Computation in generalised probabilisitic theories

2015 New J. Phys. 17 083001

(http://iopscience.iop.org/1367-2630/17/8/083001)

New Journal of Physics
The open access journal at the forefront of physics

Deutsche Physikalische Gesellschaft DPG
IOP Institute of Physics

Published in partnership with: Deutsche Physikalische Gesellschaft and the Institute of Physics

OPEN ACCESS

RECEIVED
7 March 2015

REVISED
2 June 2015

ACCEPTED FOR PUBLICATION
10 June 2015

PUBLISHED
3 August 2015

PAPER

Computation in generalised probabilisitic theories

Ciarán M Lee and Jonathan Barrett

University of Oxford, Department of Computer Science, Wolfson Building, Parks Road, Oxford OX1 3QD, UK

E-mail: ciaran.lee@cs.ox.ac.uk

Keywords: quantum computation, generalised probabilisitic theories, foundations of quantum theory

Abstract

From the general difficulty of simulating quantum systems using classical systems, and in particular the existence of an efficient quantum algorithm for factoring, it is likely that quantum computation is intrinsically more powerful than classical computation. At present, the best upper bound known for the power of quantum computation is that **BQP** $\subseteq$ **AWPP**, where **AWPP** is a classical complexity class (known to be included in **PP**, hence **PSPACE**). This work investigates limits on computational power that are imposed by simple physical, or information theoretic, principles. To this end, we define a circuit-based model of computation in a class of operationally-defined theories more general than quantum theory, and ask: what is the minimal set of physical assumptions under which the above inclusions still hold? We show that given only an assumption of tomographic locality (roughly, that multipartite states and transformations can be characterized by local measurements), efficient computations are contained in **AWPP**. This inclusion still holds even without assuming a basic notion of causality (where the notion is, roughly, that probabilities for outcomes cannot depend on future measurement choices). Following Aaronson, we extend the computational model by allowing post-selection on measurement outcomes. Aaronson showed that the corresponding quantum complexity class, **PostBQP**, is equal to **PP**. Given only the assumption of tomographic locality, the inclusion in **PP** still holds for post-selected computation in general theories. Hence in a world with post-selection, quantum theory is optimal for computation in the space of all operational theories. We then consider whether one can obtain relativized complexity results for general theories. It is not obvious how to define a sensible notion of a computational oracle in the general framework that reduces to the standard notion in the quantum case. Nevertheless, it is possible to define computation relative to a 'classical oracle'. Then, we show there exists a classical oracle relative to which efficient computation in any theory satisfying the causality assumption does not include **NP**.

1. Introduction

Quantum theory offers dramatic new advantages for various information theoretic tasks [1]. This raises the general question of what broad relationships exist between physical principles, which a theory like quantum theory may or may not satisfy, and information theoretic advantages. Much progress has already been made in understanding the connections between physical principles and some tasks, such as cryptography and communication complexity problems. It is now known that the degree of non-locality in a theory is related to its ability to solve communication complexity problems [2] and to its ability to perform super-dense coding, teleportation and entanglement swapping [3]. Teleportation and no-broadcasting are now better understood than they were when investigated solely from the viewpoint of quantum theory [4, 5]. Cryptographic protocols have been developed whose security relies not on aspects of the quantum formalism, but on general physical principles. For example, device-independent key distribution schemes have been developed that are secure against attacks by post-quantum eavesdroppers limited only by the no-signalling principle [6].

By comparison, relatively little has been learned about the connections between physical principles and computation. It was shown in [7] that a maximally non-local theory has no non-trivial reversible dynamics and, thus, any reversible computation in such a theory can be efficiently simulated on a classical computer. Aside from this result, most previous investigations into computation beyond the usual quantum formalism have centred around non-standard theories involving modifications of quantum theory. These theories often appear to have immense computational power and entail unreasonable physical consequences. For example, non-linear quantum theory appears to be able to solve **NP**-complete problems in polynomial time [8], as does quantum theory in the presence of closed timelike curves [9, 40]. Aaronson has considered other modifications of quantum theory, such as a hidden variable model in which the history of hidden states can be read out by the observer [11], and these have also been shown to entail computational speedups over the usual quantum formalism.

This work considers computation in a framework suitable for describing essentially arbitrary operational theories, where an operational theory specifies a set of laboratory devices that can be connected together in different ways, and assigns probabilities to experimental outcomes. Theories within this framework can be described that are different from classical or quantum theories, but which nonetheless make good operational sense and do not involve peculiarities like closed timelike curves. The framework, described in section 2 suggests a natural model of computation, analogous to the classical and quantum circuit models, described in section 3.

The strongest known non-relativized upper bound for the power of quantum computation is that the class **BQP** of problems efficiently solvable by a quantum computer is contained in the classical complexity class **AWPP**. The class **AWPP** has a slightly obscure definition, but is well known to be contained in **PP**, hence **PSPACE**. Section 3.4 shows that the same result holds for any theory in the operational framework that satisfies the principle of tomographic locality, where this means, roughly, that transformations can be completely characterized by product states and effects. That is, if the complexity class of problems that can be efficiently solved by a specific theory **G** is denoted schematically **BGP**, then for tomographically local theories, **BGP** $\subseteq$ **AWPP**. Once suitable definitions are in place, the proof is essentially the same as the proof for the quantum case: the idea is that this proof can be cast in a theory-independent manner, and be seen to follow from a very minimal set of assumptions on the structure of a physical theory. In fact, the containment **BGP** $\subseteq$ **AWPP** still holds even in the absence of a basic principle of causality (which, if it does hold, ensures that there can be no signalling from future to past).

It was suggested in [14] that quantum theory achieves, in some sense, an optimal balance between its set of states and its dynamics, and that this balance entails that quantum theory is powerful for computation by comparison with most theories in the space of operational theories. Although the status of this suggestion is unknown, it turns out to be exactly correct in the context of a world allowing post-selection of measurement outcomes. Aaronson showed that the class of problems efficiently solvable by a quantum computer with the ability to post-select measurement outcomes is equal to the class **PP** [10]. Section 4 extends the idea of computation with post-selection to general theories, and shows that given (as always) tomographic locality, problems efficiently solvable by any theory with post-selection are contained in **PP**. In other words: any problem efficiently solvable in a tomographically local theory with post-selection, is also efficiently solvable by a quantum computer with post-selection.

Finally, oracles play a special role in quantum computation, forming the basis of most known computational speed-ups over classical computation. Section 5 discusses the problem of defining a sensible notion of oracle in the general framework, that reduces to the standard definition in quantum theory. This problem may not have a solution that is completely general, hence we introduce instead a notion of 'classical oracle' that can be defined in any theory that satisfies the causality principle. There then exists a classical oracle such that relative to this oracle, **NP** is not contained in **BGP** for any theory **G** satisfying tomographic locality and causality .

2. The framework

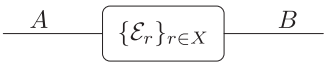
We will work in the circuit framework for generalised probabilistic theories developed by Hardy in [15, 16] and Chiribella, D'Ariano and Perinotti in [12, 13]. The presentation here is most similar to that of Chiribella *et al*.

2.1. Tests and circuits

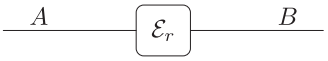
The idea of a generalised probabilistic theory is that a set of physical, or laboratory, devices is specified, which can be connected together in different ways, such that the theory will give probabilities for different outcomes. Such theories take *tests* as their primitive notions, where a test can be thought of as corresponding to a physical device with input ports, output ports, and a classical pointer. Whenever the test is applied, the pointer ends up in one of a number of positions indicating a classical outcome. Input and output ports are typed, with types given by labels A, B, C .... As discussed in more detail below, tests can be composed both sequentially and in parallel, and when

tests are composed sequentially, types must match: the output ports of the first device must have the same types as the corresponding input ports of the second.

Suppose that for a particular test, the classical outcome r takes values in a set X. We shall assume throughout that $|X|$ is finite. A test $\mathcal{E}$, with specified input and output types, then defines a set of *events*, one for each classical outcome, $\{\mathcal{E}_r\}_{r\in X}$. With an input port of type A and an output port of type B, for example, the test can be represented diagrammatically as



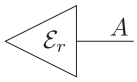
and a specific event as



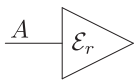
A test is *deterministic* if its outcome set X is the singleton set.

Although tests, with input and output ports, and a pointer, form the primitives of the operational theory, it is also useful to introduce a notion of *physical system*. A system may be thought of as passing between the output port of a device, and the input port of the next, and has the same type as the ports. In other words, in the diagrams above and below, systems correspond to wires. Given two systems of types A and B, we can form a *composite system* of type AB. Operationally, a test with input system AB corresponds to a physical device with a set of input ports labelled by A and a disjoint set of input ports labelled by B.

A test with no input ports corresponds to a preparation of a system—more precisely, such a test corresponds to a set of preparations, with the classical pointer indexing which preparation actually occurs. Such a test can be represented diagrammatically as:



A test with no output ports corresponds to a measurement (that destroys or discards the system), with the classical pointer indexing the measurement outcome. Diagrammatically, such a test can be written:

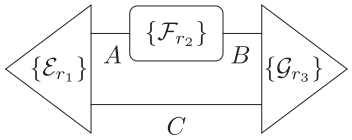


Both tests and events can be composed in sequence and in parallel. If $\{\mathcal{E}_{r_1}\}_{r_1\in X_1}$ is a test from system A to B and $\{\mathcal{U}_{r_2}\}_{r_2\in X_2}$ is a test from system B to C, then their sequential composition is a test from A to C with outcomes $(r_1, r_2) \in X_1 \times X_2$ and events $\{\mathcal{U}_{r_2} \circ \mathcal{E}_{r_1}\}_{(r_1,r_2)\in X_1\times X_2}$. Similarly, if $\{\mathcal{E}_{r_1}\}_{r_1\in X_1}$ is a test from system A to B and $\{\mathcal{U}_{r_2}\}_{r_2\in X_2}$ is a test from system C to D, then their parallel composition is a test from the composite system AC to the composite system BD with outcomes $(r_1, r_2) \in X_1 \times X_2$ and events $\{\mathcal{U}_{r_2} \otimes \mathcal{E}_{r_1}\}_{(r_1,r_2)\in X_1\times X_2}$. Sequential and parallel composition satisfy

$$\left(\mathcal{U}_{r_3} \otimes \mathcal{E}_{r_4}\right) \circ \left(\mathcal{F}_{r_1} \otimes \mathcal{K}_{r_2}\right) = \left(\mathcal{U}_{r_3} \circ \mathcal{F}_{r_1}\right) \otimes \left(\mathcal{E}_{r_4} \circ \mathcal{K}_{r_2}\right),$$

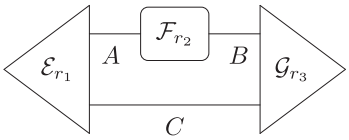
for every $\mathcal{U}_{r_3}$, $\mathcal{E}_{r_4}$, $\mathcal{F}_{r_1}$, $\mathcal{K}_{r_2}$ with the property that the output of $\mathcal{F}_{r_1}$ (respectively, $\mathcal{K}_{r_2}$) matches the input of $\mathcal{U}_{r_3}$ (respectively, $\mathcal{E}_{r_4}$). A generalised probabilistic theory specifies a set of tests, closed under sequential and parallel composition.

A circuit in a generalised probabilistic theory corresponds to a number of tests, connected in sequence and in parallel, such that there are no unconnected ports (i.e., no dangling input or output wires), and no cycles[1]. For example:



A specific outcome of the above circuit corresponds to a particular classical outcome for each of the tests, i.e., to a collection of events, connected in sequence and in parallel:

[1] Connected sets of tests with dangling wires may be called *open circuits*, but this work has no need to consider open circuits, so we use the term *circuit* throughout to refer to a closed circuit.



(2.1.1)

2.2. Probabilistic structure

So far, we have described the operational part of a generalised probabilistic theory, but not the probabilistic part. In addition to specifying a set of tests, hence sets of circuits and circuit outcomes, a probabilistic theory should assign probabilities to circuit outcomes. In a generalised probabilistic theory, every outcome of a circuit is assigned a probability $P(r_1 r_2 \ldots r_n)$, understood as the joint probability of outcomes $r_1, \ldots, r_n$ for the individual tests occurring on a single run. The joint probabilities satisfy $\sum_{r_1 r_2 \ldots r_n} P(r_1 r_2 \ldots r_n) = 1$. A further constraint is that probabilities for unconnected, i.e., independent, circuits factorize. This means that for events $\mathcal{E}_{r_1 r_2 \ldots r_m}$ and $\mathcal{F}_{s_1 s_2 \ldots s_n}$, each of which corresponds to the outcome of a closed circuit, probabilities assigned to the composite events $\mathcal{E}_{r_1 r_2 \ldots r_m} \otimes \mathcal{F}_{s_1 s_2 \ldots s_n}$, and $\mathcal{E}_{r_1 r_2 \ldots r_m} \circ \mathcal{F}_{s_1 s_2 \ldots s_n}$, each satisfy $P(r_1 \ldots r_m, s_1 \ldots s_n) = P(r_1 \ldots r_m) P(s_1 \ldots s_n)$.

The introduction of probabilities into the theory induces linear structure that will be crucial in what follows. Consider two events $\mathcal{E}_0$ and $\mathcal{E}_1$, whose input and output ports have matching types. Suppose that for every closed circuit, and every outcome of the circuit, replacing $\mathcal{E}_0$ with $\mathcal{E}_1$ does not change the probability of the outcome. In this case, $\mathcal{E}_0$ and $\mathcal{E}_1$ are *equivalent*. The events $\mathcal{E}_0$ and $\mathcal{E}_1$ may be easily distinguished operationally by the fact that the corresponding physical devices look quite different, but there is no distinction between $\mathcal{E}_0$ and $\mathcal{E}_1$ from the point of view of the probabilistic predictions of the theory. We refer to the equivalence classes of events formed in this way as *transformations*. The following will mostly be concerned with transformations, rather than the underlying primitive events. Transformations with no input ports we will sometimes call *states*, and transformations with no output ports, *effects*. For system types A and B, the sets of transformations from A to B, states on A and effects on B are denoted $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$, $\mathbf{St}(\mathbf{A})$, and $\mathbf{Eff}(\mathbf{B})$ respectively.

Quantum theory provides a specific example of a theory that can be described in this framework. A system is associated with a complex Hilbert space, with the type of the system given by the dimension of the Hilbert space. States and effects are associated with positive operators, and transformations are associated with trace non-increasing completely positive maps. A test with no input ports corresponds to what is sometimes called a 'random source of quantum states', and is associated with positive operators $\{\rho_r\}$ such that $\sum_r \mathrm{Tr}(\rho_r) = 1$. When the test is performed, the probability that the classical pointer takes position r is given by $\mathrm{Tr}(\rho_r)$, and the quantum state that is prepared, conditioned on the pointer reading being r, is the normalized operator $\rho_r / \mathrm{Tr}(\rho_r)$. A test with no output ports is associated with a positive operator-valued measurement, that is a set of positive operators $\{E_i\}$ satisfying $\sum_i E_i = \mathbb{I}$. A test with both input and output ports is associated with a *quantum instrument*, that is a set of trace non-increasing completely positive maps, one for each value of the pointer reading r, that sum to a trace-preserving map. Given these associations, the standard rules of quantum theory allow the probability to be calculated for any circuit outcome.

Returning to the general framework, it is convenient to use the 'Dirac-like' notation $|\sigma_r)_A$ to represent a state of system type A, and ${}_A(\lambda_r|$ to represent an effect on system type A, so that if the state $|\sigma_{r_1})_A$ is followed by the effect ${}_A(\lambda_{r_2}|$, the joint probability of obtaining outcome r_1 for the preparation and outcome r_2 for the measurement is given by

$$ {}_A(\lambda_{r_2}|\sigma_{r_1})_A := P(r_1, r_2). $$

In the following, we shall sometimes drop the input/output type label. A state $|\sigma_{r_1})_A$ can be identified with a function from effects on A into probabilities, such that

$$ {}_A(\lambda_{r_2}| \mapsto {}_A(\lambda_{r_2}|\sigma_{r_1})_A. $$

Since one can take linear combinations of functions, the set of states $\mathbf{St}(\mathbf{A})$ can be extended to a real vector space, which we denote $\mathbf{V_A}$. In quantum theory, for example, states are positive operators, which span the real vector space $\mathbf{V_A}$ of Hermitian operators.

Similarly, an effect ${}_A(\lambda_{r_2}|$ can be identified with a function from preparation events to probabilities:

$$ |\sigma_{r_1})_A \mapsto {}_A(\lambda_{r_2}|\sigma_{r_1})_A, $$

and the set of effects $\mathbf{Eff}(\mathbf{A})$ can be extended to a real vector space $\mathbf{V^A}$. A more general kind of transformation, from (possibly composite) system type A to (possibly composite) system type B, defines a function into probabilities, where the domain is now circuit fragments with the property that there are unconnected input and output ports, such that adding in a transformation of this type results in a closed circuit. Again, this means that the set of transformations $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$ can be extended to a real vector space, denoted $\mathbf{V_B^A}$.

Throughout the paper, we adopt

Assumption 1. For every pair of system types A and B, and every transformation from A to B, $\mathbf{V}_{\mathbf{B}}^{\mathbf{A}}$ is finite dimensional.

As a consequence, the vector space generated by effects on a system can be regarded as dual to the space of states, and vice versa: $V^A = (V_A)^\star$ and $V_A = (V^A)^\star$. In other works on generalised probabilistic theories, it is quite often assumed that the sets $\mathbf{Transf}(\mathbf{A}, \mathbf{B})$, $\mathbf{St}(\mathbf{A})$, and $\mathbf{Eff}(\mathbf{B})$ are convex subsets of the corresponding vector spaces, the idea being that probabilistic mixtures of allowed transformations should also be allowed transformations. This work, however, doesn't need this assumption: the main constraints on sets of transformations, states and effects are closure under sequential and parallel composition.

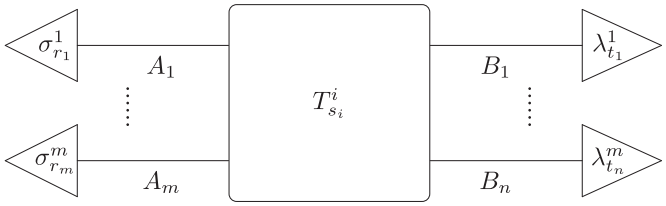
2.3. Tomographic locality

Every transformation T_s from A to B induces a linear map from $\mathbf{V_A}$ to $\mathbf{V_B}$, uniquely defined by

$$|\sigma_r)_A \in \mathbf{St}(\mathbf{A}) \mapsto T_s|\sigma_r)_A \in \mathbf{St}(\mathbf{B}), \tag{2.3.1}$$

where $T_s|\sigma_r)_A$ is the state of type B, corresponding to composition of T_s with $|\sigma_r)_A$. Without further assumptions, however, this map is in general *not* sufficient to specify the transformation T_s. To see this, consider the situation in which the transformation T_s is applied to one half of a bipartite state $|\sigma)_{AC}$. The composition defines a bipartite state of type BC, which can be schematically represented $|\sigma')_{BC} = (T_s \otimes I_C)|\sigma)_{AC}$, with I_C understood as an identity transformation (or the absence of any transformation) on system C. The action of T_s on bipartite states of type AC induces a linear map from $\mathbf{V_{AC}}$ to $\mathbf{V_{BC}}$. In general, however, there need be no simple relationship between this map, and the map above from $\mathbf{V_A}$ to $\mathbf{V_B}$. Indeed, there need not be any simple relationship between the vector space $\mathbf{V_{AC}}$ and the vector spaces for the individual systems, $\mathbf{V_A}$ and $\mathbf{V_C}$. For each possible system type C, this structure is ultimately specified by the theory, via the assignments of probabilities to circuit outcomes[2].

The representation of transformations in a generalised probabilistic theory is greatly simplified by the assumption of *tomographic locality*. A theory satisfies tomographic locality if every transformation can be fully characterized by local process tomography. More formally, consider transformations $T^1_{s_1}$ and $T^2_{s_2}$, both of which have input type $A_1 \cdots A_m$ and output type $B_1 \cdots B_n$. Consider circuit outcomes of the form



with corresponding probability $P^i(r_1 \ldots r_m, t_1 \ldots t_n, s_i)$, where $i \in \{1, 2\}$. Tomographic locality states that for all transformations $T^1_{s_1}$ and $T^2_{s_2}$ with matching input and output types, if

$$P^1(r_1 \ldots r_m, t_1 \ldots t_n, s_1) = P^2(r_1 \ldots r_m, t_1 \ldots t_n, s_2) \quad \forall\ |\sigma^1_{r_1}), \ldots, |\sigma^n_{r_m}), (\lambda^1_{t_1}|, \ldots, (\lambda^n_{t_n}|$$

then

$$T^1_{s_1} = T^2_{s_2}.$$

The whole of the rest of this work adopts

Assumption 2. Tomographic locality is satisfied.

A consequence of tomographic locality is that for a transformation with input type AB and output type CD, the corresponding real vector space has the form [12–14] ,

$$\mathbf{V}_{\mathbf{CD}}^{\mathbf{AB}} \cong \mathbf{V}^{\mathbf{A}} \otimes \mathbf{V}^{\mathbf{B}} \otimes \mathbf{V_C} \otimes \mathbf{V_D},$$

where $\otimes$ here denotes the ordinary vector space tensor product (as opposed to the symbolic $\otimes$ used above to denote parallel composition). In particular, for a bipartite state of type AC, the corresponding vector space $\mathbf{V_{AC}} \cong \mathbf{V_A} \otimes \mathbf{V_C}$. Furthermore, a transformation $T_s \in \mathbf{Transf}(\mathbf{A}, \mathbf{B})$ is completely specified by its action on $\mathbf{St}(\mathbf{A})$, hence T_s can be identified with the linear map defined by equation (2.3.1). When T_s acts on part of a

[2] The operational content of assumption 1 is that there does at least exist a finite set of system types C, such that specification of the action of $T_s \otimes I_C$ on $\mathbf{V_{AC}}$ for each of the system types in this finite set is sufficient to characterize T_s.

bipartite state of type AC, the induced linear map $\mathbf{V_{AC}} \to \mathbf{V_{BC}}$ is given by $T_s \otimes I_C$, where again, the symbol $\otimes$ represents the ordinary vector space tensor product, and I_C is now the identity operator on the vector space $\mathbf{V_C}$. In view of assumptions 1 and 2, the symbol $\otimes$ will from here on denote the ordinary tensor product of finite dimensional vector spaces.

Fixing a basis for each system type, a transformation T with input AB and output CD can be written as a matrix

$$T = \sum_{i,j,k,l} M_{ij,kl} \left(\alpha_i^A \otimes \alpha_j^B \otimes \alpha_k^C \otimes \alpha_l^D \right),$$

where $M_{ij,kl} \in \mathbb{R}$, $\{\alpha_i^A\}$, $\{\alpha_j^B\}$ are bases for $\mathbf{V^A}$ and $\mathbf{V^B}$ respectively, and $\{\alpha_l^C\}$, $\{\alpha_m^D\}$ are bases for $\mathbf{V_C}$ and $\mathbf{V_D}$ respectively. The probability associated with a circuit outcome, e.g., of the form of figure (2.1.1), can be written

$$M^3_{r_3} . \left(M^2_{r_2} \otimes I_C \right) . M^1_{r_1},$$

where $M^1_{r_1}$ (a column vector) is the matrix form of the transformation corresponding to the event $\mathcal{E}_{r_1}$, $M^2_{r_2}$ corresponds to $\mathcal{F}_{r_2}$, and $M^3_{r_3}$ (a row vector) corresponds to $\mathcal{G}_{r_3}$.

2.4. Causality

A nice feature of the Pavia-Hardy framework we have described is that a basic assumption of causality is not implicit, but can be articulated explicitly and theories considered that do not satisfy this assumption. A generalised probabilistic theory is said to be *causal* if the marginal probability of a preparation event is independent of the choice of which measurement follows the preparation. More formally, if $\{|\sigma_i)\}_{i\in X} \subset \mathbf{St(A)}$ are the states corresponding to a preparation test, consider the probability of outcome i, given that a subsequent measurement $\mathcal{E}$ corresponds to a set of effects $\{(\lambda_j|\}_{j\in Y}$:

$$P(i|\mathcal{E}) := \sum_{j\in Y} (\lambda_j|\sigma_i).$$

The theory is causal if for any system type A, any preparation test with outcome i, and any pair of measurements, $\mathcal{E}$ and $\mathcal{F}$, with input type A,

$$P(i|\mathcal{E}) = P(i|\mathcal{F}).$$

Note that the causality assumption is logically independent from tomographic locality: generalised probabilistic theories satisfying one or both or neither can be defined.

If circuits are thought of as having a temporal order, with tests later in the sequence occurring at a later time than tests earlier in the sequence, then the assumption of causality captures the intuitive notion of *no signalling from the future*. It was shown in [12] that a generalised probabilistic theory is causal if and only if for every system type A, there is a unique deterministic effect $_A(u|$. In this case, a measurement, with corresponding effects $\{(\lambda_j|\}_{j\in Y}$, satisfies $\sum_j (\lambda_j| = (u|$. A state $|\sigma)$ is normalized if and only if $(u|\sigma) = 1$. The causality assumption also implies [12] a *no-signalling* principle for the states of the theory. That is, in a causal theory, if a test is performed on the A part of a composite system of type AB, then it is not possible to get information about which test was performed by only performing a test on the B part. (For an interesting extension of this idea to arbitrary causal networks, corresponding to circuits in the Pavia-Hardy framework, see [17].)

Although the idea of *no-signalling from the future* seems intuitive, there is nothing obviously pathological about generalised probabilistic theories that do not satisfy the causality assumption, as long as one does not try to define adaptive circuits, wherein a choice of later test can depend on an earlier outcome. Indeed there is nothing about the framework as it stands that forces an interpretation of the circuits described as a sequence of tests applied in a temporal order matching the order of tests in the circuit. Perhaps an entire closed circuit is set up in advance, and the pointers attain their final resting positions together, when a 'go' button is pressed. Remarkably, the majority of the results derived in this work do not require the causality assumption, hence: *except where explicitly stated, causality is not assumed in what follows*.

2.5. Examples

As already noted, quantum theory can be formulated as a generalised probabilistic theory in the above framework, with finite dimensional quantum theory satisfying assumption 1. Quantum theory satisfies the causality assumption, as the probability of an event cannot depend on the choice of a measurement that is subsequently performed on the system. For a system associated with Hilbert space H, the unique deterministic effect, guaranteed to exist in a theory satisfying the causality assumption, is simply the identity operator $\mathbb{I}$ on H. For a system of type A, the vector space $\mathbf{V_A}$ is the real vector space of Hermitian operators, spanned by the density matrices. It is well known that quantum theory satisfies the assumption of tomographic locality. This follows from the way in which systems combine to form composite systems: a joint state is a positive operator acting on

the tensor product of the Hilbert spaces associated with the individual systems. One can then check that the real vector spaces of Hermitian operators satisfy $\mathbf{V_{AB}} \cong \mathbf{V_A} \otimes \mathbf{V_B}$.

The framework presented is also general enough to accommodate the basic classical theory of finite dimensional probability distributions and stochastic processes, as well as probabilistic theories different from either quantum or classical theory. The latter include 'box world' [3, 14], a causal theory allowing for arbitrarily strong nonlocal correlations, such as the PR box correlations of Popescu and Rohrlich [18] that maximally violate the CHSH inequality. Quantum theory defined over real, rather than complex, Hilbert spaces supplies an example of a theory that does not satisfy tomographic locality. See also [19] for an explicit construction that does not satisfy the causality assumption.

3. Computation in generalised probabilistic theories

3.1. Uniform circuits

The last section showed that in a generalised probabilistic theory, one can draw circuits representing the connections of physical devices in an experiment, and the specific events that took place in the experiment. These circuits provide a natural model of computation, based on the classical and quantum circuit models. A good notion of *efficient* computation needs a definition of a *uniform family of circuits* in a generalised probabilistic theory.

In the standard, classical or quantum, circuit model, a circuit family $\{C_n\} = \{C_1, C_2, ...\}$ consists of a sequence of circuits, each indexed by a positive integer n, denoting the input system size, where C_n is the circuit corresponding to a problem instance of size n. In a poly-size circuit family, the number of gates in C_n is bounded by a polynomial in n, and the circuit family is uniform if a Turing machine can output a description of C_n in time bounded by a polynomial in n.

In a generalised probabilistic theory, there is no reason to assume that a circuit must have the form of a number of gates acting on some input, where the input preparation encodes the problem instance—recall that we do not necessarily assume that the generalised probabilistic theory satisfies the causality assumption, in which case a circuit does not have a preferred direction. Instead, we allow the entire circuit to encode the problem instance, defining a circuit family as a set $\{C_x\}$ such that each circuit is indexed by a classical string $x = x_1 x_2 ... x_n$. A circuit family is poly-size if the number of gates is bounded by a polynomial in $|x|$. For a particular generalised probabilistic theory it might not be the case that bipartite and single system transformations together are universal for computation, as they are in classical and quantum computation. Hence for any k, l, a circuit might involve gates with k input systems and l output systems. In general, it might be the case that no finite gate set is universal for computation. Nonetheless, we will impose as a requirement of uniformity that any uniform circuit family is associated with a finite gate set[3], such that each circuit in the family is built from elements of that set. It follows that the number of distinct system types appearing in a uniform circuit family is also finite.

A further requirement for a circuit family to be uniform takes the form of a constraint on the entries of the matrices representing the transformations that appear in the finite gate set—otherwise, it may be possible to smuggle hard to compute quantities into the computation. There must exist some fixed choice of basis of $\mathbf{V_A}$ for each system A, such that a Turing machine can efficiently compute approximations to the entries of the matrices relative to these bases. We require that for any matrix entry $(M)_{ij}$, and any ϵ, a Turing machine can output a rational number, within ϵ of $(M)_{ij}$, in time bounded by a polynomial in $\log(1/\epsilon)$. This is physically reasonable, since gates are supposed to represent operational devices, and it makes sense to assume that an experimenter with access to devices governed by some generalised probabilistic theory cannot align, or employ, them with arbitrary accuracy.

Finally, for a circuit family $\{C_x\}$ to be uniform, there must be a Turing machine that, acting on input x, outputs a classical description of C_x in time bounded by a polynomial in $|x|$.

The notion of a poly-size uniform circuit family $\{C_x\}$ can be summarized as follows:

- The number of gates in the circuit C_x is bounded by a polynomial in $|x|$.
- There is a finite gate set $\mathcal{G}$, such that each circuit in the family is built from elements of $\mathcal{G}$.
- For each type of system, there is a fixed choice of basis, relative to which transformations are associated with matrices. Given the matrix M representing (a particular outcome of) a gate in $\mathcal{G}$, a Turing machine can output a matrix $\widetilde{M}$ with rational entries, such that $|(M - \widetilde{M})_{ij}| \leqslant \epsilon$, in time polynomial in $\log(1/\epsilon)$.

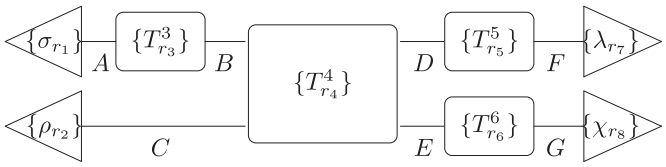
[3] For a uniformity condition where the size of the gate set grows with circuit size, see [39].

- There is a Turing machine that, acting on input $x = x_1 x_2 \ldots x_n$, outputs a classical description of C_x in time bounded by a polynomial in $|x|$.

3.2. Acceptance criterion

Now that we have defined a uniform family of circuits, we need to discuss the acceptance criterion. In quantum computation it is known that performing intermediate measurements during the computation does not increase the computational power. So, without loss of generality, all measurements can be postponed until the end of the computation. A quantum computer can be defined to accept an input string x if the outcome of a computational basis measurement on the first outcome qubit is $|0\rangle$. In a general theory, it need not be the case that all measurements can be postponed until the end of the computation without loss of generality, hence the acceptance criterion should reflect this.

The way in which a generalised probabilistic theory solves a problem might be imagined as follows. First, given the input string x, the circuit C_x is designed and built by composing gates from the fixed finite gate set sequentially and in parallel according to the description. An example of such a circuit is depicted below.



Once the circuit is built, the computation can be run. At the end of a run, each gate has a classical outcome associated with it, where the theory defines a joint probability for these outcomes. For the example above, the joint probability is given by

$$P(r_1,\ldots,r_8) = (\chi_{r_8}|(\lambda_{r_7}|(T^6_{r_6} \otimes T^5_{r_5})\,T^4_{r_4}\,(T^3_{r_3} \otimes I_C)|\rho_{r_2})|\sigma_{r_1}).$$

Denoting the string of observed outcomes by $z = r_1 \ldots r_8$, the final output of the computation will be given by a function $a(z) \in \{0,1\}$, where there must exist a Turing machine that computes a in time polynomial in the length of the input $|x|$. The probability that a computation accepts the input string x is therefore given by

$$P_x(\text{accept}) = \sum_{z\,|\,a(z)=0} P(z),$$

where the sum ranges over all possible outcome strings of the circuit C_x.

3.3. Efficient computation

The class of problems that can be solved efficiently in a generalised probabilistic theory can be defined as follows.

Definition 3.3.1. For a generalised probabilistic theory **G**, a language $\mathcal{L}$ is in the class **BGP** if there exists a poly-sized uniform family of circuits in **G**, and an efficient acceptance criterion, such that

- $x \in \mathcal{L}$ is accepted with probability at least $\frac{2}{3}$.
- $x \notin \mathcal{L}$ is accepted with probability at most $\frac{1}{3}$.

As ever, the choice of the constant 2/3 is arbitrary. Any fixed constant k, $1/2 < k < 1$ would serve equally well[4].

For a specified **G**, the class **BGP** is the natural analogue of **BPP** for probabilistic classical computation, and **BQP** for quantum computation. Indeed, **BGP** reduces to **BPP** or **BQP** in the case that the theory **G** is in fact the classical or quantum theory. See, e.g., [20] for a proof that quantum circuits with mixed states and CP maps are equivalent in computational power to standard quantum circuits with pure states and unitary transformations.

Note that the way in which the acceptance criterion is defined implies that $\mathbf{P} \subseteq \mathbf{BGP}$, for (almost) every generalised probabilistic theory **G**. This is a consequence of the fact that the final output is a function $a(z)$ of the string of observed events, and the only constraint on a is that it can be efficiently computed by a Turing machine. Degenerate cases provide exceptions to this—consider, e.g., any theory such that all transformations are

[4] Note that each uniform circuit (with an efficient acceptance condition) defines a random variable that maps circuit outcomes to the set {**accept**, **reject**} and so one can regard multiple repetitions of a computation as a collection of i.i.d random variables (independence follows from the definition of the probabilistic structure; specifically that the sequential or parallel composition of two events corresponding to outcomes of closed circuits define independent probability distributions). This fact is independent of the form of a particular theory and so holds true for all theories in the framework. Taking this fact in conjunction with the definition of **BGP** and applying the Chernoff bound provides the required result. See [1, p 154] for more discussion of the quantum case.

deterministic, i.e., the outcome set of any circuit is the singleton set. One could remove these degenerate cases by generalizing the acceptance function $a(.)$ so that it depend on both the outcome string z and the input string x. Of course, the fact that **P** $\subseteq$ **BGP** does not have much to do with the intrinsic computational power of a GPT, but is an artefact of the acceptance criterion—it might be interesting to weaken this criterion so that computation in theories intrinsically weaker than classical can be explored.

3.4. Upper bounds on computational power

Using the above definitions of uniform circuit families, and acceptance of an input, the following upper bound on the computational power of any generalised probabilistic theory can be obtained. The main assumption—in addition to those involved in uniformity—is that tomographic locality holds. The result does not require the causality assumption.

Theorem 3.4.1. *For any generalised probabilistic theory* **G** *satisfying tomographic locality,* **BGP** $\subseteq$ **AWPP** $\subseteq$ **PP** $\subseteq$ **PSPACE**.

Here, **PSPACE** consists of those problems that, roughly speaking, can be solved by a classical computer using a polynomial amount of memory. **PP** stands for Probabilistic Polynomial time, which roughly speaking, contains those problems that can be solved by a probabilistic classical computer that must get the answer right with probability $>1/2$. The probability does not need to be bounded away from $1/2$, indeed may be greater than $1/2$ only by an exponentially small amount, hence **PP** contains problems that are not thought to be efficiently solvable by a classical random computer. **AWPP** stands for Almost Wide Probabilistic Polynomial time, and it is known that **AWPP** $\subseteq$ **PP**. The best known upper bound for the class of efficient quantum computations similarly states that **BQP** $\subseteq$ **AWPP**.

To define the class **AWPP**, the notion of a **GapP** function must be introduced. Given a polynomial-time non-deterministic Turing machine M and input string x, denote by $M_{\mathrm{acc}}(x)$ the number of accepting computation paths of M given input x, and by $M_{\mathrm{rej}}(x)$ the number of rejecting computation paths of M given x. A function $f\colon \{0,1\}^{*} \to \mathbb{Z}$ is a **GapP** function if there exists a polynomial-time non-deterministic Turing machine M such that $f(x) = M_{\mathrm{acc}}(x) - M_{\mathrm{rej}}(x)$ for all input strings x. The class **AWPP** can be defined as follows [35].

Definition 3.4.2. The class **AWPP** consists of those languages $\mathcal{L}$ such that there exists a **GapP** function f, and a polynomial r such that

- If $x \in \mathcal{L}$ then $2/3 \leqslant f(x)/2^{r(|x|)} \leqslant 1$,
- if $x \notin \mathcal{L}$ then $0 \leqslant f(x)/2^{r(|x|)} \leqslant 1/3$.

Once the appropriate definitions for generalised probabilistic theories are in place, the proof of theorem 3.4.1 is a fairly straightforward extension of similar proofs for the quantum case, and is presented in appendix B.

Although formal proofs are relegated to appendices, it is useful to sketch the proof that **BGP** $\subseteq$ **PSPACE** in order to provide intuition about how the physical principles underlying generalised probabilistic theories lead to computational bounds.

Sketch proof. Consider a general circuit C_T, with $q(|T|)$ gates. Tensoring these gates with identity transformations on systems on which they do not act, and padding them with rows and columns of zeros, results in a sequence of square matrices $M^{r_q,q},\ldots,M^{r_1,1}$, where $M^{r_n,n}$ is the matrix representing the r_nth outcome of the nth gate. This can be done in such a way that the probability for outcome $z = r_1 \ldots r_q$, is given by

$$b^{\mathrm{T}} M^{r_q,q} \cdots M^{r_2,2} M^{r_1,1} b = \sum_{\{i_1,\ldots,i_{q-1}\}} M^{r_q,q}_{1 i_{q-1}} \cdots M^{r_2,2}_{i_2 i_1} M^{r_1,1}_{i_1 1},$$

where b is the vector $b = (1, 0, \ldots, 0)$ and b^{T} is its transpose. The output probability is a sum of exponentially many terms, but each term is a product of polynomially many numbers, each of which can be efficiently calculated. So a classical Turing machine can calculate each term in the sum, one after the next, keeping a running total. This requires only polynomial-sized memory. □

This proof relies on the ability to decompose the acceptance probability of the computation in a form reminiscent of a (discrete) Feynman path integral. This is a consequence of the fact that transformations in a generalised probabilistic theory are linear, and thus have a matrix representation. It is pertinent then to ask where this linearity comes from. When we introduced generalised probabilistic theories in section 2.1, we associated states (respectively, effects) with functions taking effects (respectively, states) to probabilities. As one

can take linear combinations of such functions, this induces a linear structure on the set of states (respectively, effects). Thus the linear structure of generalised probabilistic theories arises from the requirement that a physical theory should be able to give probabilistic predictions about the occurrence of possible outcomes.

Aside from linearity, a further requirement of the proof is the ability to compute efficiently the entries in the matrices representing the transformations applied in parallel in a specific circuit. Section 2.3 noted that in a theory satisfying tomographic locality, a transformation $\mathcal{E} \in \mathbf{Transf}(\mathbf{A}, \mathbf{B})$ is completely specified by its action on $\mathbf{St}(\mathbf{A})$, and so the matrix representing transformations applied in parallel can be easily calculated by taking the tensor product of the matrices representing each individual transformation. This is not the case in a theory without tomographic locality, where the tensor product structure disappears. If a transformation from **A** to **B** acts on one half of a system **AC**, there may be no simple way to relate the linear map $\mathrm{St}(\mathbf{AC}) \rightarrow \mathrm{St}(\mathbf{BC})$ to the action of the transformation when it is applied to a system **A** on its own, or indeed to a joint system **AC**′. There may therefore be no efficient way of computing matrix elements corresponding to a transformation considered as part of a circuit of arbitrary size. An interesting direction for future work might be to weaken the assumption of tomographic locality such that the results still go through. Real Hilbert space quantum theory provides an example of a theory without tomographic locality for which the above bounds hold, since there is an efficient way of calculating relevant matrix entries.

4. Post-selection and generalised probbilisitic theories

In [10] Aaronson introduced the notion of *post-selected* quantum circuits. These are quantum circuits which, in addition to having a specified qubit, on which a computational basis measurement will be made to provide the outcome, have an additional qubit on which a measurement can be performed such that we can post-select on the outcome. Instead of sampling the measurement result r directly from the computational outcome qubit according to the distribution $P(r)$, only those runs of the computation are counted for which a measurement on the post-selected qubit yields the outcome $s = 0$. The outcome distribution for the computation is taken to be the conditional distribution $P(r|s = 0)$. An extra technical condition is needed, which is that there exists a constant D and polynomial w such that $P(S = 0) \geqslant 1/D^{w(|x|)}$, i.e., we can only post-select on at most exponentially-unlikely outcomes[5].

Definition 4.0.3. A language $\mathcal{L}$ is in the class **PostBQP** if there is a polynomially-sized uniform quantum circuit family, where each circuit has a computational outcome qubit and a post-selected qubit, such that when computational basis measurements are performed on these qubits, with respective outcomes r and s,

- There exists a constant D and polynomial w such that $P(s = 0) \geqslant 1/D^{w(|x|)}$
- If $x \in \mathcal{L}$ then $P(r = 0|s = 0) \geqslant \frac{2}{3}$
- If $x \notin \mathcal{L}$ then $P(r = 0|s = 0) \leqslant \frac{1}{3}$

Aaronson showed in [10] that **PostBQP** = **PP**. Combining this with theorem 3.4.1 gives

Theorem 4.0.4. *For any generalised probabilistic theory* **G**, **BGP** $\subseteq$ **PostBQP**.

Roughly speaking, a post-selecting quantum computer can simulate computation in any other theory satisfying tomographic locality. One can also define a notion of generalised circuits with post-selection on at most exponentially-unlikely outcomes. These are poly-sized uniform circuits in a generalised probabilistic theory, where the probability of acceptance is conditioned on the circuit outcome z lying in a (polytime computable) subset of all possible values of z. Defining the class **PostBGP** in the obvious way, one then obtains

Theorem 4.0.5. *For any generalised probabilistic theory* **G**, **PostBGP** $\subseteq$ **PP**.

The proof is in appendix C. Combining this with Aaronson's result yields

[5] This extra condition was missing from Aaronson's original paper on **PostBQP**, but is needed for the definition of **PostBQP** to be independent of a choice of quantum gate set; see section 2.5 of [21]. We thank Scott Aaronson for some very interesting discussions concerning this point.

Corollary 4.0.6. *For any generalised probabilistic theory* **G**, **PostBGP** $\subseteq$ **PostBQP**.

So, in a world in which we can post-select on at most exponentially-unlikely events, quantum theory is optimal for computation in the space of all tomographically local theories. Note that the class of problems efficiently solvable on a probabilistic classical computer with the power of post-selection is unlikely to be as large as **PP**: it was shown in [22] that if this class, denoted $\mathbf{BPP}_{\mathbf{path}}$, is equal to **PP**, then the polynomial hierarchy collapses to the third level.

It was suggested in [14] (see also [31]) that quantum theory in some sense achieves an optimal balance between the sets of available states and dynamics, in such a way that quantum theory is optimal, or at least powerful, for computation relative to the class of generalised probabilistic theories. It is interesting to ask whether corollary 4.0.6 can be seen as *evidence* in favour of this idea. The following considerations show that caution is needed. Consider, for example, the class **IQP** [22], of restricted quantum computations where the only gates allowed in a circuit are diagonal in the $\{|+\rangle, |-\rangle\}$ basis. Clearly **IQP** $\subseteq$ **BQP**, but it is unlikely that **BQP** $\subseteq$ **IQP**. However, it was shown in [22] that **PostIQP** = **PP** = **PostBQP**. Alternatively, consider the class of restricted quantum computations $\mathbf{DQC_k}$, discussed in [23], known as the *one clean qubit model*, where the inputs to each circuit are restricted to be one pure qubit with as many maximally mixed qubits as desired. At the end of the computation, k qubits are measured in the computational basis. Clearly, $\mathbf{DQC_k} \subseteq$ **BQP**, but again, $\mathbf{DQC_k}$ is not believed to be universal for quantum computation[6].It was shown in [23] that $\mathbf{PostDQC_k}$ = **PP** = **PostBQP** for $k \geqslant 3$. So, while **PostBQP** $\subseteq$ $\mathbf{PostDQC_k}$, under reasonable assumptions [24] it is not the case that **BQP** $\subseteq$ $\mathbf{DQC_k}$.

5. Oracles

In classical computation, an *oracle* is a total function O: $\mathbb{N} \to \{0,1\}$. A number x is said to be in an oracle O if $O(x) = 1$, hence oracles can decide membership in a language. Let **C** and **B** be complexity classes, then $\mathbf{C^B}$ denotes the class **C** with an oracle for **B** (see [25] for formal definitions). We can think of $\mathbf{C^B}$ as the class of languages decided by a computation which is subject to the restrictions and acceptance criteria of **C**, but allowing an extra kind of computational step: an oracle for any desired language $\mathcal{L} \in \mathbf{B}$ that may be queried at any stage in the course of the computation, with each such query counting as a single computational step. That is, bit strings may be generated at any stage of the computation and presented to the oracle, which in a single step, returns the information of whether the bit string is in $\mathcal{L}$ or not. Given two complexity classes, $\mathbf{C}_1$ and $\mathbf{C}_2$, we say the relation[7] $\mathbf{C}_1 = \mathbf{C}_2$ holds relative to the oracle **B**, if $\mathbf{C}_1^{\mathbf{B}} = \mathbf{C}_2^{\mathbf{B}}$. Such a result is referred to as a *relativized separation* result.

Oracles play a special role in quantum computation, forming the basis of most known computational speed ups over classical computation [1]. In quantum computation, oracle queries are represented by a family $\{R_n\}$ of quantum gates, one for each query length. Each R_n is a unitary transformation acting on $n+1$ qubits, whose effect on the computational basis is given by

$$R_n\,|x,\,a\rangle = |x,\,a \oplus A(x)\rangle$$

for all $x \in \{0,1\}^m$ and $a \in \{0,1\}$, where A is some Boolean function that represents the specific oracle under consideration. One could also consider more general oracles that, when queried, apply some general unitary transformation to the query state, but here, we only consider oracles that compute Boolean functions. In the state vector formalism of quantum theory, the action of a unitary oracle is defined on a maximal set of pure and perfectly distinguishable states, namely the computational basis. Linearly extending this to all states in the Hilbert space uniquely defines the action of the oracle on any state.

As pointed out to us by Howard Barnum [26], the situation for generalised probabilistic theories is more subtle. Consider, for example, the density matrix formulation of quantum theory, and suppose that oracle queries correspond to a family of trace-preserving completely-positive maps $\{\mathcal{E}_n\}$. Analogously to the state vector formalism, define the action of the oracle on a maximal set of pure and perfectly distinguishable states $\{\rho_i\}_{i=1}^N$, where each ρ_i is a density matrix, by

$$\mathcal{E}_n\left(\rho_x \otimes \rho_a\right) = \rho_x \otimes \rho_{a \oplus A(x)}, \tag{5.0.1}$$

where $\rho_x = \rho_{x_1} \otimes \cdots \otimes \rho_{x_n}$ and A is the function computed by the oracle. Note that

$$\rho_x \otimes \rho_a \to \rho_x \otimes \rho_{a \oplus A(x)} \quad \Longleftrightarrow \quad |x,\,a\rangle \to \mathrm{e}^{\mathrm{i}\phi(x,a)}\,|x,\,a \oplus A(x)\rangle,$$

[6] In fact, under reasonable assumptions, $\mathbf{DQC_k}$ is provably not universal for quantum computation [24].

[7] The = can be replaced with $\neq$, $\subseteq$ or $\supseteq$ equally well.

where $a = 1,..., N$ and $e^{i\phi(x,a)}$ is some phase factor that depends on the query state. Now, in addition to being able to compute the function A, a quantum computer with access to the oracle may also acquire information about the function ϕ, which may be hard to compute [27]. The usual definition of a quantum oracle therefore prevents 'sneaking in information' through phase factors.

In generalised probabilistic theories (with sufficient distinguishable states), it is easy to produce a definition of an oracle analogous to that of equation (5.0.1). But for a system type A, a maximal set of pure and perfectly distinguishable states does not in general span the vector space $\mathbf{V_A}$. Hence the action of an oracle on such a set of states will not, in general, uniquely define its action on an arbitrary state in the state space. It is then not clear what extra condition must be placed on the oracle, first to define its action on arbitrary input states, and second to prevent non-trivial information being obtained through its action on non-basis input states (perhaps via a generalised notion of phase [28]).

Rather than attempt to solve this problem, we will instead consider a notion of 'classical oracle' that can be defined in any generalised probabilistic theory that satisfies the causality assumption of section 2.4. The causality assumption allows the construction of adaptive circuits without paradox (see [12] for a more thorough discussion of the causality assumption, adaptive circuits, and conditioned transformations). In an adaptive circuit, the choice of which test to perform can depend on the outcomes $r_1, \ldots, r_k$ of previous tests in the circuit. An oracle $A : \mathbb{N} \to \{0,1\}$ defines an extra gate that can be used in a computation in addition to those of the finite gate set, but with input and output that are classical wires, rather than being typed as with the gates intrinsic to the theory. The input to the oracle is a function $f(r_1, \ldots, r_k)$ of the outcomes of tests that appear in the circuit prior to the use of the oracle. The design of that portion of the circuit that is subsequent to the oracle can depend on the output $A(f)$ of the oracle. An oracle can be used in this way an unlimited number of times in a circuit, with each use counting as one gate. The uniformity condition must be extended, so that for each use of the oracle in a circuit, the input $f(r_1, \ldots, r_k)$, and the design of the circuit subsequent to the oracle, are computable in poly-time by a Turing machine with access to an oracle for $\mathbf{A}$. The acceptance criterion can also be extended so that for a circuit outcome z, the function $a(z)$ is computable in poly-time by a Turing machine with access to an oracle for $\mathbf{A}$.

Definition 5.0.7. For each causal generalised probabilistic theory $\mathbf{G}$, a language $\mathcal{L}$ is in the class $\mathbf{BGP}^{\mathbf{A}}_{\mathrm{cl}}$ if there exists a poly-size uniform family of circuits with access to the classical oracle $\mathbf{A}$, and an efficient acceptance condition, such that

- $x \in \mathcal{L}$ is accepted with probability at least $\frac{2}{3}$.
- $x \notin \mathcal{L}$ is accepted with probability at most $\frac{1}{3}$

We can use the notion of classical oracle to obtain the following relativized separation result.

Theorem 5.0.8. *There exists a classical oracle* $\mathbf{A}$ *such that for any causal generalised probabilistic theory* $\mathbf{G}$, $\mathbf{NP}^{\mathbf{A}} \not\subseteq \mathbf{BGP}^{\mathbf{A}}_{\mathrm{cl}}$.

The proof is in appendix D. This generalizes the results of [30] from quantum theory to causal generalised probabilistic theories that satisfy tomographic locality. The result proved in the appendix is actually stronger: there exists a classical oracle $\mathbf{A}$ such that for any causal generalised probabilistic theory $\mathbf{G}$ that satisfies tomographic locality, the polynomial time hierarchy is infinite and $\mathbf{BGP}^{\mathbf{A}}_{\mathrm{cl}} \subseteq \mathbf{P}^{\mathbf{A}}$. The oracle in question is the same oracle that was used by Fortnow and Rogers in [30].

6. Discussion and conclusion

This work has investigated the relationship between computation and physical principles. Using the circuit framework approach to generalised probabilistic theories, introduced by Hardy in [15, 16] and Chiribella, D'Ariano and Perinotti in [12, 13], the computational power of theories formulated in operational terms can be investigated, along with the role played by simple information-theoretic or physical principles that a theory may or may not satisfy. A rigorous model of computation can be defined that allows a definition of the complexity class of problems efficiently solvable by a specific theory. The strongest known inclusion for the quantum case, $\mathbf{BQP} \subseteq \mathbf{AWPP}$, which implies $\mathbf{BQP} \subseteq \mathbf{PP} \subseteq \mathbf{PSPACE}$, still holds in any theory satisfying tomographic locality, and it is notable that this includes even those theories that violate the causality principle. Combining these results with a result of Aaronson's, it follows that any problem efficiently solvable in a theory satisfying tomographic

locality can also be solved efficiently by a post-selecting quantum computer. In fact, one can say something stronger: any problem efficiently solvable with post-selection in a theory satisfying tomographic locality can also be solved efficiently by a post-selecting quantum computer. Roughly speaking, then, in a world with post-selection, quantum theory is optimal for computation in the space of all tomographically local theories.

We discussed the problem of defining a computational oracle for an arbitrary theory. In general, this problem may have no good solution, if it is required that the definition of an oracle reduce to the standard definition in the quantum case. Nonetheless, a notion of 'classical oracle' can be defined in any theory that satisfies the causality principle, and for such theories there exists a classical oracle relative to which **NP** is not contained in **BGP**. It is plausible that there is an interesting subclass of theories, for which a notion of oracle can be defined that admits 'superposition' of inputs, and reduces to the standard definition in the quantum case. If so, then for these theories, the solution of the 'subroutine problem' of [29] might serve as an interesting computational principle that could rule out certain theories, potentially providing a new principle from which quantum theory can be derived.

An open question is to establish tighter bounds on the power of general theories. Even with tomographic locality assumed, there is a lot of freedom in the construction of a generalised theory. Is there an explicit construction that solves a hard problem, that is, a problem at least thought to be hard for quantum computers? Even better, can we describe a complexity class, potentially larger than **BQP**, and an explicit construction of a general theory **G**, such that this class is contained in **BGP**? It would be interesting to determine whether violation of the causality principle can confer extra computational power. An initial thought is that there could be a non-causal theory that can efficiently solve **NP**-complete problems. Given that the inclusion **BGP** $\subseteq$ **AWPP** holds even for non-causal (tomographically local) theories, however, this can only be the case if **NP** is contained in **AWPP**. At present, this is unknown, and establishing the question either way would constitute a major advance in complexity theory. Still, it would be interesting if the violation of causality enabled the efficient solution of other problems, thought to be hard for quantum computers, but known to be in **AWPP**.

Finally, although our main results do not require the causality principle, we have nonetheless been considering circuits in which gates appear in a fixed structure. It would be interesting to investigate the computational power of theories in which there is no such definite structure. Frameworks for describing situations with indefinite causal structure have been defined with the aim of discussing aspects of quantum gravity [32, 33]. Some preliminary remarks on the computational power of such theories were given in [33, 41] and a specific query complexity problem that can be solved with fewer queries on a quantum computer in which the gates do not appear in a fixed order than on a standard quantum computer was presented in [42].

Acknowledgments

The authors thank Howard Barnum for helpful discussions on the definition of oracles in GPTs. CML thanks Lance Fortnow for a useful email correspondence about results presented in [30] and Scott Aaronson for a useful email correspondence about [10]. CML also thanks John Selby, John-Mark Allen, Matty Hoban and Ray Lal for helpful discussions and John Selby and John-Mark Allen for proof reading a draft of the current paper. This work was supported by the FQXi Large Grants *Time and the structure of quantum theory*, and *Thermodynamic versus information theoretic entropies in probabilistic theories*. This work was supported by the CHIST-ERA DIQIP project.

Appendix A. Approximate circuit families

Consider a poly-size uniform circuit family $\{C_x\}$, defined over a finite gate set $\mathcal{G}$. Each gate in $\mathcal{G}$ corresponds to some finite set of transformations, one for each classical outcome of the gate. From the uniformity condition, the entries of the matrices representing these transformations can be calculated to accuracy ϵ in time poly$(\log(1/\epsilon))$. With $\epsilon(|x|)$ a function of the input size, consider a family $\{\widetilde{C}_x\}$ of approximations to the original circuits, where matrix elements are replaced by rational numbers within $\epsilon(|x|)$ of the original matrix elements. Call $\{\widetilde{C}_x\}$ an $\epsilon(|x|)$-approximation to $\{C_x\}$. The following result shows that $\{\widetilde{C}_x\}$ can simulate $\{C_x\}$, to an accuracy dependent on $\epsilon(|x|)$.

Proposition A.0.9. *Let $\{C_x\}$ be a uniform circuit family, with the number of gates in* C_x *bounded by a polynomial $q(|x|)$. Let $\{\widetilde{C}_x\}$ be an $\epsilon(|x|)$-approximation to $\{C_x\}$, with $\epsilon(|x|) \leqslant 1$. If the circuit $C_T \in \{C_x\}$ gives an outcome sequence* z *with probability* P(z)*, then the circuit $\widetilde{C}_T \in \{\widetilde{C}_x\}$ gives outcome sequence* z *with amplitude $\widetilde{P}(z)$ such that*

$$\left|P(z) - \widetilde{P}(z)\right| \leqslant D^{q(|T|)-1} q(|T|)\epsilon(|T|)N,$$

where N *and* D *are constants depending on the gate set* $\mathcal{G}$.

The word *amplitude* here should not be confused with the complex amplitudes of quantum theory. It is used for the real-valued quantity which approximates an outcome probability for the original circuit family, and is used rather than the term *probability*, because this quantity can be (slightly) less than 0 or (slightly) greater than 1. (The approximating circuit family is a mathematical construction that need not correspond precisely to a valid circuit family in the theory.) This proposition will be useful in the main proofs, since if $\{C_x\}$ is a circuit family that decides some language $\mathcal{L}$ in **BGP**, it follows that a $\frac{1}{12q(|x|)D^{q(|x|)-1}N}$-approximation to $\{C_x\}$ will accept a string $x \in \mathcal{L}$ with amplitude at least 7/12, and will accept a string $x \notin \mathcal{L}$ with amplitude at most 5/12, hence the success amplitude is still bounded away from 1/2. The uniformity condition ensures that such an $\epsilon(|x|)$ -approximation can be constructed in time polynomial in $|x|$.

In order to prove the proposition, two lemmas will be helpful.

Lemma A.0.10. *Let* M *be a real* $n \times m$ *matrix such that for each entry,* m_{ij}, *we have that* $|m_{ij}| \leqslant \epsilon$, *for* $\epsilon > 0$. *Then*

$$\| M \|_{\mathrm{op}} \leqslant nm\epsilon,$$

where $\|.\|_{\mathrm{op}}$ *is the operator norm.*

Proof. Let M_i be the ith row of M. Then

$$|M_i|_E = \sqrt{\sum_{j=1}^{m} m_{ij}^2} \leqslant \sum_{j=1}^{m} \left| m_{ij} \right| \leqslant \epsilon m,$$

where $|.|_E$ is the Euclidean norm, hence

$$|Mv|_E \leqslant \sum_{i=1}^{n} |M_i v| \leqslant \sum_{i=1}^{n} \epsilon m = nm\epsilon,$$

for $|v| = 1$, where the second inequality follows from the Cauchy–Schwarz inequality. Thus $\| M \|_{\mathrm{op}} \leqslant nm\epsilon$.! □

Lemma A.0.11. *Let* $\{M_i\}_{i=1}^{\mathsf{T}}$ *and* $\{\widetilde{M}_i\}_{i=1}^{\mathsf{T}}$ *be two sets of matrices. Then the* T*-fold product of these matrices satisfies*

$$\| M_T \ldots M_1 - \widetilde{M}_T \ldots \widetilde{M}_1 \|_{\mathrm{op}} \leqslant D^{T-1} \sum_{i=1}^{T} \| M_i - \widetilde{M}_i \|_{\mathrm{op}},$$

where $D = \max\{\| M_1 \|_{\mathrm{op}}, \ldots, \| M_T \|_{\mathrm{op}}, \| \widetilde{M}_1 \|_{\mathrm{op}}, \ldots, \| \widetilde{M}_T \|_{\mathrm{op}}\}$.

Proof. Consider the case of $T = 2$. With $|v| = 1$,

$$\begin{aligned} &|(M_2 M_1 - \widetilde{M}_2 \widetilde{M}_1) v|_E \\ &= |(M_2 M_1 - \widetilde{M}_2 M_1) v + (\widetilde{M}_2 M_1 - \widetilde{M}_2 \widetilde{M}_1) v|_E \\ &\leqslant |(M_2 - \widetilde{M}_2) M_1 v|_E + |\widetilde{M}_2 (M_1 - \widetilde{M}_1) v|_E \\ &\leqslant \| M_2 - \widetilde{M}_2 \|_{\mathrm{op}} \| M_1 \|_{\mathrm{op}} + \| \widetilde{M}_2 \|_{\mathrm{op}} \| M_1 - \widetilde{M}_1 \|_{\mathrm{op}}. \end{aligned}$$

Thus

$$\| M_2 M_1 - \widetilde{M}_2 \widetilde{M}_1 \|_{\mathrm{op}} \leqslant D \| M_1 - \widetilde{M}_1 \|_{\mathrm{op}} + D \| M_2 - \widetilde{M}_2 \|_{\mathrm{op}}$$

The result follows from induction on T. □

We can now prove proposition A.0.9.

Proof. A particular outcome sequence of the circuit $C_T \in \{C_x\}$ corresponds to a sequence of matrices $\mathcal{G}^{r_1,1}, \ldots, \mathcal{G}^{r_q,q}$, where $\mathcal{G}^{r_i,i}$ represents the r_ith outcome of the ith gate in C_T. Note that states and effects are included in this sequence. Tensoring these gates with identity transformations on systems on which they do not act and padding the corresponding matrices with rows and columns of zeros results in a sequence of square matrices $M^{r_q,q}, \ldots, M^{r_1,1}$ such that

$$P(z) = P(r_1, \ldots, r_q) = b^{\mathsf{T}}. M^{r_q,q} \ldots M^{r_1,1}. b,$$

where b is the vector $(1,\ 0,\ldots,0)$ and b^{T} is its transpose. Similarly for $\widetilde{\mathcal{G}}^{r_1,1},\ldots,\widetilde{\mathcal{G}}^{r_q,q}$, so that

$$\widetilde{P}(z) = \widetilde{P}(r_1,\ \ldots,\ r_q) = b^{\mathsf{T}}.\ \widetilde{M}^{r_q,q}\ldots\widetilde{M}^{r_1,1}.\ b.$$

Note that $\|M^{r_i,i}\|_{\mathrm{op}} \leqslant \|\mathcal{G}^{r_i,i}\|_{\mathrm{op}}$ and $\|\widetilde{M}^{r_i,i}\|_{\mathrm{op}} \leqslant \|\widetilde{\mathcal{G}}^{r_i,i}\|_{\mathrm{op}}$, for all i. Therefore,

$$\begin{aligned}|P(z) - \widetilde{P}(z)| &= |b^{\mathsf{T}}(M^{r_q,q}\ldots M^{r_1,1} - \widetilde{M}^{r_q,q}\ldots\widetilde{M}^{r_1,1})b| \\ &\leqslant |b^{\mathsf{T}}|_E|(M^{r_q,q}\ldots M^{r_1,1} - \widetilde{M}^{r_q,q}\ldots\widetilde{M}^{r_1,1})b\ |_E \\ &\leqslant D'^{q(|T|)-1}\sum_{n=1}^{q}\|M^{r_n,n} - \widetilde{M}^{r_n,n}\|_{\mathrm{op}} \leqslant D'^{q(|T|)-1}q(|T|)N\epsilon(|T|),\end{aligned}$$

where if $n_i m_i$ is the size of the matrix $\mathcal{G}^{r_i,i}$, then

$$N = \max\{n_q m_q,\ldots,n_1 m_1\},$$

and

$$D' = \max\left\{\|\mathcal{G}^{r_1,1}\|_{\mathrm{op}},\ldots,\|\mathcal{G}^{r_q,q}\|_{\mathrm{op}},\ \|\widetilde{\mathcal{G}}^{r_1,1}\|_{\mathrm{op}},\ldots,\|\widetilde{\mathcal{G}}^{r_q,q}\|_{\mathrm{op}}\right\}.$$

Note that, as circuits are built from finite gate sets, N is a constant. The first inequality follows from the Cauchy–Schwarz inequality, the second from that fact that $|b^{\mathsf{T}}| = 1$ and lemma A.0.11, the third from lemma A.0.10, the fact that the sum has $q(|T|)$ entries and the fact that, as $\widetilde{C}_T$ is an ϵ-approximation of C_T, the matrix $M^{r_i,i} - \widetilde{M}^{r_i,i}$ has entries satisfying $|m_{ij} - \widetilde{m}_{ij}| \leqslant \epsilon$.

The reverse triangle inequality gives

$$\|\widetilde{\mathcal{G}}^{r_i,i}\|_{\mathrm{op}} - \|\mathcal{G}^{r_i,i}\|_{\mathrm{op}} \leqslant \|\widetilde{\mathcal{G}}^{r_i,i} - \mathcal{G}^{r_i,i}\|_{\mathrm{op}} \leqslant N\epsilon(|T|).$$

With $\epsilon(|T|) \leqslant 1$, and

$$D'' = \max\left\{\|\mathcal{G}^{r_1,1}\|_{\mathrm{op}},\ldots,\|\mathcal{G}^{r_q,q}\|_{\mathrm{op}}\right\},$$

we have $D' \leqslant D \equiv D'' + N$, which completes the proof. □

Appendix B. Proof of theorem 3.4.1

One method of proving theorem 3.4.1 is to use **GapP** functions. **GapP** functions were first studied in the context of quantum computation by Fortnow and Rogers in [30], where, among other things, they showed that **BQP** $\subseteq$ **AWPP**. A good discussion on **GapP** functions can be found in Watrous's survey of quantum complexity theory [34]. Proofs in this section are modifications and generalizations of proofs presented in [25, 30, 34].

Given a polynomial-time non-deterministic Turing machine M and input string x, denote by $M_{\mathrm{acc}}(x)$ the number of accepting computation paths of M given input x, and by $M_{\mathrm{rej}}(x)$ the number of rejecting computation paths of M given x. A function $f\colon \{0,1\}^* \to \mathbb{Z}$ is a **GapP** function if there exists a polynomial-time non-deterministic Turing machine M such that $f(x) = M_{\mathrm{acc}}(x) - M_{\mathrm{rej}}(x)$ for all input strings x.

Many complexity classes can be described in terms of **GapP** functions. For example the class **PP** can be defined as those languages $\mathcal{L}$ such that, for some **GapP** function f and any input string x, if $x \in \mathcal{L}$ then $f(x) > 0$ but if $x \notin \mathcal{L}$ then $f(x) \leqslant 0$. A useful class of **GapP** functions is provided by the following theorem.

Theorem B.0.12. *Any function $f\colon \{0,1\}^* \to \mathbb{Z}$ that can be computed in poly-time by a Turing machine is a* **GapP** *function.* □

For a proof, see [25, p 237].

The notation $\langle x,\ y\rangle$ denotes the pairing function [30]: that is, a poly-time computable function that maps the pair of strings x and y bijectively to the set of finite length strings $\{0,1\}^*$ such that, given $\langle x,\ y\rangle$, both x and y can be extracted in poly-time. The following proposition gives slight generalizations of standard closure properties of **GapP** functions.

Proposition B.0.13. *For a polynomial* q *and* **GapP** *function f, let $h\colon \{0,1\}^* \to \mathbb{Z}$ be defined for all $x \in \{0,1\}^*$ by*

$$h(x) = \sum_{\substack{|y|\leqslant q(|x|)\\ y\in L_x}} f(\langle x,\ y\rangle),$$

where $\mathrm{L_x}$ *is some set (that may depend on* x*) with the property that membership of* y *in* $\mathrm{L_x}$ *can be determined in time polynomial in* $|x|$*. Then* h *is a* **GapP** *function.*

Now let g*:* $\{0,1\}^\star \to \mathbb{Z}$ *be defined for all* $x \in \{0,1\}^\star$ *by*

$$g(x) = \prod_{\substack{1 \leqslant i \leqslant q(|x|) \\ i \in L_x}} f(\langle x, i\rangle),$$

where the symbol i *appearing as the second argument on the pairing is a binary encoding of* i *and* $\mathrm{L_x}$ *is some set with the property that membership of* i *in* $\mathrm{L_x}$ *can be determined in time polynomial in* $|x|$*. Then* g *is also a* **GapP** *function.*

Proof. We will prove the first statement only as the second statement follows from a similar generalization of a standard argument. Let $f(x) = M_{\mathrm{acc}}(x) - M_{\mathrm{rej}}(x)$ for some non-deterministic poly-time Turing machine, M.

Let N be a non-deterministic poly-time Turing machine that, on input $x \in \{0,1\}^\star$, guesses a string y of length $\leqslant q(|x|)$, decides whether y is in L_x, and

- If $y \in L_x$, simulates M on input $\langle x, y\rangle$.
- If $y \notin L_x$, guesses a bit b and accepts if and only if $b = 0$.

N runs in poly-time, and for every $x \in \{0,1\}^\star$, $N_{\mathrm{acc}}(x) - N_{\mathrm{rej}}(x) = h(x)$, hence h is a **GapP** function. □

For the rest of this section, assume that the pairing function is used whenever a function has two or more arguments. **GapP** functions are intimately related to computation in generalised probabilistic theories, as the following result shows.

Theorem B.0.14. *Let* $\{C_x\}$ *be a poly-size uniform family of circuits in a generalised probabilistic theory. Then for any polynomial* w *and constant* D*, there exists a function* $\epsilon(|x|) \leqslant 1/D^{w(|x|)}$*, and an* $\epsilon(|x|)$*-approximation* $\{\widetilde{C_x}\}$ *to* $\{C_x\}$*, such that the amplitude for acceptance*[8] *of a circuit* $\widetilde{C_T} \in \{\widetilde{C_x}\}$ *is given by*

$$\widetilde{P}_T(\mathrm{accept}) = \frac{f(T)}{2^{p(|T|)}},$$

where f is a **GapP** *function and* $p(|T|)$ *is a polynomial in the size of the input string.*

Proof. It follows from the uniformity condition that for any polynomial w, there is an $\epsilon(|x|)$-approximation $\{\widetilde{C_x}\}$ to $\{C_x\}$, with $\epsilon(|x|) \leqslant 1/D^{w(|x|)}$, such that the entries in the matrices representing gates in the circuit $\widetilde{C_T} \in \{\widetilde{C_x}\}$ have rational entries, and can be computed in time polynomial in $|T|$. Furthermore, the rational entries can be taken to have the form $c/2^d$, with $c \in \mathbb{Z}$, $d \in \mathbb{N}$, and d a polynomial function of $|T|$. Padding circuits with identity gates if necessary, assume that the number of gates in the circuit $\widetilde{C_T}$ is given by a polynomial function $q(|T|)$. A particular outcome of the circuit corresponds to matrices $\widetilde{\mathcal{G}}^{r_1,1}$,..., $\widetilde{\mathcal{G}}^{r_q,q}$, where $\widetilde{\mathcal{G}}^{r_i,i}$ represents the r_ith outcome of the ith gate in $\widetilde{C_T}$. States and effects are included in this sequence.

By tensoring these gates with identity transformations on systems on which they do not act and padding the corresponding matrices with rows and columns of zeros, we can obtain a sequence of square matrices $\widetilde{\mathcal{M}}^{r_1,1}, \ldots, \widetilde{\mathcal{M}}^{r_q,q}$, such that (i) rows and columns of these matrices are indexed by bit strings of length $y(|T|)$, with $y(|T|)$ a polynomial function, and (ii) the amplitude of outcome $z = r_1, \ldots, r_q$ is given by

$$b^{\mathsf{T}}.\,\widetilde{\mathcal{M}}^{r_q,q}\cdots\widetilde{\mathcal{M}}^{r_1,1}.\,b,$$

where b is the vector $(1, 0, \ldots, 0)$ and b^{T} is its transpose. Note that for each $\widetilde{\mathcal{M}}^{r_i,i}$, the matrix $2^d\widetilde{\mathcal{M}}^{r_i,i}$ has integer entries.

Consider the function $h : \{0,1\}^\star \to \mathbb{Z}$ given by

$$h\left(T, r_1, \ldots, r_q, n, i_0,\ldots, i_q\right) = M^{r_n,n}_{i_n i_{n-1}},$$

where $i_0, \ldots, i_q$ are bit strings of length $y(|T|)$, and $M^{r_n,n}_{i_n i_{n-1}}$ is the $i_n i_{n-1}$ entry of the matrix $2^d\widetilde{\mathcal{M}}^{r_n,n}$. By the uniformity condition, these matrix entries can be calculated in polynomial time by a Turing machine, so by theorem B.0.12, h is a **GapP** function.

[8] Note that, as $\{\widetilde{C_x}\}$ is a mathematical construction, it need not correspond to a valid circuit family in the theory and so cannot be said to accept or reject an input string. However, for ease of notation, we will say an approximating circuit 'accepts' an input string if $a(z) = 0$ where z is the outcome sequence of that approximating circuit, and 'rejects' the input string otherwise.

The amplitude for outcome $z = r_1 \ldots r_q$ is given by

$$\begin{aligned}\widetilde{P}(z) &= \frac{1}{2^{dq}} \sum_{\{i_1,\ldots,i_{q-1}\}} M^{r_q,q}_{1 i_{q-1}} \ldots M^{r_2,2}_{i_2 i_1} M^{r_1,1}_{i_1 1}, \\ &= \frac{1}{2^{dq}} \sum_{\{i_1,\ldots,i_{q-1}\}} \prod_{1 \leqslant n \leqslant q} h\left(T, r_1, \ldots, r_q, n, i_0 = 1, i_1,\ldots, i_{q-1}, i_q = 1\right), \\ &= \frac{1}{2^{dq}} \sum_{\{i_1,\ldots,i_{q-1}\}} g\left(T, r_1, \ldots, r_q, i_1,\ldots, i_{q-1}\right), \\ &= \frac{f'(T, z)}{2^{dq}},\end{aligned}$$

where g is a **GapP** function by proposition B.0.13, hence f' is a **GapP** function by another application of proposition B.0.13.

The amplitude for the circuit $\widetilde{C_T}$ to accept is given by

$$\widetilde{P}_T(\text{accept}) = \sum_{a(z)=0} \widetilde{P}_T(z) = \sum_{a(z)=0} \frac{f'(T, z)}{2^{dq}},$$

where $a(z)$ is the function that determines if z is an accepting or rejecting outcome. By the uniformity condition, $a(z)$ can be calculated in polynomial time by a Turing machine, hence proposition B.0.13 gives

$$\widetilde{P}_T(\text{accept}) = \frac{f(T)}{2^{p(|T|)}},$$

where f is a **GapP** function and $d(|T|)q(|T|) = p(|T|)$ is a polynomial that takes values in $\mathbb{N}$. □

The class **AWPP** time can be defined [35] as follows.

Definition B.0.15. The class **AWPP** consists of those languages $\mathcal{L}$ such that there exists a **GapP** function f, and a polynomial r such that

- If $x \in \mathcal{L}$ then $2/3 \leqslant f(x)/2^{r(|x|)} \leqslant 1$.
- If $x \notin \mathcal{L}$ then $0 \leqslant f(x)/2^{r(|x|)} \leqslant 1/3$.

The $1/3 - 2/3$ separation can be replaced by any constant, positive, separation [35].

Theorem B.0.16. *For any generalised probabilistic theory* **G**, **BGP** $\subseteq$ **AWPP**.

Proof. If a language $\mathcal{L} \in$ **BGP**, then there is a poly-size uniform circuit family $\{C_x\}$ such that $P_x(\text{accept}) \geqslant 2/3$ if $x \in \mathcal{L}$, and $P_x(\text{accept}) \leqslant 1/3$ if $x \notin \mathcal{L}$. Assume that for all x, $1/10 \leqslant P_x(\text{accept}) \leqslant 9/10$.[9] By theorem B.0.14, there is an $\epsilon(|x|)$-approximation to $\{C_x\}$ such that the amplitudes determined by the approximating family satisfy

$$\widetilde{P}_x(\text{accept}) = \frac{f(x)}{2^{p(|x|)}},$$

with f a **GapP** function. Furthermore, for any polynomial w, $\epsilon(|x|)$ can be chosen so that $\epsilon(|x|) \leqslant 1/D^{w(|x|)}$. Hence by proposition A.0.9, $\epsilon(|x|)$ can be chosen small enough that $\widetilde{P}_x(\text{accept}) \geqslant 7/12$ if $x \in \mathcal{L}$ and $\widetilde{P}_x(\text{accept}) \leqslant 5/12$ if $x \notin \mathcal{L}$, and for all x, $0 \leqslant \widetilde{P}_x(\text{accept}) \leqslant 1$. Taking $p(|x|)$ to be the function $r(|x|)$ in definition B.0.15 and noting that $5/12 - 7/12$ is a constant, positive, separation, gives the result. □

It is well known that **AWPP** $\subseteq$ **PP** $\subseteq$ **PSPACE** (see, for example, [36] and references therein).

Appendix C. Proof of theorem 4.0.5

An alternate definition of the class **PP** can be stated [35, 37] as follows.

[9] This can be ensured, if necessary, by considering the circuit C_T to be carried out in parallel with a biased coin toss. With probability 1/5, the coin is tails, in which case the output of the circuit is ignored, and acceptance/rejection are returned with probability 1/2 each. Taken together, these circuits and coin tosses define a modified circuit family $\{C'_x\}$, and in the following, approximating circuit families can be assumed to be defined relative to $\{C'_x\}$.

Definition C.0.17. The class **PP** consists of those languages $\mathcal{L}$ such that there exist **GapP** functions f and h so that for all x

- If $x \in \mathcal{L}$ then $2/3 \leqslant f(x)/h(x) \leqslant 1$.
- If $x \notin \mathcal{L}$ then $0 \leqslant f(x)/h(x) \leqslant 1/3$.

The $1/3 - 2/3$ separation can be replaced by any constant, positive, separation [35].

In order to prove theorem 4.0.5, consider a uniform family of circuits $\{C_x\}$ in the generalised probabilistic theory **G**. Let S_T be a subset of the possible outcomes of the circuit C_T, with respect to which post-selection is defined, so that $P_T(\text{accept}|S_T) \geqslant 2/3$ for $T \in \mathcal{L}$ and $\leqslant 1/3$ for $T \notin \mathcal{L}$. As in the proof of theorem 3.4.1, assume that these probabilities are also bounded away from 0 and 1 so that for all T, $1/10 \leqslant P_T(\text{accept}|S_T) \leqslant 9/10$.[10]

By theorem B.0.14, there is an $\epsilon(|x|)$-approximation to $\{C_x\}$ such that, in the approximating family, the joint amplitude to accept the computation and have an outcome from the set S_T is

$$\widetilde{P}_T(\text{accept}, S_T) = \frac{f(T)}{2^{p(|T|)}},$$

with f a **GapP** function. Similarly,

$$\widetilde{P}_T(S_T) = \frac{g(T)}{2^{q(|T|)}},$$

with g a **GapP** function and q a polynomial. Furthermore, for any polynomial w and constant D, $\epsilon(|x|)$ can be chosen so that $\epsilon(|x|) \leqslant 1/D^{w(|x|)}$. Hence by proposition A.0.9 and the fact that we are post-selecting on at most exponentially-unlikely outcomes, $\epsilon(|x|)$ can be chosen small enough that for the approximating circuit family, $\widetilde{P}_T(S_T) > 0$. This means that for the approximating circuit family, the conditional

$$\widetilde{P}_T(\text{accept}|S_T) = \frac{\widetilde{P}_T(\text{accept}, S_T)}{\widetilde{P}_T(S_T)},$$

is well defined. Furthermore, $\epsilon(|x|)$ can be chosen small enough that $\widetilde{P}_T(\text{accept}|S_T) \geqslant 7/12$ if $x \in \mathcal{L}$, $\widetilde{P}_T(\text{accept}|S_T) \leqslant 5/12$ if $x \notin \mathcal{L}$, and using the assumption that the original circuit family probabilities are bounded away from 0 and 1, the approximating amplitudes satisfy $0 \leqslant \widetilde{P}_T(\text{accept}|S_T) \leqslant 1$.

Now,

$$\widetilde{P}_T(\text{accept}|S_T) = \frac{2^{q(|T|)}f(T)}{2^{p(|T|)}g(T)} = \frac{l(T)}{h(T)},$$

where $h(T) = 2^{p(|T|)}g(T)$ and $l(x) = 2^{q(|T|)}f(T)$ are **GapP** functions. This follows from theorem B.0.12, proposition B.0.13, and the fact that both p and q are polynomials taking values in $\mathbb{N}$. The result follows.

Appendix D. Proof of theorem 5.0.8

Denote by **PH** the polynomial time hierarchy: the union of an infinite hierarchy of classes $\boldsymbol{\Sigma}_k$, $\boldsymbol{\Delta}_k$ and $\boldsymbol{\Pi}_k$ for $k \in \mathbb{N}$, where $\boldsymbol{\Sigma}_0 = \boldsymbol{\Delta}_0 = \boldsymbol{\Pi}_0 = \mathbf{P}$ and $\boldsymbol{\Sigma}_{k+1} = \mathbf{NP}^{\boldsymbol{\Sigma}_k}$, $\boldsymbol{\Delta}_{k+1} = \mathbf{P}^{\boldsymbol{\Sigma}_k}$ and $\boldsymbol{\Pi}_{k+1} = \mathbf{coNP}^{\boldsymbol{\Sigma}_k}$. The polynomial time hierarchy is a natural way of classifying the complexity of problems beyond the class **NP**. It is a strongly held belief in computer science that **NP** includes non-polynomial-time problems.

Theorem 5.0.8 is a corollary of two results, the first of which is due to [36] and [38]:

Theorem D.0.18. *There exists an oracle* **A** *such that* $\mathbf{P}^{\mathbf{A}} = \mathbf{AWPP}^{\mathbf{A}}$ *and the polynomial time hierarchy is infinite.*

The second is that theorem B.0.16 *relativizes*.

Theorem D.0.19. *For any classical oracle* **A** *we have that* $\mathbf{BGP}^{\mathbf{A}}_{\mathrm{cl}} \subseteq \mathbf{AWPP}^{\mathbf{A}}$ *for any causal* **G**.

Proof. Given the uniformity condition for circuit families with an oracle, entries in the matrices representing gates in a circuit are all computable in polynomial time by a Turing machine with access to the oracle **A**. Thus the proof of theorem B.0.14 goes through essentially unchanged, except that in this case the conclusion is that

[10] This can be done, as before, by the introduction of a biased coin parallel to the circuit. If the circuit outcome is in S_T and the coin is heads, then accept or reject, depending on the circuit outcome. If the outcome is in S_T and the coin is tails then accept or reject with probability 1/2 each.

the acceptance amplitude is

$$\widetilde{P}_x(\text{accept}) = \frac{f(x)}{2^{p(|x|)}},$$

where $p(|x|)$ is a polynomial function of the size of the input and f is a $\mathbf{GapP}^{\mathbf{A}}$ function. A $\mathbf{GapP}^{\mathbf{A}}$ function is defined in a similar fashion to a **GapP** function, except instead of counting the difference between the number of accepting and rejecting paths for any input into a non-deterministic Turing machine, $\mathbf{GapP}^{\mathbf{A}}$ functions count the difference between the number of accepting and rejecting paths for any input into a non-deterministic Turing machine with access to the oracle **A**. $\mathbf{AWPP}^{\mathbf{A}}$ can be defined with respect to $\mathbf{GapP}^{\mathbf{A}}$ functions by just replacing every mention of **GapP** functions with $\mathbf{GapP}^{\mathbf{A}}$ functions in definition B.0.15. Thus the proof that $\mathbf{BGP}^{\mathbf{A}}_{\mathrm{cl}} \subseteq \mathbf{AWPP}^{\mathbf{A}}$, for any causal GPT and oracle **A**, goes through exactly the same as the proof of theorem B.0.16. □

Hence we obtain

Theorem D.0.20. *There exists a classical oracle* **A** *relative to which* $\mathbf{BGP}^{\mathbf{A}}_{\mathrm{cl}} \subseteq \mathbf{P}^{\mathbf{A}}$, *for all causal* **G**, *and the polynomial time hierarchy is infinite.*

This implies that there exists a classical oracle relative to which **NP** is not contained in **BGP**, for any causal theory **G** satisfying tomographic locality. This generalizes the results of [30] from quantum theory to general theories.

References

[1] Nielsen M A and Chuang I L 2000 *Quantum Computation and Quanum Information* (Cambridge: Cambridge University Press)
[2] van Dam W 2005 Implausible consequences of superstrong nonlocality (arXiv:quant-ph/0501159)
[3] Short A J and Barrett J 2010 Strong nonlocality: a trade-off between states and measurements *New J. Phys.* **12** 033034
[4] Barnum H, Barrett J, Leifer M and Wilce A 2008 Teleportation in general probabilistic theories (arXiv:0805.3553)
[5] Barnum H, Barrett J, Leifer M and Wilce A 2007 A generalised no-braodcasting theorem *Phys. Rev. Lett.* **99** 240501
[6] Barrett J, Hardy L and Kent A 2005 No signalling and quantum key distribution *Phys. Rev. Lett.* **95** 010503
[7] Gross D, Mueller M, Colbeck R and Dahlsten O 2010 All reversible dynamics in maximal non-local theories are trivial *Phys. Rev. Lett.* **104** 080402
[8] Abrams D S and Lloyd S 1998 Nonlinear quantum mechanics implies polynomial-time solution for NP-complete problems *Phys. Rev. Lett.* **81** 3992–5
[9] Bacon D 2004 Quantum computational complexity in the presence of closed timelike curves *Phys. Rev.* A **70** 032309
[10] Aaronson S 2004 Quantum computing, postselection and probabilistic polynomial time (arXiv:quant-ph/0412187v1)
[11] Aaronson S 2004 Quantum computing and hidden variables: II. the complexity of sampling histories (arXiv:quant-ph/0408119)
[12] Chiribella G, D'Ariano G M and Perinotti P 2010 Probabilistic theories with purification *Phys. Rev.* A **81** 062348
[13] Chiribella G, D'Ariano G M and Perinotti P 2011 Informational derivavtion of quantum theory *Phys. Rev.* A **84** 012311
[14] Barrett J 2007 Information processing in generalised probabilitic theories *Phys. Rev.* A **75** 032304
[15] Hardy L 2001 Quantum theory from five reasonable axioms (arXiv:quant-ph/0101012v4)
[16] Hardy L 2011 Reformulating and reconstructing quantum theory (arXiv:1104.2066)
[17] Henson J, Lal R and Pusey M 2014 Theory-independent limits on correlations from generalised Bayesian networks *New J. Phys.* **16** 113043
[18] Popescu S and Rohrlich D 1994 Quantum nonlocality as an axiom *Found. Phys.* **24** 379–85
[19] D'Ariano G, Manessi F and Perinotti P 2013 Determinism without causality (arXiv:1301.7578)
[20] Aharonov D, Kitaev A and Nisan N 1998 Quantum circuits with mixed states (arXiv:quant-ph/9806029)
[21] Kuperberg G 2014 How hard is it to approximate the Jones polynomial? (arXiv:0908.0512v2)
[22] Jozsa R, Shepherd D and Bremner M 2010 Classical simulation of commuting qauntum computations implies collapse of the polynomial hierarchy (arXiv:1005.1407)
[23] Fitzsimons J, Morimae T and Fujii K 2014 On the hardness of classically simulating the one clean qubit model *Phys. Rev. Lett.* **112** 130502
[24] Ambainis A, Schuman L and Vazirani U 2000 Computing with highly mixed states (arXiv:quant-ph/0003136)
[25] Hemaspaandra L and Ogihara M 2002 *The Complexity Theory Companion* (Berlin: Springer)
[26] Barnum H 2014 private communication
[27] Machta J 1998 Phase information in quantum oracle computing (arXiv:quant-ph/9805022)
[28] Garner A, Dahlsten O, Nakata Y, Murao M and Vedral V 2013 A generl framework for phase and interference *New J. Phys.* **15** 093044
[29] Bennett C, Bernstein E, Brassard G and Vazirani U 1997 Strengths and weakneses of quantum computing (arXiv:quant-ph/9701001)
[30] Fortnow L and Rogers J 1998 Complexity limitations on quantum computation (arXiv:cs/9811023)
[31] Mueller M and Ududec C 2012 The structure of reversible computation determines the self-duality of quantum theory *Phy. Rev. Lett.* **108** 130401
[32] Hardy L 2005 Probability theories with dynamical causal structure: a new framework for quantum gravity (arXiv:gr-qc/0509120)
[33] Hardy L 2007 Quantum gravity computers: on the theory of computation with indefinite causal structure (arXiv:quant-ph/0701019)
[34] Watrous J 2008 Quantum computational complexity (arXiv:0804.3401)
[35] Fenner S 2003 PP-lowness and simple definition of AWPP *Theory of Computing Systems* vol 36, issue 2
[36] Fenner S, Fortnow L, Kurtz S and Li L 1993 An oracle builders toolkit *Proc. 8th IEEE Structure in Complexity Theory Conf.*
[37] Li L 1993 On the counting functions *PhD Thesis* (University of Chicago)
[38] Yao A 1985 Separating the polynomial time hierarchy by oracles: I. *Proc. 26th IEEE FOCS*

[39] de Beaudrap N 2014 On computation with probabilities modulo k (arXiv:1405.7381)
[40] Allen J 2014 Treating time travel quantum mechanically *Phys. Rev.* A **90** 042107
[41] Chiribella G, D'Ariano G, Perinotti P and Valiron B 2013 Quantum computations without definite causal structure *Phys. Rev.* A **88** 022318
[42] Araújo M, Costa F and Brukner C 2014 Computational advantage from quantum-controlled ordering of gates *Phys. Rev. Lett.* **113** 250402