PROCEEDINGS A

rspa.royalsocietypublishing.org

Research

Cite this article: Lee CM, Hoban MJ. 2016 Bounds on the power of proofs and advice in general physical theories. *Proc. R. Soc. A* **472**: 20160076.
http://dx.doi.org/10.1098/rspa.2016.0076

Received: 1 February 2016
Accepted: 3 May 2016

Subject Areas:
quantum physics, theory of computing

Keywords:
generalized probabilistic theories, quantum computing, computational complexity

Author for correspondence:
Matty J. Hoban
e-mail: matty.hoban@cs.ox.ac.uk

THE ROYAL SOCIETY
PUBLISHING

Bounds on the power of proofs and advice in general physical theories

Ciarán M. Lee and Matty J. Hoban

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

MJH, 0000-0001-9765-0373

Quantum theory presents us with the tools for computational and communication advantages over classical theory. One approach to uncovering the source of these advantages is to determine how computation and communication power vary as quantum theory is replaced by other operationally defined theories from a broad framework of such theories. Such investigations may reveal some of the key physical features required for powerful computation and communication. In this paper, we investigate how simple physical principles bound the power of two different computational paradigms which combine computation and communication in a non-trivial fashion: computation with advice and interactive proof systems. We show that the existence of non-trivial dynamics in a theory implies a bound on the power of computation with advice. Moreover, we provide an explicit example of a theory with no non-trivial dynamics in which the power of computation with advice is unbounded. Finally, we show that the power of simple interactive proof systems in theories where local measurements suffice for tomography is non-trivially bounded. This result provides a proof that **QMA** is contained in **PP**, which does not make use of any uniquely quantum structure—such as the fact that observables correspond to self-adjoint operators—and thus may be of independent interest.

1. Introduction

(a) Motivation

Since the mid-1980s there has been growing evidence that quantum theory offers dramatic advantages in both

computation and communication problems [1–5]. In particular, the existence of an efficient quantum algorithm for factoring [2] and of a communication problem for which quantum theory requires exponentially less communication to solve [4] has challenged classical conceptions of what problems are efficiently solvable in our physical world.

Much recent work has been concerned with uncovering the source of this quantum advantage [6–15]. One approach to this problem is to view quantum theory in the context of a framework general enough to accommodate essentially any operationally defined theory [16,17]. While most of these theories may not correspond to descriptions of our physical world, they nevertheless make good operational sense and allow one to systematically assess how computation and communication power depend on the underlying physical theory. Determining how computation and communication power vary as quantum theory is replaced by other operationally defined theories may reveal some of the key physical features required for powerful computation and communication.

More generally, within this framework, one can identify physical principles that theories may or may not satisfy, such as causality (no signalling from future to past) or tomographic locality (local measurements suffice for tomography of joint states). It has recently been shown[1] [16] that for any theory satisfying tomographic locality, whether or not causality is satisfied, computational problems that can be solved efficiently are contained in the classical complexity class **AWPP**—a fact first proved in the quantum case by Fortnow & Rogers [20].

In this paper, we investigate how simple physical principles bound the power of two different computational paradigms which combine both computation and communication in a non-trivial fashion: computation with advice and interactive proof systems. These are standard tools in computational complexity and one can view our work as methodically exploring the impact of general physical theories upon these tools (further expanding upon the work in [16]).

(b) Overview of the results

Computation with advice considers the situation where an efficient computer is supplemented with extra information, or *advice*, which, in classical computation, takes the form of a bit string and, in quantum computation, takes the form of a quantum state. The usefulness of this computational paradigm is that no so-called uniformity constraints are placed on the string or state embodying the advice—as is usually the case when one considers efficient computation—and so one can attempt to encode solutions to hard problems in the advice. Aaronson was among[2] the first to study and set bounds on the power of quantum computation with (quantum) advice [22]. His primary motivation was a desire to investigate the question "How many classical bits can 'really' be encoded into n qubits?" from a complexity theoretic point of view.

Aaronson [22] noted that quantum advice is quite closely related to quantum one-way communication,[3] as one can think of an advice state as a one-way message sent to an algorithm by a benevolent 'advisor'. The class of decision problems which can be efficiently solved on a quantum computer with access to a quantum advice state is denoted **BQP/qpoly**, and Aaronson [22] showed that $\mathbf{BQP/qpoly} \subseteq \mathbf{PP/poly}$. Based on the relation between quantum advice and quantum one-way communication, the size of the class **BQP/qpoly** can, in some sense, be thought of as a measure of prowess in communication tasks, or, intuitively speaking, as a measure of how much 'useful' information can be stored in a quantum state.

If the computational power of a general theory can be considered a measure of the richness of its dynamics, then the increase in computational power when supplemented with advice can

[1]Other investigations linking physical principles to computation can be found in [16–19].

[2]Quantum computation with advice was first defined and studied in [21].

[3]Quantum one-way communication can be described as follows: Alice has an n-bit string x, Bob has an m-bit string y and together they wish to evaluate $f(x,y)$, where $f:\{0,1\}^n \times \{0,1\}^m \to \{0,1\}$ is a Boolean function. After examining her input $x = x_1 \cdots x_n$, Alice can send a single quantum message ρ_x to Bob, whereupon Bob, after examining his input $y = y_1 \cdots y_m$, can choose some basis in which to measure ρ_x. He must then output a claimed value for $f(x,y)$. We are interested in how long Alice's message needs to be, for Bob to succeed with high probability on any x,y pair.

be thought of—à la Aaronson in the previous discussion—as a measure of the information that can be stored in its states. In §3, we provide rigorous definitions of the class of decision problems that can be solved in a specific operational theory when provided with a trusted advice state from that theory—which we call **BGP/gpoly** for a particular theory **G**. We show that in the theory colloquially known as 'Boxworld', which has the strongest correlations consistent with the no-signalling principle and was first discussed by Popescu & Rohrlich in [23,24], the class **BGP/gpoly** contains all decision problems and so is optimally powerful. Despite this, §4 shows that theories with a certain amount of non-trivial dynamics satisfy the same upper bound on the power of computation with advice as was discussed in the previous paragraph for quantum theory. In particular, for theories **G** with non-trivial dynamics we show that **BGP/gpoly** $\subseteq$ **PP/poly**. Boxword has no non-trivial reversible dynamics and it was shown by van Dam [25] that communication complexity tasks in Boxworld can be solved trivially. Our result shows that when a theory *has* non-trivial reversible dynamics there is a limit on its prowess in certain communication tasks—as quantified by the size of the class **BGP/gpoly**.

A key point in the above discussion is that one trusts the advice provider. That is, one trusts that the received advice contains the information the provider claims it does. In reality the provider could be malevolent and out to deceive the receiver. If one cannot trust the provider, a computer must be used to check—or *verify*—that the provided advice is correct and this verification process requires non-trivial dynamics to implement. Thus, by learning how computational complexity changes as the amount of trust we have in the provider is varied, we enter into a regime where both prowess in communication tasks and computational power—corresponding to the existence of non-trivial dynamics—are simultaneously tested.

Within theoretical computer science, untrusted advice has been formally referred to as *proofs* and has a long history within computational complexity. For example, the famous class **NP** can be described as a proof system between an efficient, deterministic, classical computer (verifier) and an all-powerful prover where the prover gives polynomially sized proofs to the verifier. Here the verifier wishes to check if this proof is the correct solution to a particular problem. In quantum computing, the corresponding complexity class to **NP** is denoted **QMA**, for quantum Merlin-Arthur. The question of what useful problems a quantum computer can solve when given a non-uniform quantum state as a proof from an untrusted source has led to surprising and beautiful connections between quantum computation and condensed matter physics [26].

In §3, we give a rigorous definition of the class of problems for which a verifier with an efficient computer from a specific theory can solve when given proof states from that theory—which we call **GMA** for a particular theory **G**. We show, in §5, that there exists a universal upper bound on **GMA** for all causal and tomographically local theories. In particular, we show that **GMA** $\subseteq$ **PP**, for all **G** satisfying tomographic locality and causality. Note that Boxworld is an example of such a theory. Some results concerning the connection between trusted advice and proof verification in general theories are given in §6.

2. The framework

(a) Operational theories

We work in the circuit framework for generalized probabilistic theories developed by Hardy [27] and Chiribella *et al.* [28,29]. The presentation here is most similar to that of Chiribella *et al*. We now provide a brief review of this framework; see [16] for a more in-depth review and an extended discussion of computation in general theories.

A theory within this framework specifies a set of laboratory devices that can be connected together in different ways to form experiments and assigns probabilities to different experimental outcomes. A laboratory device comes equipped with input ports, output ports and a classical pointer. When a device is used in an experiment, the pointer comes to rest in one of a number of positions, indicating some outcome has occurred. One can intuitively think of *physical systems* as passing between the input and output ports of the laboratory devices and these physical systems

come in different types, denoted by labels $A, B, C, \ldots$. In an experiment, these devices can be composed both sequentially and in parallel, and, when composed sequentially, types must match: the output system of the first device must be of the same type as the corresponding input system of the second.

In a general theory, one can depict the connections of devices in some experimental set-up by closed circuits. A fundamental requirement on any physical theory is that it should be able to give probabilistic predictions about the occurrence of possible outcomes. It is thus demanded that, in this framework, closed circuits define probability distributions. Given this structure, one then says that two physical devices are equivalent (from the point of view of the theory) if replacing one by the other in any closed circuit does not change the probabilities. The set of equivalence classes of devices with no input ports are referred to as *states*, devices with no output ports as *effects* and devices with both input and output ports as *transformations*.

The 'Dirac-like' notation $|s_r)_A$ is used to represent a state of system type A, where r is the outcome of the classical pointer, and ${}_A(e_r|$ to represent an effect on system type A, so that, if the effect ${}_A(e_{r_2}|$ is applied to the state $|s_{r_1})_A$, the probability of obtaining outcome r_1 on the physical device representing the state and outcome r_2 on the physical device representing the effect is

$$ {}_A(e_{r_2}|s_{r_1})_A := P(r_1, r_2). $$

The fact that closed circuits correspond to probabilities can be leveraged to show that each of the set of states, effects and transformations gives rise to a vector space and that the transformations and effects act linearly on the vector space of states. We assume in this work that all vector spaces are finite dimensional.

We can now formally define some examples of physical principles. We will first discuss the principles of *causality* and *tomographic locality* which were briefly mentioned in the introduction section.

Definition 2.1 (causality [28]). A theory is said to be *causal* if the marginal probability of preparing a state is independent of the choice of which measurement follows the preparation.

More formally, if $\{|s_i)\}_{i\in X}$ are the states corresponding to the preparation, consider the probability of outcome i, given that a subsequent measurement $\mathcal{M}$ corresponds to a set of effects $\{(e_j|\}_{j\in Y}$,

$$ P(i|\mathcal{M}) := \sum_{j\in Y} (e_j|s_i). $$

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

$$ P(i|\mathcal{M}) = P(i|\mathcal{N}). $$

One can think[4] of the causality principle as intuitively capturing the notion of *no signalling from the future*. It was shown in [28] that a 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 $\{(e_j|\}_{j\in Y}$ satisfies $\sum_j (e_j| = (u|$. A state $|s)$ is *normalized* if and only if $(u|s) = 1$. It can be shown that, without loss of generality, every state in a causal theory can be taken to be normalized [28].

Definition 2.2 (tomographic locality [27,28,30]). A theory satisfies tomographic locality if every transformation can be uniquely characterized by local process tomography.

That is, in a tomographically local theory, if two transformations with matching input and output ports give the same probabilities for all product state inputs and product effect measurements, then the transformations must be equivalent. Tomographic locality implies that the matrix corresponding to a composite transformation is just the vector space tensor product of the matrices of each individual transformation in the composite.

[4]Provided one thinks of circuits as having a temporal order, with tests later in the sequence occurring at a later time than tests earlier in the sequence.

We will now define strong symmetry, a principle, which if satisfied, guarantees the existence of a certain type of non-trivial dynamics. Before we define this principle, the following concepts must be introduced. We say the laboratory device $\{\mathcal{U}_j\}_{j\in Y}$, where j indexes the positions of the classical pointer, is a *coarse-graining* of the device $\{\mathcal{E}_i\}_{i\in X}$ if there is a disjoint partition $\{X_j\}_{j\in Y}$ of X such that $\mathcal{U}_j = \sum_{i\in X_j} \mathcal{E}_i$. That is, coarse-graining arises when some outcomes of a laboratory device are joined together. The device $\{\mathcal{E}_i\}_{i\in X}$ is said to *refine* the device $\{\mathcal{U}_j\}_{j\in Y}$. A state is *pure* if it does not arise as a *coarse-graining* of other states; a pure state is one for which we have maximal information. A state is *mixed* if it is not pure and it is *completely mixed* if any other state refines it. That is, $|c)$ is completely mixed if, for any other state $|\rho)$, there exists a non-zero probability p such that $p|\rho)$ refines $|c)$. States $\{|\sigma_i)\}_{i=1}^N$ are *perfectly distinguishable* if there exists a measurement corresponding to effects $\{(e_i|\}_{i=1}^N$, such that $(e_i|\sigma_j) = \delta_{ij}$ for all i, j.

Definition 2.3 (strong symmetry [31]). A theory satisfies *strong symmetry* if for any two n-tuples of pure and perfectly distinguishable states $\{|\rho_1), \ldots, |\rho_n)\}$ and $\{|\sigma_1), \ldots, |\sigma_n)\}$ there exists a reversible transformation T such that $T|\rho_i) = |\sigma_i)$, for $i = 1, \ldots, n$.

In §4, we will mainly be concerned with two special cases of the above principle:

(i) *Permutability*. A general theory satisfies *permutability* if for any n-tuple of pure and perfectly distinguishable states and any permutation π of this n-tuple

$$\{|\rho_1), \ldots, |\rho_n)\} \quad \text{and} \quad \{|\rho_{\pi(1)}), \ldots, |\rho_{\pi(n)})\}$$

there exists a reversible transformation T such that $T|\rho_i) = |\rho_{\pi(i)})$, for $i = 1, \ldots, n$.

(ii) *Bit-symmetry*. A theory satisfies *bit-symmetry* if for any two 2-tuples of pure and perfectly distinguishable states $\{|\rho_1), |\rho_2)\}$, $\{|\sigma_1), |\sigma_2)\}$ there exists a reversible transformation T such that $T|\rho_i) = |\sigma_i)$, for $i = 1, 2$.

Permutability is the special case of definition 2.3 where one of the sets of pure and perfectly distinguishable states is a permutation of the other. Bit-symmetry is the $n = 2$ case of definition 2.3.

Note that causality, tomographic locality and strong symmetry are all logically independent: generalized probabilistic theories satisfying any subset (including the empty subset) can be defined. For example, standard quantum theory satisfies all three, quantum theory with real amplitudes satisfies causality and strong symmetry but not tomographic locality, Boxworld satisfies causality and tomographic locality but not strong symmetry and the theory constructed in [32] does not satisfy causality.

(b) Efficient computation

To define the class of efficient computation in a general theory, we must first define the notions of a uniform circuit family and an acceptance condition for an arbitrary theory. The notion of a poly-size uniform circuit family $\{C_x\}$, which is indexed by some bit string x, can be defined as follows:

(i) The number of gates in the circuit C_x is bounded by a polynomial in $|x|$.
(ii) There is a finite[5] gate set $\mathcal{G}$, such that each circuit in the family is built from elements of $\mathcal{G}$.
(iii) 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 $\tilde{M}$ with rational entries, such that $|(M - \tilde{M})_{ij}| \leq \epsilon$, in polynomial time in $\log(1/\epsilon)$.
(iv) There is a Turing machine that, acting on input $x = x_1 x_2 \cdots x_n$, outputs a classical description of C_x in time bounded by a polynomial in $|x|$.

[5]For a uniformity condition, where the size of the gate grow with the circuit's size; see [33].

At the end of each run of the computation, each gate in the circuit has a classical outcome—corresponding to the final position of the classical pointer—associated with it, and the theory defines a joint probability for these outcomes. Denoting the string of observed outcomes by z, we define the final output of the computation to be given by a function $a(z) \in \{0, 1\}$, where there must exist a Turing machine that computes a in polynomial time in the length of the input $|x|$. We say the computation accepts an input string x if $a(z) = 0$, where z is an outcome string of the circuit C_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.

The class of problems that can be solved efficiently in a generalized probabilistic theory can now be defined.

Definition 2.4. For a generalized 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

(i) $x \in \mathcal{L}$ is accepted with probability at least $\frac{2}{3}$;
(ii) $x \notin \mathcal{L}$ is accepted with probability at most $\frac{1}{3}$.

The choice of constants $\left(\frac{2}{3}, \frac{1}{3}\right)$ is arbitrary as long as they are bounded away from $\frac{1}{2}$ by some constant.[6] For a discussion of this and the fact that the acceptance probability can be amplified as in the usual quantum case, see [16, p. 9]. Given these definitions, the following theorem was proved in [16].

Theorem 2.5. *For any generalized probabilistic theory* **G** *satisfying tomographic locality, we have*

$$\mathbf{BGP} \subseteq \mathbf{AWPP} \subseteq \mathbf{PP} \subseteq \mathbf{PSPACE}.$$

It is worth noting that, due to the computation of the acceptance of an input x, we are given polynomial deterministic classical computation 'for free'. As a result, the lower bound of $\mathbf{P} \subseteq \mathbf{BGP}$ is satisfied for all theories **G**.

One can define a notion of generalized circuits with the ability to *post-select* on most exponentially unlikely circuit outcomes. These are poly-sized uniform circuits in a general theory, where the probability of acceptance is conditioned on the circuit outcome z lying in a (poly-time computable) subset of all possible values of z.

Definition 2.6. A language $\mathcal{L}$ is in the class **PostBGP** if there is a poly-sized uniform circuit family in that theory and an efficient acceptance condition, such that

(i) there exists a constant D and polynomial w such that $P(z \in S) \geq 1/D^{w(|x|)}$,
(ii) if $x \in \mathcal{L}$ then $P_x(\text{accept}|z \in S) \geq \frac{2}{3}$, and
(iii) if $x \notin \mathcal{L}$ then $P_x(\text{accept}|z \in S) \leq \frac{1}{3}$,

where z is the circuit outcome, S is a subset of all possible circuit outcomes and $z \in S$ can be checked by a Turing machine in polynomial time in $|x|$.

Aaronson showed in [34] that **PostBQP** = **PP** and the following theorem was shown in [16].

[6]This can be further relaxed to being bounded away from $\frac{1}{2}$ by an inverse polynomial in the size of the input. For simplicity, we just consider being bounded away from $\frac{1}{2}$ by a constant.

Theorem 2.7. *For any generalized probabilistic theory* **G** *satisfying tomographic locality, we have*

$$\mathbf{PostBGP} \subseteq \mathbf{PostBQP}.$$

3. Proofs and advice

In this section, we provide generalizations of the definitions of classical (quantum) computing with advice and a type of classical (quantum) interactive proof system of the framework of general operational theories. For an overview of the classical (quantum) definitions, see appendix A (appendix B), respectively. We will assume that the reader is familiar with the definition of the familiar complexity classes **P** and **NP** as well as the formalism of quantum circuits. As with the definition of **BGP**, unless otherwise stated, the constants $\left(\frac{2}{3}, \frac{1}{3}\right)$ can be chosen arbitrarily as long as they are bounded away from $\frac{1}{2}$ by some constant.

(a) Definitions for general theories

Circuits from a uniform circuit family $\{C_x\}$ in some general theory are indexed by the string x that encodes the decision problem which the theory is attempting to solve. In defining the class of efficient computation in a theory, the family $\{C_x\}$ is taken to consist of closed circuits from that theory. This will not be the case when advice and proofs are involved; in this paradigm, one is given both the problem instance x and a proof or advice state, so the constructed circuit C_x must have open system ports into which this state can be plugged. Henceforth, we will assume that uniform circuit families consist of collections of circuits with a number of open input ports that can grow as a polynomial in $|x|$, which we call the *auxiliary register*. Note that the choice of finite gate set determines the possible system types of the auxiliary register. Given this convention, we can define efficient computation with trusted advice in a specific general theory.

Definition 3.1. For a general theory **G**, a language $\mathcal{L} \subseteq \{0,1\}^n$ is in the class **BGP/gpoly** if there exists a poly-sized uniform family of circuits $\{C_x\}$ in **G**, a set of (possibly non-uniform) states $\{\sigma_{|x|}\}_{n\geq 1}$ on a composite system of size $d(n)$ for some polynomial $d : \mathbb{N} \to \mathbb{N}$, and an efficient acceptance criterion, such that for all strings $x \in \{0,1\}^n$:

(i) if $x \in \mathcal{L}$ then C_x accepts with probability at least 2/3 given σ_n as input to the auxiliary register;
(ii) if $x \notin \mathcal{L}$ then C_x accepts with probability at most 1/3 given σ_n as input to the auxiliary register.

Here by 'composite system of size $d(n)$', we mean that the number of systems, or open ports, of the auxiliary register—into which the advice state is input—increases as $d(n)$, for d a polynomial in the input size. Since, as mentioned, there is an efficient, deterministic classical computer deciding acceptance and each state σ_n has a classical pointer associated with it, classical advice can always be encoded into these pointers (of which there can be polynomially many). Therefore, we can always give the lower bound $\mathbf{P/poly} \subseteq \mathbf{BGP/poly} \subseteq \mathbf{BGP/gpoly}$, where the suffix **/poly** denotes classical advice.

Definition 3.2. For a general theory **G**, a language $\mathcal{L} \subseteq \{0,1\}^n$ is in the class **GMA** if there exists a poly-sized uniform family of circuits $\{C_x\}$ in **G**, a polynomial $d : \mathbb{N} \to \mathbb{N}$ and an efficient acceptance criterion, such that for all strings $x \in \{0,1\}^n$:

(i) if $x \in \mathcal{L}$ then there exists a (possibly non-uniform) proof state σ on a composite system of size $d(n)$ such that C_x accepts with probability at least 2/3 given σ as input to the auxiliary register;
(ii) if $x \notin \mathcal{L}$ then C_x accepts with probability at most 1/3 given σ as input to the auxiliary system, for all states σ.

We refer the reader to appendix A for the definitions of computation with advice and proofs in the case of classical and quantum theory as we will make reference to these complexity classes throughout the paper. Informally, for the specific case of quantum theory, the **G** in the nomenclature should be replaced with **Q** and /**gpoly** is replaced with /**qpoly**.

The existential quantifiers in the above definition of **GMA** rigorously capture the notion of a circuit having to 'verify' the proof. Note also that advice states can only depend on the size of the input, whereas proofs can, in general, be dependent on the inputs themselves. The amplification procedure of Kitaev & Watrous [35] that achieves exponential separation for the acceptance and rejection probabilities in **QMA**, at the expense of a polynomial increase in the size of the witness state, can be adapted in a straightforward fashion to provide a similar amplification procedure for **GMA**, for arbitrary **G**. Note that **BGP** $\subseteq$ **GMA** follows straightforwardly from the definitions. Also, via the same arguments given to lower bound the class **BGP**/**gpoly**, we can always give the lower bound **NP** $\subseteq$ **GMA**.

It was proved in [35] that **QMA** $\subseteq$ **PP**, and this was improved in [36] to **QMA** $\subseteq$ $\mathbf{A_0PP}$ (see also [37]). Aaronson & Drucker [38] have shown the following remarkable relation between these two classes:

$$\mathbf{BQP/qpoly} \subseteq \mathbf{QMA/poly}.$$

This says that one can always replace (poly-size) quantum advice by (poly-size) classical advice, together with a (poly-size) quantum proof.[7] Intuitively, this relation can be summed up as follows: one can always simulate an arbitrary quantum state ρ on all small circuits, using a different state $\tilde{\rho}$ that is easy to recognize.[8] In §5, we investigate whether this relation holds for general operational theories.

(b) Example: Boxworld

We now look at Boxworld with respect to our definitions of proofs and advice in general physical theories. Towards the end, we provide a brief definition of Boxworld (see, for example, [39] for a more in-depth discussion). For a given single system A in Boxworld, there are two choices of binary-outcome measurements, $\{_A(x_a|\}$ for $x, a \in \{0, 1\}$. Here x is the bit denoting the two possible choices of measurement and a is the bit denoting the two possible outcomes of the chosen measurement, i.e the two measurements on system A are $\{_A(0_0|, _A(0_1|\}$ and $\{_A(1_0|, _A(1_1|\}$. States and measurements in this theory can produce correlations associated with the so-called Popescu–Rohrlich non-local box [23]. That is, for a bipartite system AB, there exist states $|\rho_{PR})_{AB}$ such that

$$(x_a|(y_b|\rho_{PR})_{AB} = \begin{cases} \frac{1}{2}, & \text{if } a \oplus b = xy, \\ 0, & \text{otherwise,} \end{cases}$$

where $\oplus$ represents addition modulo 2. These correlations can be extended to an n-partite system where now, for the jth party, $x_j \in \{0, 1\}$ and $a_j \in \{0, 1\}$ are the choice of measurement and its outcome, respectively. There exists a state $|\rho_f)$ and effects $\{_j(x_j, a_j|\}$ for all j parties that produce the probabilities [8,40]

$$(x_1, a_1|(x_2, a_2| \cdots (x_n, a_n|\rho_f) = \begin{cases} \dfrac{1}{2^{n-1}}, & \text{if } \bigoplus_{j=1}^{n} a_j = f(x), \\ 0, & \text{otherwise,} \end{cases}$$

where $\bigoplus$ represents summation modulo 2 and $f : \{0, 1\}^n \to \{0, 1\}$ is any Boolean function from the bit-string x with elements x_j. Therefore, if the state $|\rho_f)$ is prepared and local measurements described by effects $(x_j, a_j|$ made, a classical computer can compute the parity of all outcomes a_j and so we deterministically obtain the evaluation of Boolean function $f(x)$. This relatively straightforward observation gives us the following result.

[7]Note that advice can encode solutions to even undecidable problems; any upper bound on an advice class will be another advice class.

[8]One can even take $\tilde{\rho}$ to be the ground state of a local Hamiltonian [38].

Theorem 3.3. *There exists a generalized probabilistic theory* **G** *satisfying causality and tomographic locality, which satisfies* **BGP/gpoly** = **ALL**, *where* **ALL** *is the class of all decision problems.*

Proof. Clearly, **BGP/gpoly** ⊆ **ALL** is trivially true for Boxworld. The states $|\rho_f)$ can be used as advice states and, as all decision problems can be represented by Boolean functions, it follows that **ALL** ⊆ **BGP/gpoly**. ■

Note that the above proof still goes through if we insist that Boxworld only has reversible dynamics as the proof only requires the ability to prepare and measure states. If one considers the class **GMA** for Boxworld with only reversible transformations then we have **GMA** ⊆ **MA**, as all reversible dynamics are trivial in this theory and can thus be simulated classically [41–43]. By trivial, we mean that the circuits in Boxworld only consist of making the local 'fiducial' measurements $\{_j(x_j, a_j|\}$ on a state and performing classical post-processing on the outcomes. This process can be simulated by the prover giving the verifier the classical string of measurement outcomes similar to the approach of lemma 2 in [44]. That is, while poly-size advice states in Boxworld can encode any Boolean function, the theory has no non-trivial dynamics to efficiently verify that this function is encoded in the state if the prover cannot be trusted.

4. Consequences of non-trivial dynamics for computation

In §4a, we show that the existence of non-trivial dynamics implies that computation in that theory is at least as powerful as probabilistic classical computation: **BPP** ⊆ **BGP**. Hence non-trivial dynamics implies non-trivial computational power. Furthermore, in §4b, we show that the existence of non-trivial dynamics implies a bound on the amount of 'useful' information—quantified by the size of the class **BGP/gpoly**—that can be stored in general states.

(a) Powerful computation from non-trivial dynamics

Definition 4.1. A theory is said to be *non-classical* if, for at least one n-tuple of pure and perfectly distinguishable states $\{|\sigma_i)\}_{i=1}^N$, there exists a pure state $|y)$ such that $(e_i|y) = p_i$ for $0 < p_i < 1$, for all i, where $\{(e_i|\}_{i=1}^N$ is the measurement that distinguishes the $\{|\sigma_i)\}_{i=1}^N$.

Before we present our result, we emphasize that the result is highlighting the *intrinsic* computational power in a theory. As previously mentioned, in our framework we already have a classical computer that processes experimental data and, if a circuit in a theory **G** can produce random numbers, we can easily achieve the complexity class **BPP**. By talking about intrinsic computational power, we imagine reducing the power of our classical computer to perform extremely simple, non-universal classical computation. For example, the classical computer in deciding the output of the computation could only output the classical counter-value on one of the measurements. Our result then shows that theories with a certain amount of non-trivial dynamics still decide any problem in **BPP**.

Theorem 4.2. *Let* **G** *be a causal, non-classical theory with at least two pure and distinguishable states that satisfies permutability. Then* **BPP** ⊆ **BGP**.

Proof. For **BPP** ⊆ **BGP**, it is sufficient to show two things: that transformations of the general theory can simulate the action of any reversible Boolean function $f : \{0,1\}^n \to \{0,1\}^n$, and that it is possible to prepare a source of random bits. First, bit strings $x = x_1 \cdots x_n$ can be represented by perfectly distinguishable pure states $|x) = |x_1) \otimes \cdots \otimes |x_n)$. Then, the first condition follows from permutability: since $\{|f(0\cdots0)), \ldots, |f(1\cdots1))\}$ is a permutation of the tuple of pure and perfectly distinguishable states $\{|0\cdots0), \ldots, |1\cdots1)\}$, there must exist a reversible transformation T_f such that $T_f|x) = |f(x))$.

For the second condition, it suffices if there are circuits that can generate random bits. Consider the two pure and perfectly distinguishable states $|0)$ and $|1)$. Let $\{(e_0|, (e_1|\}$ be a measurement that distinguishes them, that is, $(e_i|j) = \delta_{ij}$, for $i, j = 0, 1$. Non-classicality implies that there exists some

pure state $|y) \notin \{|0), |1)\}$ such that $(e_0|y) = p$ and $(e_1|y) = 1 - p$, with $0 < p < 1$. Probabilities of $\frac{1}{2}$ can be generated by preparing two copies of $|y)$, implementing the measurement on each in parallel and assigning a value $y = 0$ or 1 to the outcomes 01 and 10, respectively.[9] ■

(b) Bounds on computation with advice in physical theories

Recall that a state is *mixed* if it is not pure and it is *completely mixed* if any other state refines it. That is, $|c)$ is completely mixed if, for any other state $|\rho)$, there exists a non-zero probability p such that $p|\rho)$ refines $|c)$. Intuitively, one should be able to efficiently prepare a completely mixed state on a computer in any general theory. This follows because the completely mixed state can be prepared by performing any uniform state preparation and 'forgetting' the outcome. Henceforth, we shall assume that the completely mixed state—if it exists—is uniform.

Recall the definition of a bit-symmetry from §2. In any bit-symmetric theory with at least two pure and distinguishable states, it can be shown [45] that the group of reversible transformations acts *transitively* on the set of pure states. That is, given any two pure states $|\rho), |\sigma)$, there exists a reversible transformation T such that $T|\rho) = |\sigma)$. This fact can be used [28,29] to prove the existence of a completely mixed state as the unique state—for a given system type—that is invariant under all reversible transformations.

Bit-symmetry is a powerful principle and has many useful consequences. Two more of which are:

(i) Every bit-symmetric theory is *self-dual* [45]. That is, to every pure state $|\rho_e)$ there is associated a unique pure effect $(e_\rho|$, and vice versa.[10] This association is achieved via an inner product $[.,.]$, on the real vector space V generated by the set of states, as: $(e_\rho|\sigma) = [|\rho_e), |\sigma)]$, for all states $|\sigma)$. Note that $[|\rho), |\rho)] = 1$ for all pure states $|\rho)$.

(ii) Let $\||v)\|_{\text{phy}} = 2\max_{(e|} |(e|v)|$ and $\|v\|_E = \sqrt{[v, v]}$, for v an arbitrary vector in V. The norm $\||\rho) - |\sigma)\|_{\text{phy}}$ has a natural operational interpretation as the distinguishability of $|\rho)$ and $|\sigma)$. Bit-symmetry implies [46] that $\||\rho) - |\sigma)\|_{\text{phy}} \leq c\||\rho) - |\sigma)\|_E$, where $c = \||c)\|_E$ for $|c)$ the completely mixed state.

Using the above facts, we now prove a version of the 'as good as new lemma'[11]—discussed in the quantum case in [22]—for all bit-symmetric theories. Before we state this lemma, we need to briefly introduce a notion of post-measurement state update rule for bit-symmetric theories. In this work, applying a measurement to a state corresponds to a closed circuit—that is, a probability. However, to discuss post-measurement states, this must be generalized slightly. A measurement will henceforth correspond to a laboratory device from some input state to the output post-measurement state, where the classical pointer denotes the outcome of the measurement. Consider the measurement $\{(i|\}$, consisting of pure effects $(i|$, and apply it to some state $|\rho)$. On observing outcome i, the state $|\rho)$ is updated to $|\rho_i)/(u|\rho_i)$, where $|\rho_i)$ is the unique pure state associated with $(i|$. This state update rule satisfies a natural *repeatability* condition: any state yielding outcome i with unit probability is left invariant by the update rule, thus repeated measurements always yield the same result. See Pfister & Wehner [49] for a more in-depth discussion of state update rules in general theories.

Lemma 4.3. *Given a two-outcome measurement, consisting of the pure effects* $\{(0|, (1|\}$ *and a state* $|\rho)$ *such that* $(0|\rho) = 1 - \epsilon$, *for* $\epsilon \geq 0$, *the post-measurement state on observing outcome* 0 *satisfies*

$$\||\rho) - |\rho_0)\|_{\text{phy}} \leq c\sqrt{2\epsilon},$$

[9]This argument is based on von Neumann's argument for turning two copies of a biased coin into one copy of an unbiased coin.

[10]The proof of this fact requires two further technical assumptions, both implicit in [45]. These are that the group of reversible transformations must be compact and every mathematically allowed effect is physical.

[11]Also called the 'gentle measurement lemma', which was independently proved by Winter [47] and improved by Ogawa & Nagaoka [48].

where $c = \||c)\|_E$ is the completely mixed state, in all bit-symmetric theories.

Proof. Recall in a bit-symmetric theory that $\|\|\rho) - |\sigma)\|_{\text{phy}} \leq c\|\|\rho) - |\sigma)\|_E$. We thus have

$$\begin{aligned}\|\|\rho) - |\rho_0)\|_{\text{phy}} &\leq c\|\|\rho) - |\rho_0)\|_E \\ &= c\sqrt{[|\rho) - |\rho_0), |\rho) - |\rho_0)]} \\ &\leq c\sqrt{2 - [|\rho), |\rho_0)] - [|\rho_0), |\rho)]} \\ &= c\sqrt{2 - 2[|\rho_0), |\rho)]} \\ &= c\sqrt{2 - 2(0|\rho)} = c\sqrt{2\epsilon}.\end{aligned}$$

The first line follows from the definition of $\|.\|_E$, the second from the fact that $\||\sigma)\|_E \leq 1$ for all $|\sigma)$, the third from the symmetry of the inner product $[.,.]$ and the last from the definition of self-duality. ∎

The above lemma states that if one outcome of a two-outcome measurement occurs with high probability on some state, then the post-measurement state after getting that outcome is 'close' to the original state. We are now in a position to state the main result of this section. Before we do, let us fix the accepting criterion for computation with advice and make the simplifying assumption that the accept/reject measurement consists of pure effects.

Theorem 4.4. *Any causal, bit-symmetric, tomographically local theory* **G** *with at least two pure and distinguishable states satisfies*

$$\textbf{BGP/gpoly} \subseteq \textbf{PostBGP/poly} \subseteq \textbf{PP/poly}.$$

The above theorem states that, in theories with non-trivial dynamics, there is a bound to how much useful information one can extract from any state. This result provides evidence for the existence of a trade-off between states and dynamics and can be seen as a natural converse to the results in [41–43]. Our proof is a slight variation of the original proof in the quantum case, due to Aaronson [22].

Proof. Begin by amplifying the success probability of **BGP/gpoly** on input x from $\frac{2}{3}$ to $1 - \frac{1}{2}^{q(|x|)}$. This is achieved by running a polynomial number of copies of the circuit C_x, in parallel, and taking the majority answer. Note that in this amplification scheme the total advice state is the (vector space) tensor product of advice states for each individual circuit. Recall that the completely mixed state $|c)$ is assumed to satisfy uniformity and that there exists a non-zero probability p such that $p|\sigma)$ is a refinement of $|c)$, for any $|\sigma)$. Uniformity implies that p can be well approximated by some rational $c/d^{w(|x|)}$, for c an integer and d a polynomial in the size of the input x (see the proof of theorem 14 in appendix B of [16] for a more in-depth discussion of uniformity).

Given any language $\mathcal{L} \in \textbf{BGP/gpoly}$ we now construct a **PostBGP/poly** algorithm that decides $\mathcal{L}$. Given some x, use the completely mixed state as the advice to the circuit C_x. Now, from the definition of **BGP/gpoly**, if $|c)$ cannot be used as advice to determine $x \in \mathcal{L}$, the circuit accepts with probability less than $\frac{1}{3}$. Consider the post-measurement state $|c')$ of the auxiliary register after running C_x with advice $|c)$ *post-selecting* on the event that we succeeded in outputting the correct answer. If $|c')$ cannot be used as advice for all inputs, there exists some x' such that $C_{x'}$ succeeds with probability less than $\frac{1}{3}$. As before, consider the post-measurement state of the auxiliary register after running $C_{x'}$ with advice $|c')$ post-selecting on outputting the correct answer. Continue in this fashion for some $t(|x|)$ stages, t a polynomial. Successful post-selection is guaranteed as the actual advice state refines $|c)$ with probability $c/d^{w(|x|)}$.

If, at any iteration of this process, we cannot find an x to move forward, we must be holding a state that works as advice for every input, and we can use it to run C_z on any input z, succeeding with high probability. Thus, if the process halts after a polynomial number of iterations, we are done.

If the correct advice state $|\sigma\rangle$ had been used in the computation, lemma 4.3 would imply the post-measurement state on observing that the accept outcome, $|\sigma_{\text{acc}}\rangle$, would—under the simplifying assumption that the accept/reject measurement consists of pure effects—satisfy

$$\| |\sigma\rangle - |\sigma_{\text{acc}}\rangle \|_{\text{phy}} \leq c\sqrt{\frac{1}{2^{q(|x|)-1}}}.$$

As the completely mixed state $|c\rangle$ is uniform, it follows that $c = \| |c\rangle \|_E \leq O(2^{m(|x|)})$ for m a polynomial. Therefore, $c/\sqrt{2^{q(|x|)-1}} = o(1)$. We thus have

$$\| |\sigma\rangle - |\sigma_{\text{acc}}\rangle \|_{\text{phy}} \leq o(1).$$

Therefore, on each iteration of the above process, the correct answer is output with probability

$$\frac{c}{d^{w(|x|)}}(1 - o(1)).$$

This process has been designed so that the probability that $|c\rangle$ can be re-used on each iteration and succeed at each stage is at most $\frac{1}{3}^{t(|x|)}$. Therefore, we have that

$$\frac{c}{d^{w(|x|)}}(1 - o(1)) \leq \frac{1}{3}^{t(|x|)}.$$

Thus, $t(|x|) \leq O(w(|x|))$ and we are done.

There thus exists a polynomial number of $x_1, \ldots, x_t$ such that, if $|a\rangle$ is the post-measurement state after we start with $|c\rangle$ and post-select on succeeding on each x_i in turn, $|a\rangle$ is a good advice state for every string z. Provide the algorithm with this sequence of classical strings, along with the correct outcomes $b, \ldots, b_t$ for each of them. The algorithm then prepares $|c\rangle$, uses it as advice and post-selects on getting outcomes $b, \ldots, b_t$. After this process, we obtain the state $|a\rangle$ and so all languages that can be decided in **BGP**/**gpoly** can also be decided in **PostBGP**/**poly** and thus, by tomographic locality and theorem 2.7, in **PP**/**poly**. ■

5. Bounds on the power of proofs in physical theories

In this section, we will put a non-trivial bound on **GMA**. To state our result, the notion of a **GapP** function must be introduced. Given a poly-time non-deterministic Turing machine n and input string x, let $N_{\text{acc}}(x)$ be the number of accepting computation paths of N given input x, and $N_{rej}(x)$ the number of rejecting computation paths of N given x. A function $f : \{0,1\}^* \to \mathbb{Z}$ is a **GapP** function if there exists a polynomial-time non-deterministic Turing machine N such that $f(x) = N_{\text{acc}}(x) - N_{rej}(x)$, for all input strings x. We can now define the class $\mathbf{A_0PP}$.

Definition 5.1. A language $\mathcal{L}$ is in the class $\mathbf{A_0PP}$ if and only if there exists a **GapP** function f and an efficiently computable function T such that

(i) for all $x \in \mathcal{L}$ $f(x) \geq T(x)$ and
(ii) for all $x \notin \mathcal{L}$ we have $0 \leq f(x) \leq \frac{1}{2}T(x)$.

It has been shown that the above class is contained in **PP**. Fix the efficient acceptance condition for proof verification so that, in all uniform circuits, the measurement applied at the end of the computation to the auxiliary register consists of only unit effects. We make this choice to move closer to the standard quantum acceptance condition. We also make the simplifying assumption—routinely made in the literature—that all mathematically allowed states are physically allowed. That is, all vectors whose inner product with any effect is in $[0, 1]$ correspond to physical states.

Theorem 5.2. *For any generalized probabilistic theory* **G** *satisfying causality, tomographic locality and the assumption that all mathematically allowed states are physically allowed, we have that*

$$\mathbf{GMA} \subseteq \mathbf{A_0PP} \subseteq \mathbf{PP}.$$

Proof. Recall that any matrix M has a singular value decomposition given by $M = UDV^{\mathrm{T}}$, where U, V are unitary (orthogonal if the matrix is real) matrices, V^{T} is the transpose of V and D is a diagonal matrix. The diagonal entries of D are all non-negative real numbers and are called the *singular values* of the matrix M. Note that the eigenvalues of the matrix $M^{\mathrm{T}}M = VD^{\mathrm{T}}DV^{\mathrm{T}}$ are the squares of the singular values of M.

Let M_x be the matrix representation of the uniform circuit, including states and effects on the non-auxiliary register, on input x, $(u|$ be the (tensor product) of unit effects applied on the auxiliary register and $|\rho)$ be any arbitrary state (which can be non-uniform) input to the auxiliary register. Without loss of generality, one can pad this matrix (and row and column vector) with rows and columns of zeros to ensure it is square. The probability that the circuit accepts the string x is given by $(u|M_x|\rho)$. It will now be shown that this probability is upper bounded by the largest singular value of the matrix M_x. Consider the following:

$$(u|M_x|\rho) = (u|UDV^{\mathrm{T}}|\rho) \leq \sigma_{\max}(u|UV^{\mathrm{T}}|\rho),$$

where $\sigma_{\max}$ is the largest singular value of M_x. Now UV^{T} is a unitary matrix and so can be decomposed as follows: $UV^{\mathrm{T}} = WD'W^{\mathrm{T}}$, where W is another unitary matrix and D' is a diagonal matrix consisting of the eigenvalues of UV^{T}; recall that these eigenvalues all have absolute value 1. Thus,

$$(u|M_x|\rho) \leq \sigma_{\max}(u|WD'W^{\mathrm{T}}|\rho) \leq \sigma_{\max}(u|\rho) \leq \sigma_{\max},$$

where the second inequality follows from the fact that the entries of D' have absolute value 1 and that W is unitary and the third inequality follows as $(u|\rho) \leq 1$.

Now as the squares of the singular values are the eigenvalues of the (positive definite) matrix $M_x^{\mathrm{T}}M_x$, we have that

$$(\sigma_{\max}^2)^d \leq \mathrm{Tr}((M_x^{\mathrm{T}}M_x)^d) \leq 2^n(\sigma_{\max}^2)^d,$$

where 2^n is the number of entries on the diagonal of $M_x^{\mathrm{T}}M_x$, n is a polynomial in $|x|$ and d is an arbitrary natural number. Let d be a polynomial in $|x|$ that takes values in the natural numbers and assume without loss of generality that it grows faster than the polynomial n; we will need this requirement later.

The matrix M_x satisfies the uniformity condition, and it was shown in [16] that the entries of all such matrices are **GapP** functions. By the closure properties of **GapP** (again see appendix B of [16]) functions the entries in the matrix $(M_x^{\mathrm{T}}M_x)^d$ are also **GapP** functions. Using an argument similar to that in [36], $\mathrm{Tr}((M_x^{\mathrm{T}}M_x)^d)$ can be straightforwardly shown to be a **GapP** function, denoted by $f(x)$. So, from the definition of **GMA**, we have that $f(x) \geq \sigma_{\max}^{2d} \geq \left(\frac{2}{3}\right)^{2d}$, for all x in the language.

Now the vector that achieves the bound of $\sigma_{\max}$ is the right singular vector of M_x with singular value $\sigma_{\max}$, which we denote by $|\sigma)$. If this vector is a physical state then we are done, as it follows from the definition of **GMA** and an argument similar to the one above that $f(x) \leq \frac{1}{2}\left(\frac{2}{3}\right)^{2d}$ for all x not in the language. If this vector is not a physical state then we have a bit more work to do.

Towards the end, consider the following. We are free to re-parametrize (e.g. [50, p. 7]) the set of states by an affine transformation $\phi : \mathbb{R}^m \to \mathbb{R}^m$, where $\mathbb{R}^m$ is the (smallest) real vector space that contains the set of states, as follows:

$$|\rho) \to |\tilde{\rho}) = \phi|\rho), \quad (a| \to (\tilde{a}| = (a|\phi^{-1}$$
$$\text{and} \quad M_x \to \tilde{M}_x = \phi M_x \phi^{-1},$$

as this does not change the probabilities, i.e. $(a|M_x|\rho) = (\tilde{a}|\tilde{M}_x|\tilde{\rho})$. Now, as an affine transformation is just a translation followed by a scaling, choose ϕ so that the Euclidean unit ball is contained in the re-parametrized state space (just translate the original state space and scale it appropriately to ensure this, noting that translations and scaling are reversible). As the singular vectors of every matrix are unit vectors, without loss of generality they are contained in this unit ball. The assumption that all mathematically allowed states are physically allowed ensures that these singular vectors are physical states. Thus $\sigma_{\max} = (\tilde{u}|\tilde{M}_x|\sigma)$, where $(\tilde{u}|$ is the unique deterministic

effect. The causality principle ensures that any state $|\tilde{s})$ can be scaled so that $(\tilde{u}|\tilde{s}) = 1$ (e.g. [28]). So for x not in the language we have $\sigma_{\max} = (\tilde{u}|\tilde{M}_x|\sigma) \leq \frac{1}{3}$.

It follows that

$$f(x) \leq 2^n \sigma_{\max}^{2d} \leq 2^n \left(\frac{1}{3}\right)^{2d} \leq \frac{1}{2}\left(\frac{2}{3}\right)^{2d},$$

where the first inequality follows from $\mathrm{Tr}((M_x^{\mathrm{T}} M_x)^d) \leq 2^n (\sigma_{\max}^2)^d$ and the last inequality follows from the fact that, for d increasing sufficiently faster than n, we have $2^{n+1} \leq 4^d$.

Thus, for a language $\mathcal{L}$ in **GMA** we have

(i) for all $x \in \mathcal{L}$ there exists a **GapP** function f such that $f(x) \geq (\frac{2}{3})^{2d}$ and
(ii) for all $x \notin \mathcal{L}$ we have $f(x) \leq \frac{1}{2}(\frac{2}{3})^{2d}$,

and so we have that **GMA** $\subseteq$ **A$_0$PP**. ■

6. Relating proofs and advice?

The relation **BQP/qpoly** $\subseteq$ **QMA/poly**, discussed in §3, captures an intriguing feature of proofs and advice in quantum theory: one can always replace quantum advice with classical advice together with a quantum proof. Here we study the relation

$$\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly}, \tag{6.1}$$

in general theories. Note that the relation is satisfied in classical computation:

$$\mathbf{BPP/rpoly} = \mathbf{P/poly} \subseteq \mathbf{NP/poly} \subseteq \mathbf{MA/poly},$$

where **BPP/rpoly** = **P/poly** was shown in [51,52]. Clearly, the relation in (6.1) is then not uniquely satisfied by quantum theory, but one could ask whether quantum theory is the most computationally powerful theory in which (6.1) is satisfied.

Using these observations as motivation, we obtain the following corollary of theorem 5.2.

Corollary 6.1. *There exist general theories* **G** *satisfying tomographic locality and causality such that* **BGP/gpoly** $\not\subseteq$ **GMA/poly**.

Proof. Firstly, we can use theorem 5.2 to conclude that **GMA/poly** $\subseteq$ **PP/poly** and by a counting argument **PP/poly** is strictly contained in **ALL**. From theorem 3.3, there exists a theory **G** such that **ALL** = **BGP/gpoly** and so we do not have **BGP/gpoly** $\subseteq$ **GMA/poly** $\subseteq$ **PP/poly** for this theory. ■

Motivated by the above corollary we can say something non-trivial about theories where **BGP/gpoly** $\not\subseteq$ **GMA/poly**. Consider the case of using a polynomially sized circuit from a specific theory, built from any fixed gate set in that theory, to prepare an arbitrary, but polynomially large, state in the theory. Given this set-up, we can prove the following result.

Theorem 6.2. *In any general theory* **G** *with*

$$\mathbf{BGP/gpoly} \not\subseteq \mathbf{GMA/poly},$$

there exist states (*of polynomial size*) *that cannot be prepared using an efficient circuit built from any gate set in the theory.*

Proof. Assume towards contradiction that all states can be prepared using an efficient circuit built from any gate set in the theory. Thus, as there must exist a classical description of each circuit, any advice state from this theory can be replaced with the *classical* advice that specifies the description of the circuit that efficiently prepares the given advice state. We thus have

$$\mathbf{BGP/gpoly} \subseteq \mathbf{BGP/poly} \subseteq \mathbf{GMA/poly},$$

which is a contradiction. There must, therefore, exist at least one state that cannot be prepared efficiently in this theory. ■

Thus, in theories that do not satisfy

$$\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly},$$

the dynamics are not rich enough to prepare the states that contain a large amount of 'useful' information. This is not to say that in theories satisfying this relation every state can be efficiently prepared, it is just that in theories violating the relation this assertion can be proved *directly* from the violation. As a side remark, within the theorem proof we have proved that **BGP/poly** is *strictly* contained in **BGP/gpoly** for theories **G** where $\mathbf{BGP/gpoly} \nsubseteq \mathbf{GMA/poly}$. It is presently unknown whether quantum advice is strictly stronger than classical advice for quantum computers.

In addition to proving $\mathbf{BGP/gpoly} \subseteq \mathbf{GMA/poly}$, Aaronson and Drucker [38] proved what they called a 'quantum Karp–Lipton' theorem. The Karp–Lipton theorem states that if $\mathbf{NP} \subseteq \mathbf{P/poly}$ then the polynomial hierarchy collapses to its second level, which is believed to be unlikely [53]. The quantum Karp–Lipton theorem states that if $\mathbf{NP} \subseteq \mathbf{BQP/qpoly}$ then the second level of the polynomial hierarchy is contained in $\mathbf{QMA}^{\mathbf{PromiseQMA}}$,[12] which is also thought to be unlikely [38]. We refer the reader to the original works for further details but we only wish to highlight that, due to theorem 3.3, there exist theories **G** where $\mathbf{NP} \subseteq \mathbf{BGP/gpoly}$ is necessarily satisfied. Therefore, we cannot obtain a 'generalized Karp–Lipton' theorem where unlikely consequences are expected from assuming $\mathbf{NP} \subseteq \mathbf{BGP/gpoly}$.

(a) Related work

Evidence for the existence of a general trade-off has also appeared in recent work which has considered theories satisfying the no-signalling condition from the point of view of interactive proofs. The Merlin-Arthur game is an example of an interactive proof. Another example is a multi-interactive prover (**MIP**) system where more than one of these all-powerful provers sends classical bit-strings to a probabilistic classical computer verifier [54]. Just as in the Merlin-Arthur game, the provers cannot be trusted. However, these provers are not permitted to communicate with one another. A quantum generalization of this is to allow the provers to share entangled quantum states. Ito and Vidick [55] have shown that, in this quantum generalization of **MIP**, it is possible for the verifier to efficiently compute problems in the class **NEXP**, which is the class of problems evaluated by a non-deterministic computer running in exponential time in the size of the input. However, recent work by Kalai *et al*. [56] has shown that if the provers share resources that satisfy only the no-signalling principle (such as Boxworld), then the problems that can be solved in such a model are actually contained in the class **EXP**. Since $\mathbf{EXP} \subseteq \mathbf{NEXP}$, in a theory with states more non-local than quantum mechanics these interactive proof systems have *less* computational power, unless $\mathbf{EXP} = \mathbf{NEXP}$.

7. Discussion and conclusion

The results in this paper provide another example where the best known upper bound on the quantum class **QMA** follows from very minimal assumptions on what constitutes an operational theory. This raises the question of whether better bounds can be derived in the quantum case by exploiting some of the structure unique to quantum theory.

While the definitions of advice and proof verification presented in this paper can be applied to any theory in the framework, they seem to intuitively encode a notion of causality. Note that, in a non-causal theory, circuits do not have any particular 'direction' and so inputting a given state at the 'start' of the computation is not the most natural situation one could consider. Instead of receiving an advice *state*, a more natural situation might be to receive an advice *circuit fragment*—consisting of either a state, transformation or measurement—which can be plugged into the circuit as it is being built. It would be interesting to determine whether this more general

[12]Here **PromiseQMA** is the same as **QMA** except there is a 'promise' on the inputs, i.e. all the inputs satisfy some property.

definition coincides with the standard one in extensions of quantum theory with indefinite causal structure [57,58].

On a final note, it would be fascinating to show if the analysis of computation in generalized probabilistic theories could say something concrete about quantum computing. In an analogous fashion, tools from quantum theory have been used to prove results in *classical* computer science; see [59] for a nice review of such results. We speculate that by understanding quantum theory better within the framework of more general theories we can use tools from the latter to prove results in the former.

Authors' contributions. C.M.L. and M.J.H. contributed equally to this work.
Competing interests. We have no competing interests.
Funding. The authors acknowledge funding from the EPSRC, University College Oxford, and the FQXi Large Grant *Thermodynamic versus information theoretic entropies in probabilistic theories*.
Acknowledgements. The authors acknowledge John H. Selby and all his charitable works, and thank Jon Barrett for discussions.

Appendix A. Classical case

The study of non-uniform classical computation begins with polynomial-sized Boolean circuits. These circuits can equivalently be viewed as Turing machines that take polynomial-sized advice bit-strings. These strings depend only on the size of the input and not the input itself. If the string were to depend on every input then we could just encode the solution to any problem for that input and be able to decide any language. The class of decision problems that are solved by a (uniform) deterministic classical computer with classical advice is denoted **P/poly**, where the suffix/**poly** denotes a classical advice bit-string.

Definition A.1. P/poly is the class of languages $\mathcal{L} \subseteq \{0,1\}^n$ for which there exists a poly-time uniform classical circuit family $\{\mathcal{C}_x\}$ and a set of bit-strings $\{y_n\}_{n\geq 1}$ of length $d(n)$ for some polynomial d, such that, for all strings $x \in \{0,1\}^n$, $x \in \mathcal{L}$ if and only if $\mathcal{C}_x$ accepts (x, y_n) as input.

As we will be considering probabilistic processes in full generality, it is worth defining the relevant class of computation with advice where processes are not deterministic. In full generality, we allow the possibility that the advice bit-strings are sampled from a probability distribution for each input size—we denote such advice as 'randomized advice' denoted by the suffix /**rpoly**. In addition to this, we allow the uniform circuits to accept inputs with some error as is normal in efficient probabilistic computation (cf. the definition of **BGP**). Therefore, the class **BPP/rpoly** of problems solved (with some error) by a (uniform) classical circuit with randomized advice can now be defined.

Definition A.2. BPP/rpoly is the class of languages $\mathcal{L} \subseteq \{0,1\}^n$ for which there exists a poly-time uniform classical circuit family $\{\mathcal{C}_x\}$ and a set of randomized advice bit-strings $\{y_n\}_{n\geq 1}$ of length $d(n)$ for some polynomial d, such that, for all strings $x \in \{0,1\}^n$,

(i) if $x \in \mathcal{L}$ then $\mathcal{C}_x$ accepts with probability at least $\frac{2}{3}$ given (x, y_n) as input;
(ii) if $x \notin \mathcal{L}$ then $\mathcal{C}_x$ accepts with probability at most $\frac{1}{3}$ given (x, y_n).

Interestingly, despite the ability to use probabilistic processes, via derandomization arguments it can be shown that **BPP/rpoly** = **P/poly** [51,52].

In the case where an efficient computer is given a proof from some untrusted provider, we have already mentioned the classical complexity class **NP** but this is not the most general class for probabilistic computation. If the efficient classical computer accepts some input with some error, then this is in the remit of Merlin-Arthur games with complexity as follows.

Definition A.3. MA is the class of languages $\mathcal{L} \subseteq \{0,1\}^n$ for which there exists a poly-time uniform classical circuit $\{\mathcal{C}_x\}$ and a polynomial d, such that, for all strings $x \in \{0,1\}^n$,

(i) if $x \in \mathcal{L}$ then there exists a proof $z \in \{0,1\}^{d(n)}$ such that $\mathcal{C}_x$ accepts with probability at least $\frac{2}{3}$ given (x, z) as input;
(ii) if $x \notin \mathcal{L}$ then $\mathcal{C}_x$ accepts with probability at most $\frac{1}{3}$ given (x, z) as input, for all proofs z.

The existential quantifiers in the above definition rigorously capture the notion of a circuit having to 'verify' the proof. It immediately follows that **NP** $\subseteq$ **MA**. This definition will also allow us to readily present the quantum analogue to this class along with its analogue for all possible general theories.

Appendix B. Quantum case

The class of decision problems that can be solved by an efficient quantum computer with quantum advice, denoted by **BQP/qpoly**, is defined as follows.

Definition B.1. BQP/qpoly is the set of languages $\mathcal{L} \subseteq \{0,1\}^n$ for which there exists a poly-time uniform quantum circuit family $\{\mathcal{Q}_x\}$ and a set of (possibly non-uniform) states $\{|\psi_n\rangle\}_{n\geq 1}$ of $d(n)$ qubits for some polynomial d, such that, for all strings $x \in \{0,1\}^n$,

(i) if $x \in \mathcal{L}$ then $\mathcal{Q}_x$ accepts with probability at least $\frac{2}{3}$ given $|x\rangle|\psi_n\rangle$ as input;
(ii) if $x \notin \mathcal{L}$ then $\mathcal{Q}_x$ accepts with probability at most $\frac{1}{3}$ given $|x\rangle|\psi_n\rangle$.

The class of decision problems for which a 'yes' outcome can be verified in quantum poly-time, with help from a poly-size quantum proof, or witness, state, denoted **QMA**, is defined as follows.

Definition B.2. QMA is the set of languages $\mathcal{L} \subseteq \{0,1\}^n$ for which there exists a poly-time uniform quantum circuit $\{\mathcal{Q}_x\}$ and a polynomial d, such that, for all strings $x \in \{0,1\}^n$,

(i) if $x \in \mathcal{L}$ then there exists a $d(n)$-qubit quantum proof $|\phi\rangle$ such that $\mathcal{Q}_x$ accepts with probability at least $\frac{2}{3}$ given $|x\rangle|\phi\rangle$ as input;
(ii) if $x \notin \mathcal{L}$ then $\mathcal{Q}_x$ accepts with probability at most $\frac{1}{3}$ given $|x\rangle|\phi\rangle$ as input, for all proofs $|\phi\rangle$.

The existential quantifiers in the above definition of **QMA** rigorously capture the notion of a quantum circuit having to 'verify' the quantum proof.

References

1. Aaronson S, Arkhipov A. 2011 The computational complexity of linear optics. In *Proc. of the 43rd Annu. ACM Symp. on Theory of Computing* (*STOC 2011*), *San Jose, CA, 6–8 June 2011*, pp. 333–342. New York, NY: ACM.
2. Shor P. 1997 Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Stat. Comput.* **26**, 1484–1509. (doi:10.1137/S0097539795293172)
3. Nielsen MA, Chuang IL. 2000 *Quantum computation and quantum information*. Cambridge, UK: Cambridge University Press.
4. Raz R. 1999 Exponential separation of quantum and classical communication complexity. In *Proc. of the 31st Annu. ACM Symp. on the Theory of Computing, Atlanta, GA, 1–4 May 1999*, p. 358. New York, NY: ACM.
5. Buhrman H, Cleve R, van Dam W. 2001 Quantum entanglement and communication complexity. *SIAM J. Comput.* **30**, 1829–1841. (doi:10.1137/S0097539797324886)
6. Howard M, Wallman J, Veitch V, Emerson J. 2014 Contextuality supplies the magic for quantum computation. *Nature* **510–351**. (doi:10.1038/nature13460)
7. Vidal G. 2003 Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.* **91**, 147902. (doi:10.1103/PhysRevLett.91.147902)
8. Hoban MJ. 2011 Generalized Bell-inequality experiments and computation. *Phys. Rev. A* **84**, 062107. (doi:10.1103/PhysRevA.84.062107)

9. Hoban MJ, Browne DE. 2011 Stronger quantum correlations with loophole-free post-selection. *Phys. Rev. Lett.* **107**, 120402. (doi:10.1103/PhysRevLett.107.120402)
10. Raussendorf R. 2013 Contextuality in measurement-based quantum computation. *Phys. Rev. A* **88**, 022322. (doi:10.1103/PhysRevA.88.022322)
11. Datta A, Shaji A, Caves C. 2008 Discord and the power of one qubit. *Phys. Rev. Lett.* **100**, 050502. (doi:10.1103/PhysRevLett.100.050502)
12. Stahlke D. 2014 Quantum interference as a resource for quantum speedup. *Phys. Rev. A* **90**, 022302. (doi:10.1103/PhysRevA.90.022302)
13. Brodutch A. 2013 Discord and quantum computational resources. *Phys. Rev. A* **88**, 022307. (doi:10.1103/PhysRevA.88.022307)
14. Van den Nest M. 2013 Universal quantum computation with little entanglement. *Phys. Rev. Lett.* **110**, 060504. (doi:10.1103/PhysRevLett.110.060504)
15. Buhrman H, Czekaj L, Grudka A, Horodecki M, Horodecki P, Markiewicz M, Speelman F, Strelchuk S. 2015 Quantum communication complexity advantage implies violation of a Bell inequality. (http://arxiv.org/abs/1502.01058)
16. Lee CM, Barrett J. 2015 Computation in generalised probabilistic theories. *New J. Phys.* **17**, 083001. (doi:10.1088/1367-2630/17/8/083001)
17. Lee CM, Selby JH. 2015 Generalised phase kick-back: a computational advantage for higher-order interference? (http://arxiv.org/abs/1510.04699)
18. Barrett J, de Beaudrap N, Hoban MJ, Lee CM. In preparation. The computational landscape of general physical theories.
19. CM Lee, JH Selby. 2015 Higher-order interference in extensions of quantum theory. (http://arxiv.org/abs/1510.03860)
20. Fortnow L, Rogers J. 1998 Complexity limitations on quantum computation. (http://arxiv.org/abs/9811023v1)
21. Nishimura H, Yamakami T. 2003 *Polynomial time quantum computation with advice*. Electronic Colloquium on Computational Complexity, TR03-059. See http://eccc.hpi-web.de/eccc-reports/2003/TR03-059/index.html.
22. Aaronson S. 2005 Limitations of quantum advice and one-way communication. *Theory Comput.* **1**, 1–28. (doi:10.4086/toc.2005.v001a001)
23. Popescu S, Rohrlich D. 1994 Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379–385. (doi:10.1007/BF02058098)
24. Popescu S. 2014 Nonlocality beyond quantum mechanics. *Nat. Phys.* **10**, 264–270. (doi:10.1038/nphys2916)
25. van Dam W. 2005 Implausible consequences of superstrong nonlocality. (http://arxiv.org/abs/0501159)
26. Aharonov D, Arad I, Vidick T. 2013 The quantum PCP conjecture. *ACM SIGACT* **44**, 47–79.
27. Hardy L. 2011 Reformulating and reconstructing quantum theory. (http://arxiv.org/abs/1104.2066v3)
28. Chiribella G, D'Ariano GM, Perinotti P. 2010 Probabilistic theories with purification. *Phys. Rev. A* **81**, 062348. (doi:10.1103/PhysRevA.81.062348)
29. Chiribella G, D'Ariano GM, Perinotti P. 2011 Informational derivation of quantum theory. *Phys. Rev. A* **84**, 012311. (doi:10.1103/PhysRevA.84.012311)
30. Barrett J. 2007 Information processing in generalized probabilistic theories. *Phys. Rev. A* **75**, 032304. (doi:10.1103/PhysRevA.75.032304)
31. Barnum H, Mueller MP, Ududec C. 2014 Higher-order interference and single system postulates for quantum theory. *New J. Phys.* **16**, 123029. (doi:10.1088/1367-2630/16/12/123029)
32. D'Ariano G, Manessi F, Perinotti P. 2013 Determinism without causality. (http://arxiv.org/abs/1301.7578)
33. de Beaudrap N. On computation with 'probabilities' modulo k. (http://arxiv.org/abs/1405.7381v2)
34. Aaronson S. 2005 Quantum computing, postselection and probabilistic polynomial time. *Proc. R. Soc. A* **461**, 3473–3482. (doi:10.1098/rspa.2005.1546)
35. Kitaev A, Watrous J. 2000 Parallelization, amplification, and exponential time simulation of quantum interactive proof system. In *Proc. of the 32nd Annual ACM Symposium on Theory of Computing* (*STOC 2000*), *Portland, OR, 21–23 May 2000*, pp. 608–617. New York, NY: ACM.
36. Vyalyi MN. 2003 *QMA=PP implies that PP contains PH*. Electronic Colloquium on Computational Complexity, Report No. 21. See eccc.hpi-web.de/report/2003/021/download/.

37. Marriott C, Watrous J. 2005 Quantum Merlin-Arthur games. *Comput. Complexity* **14**, 122–152. (doi:10.1007/s00037-005-0194-x)
38. Aaronson S, Drucker A. 2010 A full characterisation of quantum advice. In *Proc. of the 42nd Ann. ACM Symp. on Theory of Computing* (*STOC 2010*), *Cambridge, MA, 6–8 June 2010*, pp. 131–140. New York, NY: ACM.
39. Short AJ, Barrett J. 2010 String nonlocality: a trade-off between states and measurements. *New J. Phys.* **12**, 033034. (doi:10.1088/1367-2630/12/3/033034)
40. Barrett J, Pironio S. 2005 Popescu-Rohrlich correlations as a unit of nonlocality. *Phys. Rev. Lett.* **95**, 140401. (doi:10.1103/PhysRevLett.95.140401)
41. Gross D, Mueller M, Colbeck R, Dahlsten O. 2010 All reversible dynamics in maximal non-local theories are trivial. *Phys. Rev. Lett.* **104**, 080402. (doi:10.1103/PhysRevLett.104.080402)
42. Al-Safi SW, Short AJ. 2014 Reversible dynamics in strongly non-local Boxworld systems. *J. Phys. A* **47**, 325303. (doi:10.1088/1751-8113/47/32/325303)
43. Al-Safi SW, Richens J. 2015 Reversibility and the structure of the local state space. (http://arxiv.org/abs/1508.03491)
44. Bravyi S, DiVincenzo DP, Oliveira RI, Terhal BM. 2008 The complexity of stoquastic local hamiltonian problems. *Quant. Inf. Comp.* **8**, 361–385.
45. Mueller M, Ududec C. 2012 The structure of reversible computation determines the self-duality of quantum theory. *Phys. Rev. Lett.* **108**, 130401. (doi:10.1103/PhysRevLett.108.130401)
46. Mueller M, Oppenheim J, Dahlsten O. 2012 The black hole information problem beyond quantum theory. *J. High Energy Phys.* **2012**, 116. (doi:10.1007/s13130-012-4801-4)
47. Winter A. 1999 Coding theorem and strong converse for quantum channels. *IEEE Trans. Inform. Theory* **45**, 2481–2485. (doi:10.1109/18.796385)
48. Ogawa T, Nagaoka H. 2002 A new proof of the channel coding theorem via hypothesis testing in quantum information theory. (http://arxiv.org/abs/0208139)
49. Pfister C, Wehner S. 2013 If no information gain implies no disturbance, then any discrete physical theory is classical. *Nat. Commun.* **4**, 1851. (doi:10.1038/ncomms2821)
50. Mueller M, Masanes L. 2013 Three-dimensionality of space and the quantum bit: an information-theoretic approach. *New J. Phys.* **15**, 053040. (doi:10.1088/1367-2630/15/5/053040)
51. Adleman L. 1978 Two theorems on random polynomial time. In *Proc. of FOCS 1978, Ann Arbor, MI, 16–18 October 1978*, pp. 75–83. Washington, DC: IEEE Computer Society.
52. Aaronson S. 2006 QMA/qpoly is contained in PSPACE/poly: de-merlinizing quantum protocols. In *Proc. of the IEEE Conf. on Computational Complexity, Prague, Czech Republic, 16–20 July 2006*, pp. 261–273. Washington, DC: IEEE Computer Society.
53. Karp RM, Lipton RJ. 1982 Turing machines that take advice. *Enseign. Math.* **28**, 191–201.
54. Ben-Or M, Goldwasser S, Kilian J, Wigderson A. 1988 Multi-prover interactive proofs: how to remove intractability. In *Proc. of the 20th Annu. ACM Symp. on Theory of Computing* (*STOC 88*), *Chicago, IL, 2–4 May 1988*, pp. 113–131. New York, NY: ACM.
55. Ito T, Vidick T. 2012 A multi-prover interactive proof for NEXP sound against entangled provers. In *Proc. of FOCS 2012, New Brunswick, NJ, 20–23 October 2012*, pp. 243–252. Washington, DC: IEEE Computer Society.
56. Kalai YT, Raz R, Rothblum RD. 2014 How to delegate computations: the power of no-signaling proofs. In *Proc. of the 46th Annu. ACM Symp. on Theory of Computing* (*STOC 2014*), *New York, NY, 31 May–3 June 2014*. New York, NY: ACM.
57. Chiribella G, D'Ariano G, Perinotti P, Valiron B. 2013 Quantum computations without definite causal structure. *Phys. Rev. A* **88**, 022318. (doi:10.1103/PhysRevA.88.022318)
58. Araújo M, Costa F, Brukner C. 2014 Computational advantage from quantum-controlled ordering of gates. *Phys. Rev. Lett.* **113**, 250402. (doi:10.1103/PhysRevLett.113.250402)
59. Drucker A, de Wolf R. 2011 Quantum proofs for classical theorems. *Theory Comput. Grad. Surv.* **2**. (doi:10.4086/toc.gs.2011.002)