Goldsmiths Research Online

Goldsmiths Research Online (GRO)
is the institutional research repository for
Goldsmiths, University of London

Citation

Persistent URL

Versions

Goldsmiths
UNIVERSITY OF LONDON

ARTICLE   OPEN

The computational landscape of general physical theories

Jonathan Barrett [1], Niel de Beaudrap[1], Matty J. Hoban[1,2] and Ciarán M. Lee[3]

There is good evidence that quantum computers are more powerful than classical computers, and that various simple modifications of quantum theory yield computational power that is dramatically greater still. However, these modifications also violate fundamental physical principles. This raises the question of whether there exists a physical theory, allowing computation more powerful than quantum, but which still respects those fundamental physical principles. Prior work by two of us introduced this question within a suitable framework for theories that make good operational sense, and showed that in any theory satisfying tomographic locality, the class of problems that can be solved efficiently is contained in the complexity class AWPP. Here, we show that this bound is tight, in the sense that there exists a theory, satisfying tomographic locality, as well as a basic principle of causality, which can efficiently decide everything in AWPP. Hence this theory can efficiently simulate any computation in this framework, including quantum computation.

INTRODUCTION

There is ever-growing evidence that quantum computers are more powerful than classical computers.[1–4] However, an understanding of the source of this power remains elusive. Many features of quantum mechanics have been posited as the origin of this so-called "speed-up"[5–9] but the debate is far from resolved.[10–12] In recent years, one way of examining this power has been to ask how the computational power changes as features of quantum theory are altered. Beginning with the work of Abrams and Lloyd, it was shown that allowing more exotic transformations in quantum theory can result in easily solving hard problems.[13] This motivates the speculation that quantum theory is an "island" within the space of all possible theories; alter quantum mechanics and we obtain dramatic consequences.[14]

Another possibility is that our understanding of computation in possible physical theories is couched too much in the language of quantum theory. For example, it could be entirely possible to have a theory that has the same computational power as quantum theory but barely resembles it. We thus require an abstract framework in which to study the power of computation, where quantum and classical computation are special cases.

The study of operational theories provides us with a suitable framework for the study of information processing based on operational principles.[15–19] That is, we can make statements about the limits and power of information processing without referring explicitly to quantum theory. Some features thought unique to quantum theory (as opposed to classical physics) can be seen to be ubiquitous within these theories. For example, given some fundamental properties that reasonable operational theories should satisfy, a no-broadcasting theorem holds in any non-classical theory.[20] This then begs the question of what fundamental principles uniquely single out quantum physics from these myriad possibilities. Indeed, starting from various frameworks of operational theories there have been many derivations of

quantum theory from information theoretic principles (e.g. refs. [17,21,22]).

In refs. [23–26], a circuit-based model of computation is defined and studied in the context of a broad operationally-defined framework for physical theories. Informally, a theory in this framework specifies a set of laboratory devices that can be connected together to form experiments, and assigns probabilities to experimental outcomes. Whilst many such theories may not correspond to descriptions of our physical world, they nevertheless make good operational sense, and allow one to systematically assess how computational power depends on the underlying physical theory.

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). ref. [23] shows that for theories 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 bound first proved for the quantum case by Fortnow and Rogers.[27]

Reference [23] leaves open the question of whether the bound is tight, in the sense that there exists a theory that could solve all problems in AWPP. Such a theory would have computational power beyond that which we expect from quantum mechanics and could simulate any quantum computation. In this paper we resolve this open problem and show that there does indeed exist a non-quantum theory, satisfying both tomographic locality and causality, which can decide everything in AWPP. We may consider this theory as a "foil" theory, used to deepen our understanding of the limitations of quantum computers. This foil theory is constructed from a computational model using quasi-probabilities, i.e. an affine combination of weights assigned to particular events. This motivates the study of what minimal set of information principles recover the power of quantum computation.

[1]Department of Computer Science, Wolfson Building, University of Oxford, Parks Road, Oxford OX1 3QD, UK; [2]Department of Computing, Goldsmiths, University of London, New Cross, London SE14 6NW, UK and [3]Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK
Correspondence: Ciarán M. Lee (ciaran.lee@ucl.ac.uk)

RESULTS

Operational theories

The fundamental goal of any physical theory is to provide a consistent account of experimental data. This constitutes the core idea underlying the framework of operational theories,[15–18,21,23,28,29] where the primitive notions are operational in nature.

A theory in this framework specifies a set of *laboratory devices*, which can be connected together in certain ways, and assigns probabilities to different experimental outcomes. A laboratory device comes equipped with *input ports*, *output ports*, and a *classical pointer*, where roughly speaking, one may think of physical systems passing into input ports and emerging from output ports, with the pointer indicating an experimental outcome. Each input and output port has an associated *type*. We will often denote types A, B, C ..., and use X or Y to stand for generic types. Experiments correspond to circuits, which are formed by connecting output ports of devices to input ports of other devices in such a way that types match. By assumption, the circuit corresponding to a valid experiment must be acyclic, and closed, meaning that there are no unconnected input or output ports. When an experiment is run, each pointer comes to rest in a final position, with these pointer positions constituting jointly the outcome of the experiment. For any circuit corresponding to an allowed experiment, the theory must define a joint probability distribution over pointer positions for all devices in the circuit.

Laboratory devices include *preparation devices*, which have no input ports, and *measurement devices*, which have no output ports. Each use of a preparation device outputs a physical system in some particular *state*, where the state is determined by the variety of device used and the position attained by the pointer on the device on that run. A measurement device can be thought of as implementing a destructive measurement, since no system emerges, with the outcome denoted by the pointer position. Each outcome corresponds to an *effect*. Given a device with both input and output ports, a system may pass through in such a way that its state is altered. The change in the state is non-deterministic in the sense that the change applied is indicated by the position of the pointer. When a device has both input and output ports, each pointer position corresponds to a *transformation*.

For the formal development of operational theories, see for example refs. [15–18,21,23,28,29]. Here, rather than present an axiomatic derivation, we simply summarize the resulting mathematical structure.

Each system type X can be associated with a real vector space V_X, such that a state of the system is a vector in V_X. In this work, it is assumed throughout that for each type of system, V_X is finite dimensional. Types are closed under parallel composition, hence given a system of type X and a system of type Y, there is a composite system, whose type can be denoted XY. The theories that we are interested in satisfy the principle of *tomographic locality*,[15,16] which says that multipartite states can be uniquely specified by the joint probabilities for the outcomes of measurements performed locally on each component system. This implies that the vector space associated with a composite system is the tensor product of the vector spaces associated with the component systems: i.e., the vector space associated with the composite XY is $V_{XY} = V_X \otimes V_Y$. A state of the composite is a direct product state if it is of the form $|s\rangle_{XY} = |s\rangle_X \otimes |s\rangle_Y$.

A transformation, with input type X and output type Y, is a linear map from V_X to V_Y. Given a composite system of type XY, the parallel action of transformation T_X on the type X subsystem, and transformation T_Y on the type Y subsystem, is given by a transformation $T_{XY} = T_X \otimes T_Y$. An effect on a system of type X is a linear map from V_X to the real numbers, i.e., an effect is an element of the dual space. Consider a composite system of type XY, and suppose that local measurements are performed. If a
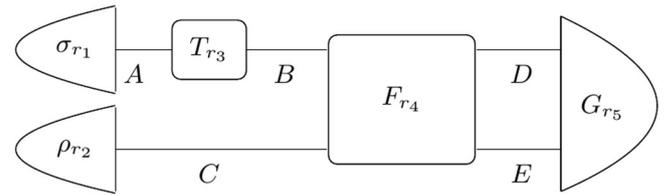
Fig. 1 Example of a closed circuit in an operational theory. A particular outcome corresponds to the pointers on the devices attaining positions $r_1, ..., r_5$

particular outcome of the measurement on the type X subsystem corresponds to an effect $_X\langle e|$, and a particular outcome of the measurement on the Y subsystem corresponds to an effect $_Y\langle e|$, then the joint outcome corresponds to an effect $_{XY}\langle e| = {_X}\langle e| \otimes {_Y}\langle e|$.

Given a closed circuit, the joint probability for observing a particular collection of final pointer positions is given by contracting the various tensors to produce a real number. For example, consider an experiment corresponding to the closed circuit of Fig. 1. Reading from left to right: systems of types A and C are prepared; there is a transformation of the A system into a system of type B; this is followed by a joint transformation of the B and C systems into systems of types D and E; finally, a joint measurement is performed. The particular outcome of the experiment shown in Fig. 1 corresponds to the pointers attaining positions $r_1, ..., r_5$. The theory assigns a probability to this outcome, given by:

$$P(r_1 \dots r_5) = \left(G_{r_5}| \circ F_{r_4} \circ (T_{r_3} \otimes I) \circ \left(|\sigma_{r_1}\rangle \otimes |\rho_{r_2}\rangle \right) \right), \tag{1}$$

where ○ denotes the action of a linear map on a vector, and I is the identity operator on V_C.

One can define a notion of *causality* for theories in this framework: the probabilities of present experiments are independent of future measurement choices. It is shown in ref. [15] that this requirement is equivalent to the existence of a unique deterministic effect for each type of system, denoted $_X\langle u|$, such that the following holds. First, for each measurement device, $\sum_e X\langle e| = {_X}\langle u|$, where the sum is over the effects corresponding to outcomes of the device. Second, the norm of a state $|s\rangle_X$ is given by $_X\langle u|s\rangle_X$, and all states satisfy $_X\langle u|s\rangle_X \leq 1$. Third, given a device with both input and output ports, the sum over the transformations corresponding to each pointer position must be a linear map that preserves the norm of the state. Note that consistent theories can be constructed with more than one deterministic effect, hence which violate causality.[30]

Finite dimensional quantum theory serves as an explicit example that illustrates the framework. Systems are associated with complex, finite dimensional Hilbert spaces, their type corresponding to the dimension of this space. States correspond to positive semi-definite operators acting on the underlying Hilbert space, with V_X being the real vector space spanned by Hermitian operators. A measurement outcome is associated with a positive operator E_X such that the corresponding effect is given by $\rho_X \rightarrow \mathrm{Tr}(E_X \rho_X)$. Quantum theory is of course causal: the unique deterministic outcome corresponds to the identity I_X, such that the positive operators for the different outcomes of the measurement must sum to I_X. The norm of a state ρ_X is $\mathrm{Tr}(\rho_X)$. Transformations correspond to trace-non-increasing completely positive maps. A device with both input and output ports corresponds to a *quantum instrument*, that is a set of trace-non-increasing completely positive (CP) maps (one for each pointer position) that sum to a trace-preserving CP map. It may be verified that tomographic locality holds in quantum theory. In particular, the state ρ_{XY} of a composite system is a Hermitian operator acting on the tensor product of the underlying Hilbert spaces; the real vector space V_{XY} spanned by such operators may be identified with $V_X \otimes V_Y$.

Theories different from quantum theory have also been studied in this framework. In the theory known as "Boxworld",[31,32] for example, the simplest non-trivial type of system has a state defined via two binary-outcome measurements, $\{(x_a|\}$, where x is a bit denoting the measurement setting and a is a bit denoting the outcome. There are four possible pure states that can be prepared. Denoting these $|z, w\rangle$, with $z, w \in \{0, 1\}$, they satisfy $(0_b|z, w) = \delta_{bw}$ for measurement setting 0, and $(1_b|z, w) = \delta_{bz}$ for measurement setting 1. Multipartite states in the theory are defined such that aritrary non-signalling correlations can be produced, including, for example, the Popescu-Rohrlich correlations that maximally violate the CHSH inequality.[31] Boxworld satisfies both tomographic locality and causality.[32]

Other interesting examples of non-quantum theories include the non-causal theory of ref. [30], and the theories investigated by ref. [33], in which the set of states of a single system corresponds to a Euclidean hyperball of dimension n. The toy theory of ref. [34] may also be described by the operational framework.

Free and non-free theories
In the usual definition of an operational theory,[15–17,21,23] a theory specifies a set of laboratory devices, from which one can build closed circuits, and assigns a probability distribution over the outcomes of each closed circuit. Any closed circuit that can be built from the laboratory devices corresponds to a valid experiment. This means that there is a significant constraint on the structure of the theory, which is that all closed circuits must give rise to a valid probability distribution over outcomes. We refer to such theories as "free" operational theories. The idea behind the terminology is that an agent is free to build an experiment corresponding to any closed circuit they like, as long as devices are composed properly, i.e., types match.

This work considers a more general definition of an operational theory, according to which a theory specifies a set of laboratory devices, and a set of *allowed* closed circuits, which may be a proper subset of the set of all closed circuits that can be built using the laboratory devices. The interpretation is that it is only the allowed closed circuits that correspond to experimental procedures that can actually be performed. The theory must assign a valid probability distribution over the outcomes of any allowed closed circuit. This definition is not unmotivated if one takes the viewpoint that a physical theory corresponds both to a consistent account of experimental data and to which experiments are implementable in principle. This is a significant generalization for the following reason. Given a closed circuit that is not in the allowed set, one may still contract the tensors associated with the device outcomes in order to produce a real number; but there is no constraint that this number has to be in the interval [0, 1].

Note that for tomographic locality to hold in a non-free theory, the set of allowed closed circuits must at least include a collection of experiments that are sufficient for local tomography to be carried out. In particular, for each state of a system of type AB, there should be allowed closed circuits involving local measurements on the subsystems such that, when the outcome probabilities for these circuits are known, the state is completely specified.

We also assume that the set of allowed closed circuits is itself closed under parallel composition, so that an experimenter may always choose to perform both of two valid experiments, independently of one another. In more detail, if C_1 is an allowed closed circuit, with outcomes $r_1, \ldots, r_k$, and C_2 is an allowed closed circuit, with outcomes $s_1, \ldots, s_l$, then the parallel composition is also an allowed closed circuit, corresponding to a valid experiment with outcomes $r_1, \ldots, r_k, s_1, \ldots, s_l$. We require that the outcome probabilities satisfy

$$P_{C_1 C_2}(r_1, \ldots, r_k, s_1, \ldots, s_l) = P_{C_1}(r_1, \ldots, r_k) P_{C_2}(s_1, \ldots, s_l), \quad (2)$$

where P_{C_1} denotes the outcome distribution that the theory assigns to the circuit C_1, similarly P_{C_2} and the circuit C_2, and $P_{C_1 C_2}$ denotes the outcome distribution for the parallel composition. Part of the reason for this assumption is that the idea of bounded-error computation makes little sense unless independent repetitions of a computation can be carried out, to verify the result, and reduce error probabilities close to zero.

Computation
The class of "yes/no" problems that a quantum computer can solve efficiently is denoted by BQP and much research has been concerned with how large this class is. At present, the tightest known upper bound is BQP $\subseteq$ AWPP,[27] where AWPP is a classical complexity class, known to be contained in PP, hence in PSPACE.[27] This class is formally defined in Methods.

In order to define efficient computation in theories belonging to the framework introduced above, we need the notion of a (polynomially sized) uniform circuit family, and a condition for a circuit to accept an input. The following definition appeared in ref. [23]. A polynomially sized uniform circuit family is a set of closed circuits $\{C_x\}$, where x ranges over finite-length bit strings and corresponds to the input to the problem, such that:

1. There is a gate set $\mathcal{G}$, consisting of laboratory devices, such that each circuit in the family is built from elements of $\mathcal{G}$.
2. The number of gates in the circuit C_x is bounded by a polynomial in $|x|$.
3. 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 $|(M - \tilde{M})_{ij}| \leq \varepsilon$, in time polynomial in $\log(1/\varepsilon)$.
4. 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 $|x|$.

This produces, for each C_x, a description of an experiment, whose devices produce classical outcomes. Denoting the string of observed outcomes by z, the final output of the computation is given by an *acceptor 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 $|x|$. We say that a run of the experiment accepts an input string x if the outcome string z of the circuit C_x satisfies $a(z) = 0$. The probability that a computation accepts the input string x is therefore given by

$$P_x(\text{accept}) = \sum_{z|a(z)=0} P(z), \quad (3)$$

where the sum ranges over all possible outcome strings z of the circuit C_x for which $a(z) = 0$.

Definition 1. For an operational theory **G**, let the class of problems that can be solved efficiently be denoted schematically **BGP**. A language $\mathcal{L}$ is in the class BGP if the set of allowed circuits defined by **G** includes a poly-sized uniform circuit family, along with an efficient acceptor, such that

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

The constants in the above definition can be chosen arbitrarily as long as they are bounded away from a half by some inverse polynomial. The following theorem was proved for free theories in

ref. [23], and follows without modification for non-free theories as well:

Theorem 1. For any operational theory **G** satisfying tomographic locality,

$$BGP \subseteq AWPP. \tag{4}$$

One might wonder if efficient quantum computation can achieve the bound of Theorem 1. In Section K we present a complexity-theoretic argument that may be considered evidence against such a possibility.

Achieving the upper bound

The main result of this work is the construction of an operational theory, satisfying causality and tomographic locality, that has *exactly* the power of this upper bound.

Theorem 2. There exists an operational theory **G**, satisfying causality and tomographic locality, such that

$$BGP = AWPP. \tag{5}$$

Hence AWPP, despite having a slightly involved definition in terms of gap functions for non-deterministic Turing machines (see Methods), can be thought of much more intuitively as the class of problems efficiently solvable by tomographically local physical theories.

An intuitive sketch of the proof of Theorem 2 is as follows (for formal definitions and proofs, see Methods). First, we show that the class AWPP is perfectly captured by a quasi-probabilistic model of computation, defined via a Turing Machine with quasi-probabilistic transition weights with the constraint that the total weight of transitions from a given state must sum to $+1$. We refer to this model as an *Affine Turing Machine*. See Fig. 2 for a schematic illustration. We then construct uniform poly-size circuits, in which the gates are certain affine transformations, which can simulate—and be simulated by—an Affine Turing Machine, and hence which also capture AWPP.

This construction results in a collection of closed circuits, which correspond to the probability that the final result of the Affine Turing Machine is "yes" or "no" on inputs of different lengths. Finally, we construct a causal (non-free) operational theory, which contains the closed circuits necessary to simulate any Affine Turing Machine amongst its set of allowed circuits, along with sufficient additional circuits to ensure that tomographic locality holds. This proves Theorem 2.
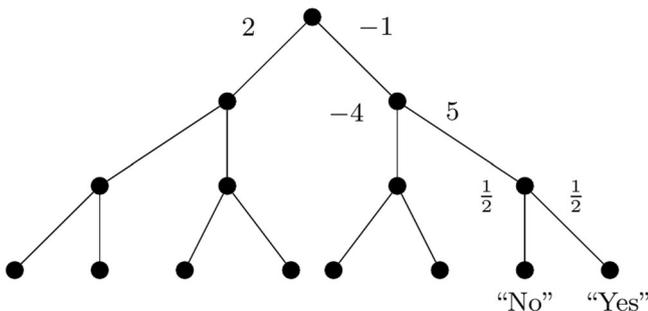
Fig. 2 Schematic illustration of an Affine Turing Machine

DISCUSSION

This work describes an operational theory, which satisfies causality and tomographic locality, such that the class of problems that can be efficiently solved by devices in that theory is exactly AWPP. This provides a converse to the results of ref. [23]. To describe this construction, we introduce a new possibility: that of a "non-free" theory, in which the possible transformations of systems are not necessarily closed under sequential and parallel composition.

An interesting feature of the AWPP-complete theory constructed in this paper is that it satisfies the principle of causality. The main result of ref. [23] was that for any theory satisfying tomographic locality, whether or not causality is satisfied, efficiently solvable computational problems are contained in AWPP. Taken together, these results show that computational circuits in any non-causal theory can always be efficiently simulated by circuits in a causal theory. Hence, in the landscape of general theories, "acausality" does not appear to be a resource for computation.

Theorem 2 is reminiscent of a result encountered when quantum correlations, obtained from measurements on entangled systems in a Bell-type experiment, are viewed in the context of the set of all non-signalling correlations.[35] Classical correlations are by definition local, and satisfy all Bell inequalities. Quantum correlations can be nonlocal, in the sense that they violate a Bell inequality, but the violation is limited by Tsirelson bounds.[36] Operational theories can be constructed that produce stronger violations than is possible with quantum systems: for example, there exists a theory colloquially known as "Boxworld"[16,31] that allows all correlations consistent with the no-signalling principle. Similarly, when the computational power of tomographically local theories is considered, classical theories can be simulated by quantum theory, and it is believed that quantum computers can solve some problems efficiently that classical computers cannot. Here, we have shown the existence of an operational theory with the strongest possible computational power, and it is unlikely that quantum computers will be able to simulate this theory. Figure 3 schematically represents this analogy between the sets of correlations satisfying the no-signalling principle, and the computational complexity classes of theories satisfying
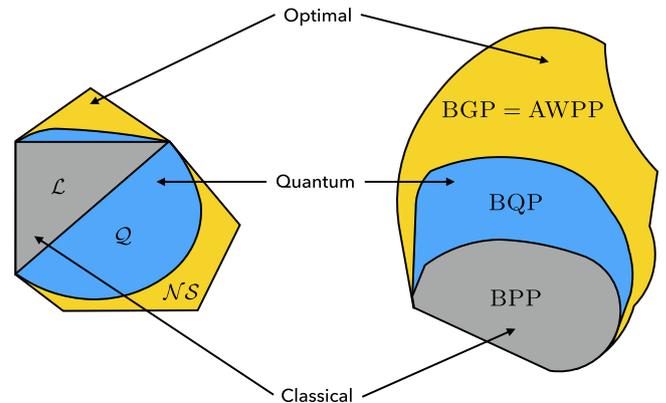
Fig. 3 Nonlocality versus computational power. The part of the figure on the left shows the set of all non-signalling correlations that could logically be obtained in a Bell test, with the sets of quantum and classical correlations strictly contained inside. The part of the figure on the right shows the computational complexity classes associated with theories that satisfy tomographic locality, with the theory presented in this paper saturating the whole of AWPP, and the classes associated with the quantum and classical theories contained in this class. Note that in the case of computation, we can only conjecture that each of these containments is strict. For the containment of classical computation within quantum, see for example[1,3] for evidence that it is strict

tomographic locality, along with the quantum and classical cases for each.

refs. [37–39] have, moreover, shown that methods employing quasi-probability distributions can simulate arbitrary non-signalling correlations. The quasi-probabilistic model of computation introduced here to build a theory with maximal computational power bears an intriguing resemblance to these approaches, providing another similarity between the set of all non-signalling correlations and the computational landscape of general theories.

Many attempts at providing reasonable physical principles that uniquely characterize the set of quantum correlations as a subset of the set of all non-signalling correlations have been made.[40–45] These principles, while not fully capturing the exact quantum boundary,[36] have deepened our understanding of quantum correlations and provided connections between physical principles and information-theoretic advantages. Insights garnered from these connections led to the development of device-independent cryptography.[46–49] So while investigating such connections has foundational interest, it has also been shown to have practical implications.

It seems prudent to ask the analogous question for the set of tomographically local theories: can the class of efficient quantum computation be characterized by some set of physical principles? Such a characterization would deepen our understanding of quantum computation and may also be of practical relevance; if one uncovers the necessary and sufficient physical requirements for universal quantum computation one could design algorithms that optimally take advantage of them. The results presented in this paper provide one with the language and tools to pose these questions in a rigorous fashion

One approach to such a characterization would be to find the minimal set of physical principles that imply the quadratic speed-up over classical computation offered by Grovers search algorithm.[50] This speed-up is optimal for quantum computers,[51–54] so any set of physical principles which imply it could be argued to capture some of the essence of quantum computation. Work in this direction has appeared in refs.,[55–58] where the quadratic lower bound to searching an unstructured database has been shown to hold for a large class of theories.

Recently, methods have been proposed that make use of quasi-probability distributions to classically estimate the output of a quantum computer.[59] These classical estimates converge on the true quantum output probabilities in a time quantified by the "negativity" of the quasi-probability distribution. The larger the negativity, the harder it is for a classical computer to estimate the output probability of a quantum computer. As we have provided an interpretation of the class AWPP in terms of quasi-probabilities, it would be interesting to determine if quantum algorithms can be constructed that estimate the output probability of this quasi-probabilistic computational model. In analogy with the classical estimation algorithms of ref. [59] the quantum algorithms may converge to the true output probability at a rate governed by the negativity of the quasi-probability distribution. Determining how hard it is for a quantum computer to simulate AWPP would provide a way to determining if quantum theory is powerful for computation in the landscape of general theories.

Finally, the distinction introduced in this paper between free and non-free theories appears to be important for the study of computation in operational theories. Indeed, it is still an open question whether there exists a free theory whose computational power equals AWPP. The important distinction between free and non-free theories is that transformations in free theories are closed under composition, implying a bound on the set of states. This need not be the case in non-free theories. Could it be the case that a quantum computer can exploit this fact and efficiently simulate computation in all tomographically-local *free* theories? If this conjecture holds, it could shed light on which physical features give rise to the quantum speed-up.

METHODS
This section contains formal definitions, and the proof of Theorem 2.

Definition of AWPP
Let Σ be a finite set of symbols, *e.g.* $\Sigma = \{0, 1\}$, and let Σ^* be the set of all finite sequences over Σ (commonly referred to as strings). For a string $x \in \Sigma^*$, we let $|x|$ denote its length. A *gap function* over Σ is a function $g : \Sigma^* \to \mathbb{Z}$; which computes the difference between the number of accepting branches and rejecting branches of some nondeterministic Turing machine $\mathbf{N}$, where $\mathbf{N}$ takes no more than $T(|x|)$ computational steps on input $|x|$ for some polynomial T on whatever input x it is given.

Fenner[60], Theorem 1.3 characterized AWPP as the class of languages $L \subseteq \Sigma^*$ for which there is a gap-function $g : \Sigma^* \to \mathbb{Z}$; and a polynomial p, such that

$$x \in L \Rightarrow \frac{2}{3} \leq \frac{g(x)}{2^{p(|x|)}} \leq 1; \tag{6A}$$

$$x \notin L \Rightarrow 0 \leq \frac{g(x)}{2^{p(|x|)}} \leq \frac{1}{3}. \tag{6B}$$

Combining this with[60], Theorem 3.1, more generally we have $L \in$ AWPP if and only if

$$x \in L \Rightarrow \frac{2}{3} \leq \frac{g(x)}{h(|x|)} \leq 1; \tag{6C}$$

$$x \notin L \Rightarrow 0 \leq \frac{g(x)}{h(|x|)} \leq \frac{1}{3}. \tag{6D}$$

for a gap-function g and *any* poly-time computable function $h : \mathbb{N} \to \mathbb{N}$. While the original definition of AWPP[61,62] further required there to exist a gap-function g and a poly-time computable function h for any polynomial $r : \mathbb{N} \to \mathbb{N}$, satisfying either $g(x)/h(|x|) \in [0, 2^{-r(|x|)}]$ or $g(x)/h(|x|) \in [1 - 2^{-r(|x|)}, 1]$, we instead use the characterizations of both Eqs. (6) and (7) in our results.

Affine turing machines
We define an *Affine Turing Machine* (AffTM) to be a non-deterministic Turing Machine, in which every transition has an associated real-valued (possibly negative) *weight*. The weights for a given machine are constant throughout the computation, and should be thought of as defined by the transition function. The weight of a given computational branch is then the product of the weights of the transitions involved. We require that for each symbol being read, the total weight of transitions from a given (non-halting) state is $+1$. In this article we consider only rational transition weights, but expect that similar results would obtain for algebraic real coefficients.

We interpret AffTMs as a model of quasi-probabilistic computation, as follows. Given an AffTM $\mathbf{M}$ whose branches all halt in in a finite number of steps, the *acceptance weight* $a_{\mathbf{M}}(x)$ of $\mathbf{M}$ on an input x is the total weight of the accepting paths on input x. An AffTM $\mathbf{M}$ is *proper* if $0 \leq a_{\mathbf{M}}(x) \leq 1$ for all inputs, and that it *decides a language L with bounded error* if furthermore $\frac{2}{3} \leq a_{\mathbf{M}}(x) \leq 1$ for $x \in L$, and $0 \leq a_{\mathbf{M}}(x) \leq \frac{1}{3}$ for $x \notin L$.

An AffTM is *efficient* if the number of computational steps it takes in any computational path on any input x is bounded by some polynomial in $|x|$. The first step towards Theorem 2 is to establish the following:

Lemma 1. The class of languages decided with bounded error by some efficient AffTM is equal to AWPP. The proof of this result is contained in following two sections.

Solving AWPP problems with an affine Turing machine
For $L \subseteq AWPP$, let $g : \Sigma^* \to \mathbb{Z}$ be a gap-function satisfying Eq. (6) for some polynomial p. Also let $\mathbf{N}$ be the non-deterministic Turing machine whose accepting/rejecting branches determine the gap-function g, and T be the polynomial bounding the number of computational steps of $\mathbf{N}$ on its input. By standard results,[61] we may require that $\mathbf{N}$ have the same number of non-deterministic transitions at each step, which we denote by $N \geq 1$, and that all computational branches of $\mathbf{N}$ have the same length on input x. We suppose that each transition of $\mathbf{N}$ is associated with some label

$\ell \in \{1, 2, \ldots, N\}$: the computational branches of **N** are then in one-to-one correspondence with sequences $\{1, 2, \ldots, N\}^{T(|x|)}$. We may then consider an AffTM **M** which simulates **N**, in the following sense:

1. 1. **M** first makes $T(|x|)$ non-deterministic transitions, writing a sequence of symbols $\beta_1, \beta_2, \ldots, \beta_{T(|x|)} \in \{0, 1, 2, \ldots, N\}$ on the tape to produce a string $\beta \in \{0, 1, 2, \ldots, N\}^{T(|x|)}$. The weights of these transitions are $+1$ for each choice $\beta_t \neq 0$, and $(1 - N)$ for each choice $\beta_t = 0$, so that the transition weights sum to $+1$.
2. In branches with one or more symbols $\beta_t = 0$, **M** transitions deterministically with weight $+1$ to a state reject. All other branches of **M** have weight $+1$ and record a string $\beta \in \{1, 2, \ldots, N\}^{T(|x|)}$ indexing some computational branch of **N**. In these branches, **M** simulates the computational branch of **N** whose transitions are indexed by β.
3. For any branch in which the simulation of **N** rejects, **M** makes a non-deterministic transition to a state dampen with weight -1, and to the reject state with weight $+2$. For the branches in which the simulation of **N** accepts, **M** transitions deterministically to dampen with weight $+1$.
4. From the state dampen, **M** makes a sequence of $p(|x|)$ non-deterministic transitions with weight $\frac{1}{2}$, in which it writes bits $\delta_1, \delta_2, \ldots, \delta_{p(|x|)}$ on the tape, thereby sampling a string $\delta \in \{0, 1\}^{p(|x|)}$ uniformly at random. If $\delta_1, \delta_2, \ldots, \delta_{p(|x|)}$, **M** transitions to an accept state; in all other branches it transitions to the reject state.

By the construction of the branch weights, **M** is an AffTM; and as the number of transitions that **M** makes is $O(T + p)$, it is efficient. By construction, the total weight of the branches which transition to the dampen state is $g(x)$; sampling the string $\delta \in \{0, 1\}^{p(|x|)}$ and rejecting unless $\delta = 11 \cdots 1$ ensures that the acceptance weight is $a_{\mathbf{M}}(x) = g(x)/2^{p(|x|)}$. By hypothesis, this is bounded between 0 and 1, is at least $\frac{2}{3}$ if $x \in L$, and is at most $\frac{1}{3}$ otherwise. Thus **M** decides L with bounded error.

Simulating an Affine Turing Machine in AWPP

Suppose that **M** is a proper and efficient AffTM which has transitions with rational weights. Let M be the common denominator of the (finite set of) transition weights of **M**, and let $T \in O(\text{poly}(n))$ be the running time of **M** on an input of length n. Let $m > 0$ be an integer, chosen such that $m \in O(T)$, and such that $2^m \geq M^T$ and $2^m \geq (|u|M)^T$ for all transition weights u of **M**. We may obtain an **AWPP** algorithm to approximately simulate **M**, as follows. We define a non-deterministic machine **N**, which simulates **M** in the following sense:

1. The machine **N** reserves some space on the tape to represent some weight $\Omega \in \mathbb{Q}$; for each branch. We call this the *recorded weight* of the branch.
2. Consider a transition made by **M**, with weight $u = U/M$. To simulate this transition, the machine **N** replaces the recorded weight Ω with $\Omega' = U\Omega$, and then simulates the actions (writing of symbols and movement of the tape head) performed by **M** in the original transition.
3. Once **N** has simulated the final transition of **M**, it non-deterministically samples a sequence of bits $a, b, c_0, c_1, \ldots, c_{m-1} \in \{0, 1\}$. If $a = 1$, we negate Ω if and only if the simulated branch is one in which **M** rejects.
4. **N** determines whether to accept or reject, treating $c_{m-1}c_{m-2} \cdots c_1 c_0$ as the binary expansion of an integer $0 \leq C < 2^m$, as follows:

- If $C \geq |\Omega|$, we reject if $b = 0$, and accept if $b = 1$.
- If $0 \leq C < |\Omega|$, we reject if $\Omega < 0$, and accept if $\Omega > 0$.

Consider the gap function $g(x)$ of the machine **N**. From Step 4, it is clear that if $C \geq |\Omega|$ in any particular branch, **N** accepts and rejects with equal measure, contributing nothing to $g(x)$. The significance of the contribution of any simulated branch of **M** is then in proportion to its recorded weight in **N**, which in absolute value is $2M^T$ times its weight in **M** (arising from the systemic failure to divide the recorded weight by M at each of the T transitions, and from the two values of b). Let $a_+(x)$ be the total weight of those accepting branches of **M** with positive weight, $a_-(x)$ be the total (absolute value of) the weight of accepting branches with negative weight; and similarly for $\rho_+(x)$ and $\rho_-(x)$ for rejecting branches of positive and negative weight. Then $a(x) := a_+(x) - a_-(x)$ is the acceptance weight and $\rho(x) := \rho_+(x) - \rho_-(x)$ is the rejection weight of **M** on input x. We decompose $g(x) = g_0(x) + g_1(x)$, where $g_0(x)$ is the contribution to the gap from branches in which $a = 0$, and $g_1(x)$ is the contribution to the gap from

branches in which $a = 1$. We then have

$$g_0(x) = 2M^T[a_+(x) + \rho_+(x) - a_-(x) - \rho_-(x)] = 2M^T, \tag{7}$$

as $a(x) + \rho(x) = 1$. In the branches where $a = 1$, the sign of the contribution from rejecting branches is negated, so that

$$g_1(x) = 2M^T[a_+(x) - \rho_+(x) - a_-(x) + \rho_-(x)]$$
$$= 2M^T[2a(x) - 1], \tag{8}$$

again using $a(x) + \rho(x) = 1$. Then $g(x) = 4M^T a(x)$, and for $h(n) = 4M^T$, we have $0 \leq g(x)/h(|x|) \leq 1$ as **M** is proper. Furthermore, if **M** decides a language L with bounded error, then either $\frac{2}{3} \leq g(x)/h(|x|) \leq 1$ or $0 \leq g(x)/h(|x|) \leq \frac{1}{3}$ according to whether $x \in L$ or $x \notin L$; then $L \in$ AWPP as well.

Constructing affine circuits

The next step towards Theorem 2 is to construct a family of circuits that can simulate a proper, efficient AffTM. The construction of the circuits is based on that used by Yao in ref. [63] to construct quantum circuits that simulate a quantum Turing Machine (and also on that of refs. [64,65] for circuits that simulate a probabilistic Turing machine). As before, let **M** be a proper and efficient AffTM with alphabet Σ, set of states Q and transition amplitudes $\delta(q, a, \tau, q', a') \in \mathbb{Q}$; with $\tau \in \{\leftarrow, \circ, \rightarrow\}$; the symbols $\leftarrow, \rightarrow$ and $\circ$ are interpreted as the tape head of the AffTM moving to the left, moving to the right, and remaining stationary. Here δ is the transition weight of **M** to change to state q', print a' on the tape and move according to τ, if the machine is currently in state q and reading a. The condition on the weights in order for **M** to be an AffTM is: $\sum_{q', a', \tau} \delta(q, a, \tau, q', a') = 1$ for all $q \in Q, a \in \Sigma$.

We may denote any configuration of the AffTM by a real basis vector

$$|s_{-t}, q_{-t}, a_{-t}, \cdots, s_i, q_i, a_i, \cdots, s_t, q_t, a_t\rangle \tag{9}$$

where the index $-t \leq i \leq t$ denotes the ith cell of the tape and t is the run time of the AffTM (there are $2t + 1$ cells, numbered from $-t$ to t). Here s_i takes on value 0 when the head is not at cell i, value 1 when it is at cell i and the transition step has not occurred and value 2 when the head has just moved according to a transition and is now at cell i. Note that we can represent s_i with two bits. The label q_i denotes the internal state of the machine at cell i, so $q_i \in Q \cup \{\emptyset\}$, where $q_i = \emptyset$ if and only if $s_i = 0$; and $a_i \in \Sigma$ denotes the alphabet character printed on cell i. It is clear that ℓ bits, where $\ell = 2 + \lceil \log(|Q| + 1) \rceil + \lceil \log(|\Sigma|) \rceil$, are required to represent the information at each cell. One can thus think of these basis vectors as being encoded by strings in $\{0, 1\}^{(2t+1)\ell}$.

The transitions made along any one branch are represented by a sequence of these vectors, where each element of the sequence is the configuration of the machine at a given moment in time. The full state of the AffTM corresponds to an affine combination of such configurations, and the evolution of the AffTM corresponds to affine transformations of these configurations in superposition. We may then simulate the AffTM by a uniform family of affine circuits.

Here, an "affine circuit" (in analogy to quantum circuits) refers to an acyclic network of "affine gates", each of which represents an affine transformation acting on real vectors. We demand that the matrices corresponding to these affine transformations have entries (with respect to the standard basis) that can be computed efficiently, i.e. in poly-time, by an ordinary Turing Machine. We also demand that the description of the circuit can be computed efficiently, and in particular that it contain only a polynomial number of gates.

A specific affine circuit in this family will correspond to the concatenation of t identical sub-circuits, which we denote by B. Each sub-circuit B simulates one time-step of the AffTM **M**. To construct these circuits, each tape cell of the AffTM is associated with $\ell = 2 + \lceil \log(|Q| + 1) \rceil + \lceil \log(|\Sigma|) \rceil$ wires in the circuit, which are sufficient to encode a tuple $(s_i, q_i, a_i) \in \{0, 1, 2\} \times (Q \cup \{\emptyset\}) \times \Sigma$, as described above. We build the sub-circuit B with 3ℓ input wires and 3ℓ output wires, constructed from copies of two gates G and I as follows. We first perform a cascading sequence of $2t-1$ copies of G (whose behaviour we describe below), with each one shifted right by ℓ wires from the preceding one. We then perform $2t+1$ copies of a gate I, in parallel, each acting on ℓ wires. The gate I acting on the ith cell changes the value of s_i with value 2 to 1 and value 1 to 2, leaving a value of $s_i = 0$ alone. It is clear that I is an affine transformation and can be built using $O(t)$ gates whose function is to implement the change in s_i for a specific i. We denote the ith instance of G as G_i. See Fig. 4 for a pictorial representation of B.

The intuitive idea behind this construction is as follows. The 3ℓ inputs to G should be thought of as describing the contents of three consecutive cells of the AffTM, including the information about the position of the
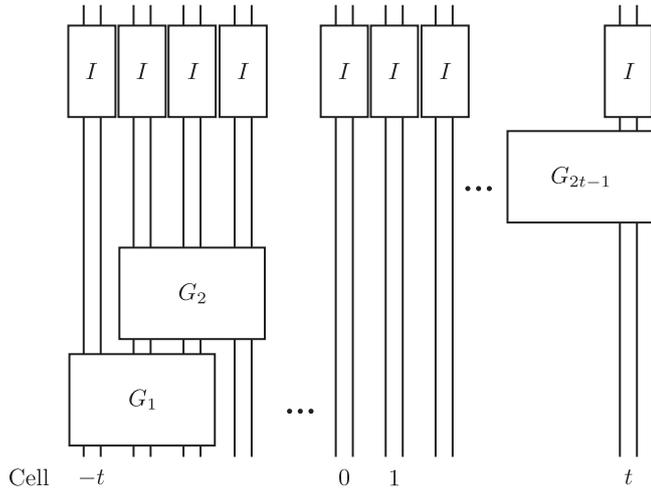
Fig. 4 Sub-circuit B simulating one transition of an affine Turing Machine **M**. This circuit acts on $2t + 1$ "cells", each consisting of ℓ wires and representing the contents of one cell of an affine Turing Machine **M**, the location of the head, and the state of **M**

head. We want G to transform the contents of these cells if the head is at the middle cell and the transition step has not occurred (i.e. $s_i = 1$ with i being the middle cell) according to how the AffTM would transform the contents. Thus we design G to act as follows:

1. For all $v = |s_{i-1}, q_{i-1}, a_{i-1}, s_i, q_i, a_i, s_{i+1}, q_{i+1}, a_{i+1})$ with $s_i \neq 1$, we have $G(v) = v$,

2. For $v' = |0, \varnothing, a_{i-1}, 1, q_i, a_i, 0, \varnothing, a_{i+1})$ we have

$$
\begin{aligned}
G(v') = &\sum_{q',a'} \delta(q_i, a_i, \leftarrow, q', a')|2, q', a_{i-1}, 0, \varnothing, a', 0, \varnothing, a_{i+1}) \\
&+ \sum_{q',a'} \delta(q_i, a_i, \circ, q', a')|0, \varnothing, a_{i-1}, 2, q', a', 0, \varnothing, a_{i+1}) \\
&+ \sum_{q',a'} \delta(q_i, a_i, \rightarrow, q', a')|0, \varnothing, a_{i-1}, 0, \varnothing, a', 2, q', a_{i+1}).
\end{aligned}
\tag{10}
$$

We can think of G as a controlled affine transformation that does nothing if the input has $s_i \neq 1$ and performs the transition step of the AffTM otherwise. (We may extend this to define $G|y) = |y)$ for any other basis state $|y)$, where $y \in \{0, 1\}^{3\ell}$ does not encode a valid tuple $(s_{i-1}, q_{i-1}, a_{i-1}, s_i, q_i, a_i, s_{i+1}, q_{i+1}, a_{i+1})$.) As the configuration of the AffTM is an affine combination of vectors encoding tuples $|s_{-t}, q_{-t}, a_{-t}, \ldots, s_t, q_t, a_t)$, and as we have defined the action of G (when tensored with the identity on cells on which it does not act) on all such vectors, extending linearly uniquely defines G's action on all configurations of the AffTM. Note that some linear combination of vectors with $s_i \neq 1$ can lead to the same output as when G is applied to a vector with $s_i = 1$, so that G may not be reversible. This may be expected, as affine transformations are not reversible in general; nor is there any requirement in the setting of operational theories to realise transformations reversibly.

We construct B using a cascading sequence of G gates, acting on the wires 1 through 3ℓ (representing cells $-t$, $-t + 1$, and $-t + 2$), then on the wires $\ell + 1$ through 4ℓ, then $2\ell + 1$ through 5ℓ, and so forth, as illustrated in Fig. 4. This in effect scans over the contents of the tape of the AffTM **M**, doing nothing in most cases but simulating one of transition of **M** on the triple whose middle cell contains the head at the beginning of the transition. The I gates then flip the value of each s_i, so that the next simulation step can be performed. In this way, B simulates one step of the AffTM.

We describe the initial state of the tape of **M** by setting $a_0 a_1 \cdots a_{n-1} = x_1 x_2 \cdots x_n$ (where $x \in \Sigma^*$ is the input of length n), and setting a_i to the blank symbol for $i < 0$ and $i > n$. We describe the initial head position of **M** by setting $s_0 = 1$ and $s_i = 0$ for $i \neq 0$; similarly we set $q_0 \in Q$ to the initial state of **M** and $q_i = \varnothing$ for $i \neq 0$. This describes the initial state $|s_{-t}, q_{-t}, a_{-t}, \ldots, s_t, q_t, a_t)$ which is the input to the affine circuit. The run time of the simulated machine is t, so by concatenating t instances of B acting on the initial state, we obtain an affine circuit simulating the entire run of **M**, producing a distribution $|\psi_x)$, which is an affine combination of basis vectors $|s'_{-t}, q'_{-t}, a'_{-t}, \ldots, s'_t, q'_t, a'_t)$ representing the final configuration of all of the branches of the AffTM.
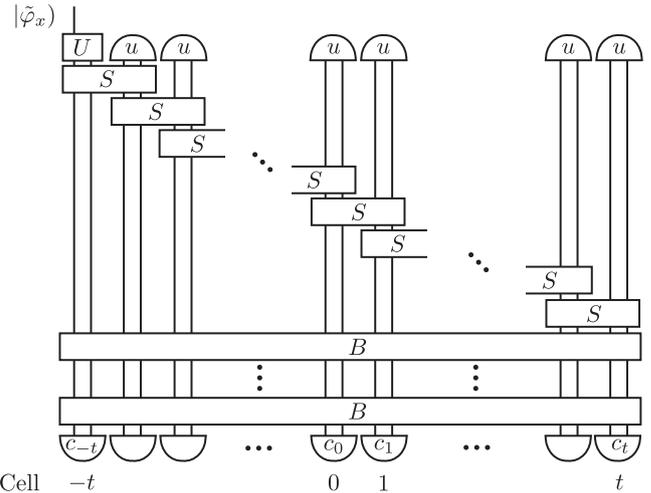
Fig. 5 Schematic illustration of an affine circuit M_n simulating an affine Turing Machine (AffTM) **M** on inputs of length n, which halts in time t. This includes t copies of the gate B, each of which decomposes as the sub-circuit illustrated in Fig. 4 and simulates one transition of **M**. The gate S serves to simulate shifting the head of the AffTM to the leftmost of a pair of cells; the cascading sequence of S gates serves to shift the head to the left-most cell in every branch of the computation, regardless of its final position when **M** halts. The preparations $|c_i)$ represent the initial configuration of the AffTM, preparing basis states $|c_i) = |s_i, q_i, a_i)$ where $s_0 = 1$ and q_0 is the AffTM initial state (and $s_i = 0$, $q_i = \varnothing$ for all $i \neq 0$), and where the symbols a_i represent the symbols written on the ith cell of the tape. The final operations $(u| = \sum_e (e|$ on each cell serves to deterministically erase the information simulating the ith cell; the operation U on the left-most cell similarly erases all information except for the distribution $|\tilde{\varphi}_x)$ representing the weights of the AffTM on the final internal states $|A)$ and $|R)$

As the position of the head in **M** in each branch may be different, we define another gate which will allow us to localise the final state of **M** in a definite subsystem. We define a gate S acting on 2ℓ wires which transforms $|0, q'_i, a'_i, 1, q'_{i+1}, a'_{i+1}) \mapsto |1, q'_{i+1}, a'_{i+1}, 0, q'_i, a'_i)$, and leaves all other basis states unchanged. By performing a cascade of S first on the wires $(2t - 1)\ell + 1$ through $(2t + 1)\ell$ (representing cells $t-1$ and t), then on $2(t - 2)\ell + 1$ through $2t\ell$ (representing cells $t-2$ and $t-1$), and so forth, each standard basis state is mapped to one of the form $|1, \bar{q}, \bar{a}, s'_0, q'_0, a'_0, \ldots)$ for some $\bar{q}$ which is either the accept state A or reject state R. Acting on $|\psi_x)$, this cascade of S gates produces a vector

$$
|\varphi_x) = |1, A)|\varphi_{A,x}) + |1, R)|\varphi_{R,x}).
\tag{11}
$$

By the conditions on the acceptance weight of **M**, the sum $w_{A,x}$ of the coefficients of $|\phi_{A,x})$ satisfies either $w_{A,x} \in [0, \frac{1}{3}]$ or $w_{A,x} \in [\frac{2}{3}, 1]$; the same holds for the sum $w_{R,x}$ of the coefficients of $|\varphi_{R,x})$. Applying the operator $_j(u| = {}_j(0| + {}_j(1|$ on all wires, except for the wires 3 through $\lceil \log(|Q| + 1) \rceil + 2$ representing the final state A or R of **M**, we then obtain a state

$$
|\tilde{\varphi}_x) = w_{A,x}|A) + w_{R,x}|R)
\tag{12}
$$

which is a distribution representing the probability with which **M** accepts x. The entire affine circuit constructed in this way is illustrated in Fig. 5.

The probability to accept is then just the factor in front of the basis state corresponding to the accepting configuration. We may thus simulate **M** by the t-fold application of B on the initial configuration, followed by the cascade of S gates and the application of unit effects described above.

A tomographically local theory

The preceding section shows how to construct a collection of affine circuits that simulate a proper, efficient AffTM. In order to prove Theorem 2, this section constructs in turn a tomographically local operational theory, which can simulate a proper, efficient AffTM. It is important that tomographic locality is satisfied in order that Theorem 2 serves as a converse to Theorem 1. As discussed in ref. [23] theories that do not satisfy tomographic locality may have additional holistic degrees of freedom pertaining to composite systems. Without further constraint, there is

nothing to stop such additional degrees of freedom enabling arbitrarily powerful computation.

It is tempting to suppose that we need only construct an operational theory that includes the affine circuits of the last section. Each of the affine circuits outputs a state given by Eq. (13), with accept and reject weights $w_{A,x}$ and $w_{R,x}$, and it follows from the premise that **M** is a *proper* AffTM that $w_{A,x}, w_{R,x} \in [0, 1]$. Hence if a circuit, representing an experiment in an operational theory, consists of the affine circuit, followed by a final measurement onto $|A\rangle$ and $|R\rangle$, the probabilities for the outcomes are at least guaranteed to be bounded by 0 and 1. Of course, closed circuits formed of arbitrary compositions of the same set of gates are not guaranteed to yield coefficients for measurement outcomes $\in [0, 1]$, hence cannot be assumed to correspond to valid experiments. For this reason, the operational theory would be a non-free theory, with the set of allowed circuits containing those necessary for the simulation of proper, efficient AffTMs, but not allowing arbitrary rearrangements of gates.

Even with the allowance of a non-free theory, however, it is not sufficient to define an operational theory as allowing exactly those circuits constructed above, along with a final accept/reject measurement. Without further structure, such a theory would simply be a theory of elaborate preparations of a 2-dimensional system, whose states define probabilities for the acceptance and rejection outcomes. Additional structure is needed for the theory to satisfy tomographic locality, in such a way that states, transformations and effects correspond to the vectors and matrices that are involved in the construction of the affine circuits.

The idea, therefore, is to allow circuits consisting of the initial part of one of the affine circuits, followed by measurements with outcomes corresponding to the basis states of each wire. If the effects were literally those dual to the basis states, this would suffice for tomographic locality; but the theory would not be well defined, because such effects would not in general yield sensible probabilities for outcomes. We therefore employ a trick, which is to allow only highly noisy versions of these measurements. If we additionally admix a small amount of noise with the final accept/reject measurement, then the theory can be shown to satisfy tomographic locality, to return sensible probabilities for measurement outcomes in all allowed circuits, and to be able to simulate a proper, efficient AffTM with bounded error. The precise construction is as follows.

Let $\{M_n\}_{n \geq 1}$ be the family of affine circuits, simulating a proper AffTM **M** on inputs of length $n \geq 1$. For each n, define types such that each wire gets a type v_n. This allows that the type of system involved can be distinct for each circuit in the family. From hereon, however, we consider a fixed n, suppressing the dependence of the type on n, and writing simply v. Define an *initial segment* of M_n to consist of any fragment that can be completed to the whole circuit M_n by the post-composition of an appropriate sequence of gates (including, as a special case, M_n itself). The closed circuits allowed by the theory consist, for each n, of an initial segment of M_n, followed by measurement devices attached to any dangling wires.

First, for any system type X, there exists a measurement device realising the trivial measurement: the device pointer has only one position, which occurs with certainty. The outcome of this device corresponds to a deterministic effect, and the outcomes of any other measurement will correspond to effects that sum to the same deterministic effect, hence the theory is causal. For systems of type v, the deterministic effect is given by

$$_v(u| = {}_v(0| + {}_v(1|. \tag{13}$$

For a composite system of type X, the deterministic effect $_X(u|$ is given by parallel composition. The deterministic effect may be appended to any dangling wire, following an initial segment of M_n.

Second, we define the measurements that enable local tomography. Define effects

$$
\begin{aligned}
_v(a_0| &= p_v\left(\,_v(0|\right) + \tfrac{(1-p_v)}{2}\left(\,_v(u|\right), \\
_v(a_1| &= p_v\left(\,_v(1|\right) + \tfrac{(1-p_v)}{2}\left(\,_v(u|\right),
\end{aligned}
\tag{14}
$$

where p_v is a parameter to which we return below. These two effects satisfy

$$_v(a_0| + {}_v(a_1| = {}_v(u|,$$

hence may correspond to the two outcomes of a binary measurement on a system corresponding to a single wire. This measurement may be appended to any dangling wire, following an initial segment of M_n.

Finally, there is the accept/reject measurement, which is a joint measurement defined on $\log(|Q| + 1)$ systems of type v. Let the unit effect for such a collection of systems be $(u| = {}_v(u|^{\otimes \log(|Q| + 1)}$, and let $(A|$ and $(R|$ denote the duals of the basis states $|A\rangle$ and $|R\rangle$, representing the accept and reject states (respectively) of the AffTM. The operational theory

will allow a noisy version of the corresponding measurement, with effects given by:

$$(e_{\text{acc}}| := q(A| + (1 - q)(R|, \tag{15}$$

$$(e_{\text{rej}}| := q(R| + (1 - q)(A|, \tag{16}$$

$$(e_{\text{none}}| := (u| - (e_{\text{acc}}| - (e_{\text{rej}}|, \tag{17}$$

with q fixed independently of n, and essentially arbitrary as long as $1 > q > 1/2$. On pain of generating a disallowed circuit, this measurement cannot be appended to an arbitrary initial segment. The measurement can only be performed following an initial segment that is almost the whole of M_n, including at least all of the S gates and the final U gate (see Fig. 5), and can only be performed on the $\log(|Q| + 1)$ systems that are the output of the U gate.

The idea of this construction is that (separately for each value of n, the size of the problem input) the parameter p_v can be chosen small enough that the measurements appearing in an allowed circuit always result in probabilities for outcomes that are bounded between 0 and 1. To see this, consider first those allowed circuits that include noisy tomographic measurements, but do not include the final accept/reject measurement. For these circuits, if $p_v = 0$ then the outcomes of the noisy tomographic measurements each occur, independently, with probability 1/2, regardless of the state. Now consider those allowed circuits that include the final accept/reject measurement, but where the final $_v(u|$ effect on one or more of the other wires has been replaced by noisy tomographic measurements. In this case, the probabilities for the accept and reject outcomes are bounded between q and $1-q$, hence strictly between 1 and 0. It follows that if $p_v = 0$, then the joint probability for either accept or reject, along with any sequence of outcomes for the tomographic measurements, is also strictly between 0 and 1. Now, in the theory under construction, there are only finitely many initial circuit fragments (in the partial construction of a single circuit on inputs of length n) on which to perform measurements. Continuity of the outcome probabilities in the effects therefore ensures that there exists a value $p_v > 0$ such that joint outcome probabilities are contained in the interval [0, 1], for all circuits that can be constructed from systems of type v. Fixing such a value of p_v results in noisy measurements that are sufficient for tomography on system v.

Given a language decided by a poly-time, proper, bounded-error AffTM, the corresponding circuit family in the operational theory will accept yes-instances and reject no-instances with probabilities $\geq (1 + q)/3$. If probabilities $\geq 2/3$ are required, they can be achieved by running several circuits in parallel. The final step in the proof of Theorem 2 is to show how to combine the preceding constructions to describe an operational theory **G** not just for a single language in AWPP, but for the entire class.

As shown above, every problem in AWPP can be solved with bounded error by a proper affine Turing machine (AffTM) which halts in polynomial time. Conversely, any poly-time proper AffTM which has an acceptance weight either $\geq \tfrac{2}{3}$ or $\leq \tfrac{1}{3}$ for all inputs, defines a language $L \in$ AWPP. We then define a theory **G** which simply contains enough devices and system types to simulate every such AffTM, and only these AffTMs. In this theory, each system type is parametrised by a (poly-time, proper, bounded-error) AffTM **M** and an input size $n \geq 1$; and each device is one of the sort described in the previous sections, also parameterised by (**M**, n). The devices $G_{\mathbf{M},n}$, $S_{\mathbf{M},n}$, $I_{\mathbf{M},n}$, and the various preparations and measurements for each system type, may then be used to construct circuits $C_{\mathbf{M},n}$ to simulate the AffTM **M** on inputs of size n; and for each such **M**, there will be a deterministic Turing machine U which can generate $C_{\mathbf{M},n}$ in poly(n) time.

To summarise: for any $L \in$ AWPP, there is a poly-time, proper AffTM **M** which decides L with bounded error, which may be simulated by an affine circuit family $\{M_n\}_{n \geq 1}$. This affine circuit family may be constructed uniformly, by the fact that it simulates an AffTM which halts in polynomial time. The family $\{M_n\}_{n \geq 1}$ may itself be simulated by a uniform circuit family $\{C_{\mathbf{M},n}\}_{n \geq 1}$ consisting of allowed experiments in the theory **G**. Then **G** is a non-free theory in which AWPP $\subseteq$ BGP. Together with Theorem 1, this concludes the proof of Theorem 2.

Promise problems

One might wonder if efficient quantum computation can achieve the bound of Theorem 1. The following complexity-theoretic argument may be considered evidence against such a possibility.

Theorem 3. If PromiseBQP = PromiseAWPP, then

$$NP \subseteq BQP \subseteq AWPP. \tag{18}$$

Here, the classes PromiseBQP and PromiseAWPP are *promise* versions of the classes BQP and AWPP, meaning that they contain *promise* rather than decision problems. A promise problem is a generalization of a decision problem, where the input is promised to belong to a subset of all possible inputs, so that there are disjoint subsets $\Pi_{ACCEPT}, \Pi_{REJECT} \subseteq \Sigma^*$ of inputs to be accepted or rejected (respectively), but which do not exhaust the set of all inputs. If an input belonging to neither Π_{ACCEPT} nor Π_{REJECT} is given to an algorithm for a certain promise problem, no requirements are placed on the output.

While, logically speaking, it could turn out that BQP = AWPP without PromiseBQP = PromiseAWPP, this seems unlikely. Indeed, problems which are often regarded as complete for BQP or AWPP, respectively, are in fact promise problems. Hence, PromiseBQP and PromiseAWPP can be loosely thought of as characterising the power of BQP and AWPP, respectively. It is also believed unlikely[51,60,66] that NP is contained in either BQP or AWPP. Hence Theorem 3 can be regarded as evidence against the assertion that the computational power of quantum theory in the promise problem setting exactly equals PromiseAWPP, and this in turn may be regarded as evidence against the possibility that BQP = AWPP.

The proof of Theorem 3 is as follows.

Proof. Recall that UNIQUE-SAT is the problem of deciding whether a given Boolean formula has exactly one satisfying truth assignment, or no satisfying assignment at all, promised that one of these is the case. It is known that UNIQUE-SAT is contained in PromiseUP, which is a subset of PromiseAWPP.[67]

The Valiant-Vazirani theorem[68] says that if one has an efficient algorithm for solving UNIQUE-SAT in conjunction with the ability to perform random reductions, then one can solve any problem in NP. More precisely, the Valiant-Vazirani theorem says the standard Boolean Satisfiability Problem SAT can be randomly reduced to UNIQUE-SAT.
. Now, if PromiseBQP = PromiseAWPP then UNIQUE−SAT ∈ PromiseBQP, so that there is a uniform family of quantum circuits that solve an instance of the promise-problem UNIQUE-SAT (with no requirements made on inputs which do not respect the promise). However, a crucial point is that, as gates in quantum theory are closed under composition (in our terminology: quantum theory is a *free* operational theory), the output of the algorithm will always result in sensible probabilities, regardless of the input. One can therefore perform the random reduction of Valiant-Vazirani in quantum theory (randomly generating an appropriate instance of SAT, and using this to generate an appropriate experiment of the sort that solves UNIQUE-SAT with bounded error), and run the algorithm many times on each input produced by the reduction to test whether it is a YES instance of UNIQUE-SAT. Performing this reduction many times enables the solution of SAT with bounded error in BQP. It then follows that NP ⊆ BQP, which using Theorem 1 gives NP ⊆ *A*WPP. □

One might wonder why the existence of a non-free theory satisfying BGP = AWPP does not immediately imply NP ⊆ AWPP. The answer is that the theory we have constructed does not necessarily allow the efficient solution of PromiseAWPP problems, since the circuits required to simulate Affine Turing Machines that only have proper behaviour on a subset of inputs are not in the allowed set defined by the theory.

One may then ask: why not construct an operational theory that *does* contain circuits corresponding to simulations of the improper Affine Turing Machines that solve PromiseAWPP problems? In this case, the Valiant-Vazirani reduction does not go through, since the reduction assumes that it is possible to at least run the computation on inputs that do not satisfy the promise; attempting this in the operational theory must be disallowed since it may result in negative probabilities. On a related note, we would argue that such a theory should be excluded on the grounds discussed at the end of the Methods section.

On computation in non-free theories
This section concludes by addressing a certain issue, which might arise with non-free theories: what if an agent can solve a hard problem (say, outside of AWPP) by simply observing whether a certain type of system exists in the universe or not? Or by simply observing whether a given circuit can be constructed or not? This would amount to a form of cheating, somewhat akin to the construction of non-uniform circuits in the classical or quantum cases. If such cheating were possible in a universe described by a non-free theory *G*, this would not contradict the claim that BGP ⊆ AWPP, which is a formal mathematical theorem. But it would undermine the significance of the claim, since the definition of BGP could not be said to accurately capture the set of problems that an agent can efficiently solve by physical actions that the agent can do.

Concerning the first of these possibilities, our answer is that we have not said anything about how difficult it is to determine whether a given type of system exists in the universe or not: we can suppose, e.g., that the universe is infinite, and that given a classical description of an Affine Turing Machine, there is no step-by-step procedure that an agent can follow to determine if a corresponding type of system exists. Hence there is no easy way for an agent to solve the (uncomputable) problem of whether a given Affine Turing Machine is proper or not.

Concerning the second possibility, if a particular type of system is employed, the theory we construct does not allow a hard problem to be solved by finding out if a circuit is allowed or not. A closed circuit is allowed if it corresponds to an implementation of the corresponding Affine Turing Machine (or an initial segment thereof, with subsequent noisy measurements), and this is easy to check with a classical computation, hence the observation that a given circuit can or cannot be constructed cannot solve any harder problem. We argue therefore that we can rule out cheating in the theory described. More generally, one might require of a non-free theory something like the following: there exists a deterministic Turing machine, such that if the input is a description of a circuit, then on the promise that all the devices in the circuit exist in the universe, the machine decides in poly time whether the circuit is allowed or not. If the input is not a valid circuit, or contains devices that do not exist, then the output is unconstrained.

Note added—While writing up the current work we became aware of the related but independent work,[69] on the characterization of AWPP.

DATA AVAILABILITY
Data sharing not applicable as no datasets were generated or analysed in the current paper.

AUTHOR CONTRIBUTIONS
All authors contributed equally to the current paper.

ADDITIONAL INFORMATION
Competing interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES
1. Aaronson, S. & Arkhipov, A. The computational complexity of linear optics. *In Proc. of the Forty-third Annual ACM Symposium on Theory of Computing (STOC 2011)*, 333–342 (San Jose, CA, USA, 2011).
2. Shor, P. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM J. Sci. Statist. Comput.* **26**, 1484 (1997).
3. Bremner, M. J., Jozsa, R. & Shepherd, D. J. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proc. R. Soc. London A* **467**, 459–472, 2126 (2010).
4. Bremner, M. J., Montanaro, A. & Shepherd, D. J. Average-case complexity versus approximate simulation of commuting quantum computations. *Phys. Rev. Lett.* **117**, 080501 (2016).
5. Howard, M., Wallman, J., Veitch, V. & Emerson, J. Contextuality supplies the magic for quantum computation. *Nature* **510**, 351 (2014).

6. Vidal, G. Efficient classical simulation of slightly entangled quantum computations. *Phys. Rev. Lett.* **91**, 147902 (2003).

7. Hoban, M. J., Wallman, J. J. & Browne, D. E. Generalised Bell inequality experiments and computation. *Phys. Rev. A* **84**, 062107 (2011).

8. Datta, A., Shaji, A. & Caves, C. Discord and the power of one qubit. *Phys. Rev. Lett.* **100**, 050502 (2008).

9. Stahlke, D. Quantum interference as a resource for quantum speedup. *Phys. Rev. A* **90**, 022302 (2014).

10. Vedral, V. The elusive source of quantum speed-up Found. *Physics* **40**, 8 (2010).

11. Brodutch, A. Discord and quantum computational resources. *Phys. Rev. A* **88**, 022307 (2013).

12. Van den Nest, M. Universal quantum computation with little entanglement. *Phys. Rev. Lett.* **110**, 060504 (2013).

13. Abrams, D. S. & Lloyd, S. Nonlinear quantum mechanics implies polynomial-time solution for NP-complete and sharp-P problems. *Phys. Rev. Lett.* **81**, 3992 (1998).

14. Aaronson, S. Quantum computing, postselection, and probabilistic polynomial-time. *Proc. R. Soc. A* **461**, 3473–3482 (2005).

15. Chiribella, G., D'Ariano, G. M. & Perinotti, P. Probabilistic theories with purification. *Phys. Rev. A* **81**, 062348 (2010).

16. Barrett, J. Information processing in generalised probabilistic theories. *Phys. Rev. A* **75**, 032304 (2007).

17. Hardy, L. Reformulating and reconstructing quantum theory. Preprint at arXiv: quant-ph/1104.2066v3 (2011).

18. Masanes, L. & Mueller, M. A derivation of quantum theory from physical requirements. *New J. Phys.* **13**, 063001 (2011).

19. Lee, C. M. & Selby, J. H. A no-go theorem for theories that decohere to quantum mechanics. *Proc. R. Soc. A* **474**, 20170732 (2018).

20. Barnum, H., Barrett, J., Leifer, M. & Wilce, A. A generalized no-braodcasting theorem. *Phys. Rev. Lett* **99**, 240501 (2007).

21. Chiribella, G., D'Ariano, G. M. & Perinotti, P. Informational derivavtion of quantum theory. *Phys. Rev. A* **84**, 012311 (2011).

22. Masanes, L., Mueller, M. P., Augusiak, R. & Perez-Garcia, D. Existence of an information unit as a postulate of quantum theory. *PNAS* **110**, 16373 (2013).

23. Lee, C. M. & Barrett, J. Computation in generalised probabilistic theories. *New J. Phys.* **17**, 083001 (2015).

24. Lee, C. M. *Bounds on Computation from Physical Principles*. DPhil. Thesis, University of Oxford (2016).

25. Lee, C. M. & Hoban, M. J. Bounds on the power of proofs and advice in general physical theories. *Proc. R. Soc. A* **472**, 2190 (2016).

26. Lee, C. M. & Hoban, M. J. The information content of systems in general physical theories. *EPTCS* **214**, 22–28 (2016).

27. Fortnow, L. and Rogers, J. Complexity limitations on quantum computation. Preprint at arXiv:cs/9811023v1 (1998).

28. Hardy, L. Quantum theory from five reasonable axioms. Preprint at arXiv:quant-ph/0101012 (2001).

29. de la Torre, G., Masanes, L., Short, A. J. & Müller, M. P. Deriving quantum theory from its local structure and reversibility. *Phys. Rev. Lett.* **109**, 090403 (2012).

30. D'Ariano, G. M., Manessi, F. & Perinotti, P. Determinism without causality. *Phys. Scr.* **T163**, 014013 (2014).

31. Popescu, S. & Rohrlich, D. Quantum nonlocality as an axiom. *Found. Phys.* **24**, 379–385 (1994).

32. Short, A. J. & Barrett, J. String nonlocality: a trade-off between states and measurements. *New J. Phys.* **12**, 033034 (2010).

33. Massar, S., Pironio, S. & Pitalúa-Garca, D. Hyperdense coding and superadditivity of classical capacities in hypersphere theories. *New J. Phys.* **17**, 113002 (2015).

34. Spekkens, R. W. In defence of the epistemic view of quantum states; a toy theory. *Phys. Rev. A* **75**, 032110 (2007).

35. Popescu, S. Nonlocality beyond quantum mechanics. *Nat. Phys.* **10**, 264–270 (2014).

36. Navascués, M., Guryanova, Y., Hoban, M. J. & Acín, A. Almost quantum correlations. *Nat. Commun.* **6**, 6288 (2015).

37. Al-Safi, S. & Short, A. Simulating all non-signalling correlations via classical or quantum theory with negative probabilities. *Phys. Rev. Lett.* **111**, 170403 (2013).

38. Oas, G., Acacio de Barros, J. & Carvalhaes, C. Exploring non-signalling polytopes with negative probability. *Phys. Scr.* **2014**, 014034 (2014).

39. Abramsky, S. & Brandenburger, A. An Operational Interpretation of Negative Probabilities and No-signalling Models. *In Horizons of the Mind. A Tribute to Prakash Panagaden*, 5 (Springer, Cham, 2014).

40. van Dam, W. Implausible consequences of superstrong nonlocality. Preprint at arXiv:quant-ph/0501159 (2005).

41. Barnum, H., Mueller, M. P. & Ududec, C. Higher-order interference and single system postulates for quantum theory. *New J. Phys.* **16**, 123029 (2014).

42. Niestegge, G. Conditional probability, three-slit experiments and the Jordan structure of quantum mechanics. *Adv. Math. Phys.* **2012**, 156573 (2012).

43. Henson, J. Bounding quantum contextuality with lack of third-order interference. *Phys. Rev. Lett.* **114**, 220403 (2015).

44. Fritz, T. et al. Local orthogonality as a multipartite principle for quantum correlations. *Nat. Commun.* **4**, 2263 (2013).

45. Pawlowski, M. et al. Information causality as a physical principle. *Nature* **461**, 1101 (2009).

46. Barrett, J., Hardy, L. & Kent, A. No signalling and quantum key distribution. *Phys. Rev. Lett.* **95**, 010503 (2005).

47. Lee, C. M. & Hoban, M. J. Towards device-independent information processing on general quantum networks. *Phys. Rev. Lett.* **120**, 020504 (2018).

48. Lee, C. M. Device-independent certification of non-classical measurements via causal models. Preprint at arXiv:1806.10895 (2018).

49. Aaronson, S. Quantum computing and hidden variables II: the complexity of sampling histories. Preprint at arXiv:quant-ph/0408119 (2004).

50. Nielsen, M. A. & Chuang, I. L. *Quantum computation and Quantum information* (Cambridge University press, 2000).

51. Bennett, C., Bernstein, E., Brassard, G., and Vazirani, U. Strengths and weaknesses of quantum computing. Preprint at arXiv:quant-ph/9701001v1 (1997).

52. Gross, D., Mueller, M., Colbeck, R. & Dahlsten, O. All reversible dynamics in maximal non-local theories are trivial. *Phys. Rev. Lett.* **104**, 080402 (2010).

53. Al-Safi, S. & Short, A. Reversible dynamics in strongly non-local boxworld systems. Preprint at arXiv:quant-ph/1312.3931 (2013).

54. Ududec, C., Barnum, H. & Emerson, J. Three slit experiments and the structure of quantum theory. *Found. Phys.* **41**, 396–405 (2011).

55. Lee, C. M. & Selby, J. H. Deriving Grover's lower bound from simple physical principles. *New J. Phys.* **18**, 093047 (2016).

56. Lee, C. M. & Selby, J. H. Generalised phase kick-back: the structure of computational algorithms from physical principles. *New J. Phys.* **18**, 033023 (2016).

57. Lee, C. M. & Selby, J. H. Higher-order interference in extension of quantum theory. *Found. Phys.* **47**, 89–112 (2017).

58. Niestegge, G. Quantum teleportation and Grover's algorithm without the wavefunction. Preprint at arXiv:1611.02926 (2016).

59. Pashayan, H. Wallman, J. J. & Bartlett, S. D. Estimating outcome probabilities of quantum circuits using quasiprobabilities. Preprint at arXiv:quant-ph/1503.07525 (2015).

60. Fenner, S. PP-lowness and simple definition of AWPP. *Theory of Comput. Syst.* **36**, 199–212 (2003).

61. Fenner, S., Fortnow, L., Kurtz, S. and Li, L. An oracle builders toolkit. *Proceedings of the 8th IEEE Structure In Complexity Theory Conference* (1993).

62. de Beaudrap, N. On computation with 'probabilities' modulo k. Preprint at arXiv: cs.CC/1405.7381v2 (2014).

63. Yao, A.C.C. Quantum circuit complexity. *In Proceedings of 1993 IEEE 34th Annual Foundations of Computer Science,* 352–361 (IEEE, 1993).

64. Savage, J.E. Computational work and time on finite functions. *J. ACM* **17**, 660–674 (1972).

65. Schnorr, C. The network complexity and turing machine complexity of finite functions. *Acta Inform.* **7**, 95–107 (1976).

66. Kobler, J., Schoning, U. & Toran, J. Graph isomorphism is low for PP. *Comput. Complex.* **2.4**, 301–330 (1992).

67. Fenner, S., Fortnow, L. & Kurtz, S. Gap-definable counting classes. *J. Comput. Syst. Sci.* **48**, 116–148 (1994).

68. Valiant, L. & Vazirani, V. NP is as easy as detecting unique solutions. *Theor. Comput. Sci.* **47**, 85–93 (1986).

69. de Campos, C. P., Stamoulis, G. & Weyland, D. A. Structured view on weighted counting with relations to counting, quantum computation and applications. Preprint at arXiv:1701.06386v1 (2017).