The computational landscape of general physical theories

Jonathan Barrett,1 Niel de Beaudrap,1 Matty J. Hoban,1,2 and Ciarán M. Lee3∗

1University of Oxford, Department of Computer Science, Wolfson Building, Parks Road, Oxford OX1 3QD, UK.
2School of Informatics, University of Edinburgh, 10 Crichton Street, Edinburgh EH8 9AB, UK
3Department of Physics and Astronomy, University College London, Gower Street, London WC1E 6BT, UK

The emergence of quantum computers has challenged long-held beliefs about what is efficiently computable given our current physical theories. However, going back to the work of Abrams and Lloyd, changing one aspect of quantum theory can result in yet more dramatic increases in computational power, as well as violations of fundamental physical principles. Here we focus on efficient computation within a framework of general physical theories that make good operational sense. In prior work, Lee and Barrett showed that in any theory satisfying the principle of tomographic locality (roughly, local measurements suffice for tomography of multipartite states) the complexity bound on efficient computation is AWPP. This bound holds independently of whether the principle of causality (roughly, no signalling from the future) is satisfied. In this work we show that this bound is tight: there exists a theory satisfying both the principles of tomographic locality and causality which can efficiently decide everything in AWPP, and in particular can simulate any efficient quantum computation. Thus the class AWPP has a natural physical interpretation: it is precisely the class of problems that can be solved efficiently in tomographically-local theories. This theory is built upon a model of computing involving Turing machines with quasi-probabilities, to wit, machines with transition weights that can be negative but sum to unity over all branches. In analogy with the study of non-local quantum correlations, this leads us to question what physical principles recover the power of quantum computing. Along this line, we give some computational complexity evidence that quantum computation does not achieve the bound of AWPP.

There is ever-growing evidence that quantum computers are more powerful than classical computers [1–3]. 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” [45–49] but the debate is far from resolved [50–52]. 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 [5]. This has even motivated the belief that quantum theory is an “island” within the space of all possible theories; alter quantum mechanics and we obtain dramatic consequences [6].

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 generalised probabilistic theories provides us with a suitable framework for the study of information processing based on operational principles [10] [12] [13] [62] [63]. That is, we can make statements about the limits and power of information processing without referring explicitly to quantum theory. For example, a no-broadcasting theorem can be proven from fundamental properties that reasonable operational probabilistic theories should satisfy [61]. Features thought unique to quantum theory (as opposed to classical physics) can be seen to be ubiquitous within these theories. This then begs the question of what fundamental principles uniquely single out quantum physics from these myriad possibilities. Indeed, starting from various frameworks of generalised probabilistic theories there have been many derivations of quantum theory from information processing principles (e.g. Refs. [11] [13] [64]).

Recently a circuit-based model of computation has been defined and studied in a broad operationally-defined framework for physical theories [21] [23] [24]. 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). An important result of [21] was to show 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 [16].

An important open problem is to determine whether
the bound of AWPP is tight for all possible theories. If there existed a theory that could decide all problems in AWPP, then it 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 which can decide everything in AWPP. We may then consider this theory as a “foil” theory, which can be used to deepen our understanding of the limitations of quantum computers. Remarkably, this foil theory is constructed from a computational model using quasi-probabilities, i.e. an affine combination of weights assigned to particular events. Furthermore, this theory satisfies both tomographic locality and causality. This then naturally motivates the study of what minimal set of information principles are needed to recover the power of quantum computation.

To outline the paper, in Sec. II we present the framework of computation in general physical theories setting, in particular defining efficient computation in such a theory, and give computational complexity bounds on this notion of computation. In Sec. III we describe the theory that attains the bound of AWPP, generating the theory from a circuit construction related to a family of computational models called Affine Turing Machines. In Sec. IV we also give some computational complexity evidence that quantum computers will not achieve AWPP. Finally, in Sec. V we conclude with some remarks about the problem of determining the computational complexity of quantum computers from basic informational principles, in analogy with the characterisation of quantum non-locality.

I. THE FRAMEWORK

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 generalised probabilistic theories \[10, 11, 13, 21\], where the primitive notions are operational in nature.

Informally, a theory in this framework specifies a set of laboratory devices, that 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”. We consider the input and output ports to correspond to physical systems. Each physical system has a particular type, denoted \( A, B, C, \ldots \). Devices may be connected by taking the output port of some device and feeding it to the input port of another device, provided that in doing so we do not introduce cycles of input/output dependencies. Moreover, the types of the output and input systems much match. An “experiment” consists of such a closed circuit of devices, in which every output port has been fed to the input port of some other device in this way. When a device is used in an experiment, the classical pointer comes to rest in one of a number of positions, indicating that a particular outcome has occurred. The theory defines a joint probability on the pointer outcomes for all of the devices in the circuit, corresponding to the outcome probabilities of the experiment represented by the closed circuit.

We elaborate on the above summary as follows. Laboratory devices can be classified into preparations, which have no input ports, transformations, which have both input and output ports, and measurements, which have no output ports. Informally, each use of a preparation device outputs a “physical system” in some particular state, determined by the variety of device used. After a given system is prepared in some state, it can pass through a transformation device which can alter the system and its state, in a possibly non-deterministic manner which is indicated by the classical pointer of the transformation. Finally, the system can enter a measurement device, on which the final resting place of the classical pointer denotes the measurement outcome.

By standard operational arguments \[10, 13\], each system gives rise to a finite dimensional real vector space \( V_A \) which contains the set of states as a subset. Moreover, transformations and single measurement outcomes—called effects—act linearly on this vector space \[12\]. Hence each state has a representation as a vector living in this real vector space, each transformation as a matrix acting on this vector space, and each effect as a vector in the dual space \[10, 11\].

We denote by \( |s \rangle_A \) the vector corresponding to a state of system \( A \). Passing a system through a transformation device results in an output state \( |s' \rangle_B = T^B_B |s \rangle_A \), where \( T^B_B \) is the matrix corresponding to the transformation. When a system of type \( B \) enters a measurement device, the dual vector representing the effect corresponding to a specific measurement outcome will be denoted \( B(e) \). The probability of preparing a system in state \( |s \rangle_A \), having transformation \( T^A_B \) applied to it, and getting measurement outcome \( B(e) \) on entering a measurement device corresponds to

\[
B(e(T^A_B |s \rangle_A))
\]

One can define a notion of causality for theories in the framework: the probabilities of present experiments are independent of future measurement choices \[10\]. This requirement is equivalent to the existence of a unique deterministic effect (meaning that the device it corresponds to only has one possible outcome) for each system, denoted \( B(u) \), such that \( \sum_e B(e) = B(u) \), where the sum is over all outcomes of a particular measurement, and \( B(u) |s \rangle_B \leq 1 \) for all states. Note however, that consistent theories can be constructed which have more than one deterministic effect and which violate causality \[14\].

In this work we assume that the vector space arising from the composite of multiple systems corresponds to the tensor product of vector spaces of the component systems. That is, \( V_{AB} = V_A \otimes V_B \). This requirement is implied by the principle of tomographic locality \[13, 12\].
which says that multipartite states can be uniquely specified by the results of local measurements on each component system. In particular, this implies that the matrix representation of a transformation corresponding to simultaneously acting with transformation \( T_B \) on subsystem \( A \) of composite system \( AB \) and transformation \( T_B \) on subsystem \( B \) is given by \( T_A \otimes T_B \), where \( \otimes \) is the vector space tensor product.

An explicit example serves to illustrate the framework. In standard finite dimensional quantum theory for instance, systems correspond to 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, effects to POVM elements with the deterministic effect being given by the identity matrix, and transformations to quantum instruments, i.e. a collection of completely positive maps summing to a completely positive and trace-preserving map. Quantum states are elements of the real vector space of Hermitian matrices with the vector spaces for distinct system types composing via the vector space tensor product.

### A. Free and non-free theories

In the standard definition of a generalised theory \([10–13 21]\), a theory specifies a set of laboratory devices from which one can construct closed circuits by connecting preparation devices to transformation and measurement devices, and assigns a probability distribution over the possible outcomes of each closed circuit. Moreover, the set of devices—and device outcomes—is closed under such sequential and parallel composition. For the purposes of this paper, we refer to such theories as "free" generalised probabilistic theories.

One can consider a modified definition of a generalised probabilistic theory in which a theory specifies a set of devices, a set of allowed closed circuits which can be built from those devices, and assigns a probability distribution over the outcomes of allowed closed circuits. Note that this modified definition is slightly more general than the one standardly discussed in the literature, as it only assigns a probability distribution to the set of allowed closed circuits specified by the theory. Our 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.

Such "non-free" theories will be the focus of the current work. Note that even in non-free theories, states are represented by vectors in a real vector space, transformations by matrices acting on this space, and effects by vectors in the dual space \([22]\). However, composing states, transformations, and effects in sequence and parallel to form closed circuits may only result in a valid probability when the closed circuit is allowed by the theory.

### B. Computation

The class of "yes/no" problems that a quantum computer can solve efficiently is denoted by \( \text{BQP} \) and much research has been concerned with how large this class is. At present, the best known upper bound is \( \text{BQP} \subseteq \text{AWPP} \), where \( \text{AWPP} \) is a (slightly obscure) classical complexity class, known to be contained in \( \text{PP} \), hence in \( \text{PSPACE} \). This class will be formally defined in Sec. III A. That is to say: a quantum computer cannot efficiently solve any problem outside \( \text{AWPP} \), but it is unknown whether a quantum computer can solve every problem in \( \text{AWPP} \).

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. A polynomially sized uniform circuit family is a set of closed circuits \( \{ C_x \} \) 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 \( |(M - \tilde{M})_{ij}| \leq \epsilon \), in time polynomial in \( \log(1/\epsilon) \).
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 a description of an experiment, whose devices produce classical outcomes (corresponding to the particular state prepared, transformation applied, or measurement outcome observed). The theory defines a joint probability for these outcomes, which we may use to perform computation as follows. Denoting the string of observed outcomes by \( z \), we define the final output of the computation to be 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

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

where the sum ranges over all possible outcome strings \( z \) of the circuit \( C_x \) for which \( a(z) = 0 \).
II. ACHIEVING THE UPPER BOUND

The main result of this work is the construction of a non-free theory, satisfying causality and tomographic locality, that has exactly the power of this upper bound.

Theorem 1. For any (free or non-free) theory $G$ satisfying tomographic locality, the following holds:

$$BGP \subseteq AWPP \subseteq PP \subseteq PSPACE.$$ 

The class $BGP$ contains problems for which an efficient quantum solution is unknown. Notable among these is the Graph Isomorphism problem, which asks for an efficient procedure to determine if two given graphs are equivalent. It is unknown whether a quantum computer can solve the Graph Isomorphism problem, and so our result provides a theory which can act as a “foil” to deepen our understanding of the limitations of quantum computers. Moreover, the promise version of $AWPP$ contains the Unique Satisfiability Problem (or UNIQUE-SAT), which asks if a given Boolean formula has either a satisfying assignment, or no satisfying assignment at all—promised one of these two cases is true. This is a very important problem whose complexity is closely related to the class of $NP$-complete problems.

We now provide an intuitive sketch of this construction, but defer the formal definitions and proofs to Sec. III. We start by introducing a quasi-probabilistic model of computation, taking the form 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 and provide a schematic illustration in Fig. 1. We show that the class of problems which can be efficiently solved by this model, with bounded error, perfectly captures the class $AWPP$. We then construct uniform poly-size circuits, in which the gates are certain affine transformations, that can simulate—and be simulated by—this affine Turing Machine, and hence $AWPP$. 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. We prove that these closed circuits correspond to closed circuits in a causal and tomographically local non-free theory, thereby proving Theorem 2.

One might wonder if efficient quantum computation can achieve the bound of Theorem 1. We now present a complexity-theoretic argument which may be considered weak evidence against such a possibility.

Theorem 3. If $\text{PromiseBQP} = \text{PromiseAWPP}$, then one necessarily has

$$NP \subseteq BQP \subseteq AWPP.$$ 

Here, the classes $\text{PromiseBQP}$ and $\text{PromiseAWPP}$ are promise versions of the classes $BQP$ and $AWPP$, meaning that they contain promise rather than decision problems. An example of a decision problem is deciding whether some bit string $x$ is in a particular language; either $x$ is in the language, or it is not. A promise problem is a generalisation of a decision problem where the input is promised to belong to a subset of all possible inputs, so that there are disjoint subsets $\Pi_{\text{ACCEPT}} \subseteq \Sigma^*$ of inputs to be accepted or rejected (respectively), but which do not exhaust the set of all inputs. If an input belonging to neither $\Pi_{\text{ACCEPT}}$ nor $\Pi_{\text{REJECT}}$ is given to an algorithm for a certain promise problem, no requirements are placed on the output, i.e. the algorithm is allowed to output anything.

While $\text{PromiseBQP} = \text{PromiseAWPP}$ may hold independently of whether $BQP = AWPP$ holds, these statements are at least conceptually related. Intuitively, if one of these statements appears unlikely, the other one should also be considered unlikely—although to a lesser degree. Indeed, problems which are complete for $BQP$ (i.e. the hardest problems in $BQP$) are in fact promise problems. Hence, $\text{PromiseBQP}$ and $\text{PromiseAWPP}$ can be loosely thought of as characterising the power of $BQP$ and $AWPP$ respectively. It is believed unlikely that $NP$ is contained in either $BQP$ or $AWPP$. 

---

**Fig. 1: Schematic illustration of an Affine Turing Machine.**
AWPP. This can, in some sense, be taken as evidence against the assertion that the computational power of quantum theory (in the promise problem setting) exactly equals PromiseAWPP.

III. PROOF OF THEOREM 2

The proof of Theorem 2 is contained in the following sections. As informally discussed in above, we introduce a quasi-probabilistic model of computation, taking the form of a Turing Machine with quasi-probabilistic transition weights, and consider the class of problems which can be efficiently solved by this model. We refer to this model as an Affine Turing Machine. We show that this model perfectly captures the class AWPP. We then show that one can construct uniform poly-size circuits, where the gates are certain affine transformations, that can simulate—and be simulated by—this (efficient) quasi-probabilistic Turing Machine. Thus, this affine circuit construction also captures the power of AWPP. We finish by showing that these circuits give rise to a non-free theory that is causal and tomographically local.

A. Definition of AWPP

Let $\Sigma$ be a finite set of symbols, e.g. $\Sigma = \{0, 1\}$, and let $\Sigma^*$ be the set of all finite sequences over $\Sigma$ (commonly referred to as strings). For a string $x \in \Sigma^*$, we let $|x|$ denote its length. A gap function over $\Sigma$ 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 $N$, where $N$ takes no more than $T(|x|)$ computational steps on input $|x|$ for some polynomial $T$ on whatever input $x$ it is given.

Fenner [17] Theorem 1.3 characterised 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

$$
    x \in L \implies \frac{2}{3} \leq \frac{g(x)}{2^p(|x|)} \leq 1; \quad (1a)
$$

$$
    x \notin L \implies 0 \leq \frac{g(x)}{2^p(|x|)} \leq \frac{1}{4}. \quad (1b)
$$

Combining this with [17] Theorem 3.1, more generally we have $L \in$ AWPP if and only if

$$
    x \in L \implies \frac{2}{3} \leq \frac{g(x)}{h(|x|)} \leq 1; \quad (2a)
$$

$$
    x \notin L \implies 0 \leq \frac{g(x)}{h(|x|)} \leq \frac{1}{4}. \quad (2b)
$$

for a gap-function $g$ and any poly-time computable function $h : \mathbb{N} \to \mathbb{N}$. While the original definition of AWPP [19] 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 characterisations of both Eqns. 1 and 2 in our results.

B. 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 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 $M$ whose branches all halt in a finite number of steps, the acceptance weight $\alpha_M(x)$ of $M$ on an input $x$ is the total weight of the accepting paths on input $x$. An AffTM $M$ is proper if $0 \leq \alpha_M(x) \leq 1$ for all inputs, and that it decides a language $L$ with bounded error if furthermore $\frac{2}{3} \leq \alpha_M(x) \leq 1$ for $x \in L$, and $0 \leq \alpha_M(x) \leq \frac{1}{4}$ 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.

1. Solving AWPP problems with an affine Turing machine

For $L \subseteq$ AWPP, let $g : \Sigma^* \to \mathbb{Z}$ be a gap-function satisfying Eqns. 1 for some polynomial $p$. Also let $N$ 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 $N$ on its input. By standard results [19], we may require that $N$ have the same number of non-deterministic transitions at each step, which we denote by $N \geq 1$, and that all computational branches of $N$ have the same length on input $x$. We suppose that each transition of $N$ is associated with some label $t \in \{1, 2, \ldots, N\}$: 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. $M$ 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 \( \beta \in \{1, 2, \ldots, N\}^{|x|} \) indexing some computational branch of \( N \). In these branches, \( M \) simulates the computational branch of \( N \) whose transitions are indexed by \( \beta \).

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 \( \rho(|x|) \) non-deterministic transitions with weight \( \frac{1}{2} \), in which it writes bits \( \delta_1, \delta_2, \ldots, \delta_{\rho(|x|)} \) on the tape, thereby sampling a string \( \delta \in \{0, 1\}^{\rho(|x|)} \) uniformly at random. If \( \delta = 11 \cdots 1 \) ensures that the acceptance weight is \( |\alpha_{M}(x)| = g(x)/2^{\rho(|x|)} \). 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.

2. 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 transition weights of \( Q \), \( T \in O(\text{poly}(n)) \) the running time of \( M \) on an input of length \( n \), and \( m > 0 \) be an integer such that \( 2^m \geq M^T \) and \( 2^m \geq (|u|M)^T \) for all transition weights \( u \) of \( M \); it follows that \( m \in O(T) \). 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 \( \Omega \in Q \) 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 \( \Omega \) 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 \( \Omega \) if and only if the simulated branch is one in which \( M \) rejects.

4. \( N \) determines whether to accept or reject, treating \( c_{m-1}c_{m-2} \cdots c_0 \) 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 \( 2^MT \) 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 \( \alpha(x) \) be the total weight of those accepting branches of \( M \) with positive weight, \( \alpha_-(x) \) 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 \( \alpha(x) := \alpha_+(x) - \alpha_-(x) \) 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

\[
g_0(x) = 2^MT \left[ \alpha_+(x) + \rho_+(x) - \alpha_-(x) - \rho_-(x) \right] = 2^MT,
\]
as \( \alpha(x) + \rho(x) = 1 \). In the branches where \( a = 1 \), the sign of the contribution from rejecting branches is negated, so that

\[
g_1(x) = 2^MT \left[ \alpha_+(x) - \rho_+(x) - \alpha_-(x) + \rho_-(x) \right] = 2^MT \left[ 2\alpha(x) - 1 \right],
\]
again using \( \alpha(x) + \rho(x) = 1 \). Then \( g(x) = 4^MT \alpha(x) \), and for \( h(n) = 4^MT \), 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 \( 0 \leq g(x)/h(|x|) \leq \frac{1}{7} \) according to whether \( x \in L \) or \( x \not\in L \); then \( L \in \text{AWPP} \) as well.

C. Constructing affine circuits

The next step towards Theorem 2 is to construct a family of circuits that can simulate an arbitrary AffTM. The final step will then be to show that the collection of all such circuit families is available in a specific non-free theory that satisfies tomographic locality and causality.

The construction of the circuits is based on that used by Yao in [44] to construct quantum circuits that simulate a quantum Turing Machine (and also on that of
for circuits that simulate a probabilistic Turing machine). As before, let $M$ be a proper and efficient AffTM with alphabet $\Sigma$, set of states $Q$ and transition amplitudes $\delta(q, a, \tau, q', a') \in \mathbb{C}$ with $\tau \in \{\leftarrow, \circ, \rightarrow\}$; the symbols $\leftarrow$, $\circ$ and $\rightarrow$ are interpreted as the tape head of the AffTM moving to the left, moving to the right, and remaining stationary. Here $\delta$ is the transition weight of $M$ to change to state $q'$, print $a'$ on the tape and move according to $\tau$, 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: $\sum_{\tau, q', a'} \delta(q, a, \tau, q', a') = 1$ for all $q \in Q$, $a \in \Sigma$.

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

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

where the index $-t \leq i \leq t$ denotes the $i$th 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 $q_i \in Q \cup \{0\}$, where $q_i = 0$ if and only if $s_i = 0$; and $a_i \in \Sigma$ denotes the alphabet character printed on cell $i$. It is clear that $\ell$ 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 $(s_i, q_i, a_i) \in \{0, 1, 2\} \times (Q \cup \{0\}) \times \Sigma$, as described above. We build the sub-circuit $B$ with $3\ell$ input wires and $3\ell$ 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 $\ell$ wires from the preceding one. We then perform $2t+1$ copies of a gate $I$, in parallel, each acting on $\ell$ wires. The gate $I$ acting on the $i$th 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 $i$th instance of $G$ as $G_i$. See Fig. 2 for a pictorial representation of $B$.

The intuitive idea behind this construction is as follows. The $3\ell$ 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 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 $v = |s_{i-1}, q_{i-1}, a_{i-1}, s_i, q_i, a_i, s_{i+1}, q_{i+1}, a_{i+1}\rangle$ with $s_i \neq 1$, we have $G(v) = v$,

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

$$G(v') = \sum_{q', a'} \delta(q_i, a_i, \leftarrow, q', a')|2, q', a_{i-1}, 0, 0, a', 0, 0, a_{i+1}\rangle + \sum_{q', a'} \delta(q_i, a_i, \circ, q', a')|0, 0, a_{i-1}, 2, q', a', 0, 0, a_{i+1}\rangle + \sum_{q', a'} \delta(q_i, a_i, \rightarrow, q', a')|0, 0, a_{i-1}, 0, 0, a', 2, q', a_{i+1}\rangle.$$

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 $y ∈ \{0, 1\}^{3ℓ}$ does not encode a valid tuple $(s_{t−1}, q_{t−1}, a_{t−1}, s_{t}, q_{t}, a_{t}, s_{t+1}, q_{t+1}, a_{t+1})$. 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 $s_i ≠ 1$ 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 GPTs, 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 $ℓ+1$ through $4ℓ$, then $2ℓ+1$ through $5ℓ$, and so forth, as illustrated in Fig. 2

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 $a_0 = 1 \cdots a_n = x_1x_2 \cdots x_n$ (where $x ∈ Σ^*$ is the input of length $n$), and setting $a_1$ 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 $i ≠ 0$; similarly we set $q_0 ∈ Q$ to the initial state of $M$ and $q_i = 0$ for $i ≠ 0$. 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 $|ψ_x⟩$, 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.

As the position of the head of $M$ 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 $|0, q_0′, q_0′, 0, q_0′, a_0⟩$ to $|1, q_0′, a_0′, 0, q_0′, a_0⟩$, and leaves all other basis states unchanged. By performing a cascade of $S$ first on the wires $(2ℓ−1)ℓ+1$ through $(2ℓ+1)ℓ$ (representing cells $−t$ and $t$), then on $2(ℓ−2)ℓ+1$ through $2ℓ$ (representing cells $t−2$ and $t−1$), and so forth, each standard basis state is mapped to one of the form $|1, q, a′, s′, q′, 0, q′, a′, 0, q′, a′, \ldots⟩$ for some $q′$ which is either the accept state $A$ or reject state $R$. Acting on $|ψ_x⟩$, this cascade of $S$ gates produces a vector $|ψ_x⟩ = |1, A⟩|ψ_{A,x}⟩ + |1, R⟩|ψ_{R,x}⟩$.

By the conditions on the acceptance weight of $M$, the sum $w_{R,x}$ of the coefficients of $|ψ_{R,x}⟩$ satisfies either $w_{A,x} ∈ [0, 1/3]$ or $w_{A,x} ∈ [2/3, 1]$; the same holds for the sum $w_{R,x}$ of the coefficients of $|ψ_{R,x}⟩$. Applying the operator $j(u) = j(0) + j(1)$ on all wires, except for the wires $3$ through $|\log(|Q| + 1)| + 2$ representing the final state $A$ or $R$ of $M$, we then obtain a state $|ψ_x⟩ = w_{A,x}A⟩ + w_{R,x}R⟩$

which is a distribution representing the probability with which $M$ accepts $x$. The entire affine circuit constructed in this way is illustrated in Fig. 3.

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.

D. A tomographically local theory

The construction in the preceding section resulted in a collection of closed (affine) circuits which correspond to
the probability that the AffTM accepts or rejects given a specific input. To work towards Theorem 2, we now show that such an affine circuit family can be simulated in a non-free theory that satisfies tomographic locality and causality.

As discussed in Sec. I, the correspondence between probabilities and closed circuits in a physical theory gives rise to a vector space structure for the states, transformations, and effects. To find a theory which "represents" an affine circuit such as the ones described above, one must ensure that the vector spaces which emerge from that theory have the same structure as that of the affine circuits, so that the states, effects and transformations in the theory correspond to the real vectors and matrices involved in the construction of the affine circuits.

This requirement is stronger than it may at first appear, for the following reason. For a proper AffTM $\mathbf{M}$, the affine circuits to simulate $\mathbf{M}$ merely represent preparations of distributions over $|A\rangle$ and $|R\rangle$, despite the fact that we may describe them as involving intermediate stages involving distributions over basis vectors $|s_{-t}, q_{-t}, a_{-t}, \ldots, s_t, q_t, a_t\rangle$ in a $2^{(2t+1)}$ dimensional vector space. Without some appropriate additional structure, a physical theory representing these affine circuits would be a theory of elaborate preparations of a 2-dimensional system, corresponding to the acceptance and rejection outcomes. To ensure that we may attribute non-trivial vector spaces to the input and output wires of each gate in the affine circuit in an appropriate way—representing the intermediate stages of the affine circuits—we could allow effects (0), (1) on each wire. However, if some of the affine transitions involve "improper" distributions (where some coefficients have negative weights), this will not produce a theory in which all outcomes result in well-defined probabilities. Furthermore, while we know that the gates $G$, $S$, and $I$ may be composed with appropriate preparations and effects to give rise to probability distributions (from the premise that $\mathbf{M}$ is a "proper" AffTM), we have no particular bounds on the coefficients of arbitrary compositions of these affine gates.

To solve this problem, the theory we construct to represent these affine circuits is taken to be a non-free theory, in which the gates are not "freely composable" in the theory but are instead limited to certain allowed compositions. At the same time, to associate each wire in a circuit with the same vector space in the physical theory as it has in the affine circuit, we introduce a collection of "noisy" measurements. These measurements, consisting of "noisy" effects that can be applied to single systems, are constructed to wash out any possible negative or super-normalised weights arising in the measurement process and to allow tomography to be performed locally on any preparation.

The allowed closed circuits $M_n$ in our non-free theory will consist of those corresponding to the closed affine circuits from the previous section which simulate AWPP computations on inputs of length $n$, and—making use of the noisy measurements—those sufficient for tomography to be performed locally on any state preparation. In our theory, preparations consist of semi-constructed versions of the circuits $M_n$, possibly followed by measurements on some of the systems.

1. System types, states, and transformations

Let $\{M_n\}_{n \geq 1}$ be the family of affine circuits, simulating a proper AffTM $\mathbf{M}$ on inputs of length $n \geq 1$, described in Sec. III C. At any specific step along one branch of the computation, the contents of cell $i$ of the AffTM tape is described by a bit string $c_i \in \{0, 1\}^t$. For each $n$, we will define types such that each wire gets a type $\nu_{n}$, each parallel composition of $\ell$ wires (corresponding to a bit string $c_i \in \{0, 1\}^\ell$ and representing one cell of the AffTM) gets a type $\nu_n$, and the parallel composition of $(2t+1)\ell$ wires (representing the memory of a single copy of the circuit $M_n$) gets a type $\nu_{tn}$. For simplicity, we will here only discuss the states, transformations, and effect for a fixed $n$; all other types will be constructed in the same manner. We will henceforth write $\nu_n$, $\nu_n$, and $\nu_{tn}$ simply as $\nu$, $A$, and $\bar{A}$.

We define vectors $\nu(d_i)$ dual to states $|c_i\rangle_A$ as follows: $\nu(d_i|c_i\rangle_A = \delta_{c_i,d_i}$ for $c_i, d_i \in \{0, 1\}^\ell$. We will use these dual vectors as the starting point for constructing noisy measurements in the following section. Given the construction of the real vectors from Sec. III C, the state corresponding to the entire contents of the tape at the very start of the AffTM computation corresponds to $|\bar{b}\rangle_{\bar{A}} = \bigotimes_{i=1}^{2t+1} |c_i\rangle_A$, for $\bar{b} \in \{0, 1\}^{(2t+1)}$. The vectors dual to these states are then $\bar{A}(|e\rangle = \bigotimes_{i=1}^{2t+1} \nu(d_i|$, for $d_i$ varying over $\{0, 1\}^\ell$.

The non-free theory we consider only allows a limited set of conceivable compositions of devices as valid circuits. Specifically, the only allowed closed circuits are those consisting of those compositions of the gates $G$, $I$, and $S$ which form an "initial segment" of the affine circuit $M_n$ defined in Sec. III C (Fig. 3), followed by measurements on any remaining unmeasured systems. We define an initial segment of $M_n$ to consist of the first $m_i$ gates to act on each wire $i \in \{1, 2, \ldots, (2t+1)\}$, where the number of gates $m_i \geq 0$ may vary for each $i$, and may include all or none of the gates acting on wire $i$. (Thus, an initial segment is one which may be completed to the whole circuit $M_n$ by post-composition of an appropriate sequence of gates.) The preparations allowed by the theory are precisely those which correspond to such initial segments of $M_n$, again possibly followed by measurements on some of the systems. The objective is to define this theory in such a way that it satisfies local tomography.

We now describe how to construct such a theory. Note that states of system $\bar{A}$ include ones of the form $B \cdots B | b \rangle \bar{A}$, 

$$t \geq T \geq 0 \text{ times}$$

where $B$ is the circuit fragment depicted in Fig. 2. As
$\mathbf{M}$ is allowed assign negative weights to certain transitions. $A(e)B\cdots B(b)\bar{A}$ may be negative. Thus effects of the form $A(e)$, for $e \in \{0,1\}^{(2^i+1)}$, may not be admissible effects on all states of system $\bar{A}$. Instead, we use versions of the effects $A(e)$ with added “noise”—consisting of $A(e)$ precomposed with a mixing operator—to wash out negative or super-normalised weights without losing the vector space structure of the individual system.

Before we introduce noisy measurements, note that the effect $\nu(u)$ on system $\nu$ given by

$$\nu(u) = \nu(0) + \nu(1)$$

is allowed, because the result of applying this effect to any affine state preserves the property of the coefficients adding to 1, and so will yield an affine state on a smaller system. By parallel composition, we may also define the effects $\nu(u)$ on any composite system $X$; in particular, it is easy to show

$$\left(x(u|T_X\right)|s\rangle = x(u|s\rangle)$$

for any affine transformation $T$ we may perform on any subsystem $X$, as $\nu(u)$ is a left $+1$-eigenvector of any such operator $T$. Thus, $x(u)$ is a deterministic effect on systems of type $X$.

2. Noisy measurements

We construct effects for the noisy measurements, so that when applied at the end of an initial segment of $M_n$, the measurement statistics are not negative values resulting from the affine transformations in the circuit. We define two effects $\nu(a_0) = \nu(0)|D_\nu$ and $\nu(a_1) = (1)|D_\nu$, in terms of a suitable stochastic $2 \times 2$ operator $D_\nu$. The role of $D_\nu$ is to provide a veil of propriety, so that for any preparation $|s\rangle$ on $k$ bits included in the theory, all of the coefficients of the matrix $(D_\nu^{\otimes k})|s\rangle$ are in the interval $[0,1]$. If we extend $x(a_i|s) = \bigotimes_{x \in X} (a_i|s)$ to define noisy effects on a subsystem $X$, it follows that $(a_i|s) \in [0,1]$ for all $e \in \{0,1\}^{|X|}$.

Our strategy is to construct $D_\nu = \frac{1-p_\nu}{2} [1 1] + p_\nu \mathbb{I}$, where $\mathbb{I}$ is the $2 \times 2$ identity matrix. As $p_\nu$ decreases, the effect of $D_\nu$ is to decrease the “bias” of the distribution over outcomes $\nu(a_0), \nu(a_1)$ for any single system of type $\nu$. The “bias” is just the difference between the two weights described by the inner products $(a_i|x)$ for $i \in \{0,1\}$, where $|x\rangle$ is a state of type $\nu$. If this bias is in the interval $[0,1]$, then the weights form a probability distribution and we can consider $\nu(a_0), \nu(a_1)$ to be a well-defined measurement acting on system $\nu$.

We prove that there exists a non-zero value of $p_\nu$ small enough to ensure that $D_\nu$ provides a veil of propriety for all allowed preparations of the system (which in our theory correspond to partial constructions of the affine circuits to simulate AffM, followed by some number of measurements on some of its subsystems). It is clear that if $p_\nu = 0$ then the bias will be 0, regardless of which state $|s\rangle_A$ we apply the measurement to and which system we measure, so that we get sensible outcome probabilities. Recall that effects are functions which take all allowed states to allowed outcome probabilities; and note that in our non-free theory, there are only finitely many valid preparations and finitely many systems (in the partial construction of a single circuit on inputs of length $n$) on which to perform a measurement. Hence, continuity of the outcome probabilities in the effects ensures that there exists a value $p_\nu > 0$ such that the bias is contained in the interval $[0,1]$, for all preparations and measured systems. Fixing such a value of $p_\nu$, results in noisy measurements that are sufficient for tomography on system $\nu$.

Thus, we obtain two distinct effects for systems of type $\nu$, which allows us to associate a non-trivial vector space structure to each wire. One can readily check that

$$\nu(a_0) + \nu(a_1) = \nu(u|D_\nu) = \nu(u).$$

Thus the deterministic effect arising from summing over noisy effects from each noisy measurement is the same as that introduced in the previous section.

3. Accepting/rejecting measurement

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, defined at the end of Sec. [III C]. The accept/reject measurement effects are defined as follows:

$$\bar{A}(e_{\text{acc}}) := \left[\nu(u)\otimes (A \otimes \nu(u)^{\otimes \log |S|})\right] \otimes \bigotimes_{i=1}^{2^t} A(u),$$

$$\bar{A}(e_{\text{rej}}) := \left[\nu(u)\otimes (R \otimes \nu(u)^{\otimes \log |S|})\right] \otimes \bigotimes_{i=1}^{2^t} A(u),$$

which apply the effect $(A)$ and $(R)$ to the appropriate part of the first cell representing the state of the AffTM $M$, and $\nu(u)$ to all other wires. We also define a third effect $\bar{A}(e_{\text{none}}) = |\bar{A}(\nu(a)) - \bar{A}(e_{\text{acc}}) - \bar{A}(e_{\text{rej}})|$ to define a set of effects which sum to $\bar{A}(u)$.

Recall that the configuration of the AffTM machine at step $t$ is such that the sum of the amplitudes of the accepting paths and the rejecting paths are both bounded within $[0,1]$, and sum to 1. Hence, $\bar{A}(e_{\text{acc}})$ and $\bar{A}(e_{\text{rej}})$ define valid effects, and the result of applying $\bar{A}(e_{\text{none}})$ at the end of the affine circuit will in fact always result in the scalar 0. Then we may define a three-valued measurement $\{\bar{A}(e_{\text{acc}}), \bar{A}(e_{\text{rej}}), \bar{A}(e_{\text{none}})\}$ whose effects sum to $\bar{A}(u)$. Performing the three-valued measurement at the end of the circuit simulating the affine circuit will yield a probability distribution $(w_{A,x}, w_{R,x}, 0)$ which indicates the result of the computation.

This measurement, and the noisy measurements that are allowed for the other initial segments of the affine circuit, are all of the allowed measurements in the theory; and in each case the sum of the effects for those measurements add to $(u)$. Then $(u)$ is the unique deterministic effect, and so the theory we are constructing is causal.
4. Allowed closed circuits

As discussed at the start of this section, the allowed closed circuits in the theory are those corresponding to initial segments of the circuits \( M_n \) simulating AffTMs—including the complete circuits \( M_n \) themselves—with noisy measurements or the deterministic effect on any remaining free wires. The set of closed circuits involving only noisy measurements ensures one can perform full local tomography of all the preparations in the theory, ensuring the theory inherits the vector space structure of the affine circuits.

This implies, among other things, that vector spaces arising in the theory have a tensor product structure; we have thus constructed a causal and tomographically local non-free theory. Furthermore, given that we can efficiently compute a time \( t \) by which \( M \) has halted, the circuit \( M_n \) may be assumed to have an efficiently computable depth, so that it is easy to determine whether a given circuit is an initial segment of \( M_n \). This yields a non-free theory to simulate \( M \).

E. An operational theory for all of AWPP

The final step in the proof of Theorem 2 is to show how to combine the preceding constructions to describe a (non-free) operational theory \( G \) not just for a single language in AWPP, but for the entire class.

As we show in Sec. IIIB 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 \( \geq \frac{2}{3} \) or \( \leq \frac{1}{3} \) for all inputs, defines a language \( L \in \text{AWPP} \). 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 parameterised by a (polynomial, 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 \( G_{M, n}, S_{M, n}, I_{M, n} \), and the various preparations and measurements for each system type, may then be used to construct circuits \( C_{M, n} \) 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 \( C_{M, n} \) in \( \text{poly}(n) \) time.

To summarise: for any \( L \in \text{AWPP} \), there is a polynomial-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 \( \{C_{M, n}\}_{n \geq 1} \) consisting of allowed experiments in the theory \( G \). Then \( G \) is a non-free theory in which \( \text{AWPP} \subseteq \text{BQP} \). Together with Theorem 1, this concludes the proof of Theorem 2.

IV. PROOF OF THEOREM 3

In this section we present a proof of Theorem 3.

Proof. Recall that UNIQUE-SAT is the problem of determining 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 \( \text{PromiseUP} \), which is a subset of \( \text{PromiseAWPP} \). The Valiant-Vazirani theorem 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 \( \text{NP} \). (More precisely, the Valiant-Vazirani theorem says the standard Boolean Satisfiability Problem SAT can be randomly reduced to UNIQUE-SAT.) Now, if \( \text{PromiseBQP} = \text{PromiseAWPP} \) then UNIQUE-SAT \( \in \text{PromiseBQP} \), so that there is a uniform family of quantum circuits which 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, 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. By performing this reduction many times, we may then solve SAT with bounded error in \( \text{BQP} \). It then follows that \( \text{NP} \subseteq \text{BQP} \), which using Theorem 1 gives \( \text{NP} \subseteq \text{AWPP} \). 

V. DISCUSSION

A. On computation in non-free theories

We have described a theory \( G \), within the framework of Generalized Probabilistic Theories, which satisfies tomographic locality and allows the solution of any \( L \in \text{AWPP} \), providing a converse to the results of [21]. 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.

Before discussing broader implications, we address an issue that might arise with non-free theories: what if an agent can solve a hard problem (say, outside of AWPP) simply by observing whether a certain type of system exists in the universe or not? Similarly: might an agent be able to solve problems outside of AWPP by simply observing whether a given circuit is an allowed experiment? This would amount to a form of ‘cheating’, similar in some ways to the construction of non-uniform circuits in
the classical or quantum cases, but is consistent with an operational account of the limits of agents in the theory. If such ‘cheating’ were possible in a universe described by a non-free theory $G$, this would not contradict the claim that $BGP \subseteq AWPP$, which is a formal mathematical theorem about uniformly specifiable families of allowed experiments. 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 operational means.

Concerning the first possibility above, of computation by testing for the existence of devices: 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, for instance, 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 a non-free theory does not admit an easy way for an agent to solve the (uncomputable!) problem of whether a given affine Turing machine is proper or not.

The second possibility, of an agent solving difficult problems by testing whether an experiment is allowed, is more problematic. It is easy to show that for any PromiseAWPP problem, there is an AffTM which may not be proper, but which halts in polynomial time and produces a bounded-error “proper” distribution for inputs which satisfy the promise; and it is easy to apply the constructions of Sections III.C and III.D to produce non-free theories to simulate such an AffTM, on those inputs for which $M$ does produce a proper distribution. Consider the problem UNIQUE-SAT $\in$ PromiseAWPP, of determining whether or not an instance of Boolean Satisfiability has a satisfying assignment, provided the promise that it has most one such assignment. It is not difficult to show that evaluating whether this promise holds, is itself an NP-complete problem. Then, if $N$ is the AffTM corresponding to the PromiseAWPP algorithm for UNIQUE-SAT, we might conceive of an enterprising experimenter who realises that they can solve NP-complete problems, by testing whether a given instance of SAT maps to an allowed experiment in the non-free theory to simulate $N$. As it is considered unlikely that NP is contained in AWPP, this raises the issue of whether $BGP \subseteq AWPP$ adequately captures the notion of efficient computation in an operational theory.

We do not have a complete account of how the experimenter can be prevented from obtaining computational power from testing the boundaries of an arbitrary non-free theory. However, for the non-free theories which we have explicitly described to solve problems in AWPP, we may show that such probing does not provide any power beyond AWPP. Consider attempts by an experimenter to test the boundaries of the AWPP-hard theory $G$ which we have constructed. A closed circuit $C$ is an allowed experiment if it simulates an initial segment of a circuit $C_{M,n}$ simulating some affine Turing Machine $M$ (or an initial segment thereof, with subsequent noisy measurements), on systems of a single explicitly specified type, which is parameterised by $(M, n)$ for some polytime proper AffTM $M$ and input-size $n \geq 1$. Whether or not $C$ is allowed is easy to check with a classical computation from $M$ and $n$. Hence the observation that a given circuit can or cannot be constructed cannot solve any harder problem. It also seems likely that a tighter definition of “a non-free theory” could be found, which explicitly prevents an agent from being able to solve difficult problems by probing the boundaries of a theory, while still permitting a theory which is AWPP-hard [65]. We leave the demonstration of such a definition as an open problem.

B. Conclusion and outlook

An interesting feature of the AWPP-complete theory $G$ constructed in this paper is that it satisfies the principle of causality. The main result of [21] 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 are viewed in the context of the set of non-signalling theories [32]. This set consists of theories satisfying the no-signalling principle, ranked according to the strength of their correlations (quantified by the violation of certain Bell Inequalities) [32]. Quantum theory is ranked above classical theory, but there exists a theory colloquially known as “boxworld” [12, 13] which has the strongest correlations consistent with the no-signalling principle and is thus ranked above quantum theory. In the current paper, we considered the set of theories satisfying tomographic locality and ranked them according to the power of efficient computation. As quantum computers can efficiently simulate classical ones, classical theory is ranked below quantum theory, but here we showed the existence of a theory with the strongest possible computational power and is hence ranked above quantum theory. In Fig. 4 we schematically represent this analogy between the sets correlations satisfying the non-signalling conditions, and the computational complexity classes of theories satisfying tomographic locality, along with the quantum and classical cases for each.

Moreover, Refs. [27–29] have 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
FIG. 4: On the left-hand-side is the schematic of the set of all non-signalling probability distributions (correlations) obtained in a Bell test, with the sets of quantum and classical correlations strictly contained inside (as is the case in general). On the right-hand-side is another schematic of 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 quantum and classical theory contained in this class. Note that we can only conjecture that each of these containments is strict, but there evidence to support this conjecture (see, for example, [1, 3] for the containment of classical computation within quantum, and Theorem 3 for the containment of quantum computation within AWPP).

and the computational landscape of general theories.

Many attempts at providing reasonable physical principles that uniquely characterise the set of quantum correlations as a subset of the set of all non-signalling correlations have been made [33, 36, 38, 39]. These principles, while not fully capturing the exact quantum boundary [57], have deepened our understanding of quantum correlations and provided connections between physical principles and information-theoretic advantages. Insights garnered from these connections have also lead to the development of Device-Independent Cryptography. 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 characterised by some set of physical principles? Such a characterisation 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 characterisation would be to find the minimal set of physical principles that imply the quadratic speed-up over classical computation offered by Grover's search algorithm [7]. This speed-up is optimal for quantum computers [40], 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 [51-57], 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 [30]. 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 [30] 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.

The theory constructed in this paper—while having the maximal computational power of any tomographically local theory—does not exhibit many physically interesting features. This feature is again reminiscent of Boxworld, which is severely restricted in its computational ability (at least for the case of reversible computation) [11, 12] as well as in its entangling dynamics [31]. Having a more refined approach to constructing theories in this landscape may allow us to investigate non-locality, or other notions of physical relevance, along with the computational power of these theories, which would deepen our understanding of how computation and other information-theoretic advantages are connected.

Finally, the distinction introduced in this paper between free and non-free theories appears to be important for the study of computation in general physical 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 were borne out, it could shed light on which physical features give rise to the quantum speed-up.

Note added — While writing up the current work we became aware of the related but independent work [60], on a similar characterisation of AWPP.
Acknowledgements

CML thanks J. Selby for useful discussions. We acknowledge support from the EPSRC, the EPSRC National Quantum Technology Hub in Networked Quantum Information Technologies, an FQXi Large Grant, the Wiener-Anspach Foundation, and the project Quantum Causal Structures, funded by the John Templeton Foundation.

[1] S. Aaronson and A. Arkhipov, The Computational Complexity of Linear Optics. Proc. of the Forty-third Annual ACM Symposium on Theory of Computing (STOC 2011), pp. 333-342, 2011.
[2] P. Shor, Polynomial-Time Algorithms for Prime Factorization and Discrete Logarithms on a Quantum Computer. SIAM J. Sci. Statist. Comput. 26, 1484, 1997.
[3] M. J. Bremner, R. Jozsa, and D. J. Shepherd, Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, rspa20100301, 2010.
[4] M. J. Bremner, A. Montanaro, and D. J. Shepherd, Average-case complexity versus approximate simulation of commuting quantum computations, Phys. Rev. Lett. 117, 080501 (2016).
[5] D. S. Abrams and S. Lloyd, Nonlinear Quantum Mechanics Implies Polynomial-Time Solution for NP-Complete and Sharp-P Problems. Phys. Rev. Lett. 81, 3992, 1998.
[6] S. Aaronson, Quantum computing, postselection, and probabilistic polynomial-time. Proc. R. Soc. A, 461, 3473-3482, 2005.
[7] M. A. Nielsen and I. L. Chuang, Quantum computation and Quantum information. Cambridge University press, 2000.
[8] J. Barrett, L. Hardy, and A. Kent, No Signalling and Quantum key Distribution. Phys. Rev. Lett 95, 010503, 2005.
[9] S. Aaronson, Quantum Computing and Hidden Variables II: The Complexity of Sampling Histories. arXiv:quant-ph/0408119, 2004.
[10] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Probabilistic theories with purification. Phys. Rev. A 81, 062348, 2010.
[11] G. Chiribella, G. M. D’Ariano, and P. Perinotti, Informational derivation of Quantum Theory. Phys. Rev. A 84, 012311, 2011.
[12] J. Barrett, Information processing in generalised probabilistic theories. Phys. Rev. A 75 No. 3, 032304, 2007.
[13] L. Hardy, Reformulating and reconstructing quantum theory. arXiv:quant-ph/1104.2066v3, 2011.
[14] G. M. D’Ariano, F. Manessi, and P. Perinotti, Determinism without causality, Phys. Scr. T163 (2014) 014013.
[15] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom. Found. Phys. Volume 24, Issue 3, pp. 379-385, 1994.
[16] L. Fortnow and J. Rogers, Complexity limitations on quantum computation. arXiv:cs.CC/9811023, 1998.
[17] S. Fenner, PP-lowness and simple definition of AWPP. Theory of Computing Systems, Volume 36, Issue 2, 2003.
[18] J. Kobler, U. Schoning, and J. Toran, Graph isomorphism is low for PP, Computational Complexity 2.4 (1992): 301-330.
[19] S. Fenner, L. Fortnow, S. Kurtz, and L. Li, An oracle builders toolkit. Proceedings of the 8th IEEE structure in complexity theory conference, 1993.
[20] N. de Beaudrap, On computation with ‘probabilities’ modulo k. arXiv:cs.CC/1405.7381v2.
[21] C. M. Lee and J. Barrett, Computation in generalised probabilistic theories. New J. Phys. 17 (2015) 083001.
[22] C. M. Lee, Bounds on computation from physical principles. DPhil Thesis, University of Oxford, 2016.
[23] C. M. Lee and M. J. Hoban, Bounds on the power of proofs and advice in general physical theories. Proc. R. Soc. A. Vol. 472. No. 2190, 2016.
[24] C. M. Lee and M. J. Hoban, The information content of systems in general physical theories. EPTCS 214, 2016, pp. 22-28.
[25] L. Valiant and V. Vazirani, NP is as easy as detecting unique solutions, Theoretical Computer Science 47: 85-93, 1986.
[26] S. Fenner, L. Fortnow, and S. Kurtz, Gap-definable counting classes, Journal of Computer and System Sciences Volume 48, Issue 1, 1994, Pages 116-148.
[27] S. W. Al-Safi and A. J. Short, Simulating all non-signalling correlations via classical or quantum theory with negative probabilities. Phys. Rev. Lett. 111, 170403, 2013.
[28] G. Oas, J. Acacio de Barros and C. Carvalhaes, Exploring non-signalling polytopes with negative probability. Phys. Scr. 014034, 2014.
[29] S. Abramsky, and A. Brandenburger, An Operational Interpretation of Negative Probabilities and No-Signalling Models, Horizons of the Mind: A Tribute to Prakash Panagaden, Springer, pages 59–75, 2014.
[30] H. Pushayan, J. J. Wallman and S. D. Bartlett, Estimating outcome probabilities of quantum circuits using quasiprobabilities. arXiv:quant-ph/1503.07525, 2015.
[31] A. J. Short and J. Barrett, String nonlocality: A trade-off between states and measurements. New J. Phys 12, 033034, 2010.
[32] S. Popescu, Nonlocality beyond quantum mechanics. Nature Physics 10, 264-270, 2014.
[33] W. van Dam, Implausible consequences of superstrong nonlocality, arXiv:quant-ph/0501159 2005.
[34] H. Barnum, M. P. Mueller and C. Ududec, Higher-order interference and single system postulates for quantum theory. New J. Phys. 16, 123029, 2014.
[35] G. Niestegge, Conditional probability, three-slit experiments and the Jordan structure of quantum mechanics. Advances in Mathematical Physics 156573, 2012.
[36] J. Henson, Bounding quantum contextuality with lack of third-order interference. Phys. Rev. Lett. 114, 220403, 2015.
[37] M. Navascués, Y. Guryanova, M. J. Hoban and A. Acín, Almost quantum correlations. Nature Communications 6, 6288, 2015.
[38] T. Fritz, A. B. Sainz, R. Augusiak, J. Bohr Brask, R.
Chaves, A. Leverrier and A. Acín, Local orthogonality as a multipartite principle for quantum correlations. Nature Communications 4, 2263, 2013.

[39] M. Pawlowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter and M. Zukowski, Information causality as a physical principle. Nature 461, 1101, 2009.

[40] C. Bennett, E. Bernstein, G. Brassard, and U. Vazirani, Strengths and weaknesses of quantum computing. arXiv:quant-ph/9701001, 1997.

[41] D. Gross, M. Mueller, R. Colbeck, and O. Dahlsten, All reversible dynamics in maximal non-local theories are trivial. Phys. Rev. Lett. 104 080402, 2010.

[42] S. W. Al-Safi and A. J. Short, Reversible dynamics in strongly non-local boxworld systems. arXiv:quant-ph/1312.3931, 2013.

[43] C. Ududec, H. Barnum and J. Emerson, Three slit experiments and the structure of quantum theory. Foundations of Physics, 41, 3, pages 396–405, 2011.

[44] A. Yao, Quantum circuit complexity. FOCS, 1993.

[45] M. Howard, J. Wallman, V. Veitch and J. Emerson, Contextuality supplies the magic for quantum computation. Nature 510, 351 (2014).

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

[47] M. J. Hoban, J. J. Wallman and D. E. Browne, Generalized Bell inequality experiments and computation. Phys. Rev. A 84, 062107 (2011).

[48] A. Datta, A. Shaji and C. Caves, Discord and the power of one qubit. Phys. Rev. Lett. 100, 050502 (2008).

[49] D. Stahlke, Quantum interference as a resource for quantum speedup. Phys. Rev. A 90, 022302, 2014.

[50] V. Vedral, The elusive source of quantum speed-up. Found. Physics, 40, 8, 2010.

[51] A. Brodutch, Discord and quantum computational resources. Phys. Rev. A 88, 022307 (2013).

[52] M. Van den Nest, Universal Quantum Computation with Little Entanglement. Phys. Rev. Lett. 110, 060504 (2013).

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

[54] C. M. Lee and J. H. Selby, Deriving Grover’s lower bound from simple physical principles. New J. Phys. 18 (2016) 093047.

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

[56] C. M. Lee and J. H. Selby, Higher-order interference in extension of quantum theory. Foundations of Physics, Volume 47, Issue 1, pp 89–112 (2017).

[57] G. Niestegge, Quantum teleportation and Grover’s algorithm without the wavefunction. arXiv:1611.02926, 2016.

[58] J. Savage, Computational work and time on finite functions. Journal of ACM, 17, 1972.

[59] C. Schnorr, The network complexity and Turing machine complexity of finite functions. Acta Informatica, 7, 1976.

[60] C. P. de Campos, G. Stamoulis, and D. Weyland, A Structured View on Weighted Counting with Relations to Counting. Quantum Computation and Applications. arXiv:1701.06386v1, 2017.

[61] H. Barnum, J. Barrett, M. Leifer, and A. Wilce, A generalized no-broadcasting theorem. Phys. Rev. Lett 99.240501, 2007.

[62] L. Masanes, M. Mueller, A derivation of quantum theory from physical requirements, New J. Phys. 13 063001 2011.

[63] C. M. Lee and J. H. Selby, A no-go theorem for theories that decohere to quantum mechanics, arXiv preprint arXiv:1701.07449, 2017.

[64] L. Masanes, M. Mueller, R. Augusiak, D. Perez-Garcia, Existence of an information unit as a postulate of quantum theory, PNAS 110 41 16373, 2013.

[65] One approach might be to equip any non-free theory with deterministic Turing machine $V$, which takes as input a description of a closed circuit. In such a description, the devices would be described using (and efficiently computed from) labels, but not all labels would necessarily correspond to devices of the theory. On any circuit description in which all labels correspond to devices, we then require $V$ to halt in polynomial time, and to decide whether the circuit is both well-typed and allowed in the theory; but for other inputs we make no commitments as to the performance of $V$. In this way we may prevent the experimenter from getting super-polynomial power from testing the boundaries of such a non-free theory, without requiring $V$ to solve any uncomputable problem.