between terms, and equivalences between propositions:

$$\begin{array}{ll} \text{swap}[x, y] = \langle y, x \rangle & U^{-1}[U[x]] = x \\ (UV)[x] = U[V[x]] & I[x] = x \\ (U \otimes V)[x, y] = \langle U[x], V[y] \rangle & C(U, V)[\text{new } 1, x] = \langle \text{new } 1, U[x] \rangle \\ \text{phase}_\theta[\text{new } 1] = \text{new } 1 & \text{pauli} - X[x] = 1 \equiv x = 0 \\ \text{pauli} - Y[x] = 1 \equiv x = 0 & \text{phase}_\theta[x] = 1 \equiv x = 1 \\ \text{new } 1 = 1 \equiv 1 & \end{array}$$

These can be given semantics in $\mathbf{FdHilb}_{\text{Un}}$ straightforwardly. It remains to be seen what conditions on an object in \mathcal{V} are required to give an interpretation to this system, and what completeness theorem (if any) can be proved for this system.

6 Natural Isomorphisms

It is interesting to consider the question of when the natural transformations α and β are isomorphisms. In the $\mathbf{FdHilb}_{\text{Un}}$ example, α and β are both isomorphisms. [13]. In the $\mathbf{Kl}(\mathcal{D})$ example, α is an isomorphism but β is not. In the $\mathbf{CStar}_{\text{PU}}^{\text{op}}$ example, β is an isomorphism but α is not.

We can extend the system so it captures the state-and-effect triangles in which β is an isomorphism as follows.

Theorem 15 (Completeness). *Add to the system the rule*

$$\frac{\vdash \phi \text{ prop} \quad \Gamma \vdash M : A \quad \Gamma \vdash N : A \quad \Delta, x : A \vdash \psi \text{ prop}}{\Gamma, \Delta \vdash [(\text{measure } \phi \mapsto M | \phi^\perp \mapsto N) / x] \psi \equiv (\phi \cdot [M/x] \psi) \oplus (\phi^\perp \cdot [N/x] \psi)}$$

If a judgement is true in every state-and-effect triangle in which α and β are natural isomorphisms, then it is derivable in this system.

I do not yet have a system that captures the state-and-effect triangles in which α is a natural isomorphism.

The case where α is an isomorphism is particularly interesting, as it is this that allows *weakest preconditions* in d'Hondt-Panangaden's sense to be defined.

Definition 16. Let P and Q be quantum predicates, and F a quantum program. Then P is a *precondition* for Q with respect to M , PFQ , iff for all density matrices ρ , $\text{tr}(P\rho) \leq \text{tr}(QF(\rho))$. P is the *weakest precondition* for Q with respect to M , $P = \text{wp}(F)(Q)$ iff P is the greatest precondition for Q w.r.t. M under the Löwner order.

The weakest precondition for Q w.r.t. F always exists and is unique [8].

Lemma 17. In the $\mathbf{FdHilb}_{\text{Un}}$ state-and-effect triangle, the weakest precondition for $Q \in PH$ with respect to $F : SK \rightarrow SH$ is $\alpha^{-1}(F \circ \alpha(P))$. The operation $\text{wp}(F)$ is therefore the effect module homomorphism $\alpha^{-1} \circ \mathbf{Conv}_M[1, F] \circ \alpha : PH \rightarrow PK$. The operation wp is therefore the natural transformation

$$\text{wp}_{PHK} = \alpha^{-1} \circ \mathbf{Conv}_M[1, -] \circ \alpha : \mathbf{Conv}_M[SK, SH] \rightarrow \mathbf{EMod}_M[PH, PK]$$

Lemma 18. Given $\Gamma \vdash M : A$ and $x : A \vdash \phi \text{ prop}$, then in the $\mathbf{FdHilb}_{\text{Un}}$ semantics:

$$\text{wp}(\llbracket \Gamma \vdash M : A \rrbracket)(\llbracket x : A \vdash \phi \text{ prop} \rrbracket) = \llbracket \Gamma \vdash [M/x] \phi \text{ prop} \rrbracket$$

7 Conclusion, Related Work and Future Work

We have presented QPEL, a syntactic system involving both terms and propositions that captures the categorical notion of 'state-and-effect triangle' which has proved to be a general setting for describing both quantum programs, and effects. It is therefore a promising candidate for a language that allows us to reason about and prove properties of quantum programs, and shows how such a logic for quantum effects might be added on top of any quantum programming language.

Baltag and Smets in a series of papers [2, 5, 4, 3, 1] describe the language QDL, Quantum Dynamic Logic. This is also a language for describing quantum programs and properties of quantum programs. Their work differs from mine because their term language is an underspecification language (as is Dynamic Logic's), and their propositions can denote all propositions expressible in classical logic, not just those that correspond to quantum effects.

d'Hondt-Panangaden [8] and Ying [16] have investigated the notion of a *quantum predicate*. Ying has given a Floyd-Hoare style logic which, given a program F written in his syntax, allows the weakest precondition of a predicate with respect to F to be calculated. Their work differs from mine because they do not give a syntax for the predicates, instead using the effects on a Hilbert space as the predicates directly.

In the future, the most important tasks are to apply the system to prove the correctness of a simple quantum program (e.g. the quantum teleportation protocol or quantum broadcasting), and to look for ways to extend the system in order to represent looping and/or recursion.

I will also try to capture the conditions that make α or β a natural isomorphism. I will investigate the conditions that a state-and-effect triangle needs to satisfy to represent the type of qubits correctly, possibly involving Selinger's notion of a Quantum Flowchart Category. I will investigate formal translations

between this system and other quantum programming languages, such as the quantum lambda calculus [15]. I will investigate which of Ying's equations on weakest preconditions [16] can be derived within our system.

Acknowledgements Thanks to Sam Staton for an early version of the rules presented here, in particular for the equations on qubits. Thanks to Sam Staton and Bart Jacobs for many helpful discussions.

References

- [1] A. Baltag & S. Smets (2004): *The Logic of Quantum Programs*. In: *QPL 2004*, pp. 39–56.
- [2] A. Baltag & S. Smets (2005): *Complete Axiomatizations for Quantum Actions*. *International Journal of Theoretical Physics* 44.
- [3] A. Baltag & S. Smets (2005): *LQP: The Dynamic Logic of Quantum Information*. *Mathematical Structures in Computer Science*.
- [4] A. Baltag & S. Smets (2011): *Quantum Logic as a Dynamic Logic*. *Synthese* 179, pp. 285–306, doi:10.1007/s11229-010-9783-6.
- [5] A. Baltag & S. Smets (2012): *The Dynamic Turn in Quantum Logic*. *Synthese* 186, pp. 753–773, doi:10.1007/s11229-011-9915-7.
- [6] N. Benton, G. Bierman, Valeria De Paiva & Martin Hyland (1993): *A Term Calculus for Intuitionistic Linear Logic*. In: *TLCA, Lecture Notes in Computer Science* 664, Springer-Verlag, pp. 75–90.
- [7] P. N. Benton (1995): *A Mixed Linear and Non-Linear Logic: Proofs, Terms and Models (Extended Abstract)*. In: *Selected Papers from the 8th International Workshop on Computer Science Logic, CSL '94*, Springer-Verlag, London, UK, UK, pp. 121–135. Available at <http://dl.acm.org/citation.cfm?id=647844.736565>.
- [8] E. d'Hondt & P. Panangaden (2006): *Quantum Weakest Preconditions*. *Math Struct in Comp Science* 16, pp. 429–451.
- [9] B. Jacobs (2010): *Convexity, duality, and effects*. In C. S. Clade & V. Sassone, editors: *IFIP Theoretical Computer Science 2010, IFIP Adv. in Inf. and Comm. Techn.* 82, Springer, Boston, pp. 1–19.
- [10] B. Jacobs (2011): *Probabilities, Distribution Monads, and Convex Categories*. *Theor. Comput. Sci.* 412(28), pp. 3323–3336.
- [11] B. Jacobs (2013): *On Block Structures in Quantum Computation*. In D. Kozen & M. Mislove, editors: *MFPS 2013, ENTCS* 298, pp. 233–255.
- [12] B. Jacobs (2014): *New Directions in Categorical Logic, for Classical, Probabilistic and Quantum Logic*.
- [13] B. Jacobs & J. Mandemaker (2013): *Relating Operator Spaces via Adjunctions*. In J. Chubb Reimann, V. Harizanov & A. Eskandarian, editors: *Logic and Algebraic Structures in Quantum Computing and Information*, Lect. Notes in Logic, Camb.
- [14] P. Selinger (2004): *Towards a Quantum Programming Language*. *Math Struct in Comp Science* 14(4), pp. 527–586.
- [15] P. Selinger & B. Valiron (2010): *Quantum Lambda Calculus*. In S. Gay & I Mackie, editors: *Semantical Techniques in Quantum Computation*, Cambridge University Press, pp. 135–172.
- [16] M. Ying (2011): *Floyd-hoare logic for quantum programs*. *ACM Trans. Program. Lang. Syst.* 33(6), p. 19. Available at <http://doi.acm.org/10.1145/2049706.2049708>.

Semantics for a Quantum Programming Language by Operator Algebras

Kenta Cho

Institute for Computing and Information Sciences (iCIS)
Radboud University Nijmegen, The Netherlands
K.Cho@cs.ru.nl, <http://www.cs.ru.nl/K.Cho/>

This paper presents a novel semantics for a quantum programming language by *operator algebras*, which are known to give a formulation for quantum theory that is alternative to the one by Hilbert spaces. We show that the opposite category of the category of W^* -algebras and normal completely positive pre-unital maps is an elementary quantum flow chart category in the sense of Selinger. As a consequence, it gives a denotational semantics for Selinger’s first-order functional quantum programming language QPL. The use of operator algebras allows us to accommodate infinite structures and to handle classical and quantum computations in a unified way.

1 Introduction

Aiming at high-level and structured description of quantum computation/information, many *quantum programming languages* have been proposed and their semantics studied [11, 42]. As one of pioneering works, in 2004 Selinger [37] proposed a first-order functional quantum programming language QPL (or QFC), and gave its denotational semantics rigorously in terms of categories. He (jointly with Valiron) successively started to study a higher-order functional quantum programming language, or a *quantum lambda calculus* [38–40]. It turned out to be challenging to give a denotational semantics for a quantum lambda calculus (with the full features, such as the ! modality and recursion). The first denotational semantics was given via Geometry of Interaction [13]. Recently, another denotational semantics was obtained from a different approach [31]. As is stated in [31, §1], the problem lies in that quantum computation is typically modelled by using finite dimensional Hilbert spaces, and hence it is difficult to model *infinite* structures in computation.

The present paper presents a novel denotational semantics for a quantum programming language by *operator algebras*. Operator algebras, specifically C^* -algebras and W^* -algebras (the latter are also known as von Neumann algebras), give an alternative formulation for quantum theory (sometimes called the *algebraic* formulation [21]). It is worth mentioning that von Neumann himself, who formulated quantum theory by Hilbert spaces [29], developed the theory of operator algebras [24–28] (some of them jointly with Murray), and later preferred the algebraic approach for quantum theory [33]. Operator algebras have been successfully used in areas such as quantum statistical mechanics [3] and quantum field theory [2, 12]. They have also been of growing importance in the area of quantum information [18]; for example, in [8], the impossibility of quantum bit commitment is (re)examined in the algebraic formalism. For the use of operator algebras in quantum computation, see ‘Related work’ below.

Contributions In this paper it is shown that the category $\mathbf{Wstar}_{\text{CP-PU}}$ of W^* -algebras and normal completely positive pre-unital maps is a \mathbf{Dcppo}_\perp -enriched symmetric monoidal category with \mathbf{Dcppo}_\perp -enriched finite products. It follows that its opposite $(\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$ is an $\omega\mathbf{Cppo}$ -enriched *elementary*

quantum flow chart category. As a consequence, it gives rise to a denotational semantics for a first-order functional quantum programming language QPL designed by Selinger [37].

Selinger himself gave a denotational semantics for QPL by the category \mathbf{Q} [37]. In comparison to his original semantics, our semantics by operator algebras has the following two advantages. First, our semantics accommodates infinite structures such as a type **nat** of natural numbers. This is because we discuss general W^* -algebras and do not restrict them to finite dimensional ones. In §8 we will see that our model can be considered as an infinite dimensional extension of Selinger’s model. Second, it is known (see e.g. [20, §5.8], [10]) that there are dualities between *commutative* C^* -algebras (resp. W^* -algebras) and certain type of topological (resp. measurable) spaces. Via the dualities, classical computations can be interpreted as maps between commutative algebras, though, in this paper (in §9), we restrict it to deterministic computation (in the category **Set**). It enables us to handle classical and quantum computations in a unified way.

One may also find an advantage in the simplicity of the semantics: a type is interpreted just as a W^* -algebra, and a program as a map of them (in opposite direction). There is no complicated construction such as used in [13] or [31], although the theory of operator algebras itself could be complicated enough.

Related work It does not seem that the use of operator algebras is very common in the context of quantum computation. Recently, there are a few such works [10, 17] involving C^* -algebras, which led the author to the present work. The use of W^* -algebras in this context appeared independently and coincidentally in Rennela’s thesis [34] and the present work (i.e. the author’s thesis [6]). Rennela also observed that the category $\mathbf{Wstar}_{\text{P-PU}}$ of W^* -algebras and normal positive pre-unital maps are **Dcppo**-enriched. Fortunately the author’s and Rennela’s works seem to be complementary to each other. Rennela’s work is mainly on the effect logic (see e.g. [15, 16]) in W^* -algebras, which was the author’s future work.

Some related technical results have appeared in [5], which studies spaces of maps (‘quantum operations’) between W^* -algebras, and maps (‘quantum supermaps’) between them. For instance, [5, Proposition 7] states (in our terms) that $\mathbf{Wstar}_{\text{CP}}(M, N)$ is bounded directed complete.

Organisation of the paper First we review complete partial orders in §2, and then review Selinger’s work on QPL in §3. In §4 we also review the basics of operator algebras, and then look at the order-theoretic perspective of operator algebras. We show in §5 that $\mathbf{Wstar}_{\text{CP-PU}}$ is a **Dcppo**_⊥-enriched symmetric monoidal category with **Dcppo**_⊥-enriched finite products and in §6 that its opposite ($\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$ is an $\omega\mathbf{Cppo}$ -enriched elementary quantum flow chart category, hence gives a denotational semantics for QPL. Section 7 discuss a duality between Selinger’s and our semantics, and in §8 it is shown that Selinger’s model is contravariantly embedded into our model. In §9 we see that our model can accommodate infinite types, and also classical computation as commutative structures. Section 10 concludes the paper with future work.

This paper is based on the author’s master thesis [6], in which one may find more information and details omitted from this paper.

2 Complete partial orders

In this section, we will briefly review the notion of complete partial orders, which is essential in domain theory [1], and is fundamental for the denotational semantics of programs with features such as recursion and loop.

Definition 2.1. A poset is

1. *directed complete* if every directed subset has a supremum;

2. *bounded directed complete* if every directed subset that is bounded from above has a supremum;
3. *ω -complete* if every ω -chain $((x_n)_{n \in \omega}$ with $x_n \leq x_{n+1}$) has a supremum;
4. *pointed* if it has a least element (denoted by \perp).

A (bounded) directed complete poset is abbreviated as a $(b)dcpo$, and an ω -complete poset as a ωcpo .

Definition 2.2. A map between posets (in 3, pointed posets) is

1. *Scott-continuous* if it preserves suprema of directed subsets;
2. *ω -(Scott-)continuous* if it preserves suprema of ω -chains;
3. *strict* if it preserves the least element.

Note that every $dcpo$ is an ωcpo , and every Scott-continuous map is ω -continuous. The next theorem is a fundamental tool to give an interpretation of a program with recursion and loop.

Theorem 2.3. *Every ω -continuous endomap f on a pointed ωcpo has a least fixed point, which is given by $\bigvee_n f^n(\perp)$.* ■

We here fix the notations of categories we use in this paper.

Definition 2.4. We denote by \mathbf{Dcppo}_\perp the category of pointed $dcpos$ and strict Scott-continuous maps, and by $\omega\mathbf{Cppo}$ the category of pointed $\omega cpos$ and ω -continuous maps.

Both categories \mathbf{Dcppo}_\perp and $\omega\mathbf{Cppo}$ have products, which are given by cartesian products of underlying sets with coordinatewise order. When we speak of \mathbf{Dcppo}_\perp - (or $\omega\mathbf{Cppo}$)-enrichment of categories, these cartesian structures are taken as monoidal structures.

3 Selinger's QPL and its semantics

In [37], Selinger proposed a quantum programming language QPL and its denotational semantics. It is a first-order functional language with loop and recursion, and is described by flow chart syntax, besides by textual syntax.¹ He showed a denotational semantics for QPL is given by the following category.

Definition 3.1 ([37, §6.6]). An *elementary quantum flow chart category* is a symmetric monoidal category (\mathbf{C}, \otimes, I) with traced finite coproducts $(\oplus, 0, \text{Tr})$ such that:

- For each $A \in \mathbf{C}$, $A \otimes (-)$ is a traced monoidal functor.
- \mathbf{C} has a distinguished object **qbit** with arrows $\iota: I \oplus I \rightarrow \mathbf{qbit}$ and $p: \mathbf{qbit} \rightarrow I \oplus I$ such that $p \circ \iota = \text{id}$.

Theorem 3.2 ([37, §6.6]). *Let \mathbf{C} be an elementary quantum flow chart category. Suppose we have an assignment η of built-in unitary operator symbols S of arity n to arrows $\eta_S: \mathbf{qbit}^{\otimes n} \rightarrow \mathbf{qbit}^{\otimes n}$ in \mathbf{C} . Then we have an interpretation of QPL programs without recursion. If \mathbf{C} is additionally $\omega\mathbf{Cppo}$ -enriched, then we can also interpret QPL programs with recursion.* ■

Selinger also gave a concrete model for QPL, constructing an $\omega\mathbf{Cppo}$ -enriched elementary quantum flow chart category \mathbf{Q} as follows.

Definition 3.3. We write \mathcal{M}_n for the set of complex $n \times n$ matrices.

¹Strictly speaking, the name QPL is reserved for the textual syntax. In the present paper we do not distinguish the two syntaxes because semantics are given in the same manner.

1. The category \mathbf{CPM}_s is defined as follows.
 - An object is a natural number.
 - An arrow $f: n \rightarrow m$ is a completely positive map $f: \mathcal{M}_n \rightarrow \mathcal{M}_m$.
2. The category \mathbf{CPM} is the finite biproduct completion of \mathbf{CPM}_s . Specifically:
 - An object is a sequence $\vec{n} = (n_1, \dots, n_k)$ of natural numbers.
 - An arrow $f: \vec{n} \rightarrow \vec{m}$ is a matrix (f_{ij}) of arrows $f_{ij}: n_j \rightarrow m_i$ in \mathbf{CPM}_s .
3. The category \mathbf{Q} is a subcategory of \mathbf{CPM} such that
 - Objects are the same as \mathbf{CPM} .
 - An arrow is $f: \vec{n} \rightarrow \vec{m}$ in \mathbf{CPM} which is trace-nonincreasing, i.e.

$$\sum_i \sum_j \text{tr}(f_{ij}(A_j)) \leq \sum_j \text{tr}(A_j)$$

for all $(A_j)_j$ with positive $A_j \in \mathcal{M}_{n_j}$.

The category \mathbf{Q} turns out to have the monoidal structure (\otimes, I) and finite coproducts $(\oplus, 0)$, and furthermore to be $\omega\mathbf{Cppo}$ -enriched. The category is also equipped with a monoidal trace wrt. $(\oplus, 0)$, which is obtained from its $\omega\mathbf{Cppo}$ -enriched structure. With an object $\mathbf{qbit} := 2$, it forms $\omega\mathbf{Cppo}$ -enriched elementary quantum flow chart category. As Selinger mentioned [37, §6.4], the construction of monoidal trace from $\omega\mathbf{Cppo}$ -enriched structure is valid for every $\omega\mathbf{Cppo}$ -enriched category with finite coproducts. That is:

Theorem 3.4.

1. Every $\omega\mathbf{Cppo}$ -enriched cocartesian category with right-strict composition (i.e. $f \circ \perp = \perp$) is traced.
2. Let \mathbf{C} and \mathbf{D} be $\omega\mathbf{Cppo}$ -enriched cocartesian categories with right-strict composition, which are traced by 1. Every $\omega\mathbf{Cppo}$ -enriched cocartesian functor between \mathbf{C} and \mathbf{D} satisfying $F\perp = \perp$ is traced. ■

Here, a cocartesian category refers to a monoidal category whose monoidal products are finite coproducts. Selinger just stated this result informally. More detailed discussion (in the dual form) is found in [6, Appendix]. We conclude the section by the following theorem, which is obtained by combining Definition 3.1 and Theorem 3.4.

Theorem 3.5. *An $\omega\mathbf{Cppo}$ -enriched symmetric monoidal category (\mathbf{C}, \otimes, I) is an ($\omega\mathbf{Cppo}$ -enriched) elementary quantum flow chart category if the following conditions hold.*

- \mathbf{C} has $\omega\mathbf{Cppo}$ -enriched finite coproducts $(\oplus, 0)$;
- the composition is right-strict;
- for each $A \in \mathbf{C}$, a functor $A \otimes (-)$ preserves finite coproducts and bottom arrows;
- \mathbf{C} has a distinguished object \mathbf{qbit} with arrows $\iota: I \oplus I \rightarrow \mathbf{qbit}$ and $p: \mathbf{qbit} \rightarrow I \oplus I$ such that $p \circ \iota = \text{id}$. ■

4 Operator algebras: C^* -algebras and W^* -algebras

4.1 C^* -algebras and W^* -algebras

The theory of operator algebras is usually concerned with C^* -algebras and W^* -algebras (the latter are often studied as von Neumann algebras). Due to limitations of space, here we just fix notations and terminology, and list the basic results. A reader who is not familiar with these topics may consult [35, 41] for the standard theory and [19, 23] for the categorical perspective. The author's thesis [6, Chapters 3–4] would help, too.

In this paper, C^* -algebras are always assumed to be unital. Note also that a ‘map’ usually refers to a *linear* map. For a C^* -algebra A , we write A_{sa} for the set of self-adjoint elements, and $[0, 1]_A := \{x \in A \mid 0 \leq x \leq 1\}$ for the “unit interval”, or the set of *effects*. We write **Cstar** for the category of C^* -algebras and bounded maps. We will denote subcategories of **Cstar**, with the same objects but different maps, by adding subscripts (separated by hyphens) to **Cstar** as follows: M for ‘multiplicative’; I for ‘involutive’; P for ‘positive’; CP for ‘completely positive’; U for ‘unital’; PU for ‘pre-unital’². For example, **Cstar**_{CP-PU} is the category of C^* -algebras and completely positive pre-unital maps³.

A W^* -algebra is a C^* -algebra that has a (necessarily unique) predual. Every W^* -algebra is equipped with the weak* topology introduced by its predual, which is called the *ultraweak* topology. A map between W^* -algebras is said to be *normal* if it is ultraweakly continuous. We write **Wstar** for the category of W^* -algebras and normal maps. We denote subcategories of **Wstar** in the same manner as **Cstar**. For example, **Wstar**_{CP-PU} is the category of W^* -algebras and normal completely positive pre-unital maps. Note that **Wstar**_{CP-PU} is a *non-full* subcategory of **Cstar**_{CP-PU}, since maps in **Wstar**_{CP-PU} are required to be normal.

Direct sums [35, Definition 1.1.5], denoted by \oplus , are defined in the same way for C^* -algebras and W^* -algebras. They form categorical products in categories **Cstar**_{M-I-U}, **Cstar**_{CP}, **Cstar**_{CP-PU}, **Wstar**_{M-I-U}, **Wstar**_{CP}, **Wstar**_{CP-PU} etc. The nullary direct sum is the zero space 0.

There are several notions of tensor products. In this paper we use *spatial C^* -tensor products* [35, Definition 1.22.8], denoted by \otimes , for C^* -algebras, and *spatial W^* -tensor products* [35, Definition 1.22.10], denoted by $\overline{\otimes}$, for W^* -algebras. Spatial C^* -tensor products make categories **Cstar**_{M-I-U}, **Cstar**_{CP}, **Cstar**_{CP-PU} etc. symmetric monoidal categories with the unit object \mathbb{C} , but not **Cstar**_P. A tensor product of positive maps can be unbounded [4, Proposition 3.5.2]. This is why we need the notion of the *complete* positivity. Similarly, spatial W^* -tensor products make categories **Wstar**_{M-I-U}, **Wstar**_{CP}, **Wstar**_{CP-PU} etc. symmetric monoidal categories.

Spatial C^* - and W^* -tensor products distribute over finite direct sums, i.e.

$$A \otimes (B \oplus C) \cong (A \otimes B) \oplus (A \otimes C) \quad , \quad A \otimes 0 \cong 0 \quad , \quad M \overline{\otimes} (N \oplus L) \cong (M \overline{\otimes} N) \oplus (M \overline{\otimes} L) \quad , \quad M \overline{\otimes} 0 \cong 0 \quad ,$$

for C^* -algebras A, B, C and W^* -algebras M, N, L . These properties seem to be known results, but to be missing in the standard literature. One can find the proofs in the author's thesis [6, §3.7, §4.6]. For our purpose, it is useful to restate them as follows.

Proposition 4.1.

1. For each C^* -algebra A , a functor $A \otimes (-)$ on **Cstar**_{CP-PU} preserves finite products.
2. For each W^* -algebra M , a functor $M \overline{\otimes} (-)$ on **Wstar**_{CP-PU} preserves finite products. ■

²A map f between C^* -algebras is said to be *pre-unital* if $f(1) \leq 1$. It seems that ‘sub-unital’, used in e.g. [34], is a more accepted term, but in this paper ‘pre-unital’ is adopted for the consistency with the author's thesis [6].

³Positive maps between C^* -algebras are automatically bounded. See e.g. [32, Proposition 2.1]

4.2 Order theory in operator algebras

Recall each C^* -algebra is equipped with a partial order \leq defined by: $a \leq b \iff 'b - a \text{ is positive}'$. Many concepts on operator algebras can be rephrased in terms of the orders. Observe, for instance, the following easy proposition.

Proposition 4.2. *A map between C^* -algebras is positive if and only if it is monotone.* ■

In fact, the orders on W^* -algebras have a nice property, called *monotone closedness*, which distinguishes W^* -algebras from C^* -algebras.

Definition 4.3. A C^* -algebra A is *monotone closed* if every norm-bounded directed subset of A_{sa} has the supremum in A_{sa} .

Proposition 4.4 ([35, Lemma 1.7.4]). *Every W^* -algebra is monotone closed.* ■

We can also rephrase the notion of normality of positive (i.e. monotone) maps between W^* -algebras.

Proposition 4.5 ([7, Corollary 46.5]). *Let $f: M \rightarrow N$ be a positive map between W^* -algebra. Then f is normal (i.e. ultraweakly continuous) if and only if it preserves the supremum of every norm-bounded directed subset of M_{sa} .* ■

Therefore, we shall say a positive map $f: A \rightarrow B$ between monotone closed C^* -algebras is *normal* if it preserves the supremum of every norm-bounded directed subset of A_{sa} . Then W^* -algebras can be characterised as follows.

Theorem 4.6 ([41, Theorem III.3.16]). *A C^* -algebra is a W^* -algebra if and only if it is monotone closed and admits sufficiently many normal positive functionals (i.e. they separate the points).* ■

Now, we shall recapture the order structures in operator algebras from a more order-theoretic (or *domain-theoretic* [1]) point of view.

Proposition 4.7. *Let A be a C^* -algebra. The following are equivalent.*

1. A is monotone closed.
2. A_{sa} is bounded directed complete.
3. $[0, 1]_A$ is directed complete.

Proof. Without loss of generality, we may assume directed subsets are bounded from below. Then, $1 \iff 2$ follows from the fact that norm-boundedness and order-boundedness coincide [6, Lemma 4.8.3].

$2 \implies 3$ is trivial. For the converse, note that A_{sa} is an ordered vector space over \mathbb{R} . Hence we can obtain the supremum of a bounded directed subset of A_{sa} by transforming it into a directed subset of $[0, 1]_A$ by shifting and scaling. ■

Consequently, for every W^* -algebras M , M_{sa} is a bdcpo, and $[0, 1]_M$ is a pointed dcpo. For normality of maps, we also have:

Proposition 4.8. *Let $f: M \rightarrow N$ be a positive map between W^* -algebra. The first two of the following are equivalent. They are also equivalent to the third when f is pre-unital.*

1. f is normal.
2. The restriction $f|_{M_{\text{sa}}}: M_{\text{sa}} \rightarrow N_{\text{sa}}$ is Scott-continuous.
3. The restriction $f|_{[0, 1]_M}: [0, 1]_M \rightarrow [0, 1]_N$ is Scott-continuous. ■

5 \mathbf{Dcppo}_\perp -enrichment of the category of W^* -algebras

In this section we will show that the category $\mathbf{Wstar}_{\text{CP-PU}}$ is \mathbf{Dcppo}_\perp -enriched. We also see that the monoidal product $(\overline{\otimes}, \mathbb{C})$ and finite products $(\oplus, 0)$ on $\mathbf{Wstar}_{\text{CP-PU}}$ are \mathbf{Dcppo}_\perp -enriched.

Definition 5.1. Let M, N be W^* -algebras. We define a partial order \sqsubseteq on $\mathbf{Wstar}_{\text{CP-PU}}(M, N)$ by

$$f \sqsubseteq g \stackrel{\text{def}}{\iff} g - f \text{ is completely positive} .$$

Proposition 5.2. For any W^* -algebras M and N , $\mathbf{Wstar}_{\text{CP-PU}}(M, N)$ with the order \sqsubseteq is a pointed dcpo.

Proof. Note first that a positive pre-unital map $f: A \rightarrow B$ between C^* -algebras restricts to $[0, 1]_A \rightarrow [0, 1]_B$, while a “linear”⁴ map $g: [0, 1]_A \rightarrow [0, 1]_B$ extends to a positive pre-unital map $A \rightarrow B$. Hence such maps correspond bijectively (cf. [10, Lemma 2]).

Let (f_i) be a monotone net in $\mathbf{Wstar}_{\text{CP-PU}}(M, N)$. For each $x \in [0, 1]_M$, $(f_i(x))$ is a monotone net in $[0, 1]_N$, which is a dcpo. Hence define $f(x) := \sup f_i(x)$. It is easy to see f is a “linear” map $[0, 1]_M \rightarrow [0, 1]_N$, so that we obtain a positive pre-unital map $f: M \rightarrow N$.

Normality of f is proved as follows. For a monotone net (x_j) in $[0, 1]_M$,

$$f\left(\sup x_j\right) = \sup_i f_i\left(\sup_j x_j\right) = \sup_i \left(\sup_j f_i(x_j)\right) = \sup_j \left(\sup_i f_i(x_j)\right) = \sup_j f(x_j) .$$

Note that we can exchange of the order of sup ([1, Proposition 2.1.12]). Therefore $f: [0, 1]_M \rightarrow [0, 1]_N$ is Scott-continuous, and $f: M \rightarrow N$ is normal by Proposition 4.8.

We can also show that f is completely positive ([6, Theorem 4.9.7]), and $f_i \sqsubseteq f$ for each i ([6, Theorem 4.9.8]). Then it follows that f is the supremum of (f_i) . It is easy to see that the zero map is the bottom of $\mathbf{Wstar}_{\text{CP-PU}}(M, N)$. ■

Theorem 5.3. The category $\mathbf{Wstar}_{\text{CP-PU}}$ is \mathbf{Dcppo}_\perp -enriched. Moreover, the composition is bi-strict, i.e. $\perp \circ f = \perp$ and $f \circ \perp = \perp$ for each arrow f .

Proof. It is straightforward to check that the composition is Scott-continuous. Bi-strictness is immediate since \perp is the zero map. ■

Theorem 5.4. Finite products in $\mathbf{Wstar}_{\text{CP-PU}}$ are \mathbf{Dcppo}_\perp -enriched.

Proof. Essentially we have only to show the tupling operation $\langle \cdot, \cdot \rangle$ is Scott-continuous and strict, which is not very difficult. ■

Proving the next theorem needs more work. The interested reader is referred to the author’s thesis [6, Theorem 4.9.17].

Theorem 5.5. The symmetric monoidal product $(\overline{\otimes}, \mathbb{C})$ on $\mathbf{Wstar}_{\text{CP-PU}}$ is \mathbf{Dcppo}_\perp -enriched. ■

Remark 5.6. It is worth noting that the category $\mathbf{Cstar}_{\text{CP-PU}}$ is never \mathbf{Dcppo}_\perp -enriched, nor $\omega\mathbf{Cppo}$ -enriched. This is because we have an order-isomorphism $\mathbf{Cstar}_{\text{CP-PU}}(\mathbb{C}, A) \cong [0, 1]_A$, whereas there exists a C^* -algebra such that $[0, 1]_A$ is not ω -complete (take $A = C([0, 1])$ for example).

⁴ $g(0) = 0$, $g(x+y) = g(x) + g(y)$ for $x+y \leq 1$, $g(rx) = rg(x)$ for $r \in [0, 1]$.

6 Semantics for QPL by W^* -algebras

We have proved that $\mathbf{Wstar}_{\text{CP-PU}}$ is an \mathbf{Dcppo}_\perp -enriched symmetric monoidal category with \mathbf{Dcppo}_\perp -enriched finite products. Now we can show the following theorem.

Theorem 6.1. *The opposite category of $\mathbf{Wstar}_{\text{CP-PU}}$ is an $\omega\mathbf{Cppo}$ -enriched elementary quantum flow chart category.*

Proof. We apply Theorem 3.5. All requirements are shown in Theorems 5.3, 5.4, 5.5 (\mathbf{Dcppo}_\perp -enrichment implies $\omega\mathbf{Cppo}$ -enrichment), and Proposition 4.1.2, except $M \overline{\otimes} \perp = \perp$ and a distinguished object **qbit** with arrows ι, p .

It is easy to see $M \overline{\otimes} \perp = \perp$. We can take **qbit** := \mathcal{M}_2 , the algebra of complex 2×2 -matrices. We define two maps ι, p by

$$\iota \left(\begin{bmatrix} x & y \\ z & w \end{bmatrix} \right) = (x, w) \ , \quad p(x, y) = \begin{bmatrix} x & 0 \\ 0 & y \end{bmatrix} \ .$$

It is straightforward to see the two maps are positive, hence completely positive by [41, Corollary IV.3.5 and Proposition IV.3.9] (notice that $\mathbb{C} \oplus \mathbb{C}$ is commutative). They are normal because they are maps between finite dimensional W^* -algebras. Moreover they are clearly unital. Therefore ι and p are arrows in $\mathbf{Wstar}_{\text{CP-PU}}$. It is clear that $\iota \circ p = \text{id}$, hence $p \circ \iota = \text{id}$ in $(\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$. ■

Every unitary operator $S: (\mathbb{C}^2)^{\otimes n} \rightarrow (\mathbb{C}^2)^{\otimes n}$ of arity n determines an arrow $\eta_S: (\mathcal{M}_2)^{\overline{\otimes} n} \rightarrow (\mathcal{M}_2)^{\overline{\otimes} n}$ in $\mathbf{Wstar}_{\text{CP-PU}}$ by

$$(\mathcal{M}_2)^{\overline{\otimes} n} \cong \mathcal{M}_{2^n} \longrightarrow \mathcal{M}_{2^n} \cong (\mathcal{M}_2)^{\overline{\otimes} n} \ , \quad x \longmapsto S^\dagger x S \ ,$$

where S is seen as a complex $2^n \times 2^n$ matrix. By Theorem 3.2, finally, we obtain a semantics for QPL by W^* -algebras.

Theorem 6.2. *The opposite category of $\mathbf{Wstar}_{\text{CP-PU}}$ gives a denotational semantics for QPL (with recursion).* ■

7 Schrödinger vs. Heisenberg picture

We have shown the category $(\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$ gives a semantics for QPL. From now on we will discuss a comparison between two semantics by Selinger's **Q** and our $(\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$.

Recall that for a Hilbert space \mathcal{H} , $\mathcal{B}(\mathcal{H})$, i.e. the set of bounded operators on \mathcal{H} , is a W^* -algebra with the predual $\mathcal{T}(\mathcal{H})$, i.e. the set of trace class operators on \mathcal{H} . For every normal map $\mathcal{E}: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{K})$, there exists a corresponding bounded map $\mathcal{E}_*: \mathcal{T}(\mathcal{K}) \rightarrow \mathcal{T}(\mathcal{H})$ between preduals. They are related in the following way:

$$\text{tr}(\mathcal{E}(S) \cdot T) = \text{tr}(S \cdot \mathcal{E}_*(T)) \tag{1}$$

for all $S \in \mathcal{B}(\mathcal{H})$ and $T \in \mathcal{T}(\mathcal{K})$. Furthermore the following hold.

Proposition 7.1 ([14, §4.1.2]). *Let \mathcal{H} and \mathcal{K} be Hilbert spaces. Suppose a normal map $\mathcal{E}: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{K})$ and a bounded map $\mathcal{E}_*: \mathcal{T}(\mathcal{K}) \rightarrow \mathcal{T}(\mathcal{H})$ related as above. Then*

1. \mathcal{E} is completely positive if and only if \mathcal{E}_* is completely positive.
2. \mathcal{E} is unital if and only if \mathcal{E}_* is trace-preserving.
3. \mathcal{E} is pre-unital if and only if \mathcal{E}_* is trace-nonincreasing. ■

Hence, a normal completely positive pre-unital map $\mathcal{E}: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{K})$, which is an arrow in $\mathbf{Wstar}_{\text{CP-PU}}$, corresponds to a completely positive trace-nonincreasing map $\mathcal{E}_*: \mathcal{T}(\mathcal{K}) \rightarrow \mathcal{T}(\mathcal{H})$, which is known as a *quantum operation* (see e.g. [30, §8.2], [14, Chap. 4]). This is the well-known duality between the Heisenberg and Schrödinger pictures: one transforms observables (i.e. self-adjoint operators), while another transforms states (i.e. density operators).

Hence, it is understood that our semantics for QPL by $\mathbf{Wstar}_{\text{CP-PU}}$ is given in the Heisenberg picture, while Selinger's semantics by \mathbf{Q} is given in the Schrödinger picture. In the words of [9], our semantics can also be thought of as the *weakest precondition* semantics. This is because a positive pre-unital map $\mathcal{E}: \mathcal{B}(\mathcal{H}) \rightarrow \mathcal{B}(\mathcal{K})$ can be restricted to a map $\mathcal{E}: \mathcal{Ef}(\mathcal{H}) \rightarrow \mathcal{Ef}(\mathcal{K})$ between their effects, where $\mathcal{Ef}(\mathcal{H}) := [0, 1]_{\mathcal{B}(\mathcal{H})}$ is the set of effects on \mathcal{H} , and coincides with the set of *predicates* in [9].

8 Embedding \mathbf{Q} into $(\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$

As seen in the previous section, the two semantics by $\mathbf{Wstar}_{\text{CP-PU}}$ and \mathbf{Q} can be considered as different viewpoints (Schrödinger vs. Heisenberg) for the same phenomena. We can state it categorically: the category \mathbf{Q} can be contravariantly embedded into $\mathbf{Wstar}_{\text{CP-PU}}$.

First we show the following embedding.

Theorem 8.1. *There is a full embedding $I: \mathbf{CPM} \rightarrow (\mathbf{Wstar}_{\text{CP}})^{\text{op}}$.* ■

Proof. Observe the following bijective correspondences.

$$\begin{array}{c}
 f: (n_1, \dots, n_k) \longrightarrow (m_1, \dots, m_l) \quad \text{in } \mathbf{CPM} \\
 \hline
 f_{ij}: n_j \longrightarrow m_i \quad \text{in } \mathbf{CPM}_s, \text{ for each } i, j \\
 \hline
 f_{ij}: \mathcal{M}_{n_j} \longrightarrow \mathcal{M}_{m_i} \quad \text{completely positive, for each } i, j \\
 \hline
 (f_{ij})^*: \mathcal{M}_{m_i} \longrightarrow \mathcal{M}_{n_j} \quad \text{completely positive, hence an arrow in } \mathbf{Wstar}_{\text{CP}}, \text{ for each } i, j \\
 \hline
 I(f): \bigoplus_{i=1}^l \mathcal{M}_{m_i} \longrightarrow \bigoplus_{j=1}^k \mathcal{M}_{n_j} \quad \text{in } \mathbf{Wstar}_{\text{CP}}
 \end{array}$$

For the third correspondence, note that the self-duality of finite dimensional spaces:

$$\mathcal{M}_{n_j} \cong \mathcal{B}(\mathbb{C}^{n_j}) \cong \mathcal{T}(\mathbb{C}^{n_j})^* \cong (\mathcal{M}_{n_j})^* .$$

The last correspondence comes from the fact finite direct sums are biproducts in $\mathbf{Wstar}_{\text{CP}}$. Hence the mapping $I(n_1, \dots, n_k) = \bigoplus_{j=1}^k \mathcal{M}_{n_j}$ defines a contravariant functor $I: \mathbf{CPM} \rightarrow (\mathbf{Wstar}_{\text{CP}})^{\text{op}}$, which is full and faithful by definition, and clearly injective on objects. ■

We use the following lemma.

Lemma 8.2. *Let \mathcal{H} be a Hilbert space. A bounded operator $T \in \mathcal{B}(\mathcal{H})$ is positive if and only if $\text{tr}(TS) \in \mathbb{R}^+ (= [0, \infty))$ for all positive $S \in \mathcal{T}(\mathcal{H})$.* ■

Theorem 8.3. *There is a full embedding $I': \mathbf{Q} \rightarrow (\mathbf{Wstar}_{\text{CP-PU}})^{\text{op}}$.*

Proof. The functor $I: \mathbf{CPM} \rightarrow (\mathbf{Wstar}_{\mathbf{CP}})^{\text{op}}$ restricts to a full and faithful functor $I': \mathbf{Q} \rightarrow (\mathbf{Wstar}_{\mathbf{CP-PU}})^{\text{op}}$ as follows.

$$\begin{aligned}
& f: \vec{n} \rightarrow \vec{m} \text{ is trace-nonincreasing (Definition 3.3.3)} \\
& \iff \sum_i \sum_j \text{tr}(f_{ij}(A_j)) \leq \sum_j \text{tr}(A_j) \text{ for all } (A_j)_j \text{ with positive } A_j \in \mathcal{M}_{n_j} \\
& \iff \sum_i \text{tr}(f_{ij}(A)) \leq \text{tr}(A) \text{ for each } A \in \mathcal{M}_{n_j}, \text{ for each } j \\
& \iff^* \sum_i \text{tr}\left(\left((f_{ij})^*(1)\right)A\right) \leq \text{tr}(A) \text{ for each } A \in \mathcal{M}_{n_j}, \text{ for each } j \\
& \iff \text{tr}\left(\left(1 - \sum_i (f_{ij})^*(1)\right)A\right) \geq 0 \text{ for each } A \in \mathcal{M}_{n_j}, \text{ for each } j \\
& \iff^{**} 1 - \sum_i (f_{ij})^*(1) \geq 0 \text{ for each } j \\
& \iff \sum_i (f_{ij})^*(1) \leq 1 \text{ for each } j \\
& \iff I(f)((1)_i) \leq (1)_j \\
& \iff I(f): \bigoplus_i \mathcal{M}_{m_i} \longrightarrow \bigoplus_j \mathcal{M}_{n_j} \text{ is pre-unital ,}
\end{aligned}$$

where \iff^* is by the equation (1) and \iff^{**} is by Lemma 8.2. ■

In fact, we can say more about the embedding. Notice that $I(\vec{n})$ is a finite dimensional W^* -algebra for each $\vec{n} \in \mathbf{CPM}$. Hence the embedding is restricted to $I_{\text{fd}}: \mathbf{CPM} \rightarrow (\mathbf{FdWstar}_{\mathbf{CP}})^{\text{op}}$, where $\mathbf{FdWstar}_{\mathbf{CP}}$ denotes the category of finite dimensional W^* -algebras and normal completely positive maps. In the same way we have an embedding $I'_{\text{fd}}: \mathbf{Q} \rightarrow (\mathbf{FdWstar}_{\mathbf{CP-PU}})^{\text{op}}$.

Theorem 8.4. *The embeddings*

$$I_{\text{fd}}: \mathbf{CPM} \rightarrow (\mathbf{FdWstar}_{\mathbf{CP}})^{\text{op}} , \quad I'_{\text{fd}}: \mathbf{Q} \rightarrow (\mathbf{FdWstar}_{\mathbf{CP-PU}})^{\text{op}}$$

give equivalences of categories:

$$\mathbf{CPM} \simeq (\mathbf{FdWstar}_{\mathbf{CP}})^{\text{op}} , \quad \mathbf{Q} \simeq (\mathbf{FdWstar}_{\mathbf{CP-PU}})^{\text{op}} .$$

Proof. [41, Theorem I.11.2] implies that I_{fd} and I'_{fd} are essentially surjective. A full, faithful and essentially surjective functor is a part of equivalence [22, Theorem IV.4.1]. ■

9 QPL with infinite types

We have defined the category $\mathbf{Wstar}_{\mathbf{CP-PU}}$ to give the denotational semantics of the language QPL. Because Selinger's category \mathbf{Q} is contravariantly embedded into $\mathbf{Wstar}_{\mathbf{CP-PU}}$, the category $\mathbf{Wstar}_{\mathbf{CP-PU}}$ can be thought of as an infinite dimensional extension of \mathbf{Q} . Working in the category $\mathbf{Wstar}_{\mathbf{CP-PU}}$ rather than \mathbf{Q} enables us to handle infinite types. For example, as Selinger suggested in [37, §7.3], a type **int** should be interpreted as $\llbracket \mathbf{int} \rrbracket = \bigoplus_{n \in \mathbb{N}} \mathbb{C} (\cong \ell^\infty(\mathbb{N}) \text{ below})$, which is indeed in $\mathbf{Wstar}_{\mathbf{CP-PU}}$, but not in \mathbf{Q} .

This observation for **int** can be generalised as follows.

Definition 9.1. For a set S and for a real number $p \geq 1$, we define

$$\ell^p(S) := \left\{ \varphi: S \rightarrow \mathbb{C} \mid \sum_{s \in S} |\varphi(s)|^p < \infty \right\} , \quad \ell^\infty(S) := \left\{ \varphi: S \rightarrow \mathbb{C} \mid \sup_{s \in S} |\varphi(s)| < \infty \right\} .$$

Proposition 9.2. *Let S and T be sets.*

1. $\ell^\infty(S)$ is a W^* -algebra with the predual $\ell^1(S)$.
2. Any function $f: S \rightarrow T$ induces a normal unital $*$ -homomorphism⁵ $\ell^\infty(f): \ell^\infty(T) \rightarrow \ell^\infty(S)$ by

⁵“ $*$ -homomorphism” is a synonym for “multiplicative involutive map”.

$$\ell^\infty(f)(\varphi) = \varphi \circ f.$$

3. *There is a (normal unital) $*$ -isomorphism: $\ell^\infty(S) \overline{\otimes} \ell^\infty(T) \cong \ell^\infty(S \times T)$.*

Proof. We will just sketch the proof of 3. Note that $\ell^2(S)$ is a Hilbert space, and $\ell^\infty(S)$ has a canonical normal unital faithful representation $\pi: \ell^\infty(S) \rightarrow \mathcal{B}(\ell^2(S))$ by $\pi(\varphi)(\psi) = \varphi\psi$ (pointwise multiplication). We can identify the spatial W^* -tensor product $\overline{\otimes}$ with the tensor product of von Neumann algebras. Note the isomorphism $\ell^2(S) \otimes \ell^2(T) \cong \ell^2(S \times T)$ of Hilbert spaces. By the identification $\mathcal{B}(\ell^2(S) \otimes \ell^2(T)) \cong \mathcal{B}(\ell^2(S \times T))$, we have an inclusion $\ell^\infty(S) \odot \ell^\infty(T) \subseteq \ell^\infty(S \times T)$.⁶ The (ultra)weak denseness of the inclusion proves $\ell^\infty(S) \overline{\otimes} \ell^\infty(T) \cong \ell^\infty(S \times T)$. ■

Corollary 9.3. *There is an embedding $\ell^\infty: \mathbf{Set} \rightarrow (\mathbf{CWstar}_{\mathbf{M-I-U}})^{\text{op}}$, from the category of sets and functions to the category of commutative W^* -algebras and unital $*$ -homomorphisms. Moreover, it maps finite products of sets to spatial W^* -tensor products.* ■

We can embed sets S_1, \dots, S_n, T and a function $f: S_1 \times \dots \times S_n \rightarrow T$ contravariantly into the category $\mathbf{CWstar}_{\mathbf{M-I-U}}$ (hence into $\mathbf{Wstar}_{\mathbf{CP-PU}}$) as:

$$\ell^\infty(f): \ell^\infty(T) \longrightarrow \ell^\infty(S_1 \times \dots \times S_n) \cong \ell^\infty(S_1) \overline{\otimes} \dots \overline{\otimes} \ell^\infty(S_n) .$$

Therefore, any classical data type and function between them interpreted in \mathbf{Set} can also be interpreted in $\mathbf{Wstar}_{\mathbf{CP-PU}}$. This enables us, for example, to build in a function symbol $f: \mathbf{nat}^{\otimes n} \rightarrow \mathbf{nat}$ interpreted by a function $\mathbb{N}^n \rightarrow \mathbb{N}$ into the language.

The inhabitation of classical data in W^* -algebras is in fact described more generally. It is known that there is the following dual equivalence of categories

$$\mathbf{CWstar}_{\mathbf{M-I-U}} \simeq \mathbf{LocMeas}^{\text{op}} \quad (2)$$

between the category of commutative W^* -algebras and normal unital $*$ -homomorphisms and the category of localisable measurable spaces and measurable functions [20, §5.8] (see also [36] and [35, Proposition 1.18.1]). Proposition 9.2 and Corollary 9.3 should be generalised for localisable measurable spaces and measurable functions. The detail will be future work.

10 Conclusions and future work

We have given a novel denotational semantics for a first-order functional quantum programming language QPL by operator algebras, specifically by W^* -algebras and normal completely positive pre-unital maps. Our model can be considered as an infinite dimensional extension of Selinger's original model \mathbf{Q} , and is more flexible model to accommodate infinite structures and to unify classical computation as commutative structures.

We believe that W^* -algebras are appropriate universe for quantum computation with good prospects. One of future work is to give a denotational semantics for a higher-order language, or a quantum lambda calculus. In an unpublished paper [19], it is reported that the symmetric monoidal category $((\mathbf{Wstar}_{\mathbf{M-I-U}})^{\text{op}}, \overline{\otimes}, \mathbb{C})$ is *closed*. We may use this result.

Another of future work is to study the detail of the duality (2) of commutative W^* -algebras and localisable measurable spaces. We might have a similar result with [10], showing that $\mathbf{Wstar}_{\mathbf{CP-PU}}$ can also accommodate probabilistic computation.

⁶Here the symbol \odot denotes the algebraic tensor product.

References

- [1] Samson Abramsky & Achim Jung (1994): *Domain Theory*. In: *Handbook of Logic in Computer Science*, III, Clarendon Press, pp. 1–168. Corrected and expanded version available online.
- [2] Huzihiro Araki (1999): *Mathematical Theory of Quantum Fields*. *International Series of Monographs on Physics* 101, Oxford University Press. Originally published in Japanese as *Ryoshiba no Suri* (Iwanami Shoten Publishers, Tokyo, 1993); translated by Ursula Carow-Watamura.
- [3] Ola Bratteli & Derek W. Robinson (1987/1997): *Operator Algebras and Quantum Statistical Mechanics (2 volumes)*, second edition. *Texts and Monographs in Physics*, Springer.
- [4] Nathaniel P. Brown & Narutaka Ozawa (2008): *C^* -Algebras and Finite-Dimensional Approximations*. *Graduate Studies in Mathematics* 88, American Mathematical Society.
- [5] Giulio Chiribella, Alessandro Toigo & Veronica Umanità (2013): *Normal Completely Positive Maps on the Space of Quantum Operations*. *Open Systems & Information Dynamics* 20(1), doi:10.1142/S1230161213500030.
- [6] Kenta Cho (2014): *Semantics for a Quantum Programming Language by Operator Algebras*. Master's thesis, The University of Tokyo. Available at <http://www-mmm.is.s.u-tokyo.ac.jp/~ckn/papers/master-thesis.pdf>.
- [7] John B. Conway (2000): *A Course in Operator Theory*. *Graduate Studies in Mathematics* 21, American Mathematical Society.
- [8] Giacomo Mauro D'Ariano, Dennis Kretschmann, Dirk Schlingemann & Reinhard F. Werner (2007): *Re-examination of quantum bit commitment: The possible and the impossible*. *Phys. Rev. A* 76(032328), doi:10.1103/PhysRevA.76.032328.
- [9] Ellie D'Hondt & Prakash Panangaden (2006): *Quantum weakest preconditions*. *Mathematical Structures in Computer Science* 16, pp. 429–451, doi:10.1017/S0960129506005251.
- [10] Robert Furber & Bart Jacobs (2013): *From Kleisli Categories to Commutative C^* -Algebras: Probabilistic Gelfand Duality*. In: *Algebra and Coalgebra in Computer Science*, *Lecture Notes in Computer Science* 8089, Springer Berlin Heidelberg, pp. 141–157, doi:10.1007/978-3-642-40206-7_12.
- [11] Simon J. Gay (2006): *Quantum programming languages: survey and bibliography*. *Mathematical Structures in Computer Science* 16, pp. 581–600, doi:10.1017/S0960129506005378.
- [12] Rudolf Haag (1996): *Local Quantum Physics: Fields, Particles, Algebras*, second edition. *Theoretical and Mathematical Physics*, Springer, doi:10.1007/978-3-642-61458-3.
- [13] Ichiro Hasuo & Naohiko Hoshino (2011): *Semantics of Higher-Order Quantum Computation via Geometry of Interaction*. *Logic in Computer Science*, Symposium on 0, pp. 237–246, doi:10.1109/LICS.2011.26.
- [14] Teiko Heinosaari & Mário Ziman (2012): *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*. Cambridge University Press.
- [15] Bart Jacobs (2012): *New Directions in Categorical Logic, for Classical, Probabilistic and Quantum Logic*. ArXiv:1205.3940v2 [math.LO].
- [16] Bart Jacobs (2013): *Measurable Spaces and Their Effect Logic*. In: *Logic in Computer Science (LICS), 2013 28th Annual IEEE/ACM Symposium on*, pp. 83–92, doi:10.1109/LICS.2013.13.
- [17] Bart Jacobs (2013): *On Block Structures in Quantum Computation*. *Electronic Notes in Theoretical Computer Science* 298, pp. 233–255, doi:10.1016/j.entcs.2013.09.016.
- [18] Michael Keyl (2002): *Fundamentals of quantum information theory*. *Physics Reports* 369(5), pp. 431–548, doi:10.1016/S0370-1573(02)00266-1.
- [19] Andre Kornell (2012): *Quantum Collections*. ArXiv:1202.2994v1 [math.OA].
- [20] Ryszard Paweł Kostecki (2013): *W^* -algebras and noncommutative integration*. ArXiv:1307.4818v2 [math.OA].

- [21] Nicolaas P. Landsman (2009): *Algebraic Quantum Mechanics*. In Daniel Greenberger, Klaus Hentschel & Friedel Weinert, editors: *Compendium of Quantum Physics*, Springer Berlin Heidelberg, pp. 6–10, doi:10.1007/978-3-540-70626-7_3.
- [22] Saunders Mac Lane (1998): *Categories for the Working Mathematician*, second edition. *Graduate Texts in Mathematics* 5, Springer.
- [23] Ralf Meyer (2008): *Categorical aspects of bivariant K-theory*. In: *K-Theory and Noncommutative Geometry*, EMS Series of Congress Reports, European Mathematical Society, pp. 1–39, doi:10.4171/060-1/1. ArXiv:math/0702145 [math.KT].
- [24] Francis Joseph Murray & John von Neumann (1936): *On Rings of Operators*. *Annals of Mathematics* 37(1), pp. 116–229.
- [25] Francis Joseph Murray & John von Neumann (1937): *On Rings of Operators. II*. *Transactions of the American Mathematical Society* 41(2), pp. 208–248.
- [26] Francis Joseph Murray & John von Neumann (1943): *On Rings of Operators. IV*. *Annals of Mathematics* 44(4), pp. 716–808.
- [27] John von Neumann (1940): *On Rings of Operators. III*. *Annals of Mathematics* 41(1), pp. 94–161.
- [28] John von Neumann (1949): *On Rings of Operators. Reduction Theory*. *Annals of Mathematics* 50(2), pp. 401–485.
- [29] John von Neumann (1955): *Mathematical Foundations of Quantum Mechanics*. Princeton University Press. Originally published in German as *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932); translated by Robert T. Beyer.
- [30] Michael A. Nielsen & Isaac L. Chuang (2000): *Quantum Computation and Quantum Information*. Cambridge University Press.
- [31] Michele Pagani, Peter Selinger & Benoît Valiron (2014): *Applying Quantitative Semantics to Higher-order Quantum Computing*. In: *Proceedings of the 41st ACM SIGPLAN-SIGACT Symposium on Principles of Programming Languages*, ACM, pp. 647–658, doi:10.1145/2535838.2535879.
- [32] Vern Paulsen (2003): *Completely Bounded Maps and Operator Algebras*. *Cambridge Studies in Advanced Mathematics* 78, Cambridge University Press. Solutions to Exercises available online.
- [33] Miklós Rédei (1996): *Why John von Neumann did not Like the Hilbert Space formalism of quantum mechanics (and what he liked instead)*. *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics* 27(4), pp. 493–510, doi:10.1016/S1355-2198(96)00017-2.
- [34] Mathys Rennela (2013): *On operator algebras in quantum computation*. Master’s thesis, Université Paris 7 Denis Diderot. Available at <http://www.cs.ru.nl/~mathysr/papers/masterthesis.pdf>.
- [35] Shôichirô Sakai (1998): *C*-Algebras and W*-Algebras*. *Classics in Mathematics*, Springer. Reprint of the 1971 Edition.
- [36] I. E. Segal (1951): *Equivalences of Measure Spaces*. *American Journal of Mathematics* 73(2), pp. 275–313.
- [37] Peter Selinger (2004): *Towards a quantum programming language*. *Mathematical Structures in Computer Science* 14, pp. 527–586, doi:10.1017/S0960129504004256.
- [38] Peter Selinger & Benoît Valiron (2006): *A lambda calculus for quantum computation with classical control*. *Mathematical Structures in Computer Science* 16, pp. 527–552, doi:10.1017/S0960129506005238.
- [39] Peter Selinger & Benoît Valiron (2008): *On a Fully Abstract Model for a Quantum Linear Functional Language: (Extended Abstract)*. *Electronic Notes in Theoretical Computer Science* 210(0), pp. 123–137, doi:10.1016/j.entcs.2008.04.022.
- [40] Peter Selinger & Benoît Valiron (2009): *Quantum lambda calculus*. In Simon Gay & Ian Mackie, editors: *Semantic Techniques in Quantum Computation*, Cambridge University Press, pp. 135–172.
- [41] Masamichi Takesaki (2001/2003): *Theory of Operator Algebras (3 volumes)*. *Encyclopaedia of Mathematical Sciences* 124–125, 127, Springer.

- [42] Benoît Valiron (2013): *Quantum Computation: From a Programmer's Perspective*. *New Generation Computing* 31(1), pp. 1–26, doi:10.1007/s00354-012-0120-0.

A Kochen-Specker system has at least 21 vertices

(extended abstract)

Sander Uijlen and Bas Westerbaan

Institute for Computing and Information Sciences

Radboud Universiteit Nijmegen

`{suijlen,bwesterb}@cs.ru.nl`

May 12, 2014

Abstract

At the heart of the Conway's Free Will theorems and Kochen and Specker's argument against non-contextual hidden variable theories is the existence of a Kochen-Specker (KS) system: a set of points on the sphere that has no $\{0, 1\}$ -coloring such that at most one of two orthogonal points are colored 1 and of three pairwise orthogonal points exactly one is colored 1. In public lectures, Conway encouraged the search for small KS systems. At the time of writing, the smallest known KS system has 31 vectors.

Arends, Ouaknine and Wampler have shown that a KS system has at least 18 vectors, by reducing the problem to the existence of graphs with a topological embeddability and non-colorability property. The bottleneck in their search proved to be the sheer number of graphs on more than 17 vertices and deciding embeddability.

Continuing their effort, we prove a restriction on the class of graphs we need to consider and develop a more practical decision procedure for embeddability to improve the lower bound to 21.

1 Introduction

1.1 The experiment

Consider the following experiment. Shoot a deuterium atom, or any other spin-1 particle, along, say: the x-axis, through a inhomogeneous magnetic field. Depending on the direction of the magnetic field, the particle will move undisturbed or deviate.

Quantum Mechanics only predicts the probability, given the configuration of the field, whether the particle will deviate. Its probabilistic prediction has been thoroughly tested. One wonders: is there a deterministic non-contextual theory predicting the outcome of this experiment?

Kochen and Specker proved that such a theory cannot satisfy:

SPIN Axiom [5]. Given three pairwise orthogonal directions. In exactly one of the directions, the particle will not deviate.

Their argument is based on the existence of a Kochen-Specker system.

Definition 1. A **Kochen-Specker (KS) system** is a finite set of points on the sphere¹ for which each pair is not antipodal and there is no **010-coloring**. A 010-coloring is a $\{0, 1\}$ -coloring of the points such that ²

1. no pair of orthogonal points are both colored 1 and
2. of three pairwise orthogonal points exactly one is colored 1; or alternatively: they are colored 0, 1 and 0 in some order.

A point on the sphere obviously corresponds to a direction in space. Because of this, the term point, vector and direction can be used interchangeably. Antipodal points correspond to opposite vectors and these span the same direction in space.

Suppose there is a KS system and a non-contextual deterministic theory satisfying the SPIN Axiom. Then we color a point of this system 0, whenever this theory predicts that the particle will deviate if the spin is measured in the direction corresponding to that point, and 1 otherwise. Given two orthogonal points of the system, we can find a third point orthogonal to both of them. The SPIN axiom implies exactly one of them is colored 1, so they cannot both be colored 1. Similarly, given three pairwise orthogonal vectors in the system, the SPIN axiom implies exactly one of them is colored 1. Hence there would be a 010-coloring of the KS system, quod non. Therefore a deterministic non-contextual theory cannot satisfy the SPIN Axiom.

The KS system proposed by Kochen and Specker contained 117 points[7]. Penrose and Peres[10] independently found a smaller system of 33 points. The current record is the 31 point system of Conway[11, p. 197]. As pointed out by [3, 2], finding small KS systems is of both theoretical and practical interest. In public lectures, Conway himself, stressed the search for small KS systems.[9]

¹ We define KS systems to be three dimensional, as in the original proof of Kochen and Specker. Later, higher dimensional systems have been studied. See, for instance [11, p. 201].

² In other papers, like [2], the 0 and 1 are swapped; they consider 101-colorings. These colorings are of course equivalent and the difference arises from considering either squared spin measurements S^2 , or $1 - S^2$.

1.2 Overview

In [2] Arends, Ouaknine and Wampler (AOW) give a computer aided proof that a KS system must have at least 18 vectors. We improve their lower bound and show that a KS system must have at least 21 vectors.

First, in Subsection 1.3, we repeat a part of AOW's work, in particular the reduction of KS systems to graphs. The bottleneck of their search was the sheer number of graphs and the deciding whether such graphs are embeddable.

In Section 2, we improve upon their reduction, to cut down the number of graphs to consider drastically, and state the results of our main computation. Finally, in Section 3, we describe our practical embeddability test.

1.3 Kochen-Specker graphs

We follow [2] and reduce the search for Kochen-Specker systems to the search of a certain class of graphs. First note that in a Kochen-Specker system we may replace a point with its antipodal point. They are both orthogonal to the same points and hence the non-010-colorability is preserved. Therefore, we may assume antipodal points are identified on the sphere. That is: a Kochen-Specker system is a finite subset of the projective plane that is not 010-colorable.

Definition 2. Given a finite subset S of the projective plane (or equivalently, a finite subset of the northern hemisphere without equator). Define its **orthogonality graph** $G(S)$ as follows. The vertices are the points of S . Two vertices are joined by an edge, if their corresponding points are orthogonal.

Definition 3. A graph G is called **embeddable**, if it occurs as a subgraph of an orthogonality graph; that is: if there is a finite subset S of the projective plane, such that $G \leq G(S)$.

Definition 4. A graph is **010-colorable** if there is a $\{0, 1\}$ -coloring, such that

1. for each triangle there is exactly one vertex that is colored 1 and
2. adjacent vertices are not both colored 1.

Definition 5. A **Kochen-Specker graph** is a embeddable graph that is not 010-colorable.

It is an easy, but important, consequence of the definitions that:

Fact 6. *A finite subset S of the projective plane is a Kochen-Specker system, if and only if its orthogonality graph $G(S)$ is Kochen-Specker.*

To prove there is no Kochen-Specker system on 17 points, it would be sufficient to enumerate all graphs on 17 vertices and check these are not 010-colorable or not embeddable. However, this is infeasible as there are already $\sim 10^{26}$ non-isomorphic graphs on 17 points.[12] Luckily, we can restrict ourselves to certain classes of graphs.

Proposition 7 ([2]). *An embeddable graph is squarefree. That is: it does not contain the square as a subgraph.*³

³Some authors call a graph squarefree if it does not contain the square as induced subgraph. For them the complete graph on four vertices is squarefree. We follow Weisstein[15] and Sloane[13] and call a graph squarefree if it does not contain the square as subgraph. For us the complete graph on four vertices is not squarefree.

Proof. Given two non antipodal points $a \neq b$. Consider the points orthogonal to a . This is a great circle. The points orthogonal to b is a different great circle. They intersect in precisely two antipodal points. Hence, if c and d are both orthogonal to a and b , then c and d are equivalent. Therefore, an embeddable graph cannot contain a square. \square

The squarefreeness is a considerable restriction. There are only $\sim 10^{10}$ non-isomorphic squarefree graphs on 17 vertices.[13] We can restrict ourselves to connected graphs.

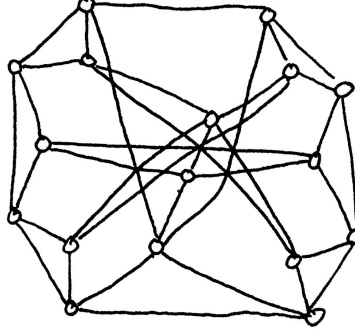
Proposition 8 ([2]). *A minimal Kochen-Specker graph is connected.*

Proof. Suppose G is a non-connected Kochen-Specker graph. Then one of its components is not 010-colorable. As a subgraph of an embeddable graph, is embeddable, this component is embeddable as well. Hence it is a smaller connected Kochen-Specker graph. \square

The gain, however, is small. There are only $\sim 10^9$ non-isomorphic squarefree graphs on 17 vertices that are not connected. In our computations, checking for connectedness required more time than would be gained by reducing the number of graphs.

We have verified the main result of [2]:

Computation 9. *There is a unique non-010-colorable squarefree connected graph on 17 or less vertices:*



It is not embeddable, as the graph in Figure 1 is an unembeddable subgraph. For our proof, see Proposition 19. Hence a Kochen-Specker system has at least 18 points.

2 An improved lower bound

Continuing the effort of Arends, Ouaknine and Wampler, we consider another restriction.

Proposition 10. *A minimal Kochen-Specker graph has minimal vertex-order three.*

Proof. Given a Kochen-Specker graph G . Suppose v is a vertex with order less than or equal 2. Let G' be G with v removed. Clearly G' is embeddable. Suppose G' is 010-colorable. Then we can extend the coloring to a coloring of G as follows. If v is adjacent to only one or no vertex, then we can color v with 0. Suppose v is adjacent to two vertices, say w and w' . If one of w or w' is colored 1, we can color v with 0. If both w and w' are colored 0, we can color v with 1. This would imply G is 010-colorable, quod non. Therefore G' is a smaller Kochen-Specker graph, which contradicts minimality. \square

There are only $\sim 10^7$ squarefree non-isomorphic graphs on 17 vertices with minimal vertex order 3. Even though Arends, Ouaknine and Wampler note this restriction once, surprisingly, they did not restrict their graph enumeration to graphs with minimal vertex order 3.

We continue with a strengthening of Proposition 8.

Proposition 11. *A minimal Kochen-Specker graph is biconnected, that is: removing any single edge leaves the graph connected.*

We need some preparation, before we can prove this Proposition.

Definition 12. Given a graph G and a vertex v of G . We say, v **has fixed color c (in G)**, if G is 010-colorable and for every 010-coloring of G , the vertex v is assigned color c .

We are interested in these graphs because of the following observation.

Lemma 13. *If there is an embeddable graph G on n vertices with a vertex with fixed color 1, then there is a Kochen-Specker graph on $2n$ vertices.*

Proof. Let G be a graph and v a vertex of G with fixed color 1. Consider two copies of the graph G . Connect the two instances of v with an edge. Call this graph G' . Clearly, G' is not 010-colorable.

We need to show G' is embeddable. Given an embedding S of G . We may assume that the point in S corresponding to v is the north pole. Furthermore, we may assume that there is no point on the x -axis, by rotating points along the north pole. Let S' be S rotated 90 degrees along the y -axis. Some points of S and S' might overlap. That is: there might be a point s in S and s' in S' that are equal or antipodal. Observe that if no points of S' and S overlap, then $S \cup S'$ is an embedding of G' .

Suppose there are points in S' and S that overlap. Note that the north pole (and south pole) is not in S' . Let S'' be S' rotated along the north pole at some angle α . There are finitely many angles such that there are overlapping points. Thus there is an angle such that $S \cup S''$ is an embedding of G' . \square

Unfortunately, these graphs are not small.

Computation 14. Let $F_n^{3,1}$ denote the number of connected graphs with minimal vertex-order 3 on n vertices with a vertex with fixed color 1. Then:

$$\frac{n}{F_n^{3,1}} \left| \begin{array}{ccc} \leq 13 & 14 & 15 \\ 0 & 4 & 59 \end{array} \right.$$

Using the methods of Section 3, we have determined these graphs unembeddable.

We need one final computational result.

Computation 15. *There are no connected graphs on less than 15 vertices with minimal vertex-order 2 with a vertex with fixed color 0.*

We are ready to prove that a minimal Kochen-Specker graph is biconnected.

Proof of Proposition 11. Suppose we are given a connected but not biconnected graph G . Then we can decompose this graph into its biconnected components. A biconnected components is a maximal subgraph that is biconnected. We can consider the graph of biconnected components $B(G)$: two biconnected components are adjacent if there is an edge between a vertex in one component and a vertex in the other. Note that there can at most be one edge between the vertices of biconnected components. $B(G)$ does not contain loops: if it would then the union of the biconnected components in the graph, would be itself biconnected. Thus $B(G)$ is a tree.

Furthermore suppose, reasoning towards contradiction, that G is a minimal Kochen-Specker graph. Now consider a leaf A in the biconnected component graph $B(G)$. That is, A is a biconnected component that is connected to only one other biconnected component. Let B be the remainder of the graph. There is exactly one pair (a, b) with a in A and b in B and a adjacent to b .

Note that A and B are 010-colorable. Suppose that a does not have fixed color 1 in A . Then there is a 010-coloring of A which assigns 0 to a . But then G is 010-colorable, quod non. Thus A is an embeddable graph with a vertex with fixed color 1. Similarly B is an embeddable graph with a vertex with fixed color 1.

Let A' denote the graph containing two copies of A connected at the copies of the vertex with fixed color 1. Using the reasoning in the proof of Lemma 13, we see A' is a Kochen-Specker graph that is not biconnected. Define B' similarly. Observe that A' and B' must have the same number of vertices as G by the minimality of G .

The graph A' has minimal vertex-order 3. Thus every vertex of A has minimal order 3 except for a , which has minimal order 2. Suppose a has minimal order 3 in A . Then by Computation 14, the graph A has at least 16 vertices. Thus A' has at least 32 vertices, quod non.

Thus a must have order 2 in A . Let $c, d \in A$ be the vertices connected to a in A . For a to have fixed color 1 in A , the vertices c and d must have fixed color 0 in A . By Computation 15, the graph A has at least $15 + 1$ vertices. Thus A' has at least 32 vertices. Contradiction. \square

We believe a minimal KS graph is also triconnected. However, we did not find a proof.

Although these restrictions are theoretically pleasing, they seem to be of little use as a practical restriction. Concerning excluding unconnected graphs:

Computation 16. *There are five non-isomorphic minimal squarefree connected graphs with minimal vertex order 3 and they have 10 vertices.*

Corollary 17. *Any unconnected squarefree graph with minimal vertex order 3 has at least 20 vertices, for it has two connected components, each with at least 10 vertices. With 20 vertices, there are exactly 25 of these.*

This justifies, at this stage, not checking for connectedness. Similarly, we believe there are very few connected but not biconnected graphs.

Now we can state our main computation.

Computation 18. *Let C_n denote the number of non-010 colorable squarefree graphs with minimal vertex order 3 on n nodes. Then:*

n	≤ 16	17	18	19	20
C_n	0	1	2	19	441

All these 463 graphs are not embeddable. See Computation 20.

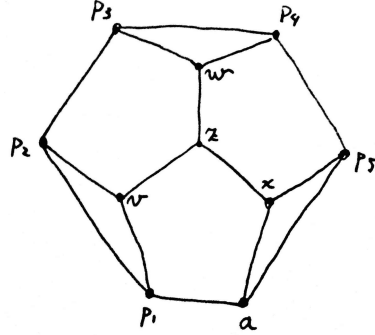
The computation took roughly a week on a 64-core Opteron 6276. It was executed as follows. We enumerated all squarefree graphs with minimal vertex order 3 on less than or equal 20 vertices, using the **geng** util of the nauty software package, which uses the isomorphism-free exhaustive generation method of McKay[8]. The output of **geng**, we passed through a custom heuristic back-tracker written in C++ to decide 010-colorability of these graphs.

3 Embeddability

Our computation has yielded a few hundred non-010-colorable graphs. If we show one of them is embeddable, we have found a new KS system. If we demonstrate all of them are not embeddable, we have proven a lower bound on the size of a minimal KS system.

In [2], Arends, Wampler and Ouaknine discuss several computer-aided methods to test embeddability of a graph. None of these methods could decide for all graphs considered, whether they were embeddable or not.

We propose a new method, which for all graphs we considered, could decide whether they were embeddable or not. First we give a pen-and-paper example.



Proposition 19. *The graph in Figure 1 is not embeddable.*

Figure 1: One of the two minimal non-embeddable graphs

Proof. Suppose it is embeddable. Consider p_1 . It is orthogonal to both a and v . a and v are not collinear, hence p_1 must be collinear to $v \times a$, the cross-product of v and a . Similarly, p_2 is collinear to $v \times p_1 = v \times (v \times a)$. Continuing in this fashion, we see that

$$a \text{ is collinear to } x \times (x \times (w \times (w \times (v \times (v \times a))))). \quad (1)$$

Now, we may assume that $z = (0, 0, 1)$ and $x = (1, 0, 0)$. Thus: $v = (v_1, v_2, 0)$; $w = (w_1, w_2, 0)$ and $a = (0, a_2, a_3)$ for some $-1 \leq v_1, v_2, w_1, w_2, a_2, a_3 \leq 1$, with $v_1^2 + v_2^2 = 1$; $w_1^2 + w_2^2 = 1$ and $a_2^2 + a_3^2 = 1$. Now, (1) becomes:

$$\begin{pmatrix} 0 \\ a_2 \\ a_3 \end{pmatrix} \text{ is collinear to } \begin{pmatrix} 0 \\ -a_2 v_1 w_2 (v_1 w_1 + v_2 w_2) \\ -a_3 (v_1^2 w_1^2 + v_1^2 w_2^2 + v_2^2 w_1^2 + v_2^2 w_2^2) \end{pmatrix}.$$

And therefore

$$\begin{aligned}
v_1 w_2 \langle v, w \rangle &= v_1 w_2 (v_1 w_1 + v_2 w_2) \\
&= v_1^2 w_1^2 + v_1^2 w_2^2 + v_2^2 w_1^2 + v_2^2 w_2^2 \\
&= (v_1^2 + v_2^2) w_1^2 + (v_1^2 + v_2^2) w_2^2 \\
&= w_1^2 + w_2^2 = 1.
\end{aligned}$$

Since v and w are not collinear, we have by Cauchy-Schwarz $|\langle v, w \rangle| < 1$. Recall $|v_1|, |w_2| \leq 1$. Thus: $|v_1 w_2 \langle v, w \rangle| < 1$. Contradiction. \square

In the previous proof, we fixed, without loss of generality, the position of a few vertices. Then we derived cross-product expressions for the remaining vertices. Finally, we find an equation relating some of the cross-product expressions and show it is unsatisfiable. We can automate this reasoning as follows.

```

while there are unassigned vertices do
    pick an unassigned vertex  $v$ ;
    assign  $V(v) = v$ ;
    mark  $v$  as free;
    while there are unassigned vertices adjacent to two different assigned
    vertices do
        pick such a vertex  $w$  adjacent to the assigned  $w_1$  and  $w_2$ ;
        assign  $V(w) = V(w_1) \times V(w_2)$ ;
        mark edges  $(v, w_1)$  and  $(v, w_2)$  as accounted for;
    end
end
for each pair of vertices  $(v_1, v_2)$  do
    if  $(v_1, v_2)$  is not an edge then
        record requirement: “ $V(v_1)$  is not collinear to  $V(v_2)$ ”;
    end
end
for each edge  $(v_1, v_2)$  not accounted for do
    record requirement: “ $V(v_1)$  is orthogonal to  $V(v_2)$ ”;
end

```

At two points in the algorithm, there is a choice which vertex to pick. Depending on the vertices chosen, the number of recorded requirements and free points may significantly vary. By considering all possible choices, one can find the one with least free points.

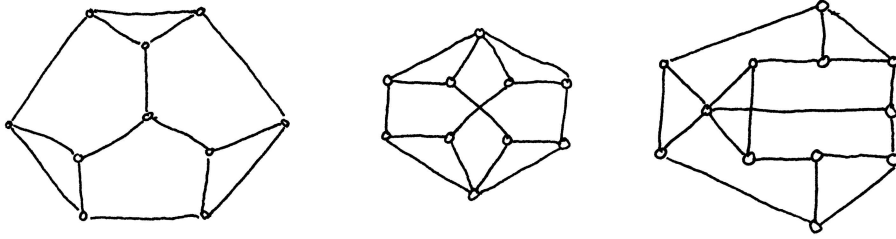
The requirements can be mechanically converted to a formal sentence in the language of the real numbers. This sentence is true if and only if the graph is embeddable. Famously, Tarski proved[14] that such sentences are decidable. His decision procedure has an impractical complexity. However, its practical value has been improved by, for instance, the method of cylindrical algebraic decomposition[4]. We have used the redlog[6] package of the reduce algebra system, which implements a variant of Tarski’s quantifier elimination. In appendix A the Reader can find the reduce script generated mechanically for the graph in Figure 1.

Different assignments give different sentences. In our tests, some assignments would yield sentences that were decided within milliseconds, whereas another assignment with less free vertices would yield a sentence that could not be decided (directly). Therefore, when determining embeddability of a graph, we try several assignments in parallel.

In this way, there were still a few (010-colorable) graphs of which we could not decide embeddability. By hand, we determined these graphs to be embeddable. We adapted the algorithm, as to guess for some assignments the position of one of the vectors. If the corresponding sentence turns out false, we know nothing. However, if the sentence is true, we know the graph is embeddable.

With this method, we have decided in an hour the embeddability of every squarefree graph with minimal vertex order three of 13 vertices or less. In particular:

Computation 20. *Every squarefree graph of minimal vertex order three that is not 010-colorable of order less than or equal to 20 contains, as a subgraph, one of the following three graphs:*



These three graphs are unembeddable. The left and middle graph are the only minimal unembeddable squarefree graph.

For the first graph, we have proven directly that it is unembeddable. See Proposition 19. For the second graph, we also have a similar direct proof. The third graph is shown to not be embeddable using our algorithm.

4 Conclusion and future research

Arends, Ouaknine and Wampler struggled with two problems: enumerating candidate graphs of less than 31 vertices and testing their embeddability. We have verified most of their computations. Then we enumerated all candidate graphs up to and including 20 vertices. Furthermore, we have proposed a new decision procedure, which was able to decide embeddability for all candidate graphs we found. Therefore, we demonstrate: a Kochen-Specker system must have at least 21 points. At the time of writing, we are computing whether there is a KS system of 21 vertices.⁴

Enumerating all candidate graphs of less than 31 vertices is computationally infeasible. To bridge the enormous gap between 21 and 31, requires a new insight. For instance: another restriction on which graphs to consider.

The Reader, interested in pursuing this line of research, is encouraged to read the master thesis[1] of Arends, in which he discusses in detail several other properties that a minimal KS system must enjoy, as well as some failed attempts.

⁴ The authors have a wager whether there is a minimal KS system of less than 25 vertices.

5 Acknowledgments

We wish to thank the following for their generous contribution to the distributed computation: the Digital Security group, Intelligent Systems group and the C&CZ service of the Radboud University; Wouter Geraedts and Jille Timmermans.

We are grateful to prof. McKay for discussing the feasibility of certain graph restrictions.

A Example reduce script

The following is (a part of) the reduce script mechanically generated to prove the non-embeddability of the graph of Figure 1. The algorithm choose a different assignment, than we did in the proof of Proposition 19. As free points it picked, in order, z , x , v and p_3 . The point w is assigned $p_3 \times z$.

```
load_package redlog;
rlset R;

procedure d(x,y);
  (first x) * (first y) +
  (second x) * (second y) +
  (third x) * (third y);

procedure k(x,y);
  {(second x)*(third y) - (third x)*(second y),
   (third x)*(first y) - (first x)*(third y),
   (first x)*(second y) - (second x)*(first y)};

v0c1 := 1; v0c2 := 0; v0c3 := 0;
v1c1 := 0; v1c2 := 1; v1c3 := 0;

v0 := {v0c1, v0c2, v0c3};
v1 := {v1c1, v1c2, v1c3};
v2 := {v2c1, v2c2, v2c3};
v3 := {v3c1, v3c2, v3c3};
v2c1 := 0;
neq0 := k(v0,k(v3,v1));

                                     (snip)

neq29 := k(k(k(k(v3,v1),v1),v2),k(k(v3,v0),v3));
phi :=
  (first neq0 neq 0 or
   second neq0 neq 0 or
   third neq0 neq 0) and

                                     (snip)

  (first neq29 neq 0 or
   second neq29 neq 0 or
   third neq29 neq 0) and
  d(v2,v0) = 0 and
  d(k(k(v3,v0),v3),k(k(k(v3,v1),v1),v2),v2)) = 0 and
  true;
rlqe ex(v3c3,
  ex(v3c2,
  ex(v3c1,
  ex(v2c3,
  ex(v2c2,phi)))));
```

References

- [1] Felix Arends. A lower bound on the size of the smallest kochen-specker vector system in three dimensions. Master's thesis, University of Oxford, 2009. <http://www.cs.ox.ac.uk/people/joel.ouaknine/download/arends09.pdf>.
- [2] Felix Arends, Joël Ouaknine, and Charles W Wampler. On searching for small kochen-specker vector systems. In *Proceedings of the 37th international conference on Graph-Theoretic Concepts in Computer Science*, pages 23–34. Springer-Verlag, 2011.
- [3] Adán Cabello. Kochen–specker theorem and experimental test on hidden variables. *International Journal of Modern Physics A*, 15(18):2813–2820, 2000.
- [4] George E Collins. Quantifier elimination for real closed fields by cylindrical algebraic decomposition. In *Quantifier elimination and cylindrical algebraic decomposition*, pages 85–121. Springer, 1998.
- [5] John H Conway and Simon Kochen. The strong free will theorem. *Notices of the AMS*, 56(2):226–232, 2009.
- [6] Andreas Dolzmann and Thomas Sturm. Redlog: Computer algebra meets computer logic. *Acm Sigsam Bulletin*, 31(2):2–9, 1997.
- [7] Simon Kochen and EP Specker. The problem of hidden variables in quantum mechanics. In *The Logico-Algebraic Approach to Quantum Mechanics*, pages 293–328. Springer, 1975.
- [8] Brendan D McKay. Isomorph-free exhaustive generation. *Journal of Algorithms*, 26(2):306–324, 1998.
- [9] Joël Ouaknine. personal communication. Attended such a lecture of Conway at the Oxford Mathematical Institute in 2005.
- [10] Asher Peres. Two simple proofs of the kochen-specker theorem. *Journal of Physics A: Mathematical and General*, 24(4):L175, 1991.
- [11] Asher Peres. *Quantum theory: concepts and methods*, volume 57. Springer, 1995.
- [12] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. <http://oeis.org/A000088>. Number of graphs on n unlabeled nodes.
- [13] N. J. A. Sloane. The On-Line Encyclopedia of Integer Sequences. <http://oeis.org/A006786>. Squarefree graphs on n vertices.
- [14] Alfred Tarski. *A decision method for elementary algebra and geometry*. Springer, 1998.
- [15] Eric W. Weisstein. Square-free graph. Last visited on may 6th 2014.