INVESTIGATIONS IN QUANTUM COMPUTING
Causality and Graph Isomorphism

Thesis by
David Eugene Beckman

Advisor
John Preskill

In Partial Fulfillment of the Requirements
for the Degree of
Doctor of Philosophy

California Institute of Technology
Pasadena, California

2004

(Defended May 14, 2004)

Acknowledgements

First and foremost, I'd like to thank my advisor John Preskill, for immensely interesting discussions and for the patience of a Buddhist in awaiting the completion of this manuscript. In addition, I'd like to thank Michael Nielsen for especially useful discussions on many topics. I'd also like to thank Dorit Aharonov and Amalavoyal Chari for helpful discussions regarding graph isomorphism; and Charlene Ahn, John Cortese, Jim Harrington, Alexei Kitaev, Andrew Landahl, Barry Simon, and Clint White for their comments and discussions regarding causality and causal operations.

Also deserving of recognition are all those who, by encouraging comments or pointed questions, helped to convince me to finally complete this thesis. These are too numerous to list here, but special mention should be made of my parents, who mostly restricted themselves to encouraging comments, and of Matt Matuszewski, who felt free to ask pointed questions. Finally, I would like to thank the management of Toyon Research Corporation for their patience and flexibility in allowing me the time to complete and defend my dissertation.

INVESTIGATIONS IN QUANTUM COMPUTING

Causality and Graph Isomorphism

by

David Eugene Beckman

In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Abstract

In this thesis I explore two different types of limits on the time complexity of quantum computation—that is, limits on how much time is required to perform a given class of quantum operations on a quantum system. Upper limits can be found by explicit construction; I explore this approach for the problem of determining whether two graphs are isomorphic. Finding lower limits, on the other hand, usually requires appeal to some fundamental principle of the operation under consideration; I use this approach to derive lower limits placed by the requirements of relativistic causality on the time required for implementation of some nonlocal quantum operations. In some situations these limits are attainable, but for other physical spacetime geometries we exhibit classes of operations which do not violate relativistic causality but which are nevertheless not implementable.

Contents

List of Figures

Chapter 1

Introduction

1.1 Why quantum computation?

Research in classical complexity theory, the branch of computer science which studies the power of algorithms and the limitations of computers, has generally followed one of two complementary approaches. The first and most obvious line of research is the development of efficient algorithms to solve classes of problems. This is the most straightforward method of determining how difficult a problem is, and, in particular, whether it can be solved with the allotted resources. But this approach, of course, only provides *upper* bounds on resource requirements; an as-yet-undiscovered algorithm for the problem may require fewer resources than the best algorithm currently known.

Finding nontrivial *lower* bounds on the resources required to solve a given problem often requires more subtlety. The theories of computability and complexity (briefly discussed in §1.2.2 below) have been developed to place limits on whether a given problem can be solved by a well-defined algorithm and, if so, what resources are required for the solution. For apparently difficult problems, in which significant effort has not produced an efficient solution algorithm, these formalisms can be a useful way of quantifying how hard the problem is. Even for tractable problems, it may be useful to know whether more efficient algorithms are possible in principle.[1]

[1] In fact, few nontrivial bounds on the difficulty of solving practical problems are known; in many cases all that can be done is to reduce a problem which seems hard to a special case of another problem that is also believed to be hard. Even this is valuable information, though. As Garey and Johnson point out [50,

The concept of a *quantum computer*, the central idea of quantum computation, arises somewhat naturally as the result of two different lines of thought somewhat analogous to these two motivations of classical computer science. The quantum computer, a computational device using the axioms of quantum mechanics in some fundamental way to perform its operations,[2] may first be thought of as a framework, based on physical theory, for describing precisely what operations it is possible for a computer to perform—that is, for determining what a computer cannot do. Second, it may be considered as a particular computational model which is seemingly difficult to simulate on a "classical" computer and hence (in a sense described in §1.2.2 below) apparently more powerful than a classical computer.

1.1.1 Physical limits to computation

Historically, the first approach was pioneered by researchers such as Landauer. As early as 1961, Landauer [69, 70] argued, using a simple physical model for a unit of memory, that the irreversible erasure of information is necessarily dissipative; that is, it necessarily results in heat generation of at least $kT \ln 2$ per erased bit. (Earlier references to the value $kT \ln 2$, such as von Neumann's [100, p. 469 in [8]], do not explicitly identify *erasure* as the fundamental heat-generating process in computation. See [70] and [16] for summaries of the historical evolution of this argument.) Landauer's original paper [69] is instructive as a demonstration of the usefulness of physical models for computation. This utility is of course demonstrated by the power of his main result, in providing a fundamental

pp. 2–4], once we know that exact solution of a given problem is "probably hard" we know that we can more profitably direct our efforts to changing the problem requirements than to solving the given problem.

[2] A "quantum computer" should be distinguished from a device which, though its size is such that quantum-mechanical perturbations of classical effects become nonnegligible in designing its fundamental components, is constructed so that the overall operation can be understood as a purely classical process. For example, an integrated circuit may have transistors whose gate region is so small that the presence or absence of a single electron may affect the transistor's switching characteristics; thus charge quantization is a nonnegligible effect. However, in a classical computer such a transistor may be driven, either at a high enough voltage or for a long enough time, that this quantization effect is negligible in the amplified output signal, and the time-averaged current flow can be approximated classically.

limit, on what a computer may do, of a very different sort than the purely-mathematical Turing limits on "computable" functions. But it is also shown in the negative by the unwieldiness of his sketched argument (later seen to rest on an implicit assumption not true for all possible computational models) that logical irreversibility, with the corresponding dissipation of heat, is *necessary* for the operation of any general-purpose computing device: With the existence of a general model for computation, such an argument might be easily formalized, and its truth and assumptions easily seen as such. (Arguments such as these, giving physical constraints on the performance of a computer, are still being made. For example, a review by Meindl *et al.* [79] collects performance limits for silicon integrated circuits; Bekenstein [13, 14] and Lloyd [75] derive ultimate architecture-independent limits on computational power—of course, the generality of these limits comes at the cost of outlandish bounds.)

Thus quantum computation, being on a firmer physical foundation than classical computation, also has the potential to place more strict fundamental limits on the power of computation than classical computer science.

It is reasonable, since nature is currently believed to be fundamentally quantum, to turn to quantum theory as the basis for a general framework for computation, and attempt to use this quantum computational model to derive limits on what a computer may compute.

1.1.2 Computational power of physics

The second approach, in which quantum effects in a computer are considered as desirable properties allowing the performance of interesting operations different from those performable on a classical computer, began with proposals such as those of Feynman [46, 47] and Deutsch [35, 36]. (These in turn were motivated by work by Fredkin and Toffoli, who discussed reversible computation from a physical point of view.) Development of new algorithms for which no classical equivalents are known has driven much of the recent interest in the field; quantum computers are likely to be much more difficult to construct and operate than their classical equivalents, so they are likely to be practically useful only if a correspondingly large gain in computational power can be achieved.

Feynman [46], noting that large quantum systems seem to be difficult to simulate on a classical computer (apparently requiring time exponential in the size of the quantum system), proposed designing a "quantum computer" made up, e.g., of a lattice of spins in which a local Hamiltonian could be tailored to match the Hamiltonian of the system under study, allowing the "quantum computer" to effortlessly simulate the system of interest: *efficient* simulation of the dynamics of arbitrary quantum systems. Deutsch [35], motivated both by Feynman's suggestion of a more powerful model of computation and by the persistence of the "Church-Turing thesis"[3] in classical computer science (suggesting to him the presence of some physical law limiting computational power: thus his motivation was apparently a combination of both of the reasons sketched here), expanded these proposals to a full definition of a quantum computer. His original proposal was a quantum analogue to the machine proposed originally by Turing [97], having a "head" (a coherent quantum analogue of a classical finite state machine) moving along a linear "tape" (a linear array of quantum memory registers, infinite in both directions), undergoing at each discrete time step a local unitary operation involving the head's state and the state of the memory register at that point on the tape. In later work Deutsch [36] displayed a different model for quantum computation, the *quantum circuit model* analogous to a description of the various "gates" used as fundamental building blocks for creating a classical circuit (where each gate represents a quantum operation acting at a specified time on a specified set of quantum registers), and gave an explicit set of gates allowing universal quantum computation. This latter model is more commonly used in practice today.

Further hints at the increased power of a quantum computer, after Feynman's comment, were provided by Deutsch and Jozsa. As an example demonstrating the utility of a quantum computer, Deutsch proposed the problem of determining, for a black-box function $f : \{0,1\} \to \{0,1\}$, the value of $f(0) \oplus f(1)$. Classically this always requires two evaluations of f; but Deutsch and Jozsa [37] showed that a quantum computer could determine $f(0) \oplus f(1)$, with certainty, with *a single* (quantum) evaluation of f.[4] Deutsch

[3] The Church-Turing thesis is the conjecture that the class of "computable" functions does not depend much on the particular definition of a computer used; this is discussed more fully in §1.2.2 below.

[4] Deutsch [35] originally showed only that the quantum computer could give the correct answer half the

and Jozsa extended this result to the more general case of determining whether a function $f : \{0, \dots, 2N-1\} \to \{0,1\}$ is balanced (having $|f^{-1}(0)| = |f^{-1}(1)| = N$) or constant. The quantum algorithm takes time[5] $\mathcal{O}(\log N)$ plus exactly two evaluations of f, while the classical algorithm takes time $\mathcal{O}(N)$ plus $N+1$ evaluations of f in the worst case (although on average it requires only time $\mathcal{O}(\log N)$ and approximately three evaluations of f)—a further suggestion that quantum computers might be exponentially more powerful than classical ones, at least for some problems.

The most famous result in the theory of quantum algorithms occurred in 1994, when Shor [91, 92] proved that a quantum computer could factor integers in "polynomial" time,[6] in contrast to the classical case in which no polynomial-time algorithm is known. (The best classical factoring algorithms known are called "superpolynomial," meaning that they grow more quickly than any polynomial in $\log N$ but more slowly than $N = \exp(\log N)$. The number-field sieve [72], the classical algorithm whose asymptotic complexity is the lowest known, requires time $\mathcal{O}\left[\exp(c(\log N)^{1/3}(\log\log N)^{2/3})\right]$.

Two years later, Grover [55, 56] demonstrated another result, less spectacular in its speed increase but far more general, for finding roots of arbitrary black-box functions in less time (measured by the number of function evaluations required) than classically possible: A quantum computer using Grover's algorithm requires only $\mathcal{O}(\sqrt{N})$ evaluations of f where the classical algorithm requires $\mathcal{O}(N)$).[7] Grover's algorithm may immediately be applied to many search problems; in particular, every problem in NP[8] can be recast

time; the other half the time it would return a "$|fail\rangle$" state, giving no information about the value of $f(0) \oplus f(1)$. This solution does not improve the mean runtime, but it does improve the best-case runtime.

[5] See §1.2.2 for a definition of $\mathcal{O}$-notation.

[6] That is, the algorithm can find nontrivial factors of a composite integer N in time polynomial in $\log N$, the length of the binary representation of N. A simple algorithm [11] suffices to perform the factoring in $\mathcal{O}\left[(\log N)^3\right]$ operations using space linear in $\log N$; more complicated algorithms can reduce the time required even further.

[7] Given an arbitrary function $f(x) : X \to \{0,1\}$, a solution $f(x) = 0$ may be found with high probability, if one exists, in at most $\mathcal{O}(\sqrt{|X|})$ evaluations of f, without knowing any further properties of f. If the solution x is known to be unique, about $\frac{\pi}{2}\sqrt{|X|}$ evaluations of f are required; classically the solution takes $|X|/2$ evaluations of f on average, so the quantum advantage is quadratic.

[8] NP is the class of nondeterministic polynomial-time algorithms, defined in §1.2.2 below.

in a form amenable to use of Grover's algorithm, by performing a search for one of the "succinct certificates" (further described in Chapter 2 below) demonstrating the solution of the problem.[9]

Thus, we can view the theory of quantum computation as a tool both for discovering limits on computational power and for improving computational power. The following chapters in this thesis explore both of these approaches. The remainder of this Introduction provides background material and notation for the later chapters: The design and notation of the quantum computers discussed here are outlined, and the subject of computational complexity is briefly reviewed.

1.2 A physical model for computation

The core idea behind the quantum computer is that it is a *physical model* for computation. That is, unlike the basically empirical or intuitive rules used in specifying the allowed fundamental operations for various classical models of computation, the quantum computer should be designed so that its states and operations correspond closely to fundamental physical laws—in particular, the laws of quantum mechanics.[10] Because this mapping is from such fundamental elements of quantum mechanics, a statement about quantum mechanics may immediately map to a statement about quantum computation, possibly providing nontrivial information about the power or limitations of quantum computation.

[9] Note, however, that this demonstrates an improvement over only the most naïve classical algorithm for solution of the problem, a brute-force search over the problem space. For many interesting problems in NP algorithms more efficient than brute-force searching are known. Grover's algorithm can sometimes be used with these classical algorithms to further reduce the complexity, but in some cases the classical algorithms do not seem to reduce easily to the Grover black-box search form. Similarly, some harder problems (not in NP) do not seem amenable to Grover reductions [84, 44].

[10] A complete physical model should thus use a complete unified field theory as basis for the computational model; this is beyond the scope of this paper and unlikely to be relevant to the functioning of experimental realizations of quantum computers in the near future. In this work we will generally assume (as is common in quantum-computing research) that *nonrelativistic* quantum mechanics is a valid approximation for the systems under consideration. We will, however, consider the effects of finite propagation speeds, for signals traveling between widely-separated subsystems, on models of allowed operations.

At the most basic level, we thus define our quantum computer by reference to the axioms of quantum mechanics [102, 86]: The **state** of a quantum computer is the quantum-mechanical state of the system, a ray in the computer's Hilbert space. An allowed **operation** of a quantum computer is a quantum-mechanical operator acting on the system's state. In the standard formalism for quantum mechanics of closed systems, the state can be represented by a vector $|\Psi\rangle$ in the Hilbert space $\mathcal{H}$ of the system; the operators can be thought of as compositions of unitary operators U and projective (or "von Neumann") measurement operators $\mathbf{M}$. The Hilbert space $\mathcal{H}$ of a quantum computer is typically finite-dimensional and often considered to be composed of a tensor product of some number n of fundamental d-dimensional Hilbert spaces $\mathcal{H}_d$ called **qudits** (**qubits** for $d = 2$), $\mathcal{H} = \mathcal{H}_d \otimes \cdots \otimes \mathcal{H}_d$. For the purposes of this paper we will not typically refer directly to the qudit Hilbert spaces but to quantum "memory registers" composed of some number of qudits and having some distingushed basis $\{|0\rangle, \ldots, |N-1\rangle\}$ (in which state preparation and measurements are easy to perform), the **computational basis** (using the usual Dirac bra-ket notation to denote states of a quantum register). The computational basis states of a quantum register will often be interpreted as radix-N numbers, just as usual in classical computation.

In general, however, there is no reason to restrict our computer to begin in such a "pure state" uncorrelated with external states; and we may be unable in practice to keep our quantum system perfectly isolated from the environment. So, more generally, the quantum computer's state can be described using the language of open quantum systems, where the state of the system is described by a density matrix ρ. In this formalism, operators are allowed to interact both with the system of interest and (either intentionally or parasitically) with the external environment. Unitary operators are replaced by more general superoperators $\$$, and generalized measurements (or positive operator-valued measures, POVMs) $\mathcal{M}$ replace projective measurement operators [88, §3.1–3.3]. Superoperators and POVMs can both be written explicitly using a **Kraus representation**,[11] a set of linear

[11] The Kraus representation, as well as being a convenient form for calculations, emphasizes a formal similarity between "unitary" evolution and "measurements" which is kept rather mysterious in the usual Copenhagen picture. A superoperator, in this language, can be thought of as a particular implementation

operators $\{M_i\}$ satisfying $\sum_i M_i^\dagger M_i = 1$. A superoperator $\$[\{M_i\}]$ with Kraus representation $\{M_i\}$ acts on a density matrix ρ as $\$[\{M_i\}](\rho) = \sum_i M_i \rho M_i^\dagger$. A POVM with this Kraus representation is a measurement with $|\{M_i\}|$ outcomes, outcome i occurring with probability $\operatorname{tr} M_i^\dagger M_i \rho$. The operation of the quantum computer will cause the initial state ρ to evolve to some final state $\$(\rho)$. If, as is usually the case when the goal is solution of a classical problem, we desire a classical answer at the end of the computation, a measurement may be performed on this final state to read out the answer.

The computer's interactions with the environment during the course of its computation can be divided into "desirable" interactions (e.g., those in which the computer is observing or interacting with some other system) and undesirable "parasitic" interactions. The latter interactions can be suppressed using quantum error correction, in which a smaller, protected subspace of the computer's full Hilbert space is used as the computer's state; the most likely parasitic interactions of the computer with the environment cause excursions of the computer's state out of this subspace which can, for a properly-designed quantum error-correcting code (see, e.g., [83, Ch. 10] and [88, Ch. 7] for introductory treatments), be measured and corrected; so such interactions need not be considered in the theoretical analysis of the quantum computer. In the former case, if the external system obeys quantum mechanics, then we may incorporate it as part of the quantum system making up the computer to remove this source of interaction with an external environment. (The caveat "if the external system obeys quantum mechanics" probably seems rather weak. However, as in classical computer science, it is sometimes interesting to consider **oracle** problems, in which a black box of unknown design—hence, possibly computing difficult or even uncomputable functions—may answer certain questions. A common goal in such problems is to find out what extra power such an oracle could provide, if it existed. Another use of oracles is to provide black-box devices that cannot be reverse-engineered. The Grover result, for example, can be stated as a quadratic speedup over the best possible classical algorithm if the function f, for which a root is to be found, is considered to be a randomly-chosen oracle function. If, on the contrary, f were given by some explicit formula, then a clever algorithm might be able to simplify the formula and find an algebraic

of a POVM in which the measurement results are ignored and discarded.

solution in less time than a brute-force search.)

So, as long as the environment is quantum-mechanical in nature, interactions between the computer and the environment may be either suppressed, by embedding the computer's state space in a protected subspace of a larger Hilbert space, or modeled, by enlarging our computer's Hilbert space to include the interacting system. Thus we can usually pretend that our quantum computer is isolated from the external environment: undergoing only unitary evolution over the course of its computation, perhaps followed by a single measurement at the end so that we can learn the result of the computation.

This can be shown in two steps. First we can convert each instance of a superoperator or POVM into a unitary operator or von Neumann measurement operator (respectively) on an augmented Hilbert space; for a superoperator or POVM with Kraus representation $\{M_i\}$ (acting on the computer's Hilbert space S), an extra quantum register K of at least $\log|\{M_i\}|$ qubits (able to hold the Kraus index i), initially in state $|0\rangle_{\mathrm{K}}$, is adjoined to the computer, and the operator (which turns out to be unitary—see [88, §3.2])

$$U[\{M_i\}] : |\psi\rangle_{\mathrm{S}} |0\rangle_{\mathrm{K}} \mapsto \sum_i \left[M_i |\psi\rangle_{\mathrm{S}}\right] |i\rangle_{\mathrm{K}} \tag{1.1}$$

is performed.[12] For a POVM, the register K is measured (in the computational basis); for a superoperator, the register K is simply discarded. We are left with a quantum computer (on an augmented Hilbert space) undergoing only unitary evolution and projective measurements. Each measurement can be thought of as correlating the external environment with the state of the system via a unitary operator integrated from an interaction Hamiltonian (see [88, §3.1]). (The measurement result may be used in constructing later

[12] We will label quantum registers with letters in a roman font; thus $|\psi\rangle_{\mathrm{X}} \in \mathcal{H}_{\mathrm{X}}$. Lowercase letters will be used for single-qubit registers; uppercase letters denote longer registers. For a register X thought of as containing natural numbers we may use the notation X_i to denote the notional qubit holding the ith "bit" of register X, the coefficient of 2^i in its binary representation. (This qubit may not exist as part of the register in all implementations: for example, if our quantum computer is based on 3-dimensional "qutrit" systems we may be using trinary representation instead of binary. But base conversions are easy to compute, so we may assume that this qubit exists without greatly affecting our estimates of the difficulty of a computation.)

operations, and it may also be of interest as part of the computer's output.) We can replace the coupling to the environment by a coupling to another new quantum register M (again, initially in the state $|0\rangle_{\mathrm{M}}$) of sufficiently large dimension; after performing this unitary operation, M, instead of the external environment, holds the measurement results. Any (unitary) operator later in the computation which was originally classically conditioned on the outcome of this measurement may be replaced by a *controlled-unitary* operator (defined in §1.2.3 below) with control register M. Finally we have a quantum system undergoing only unitary evolution. At the end of the computation, we can read out all of the measurement results by measuring the ancilla systems M added in unitarizing the measurements.

1.2.1 Reversible computation

This is already a theoretically useful result. As a simple example of its power, we can already see that, since its unitary evolution U may be perfectly undone by the evolution $U^{-1} = U^{\dagger}$, such a quantum computer is *reversible.* Thus (apart from uncorrectible errors arising from interactions with an inaccessible environment during the computer's operation) we can simulate any quantum computer with a reversible quantum computer undergoing only unitary evolution. This is a very different sort of computation than that performed by a standard classical computer. Among even the simplest sorts of operations possible on a classical computer are those, such as erasing memory or replacing two numbers with their sum, which are not reversible and which therefore cannot be performed with a unitary quantum computer. Because of this strong restriction on the allowable operations it is not immediately obvious that a reversible computer can efficiently perform all of the operations of a normal classical computer. However, as shown by Bennett [15] and Fredkin and Toffoli [49], in fact a reversible computer can simulate the action of an (irreversible) classical computer fairly efficiently.[13]

Of course, this simulation would not be difficult if our reversible computer had unlimited storage space. In this case an arbitrary computation can be performed reversibly by

[13] Exactly what is meant by "efficiency"—the notion of the *complexity* of a computation—will be discussed in §1.2.2 below.

saving all of the intermediate values which are destroyed by the irreversible computer; for example, whenever the irreversible computer performs an operation[14] $[x] \mapsto [f(x)]$, then the reversible computer, simulating this operation, performs instead[15] $U[f] : |x\rangle_{\mathrm{X}} |y\rangle_{\mathrm{Y}} \mapsto |x\rangle_{\mathrm{X}} |y \oplus f(x)\rangle_{\mathrm{Y}}$. The reversible computer, limited to performing only bijective operations, can still compute arbitrary functions by preserving the register containing the input state and writing its output to a second register. Typically the initial value of this second register is $y = 0$ so that the final state is just $|x\rangle_{\mathrm{X}} |f(x)\rangle_{\mathrm{Y}}$, but to fully define the operation we must define its action on each state $|x\rangle_{\mathrm{X}} |y\rangle_{\mathrm{Y}}$ of the computational basis. (For a quantum computer, where arbitrary superpositions of states are allowed, we can use linearity to extend this definition to the entire Hilbert space.) To maintain reversibility we store the function in the second register using some simple-to-compute bijection (here, as is common, we use addition). We will often define the reversible operation as $|x\rangle_{\mathrm{X}} |0\rangle_{\mathrm{Y}} \mapsto |x\rangle_{\mathrm{X}} |f(x)\rangle_{\mathrm{Y}}$, implicitly defining its action on states with initial output register $y \neq 0$ by addition, as above.

In a long classical computation many irreversible operations will generally be performed, and maintaining all of the intermediate results, to preserve reversibility, may become infeasible. However, Bennett found a general method for restraining this memory usage and periodically erasing (reversibly!) most of the intermediate results. Suppose we have a computer $U[f]$ reversibly calculating $f(x)$; however, in the process, it is forced to save various intermediate results $g(x)$,

$$|x\rangle_{\mathrm{X}} |0\rangle_{\mathrm{Y}} |0\rangle_{\mathrm{G}} \xmapsto{U[f]} |x\rangle_{\mathrm{X}} |f(x)\rangle_{\mathrm{Y}} |g(x)\rangle_{\mathrm{G}} \,.$$

[14] We use the notation $[x]$ to represent a register of an irreversible classical computer with value x, to stress the difference between reversible and irreversible computations.

[15] Operations are usually considered to be performed modulo the dimension N of the register's Hilbert space when necessary; thus the addition operation in $|y \oplus f(x)\rangle_{\mathrm{Y}}$ is understood as $\bmod N$ addition. N will typically be chosen implicitly to be large enough to hold all of the values of interest. Subscripts on bras and kets are used to label the various register subspaces so that the entire state of the computer need not be given for an operation acting nontrivially only on a small subsystem of the computer. To avoid ambiguity, an operation may also be labeled with the registers on which it acts nontrivially; these will, however, usually be omitted for clarity of notation.

Since $U[f]$ is reversible, we can erase the intermediate results $g(x)$ by merely performing $U[f]^{-1}$, running our reversible computer backwards. Of course, this erases the desired output $f(x)$ as well, so before doing this we should save the desired result in yet another register:

$$\begin{aligned} |x\rangle_{\mathrm{X}}\,|0\rangle_{\mathrm{Y}}\,|0\rangle_{\mathrm{G}}\,|0\rangle_{\mathrm{F}} \quad &\xmapsto{U[f]} \quad |x\rangle_{\mathrm{X}}\,|f(x)\rangle_{\mathrm{Y}}\,|g(x)\rangle_{\mathrm{G}}\,|0\rangle_{\mathrm{F}} \\ &\xmapsto{U[\oplus]_{\mathrm{Y,F}}} \quad |x\rangle_{\mathrm{X}}\,|f(x)\rangle_{\mathrm{Y}}\,|g(x)\rangle_{\mathrm{G}}\,|f(x)\rangle_{\mathrm{F}} \\ &\xmapsto{U[f]^{-1}} \quad |x\rangle_{\mathrm{X}}\,|0\rangle_{\mathrm{Y}}\,|0\rangle_{\mathrm{G}}\,|f(x)\rangle_{\mathrm{F}}\ . \end{aligned}$$

At the close of this computation, the ancillary registers Y and G have been restored to their initial states $|0\rangle$, so they can be reused in later computations. (The operation labeled $U[\oplus]$, used to copy the result into the output register F, is just another addition operation, adding the contents $f(x)$ of register Y to those of register F, initialized to 0.) This process can be repeated throughout the computation to clean up unneeded "scratch space." It can even be done recursively; for example, the computation $U[f]$ may include some subroutine $U[h]$ which cleans up its own intermediate results. It should be noted, though, that since the clean version of a subroutine requires two iterations of the messy version $U[h]$, the clean version of $U[f]$ now requires *four* iterations of $U[h]$. Thus cleaning up recursive subroutines using this process may be more time-consuming than expected, and it may be more efficient to redesign the subroutine to use less scratch space.

The methods and results described so far apply to any reversible computer; although we have been using the language and notation of quantum mechanics the results have been quoted in terms of actions on individual basis states and so are not explicitly quantum. The fundamental quantum nature of these equations only becomes explicit when we begin to consider the action of these operators on superpositions of states in the computational basis. For example, consider the operator $U[\oplus]$ used above to copy our final answer. This operator only acts as a "copy" operator *in the computational basis.* For example, the state[16] $\big[\,|x\rangle_{\mathrm{A}} + |y\rangle_{\mathrm{A}}\big]\,|0\rangle_{\mathrm{B}}$ becomes, when acted on by $U[\oplus]_{\alpha,\beta}$, $\big[\,|x\rangle_{\mathrm{A}}\,|x\rangle_{\mathrm{B}} + |y\rangle_{\mathrm{A}}\,|y\rangle_{\mathrm{B}}\big]$, which

[16] In these expressions the overall normalization factors are not shown explicitly, a difference from standard quantum-mechanics notation. In quantum computation sums over states in an orthonormal basis are

is not the same as $\left[\,|x\rangle_{\mathrm{A}} + |y\rangle_{\mathrm{A}}\right]\left[\,|x\rangle_{\mathrm{B}} + |y\rangle_{\mathrm{B}}\right]$. (The quantum *no-cloning* theorem [38, 105], in fact, states that it is impossible to make a perfect copy of an unknown quantum state. $U[\oplus]$ only makes perfect copies of states in a known basis, which does not violate the theorem.)

1.2.2 Computational complexity

Although this formal equivalence between quantum-mechanical axioms and elements of quantum computation already allows some useful discussions, it is not quite enough for a fully satisfactory definition of a quantum computer; our descriptions so far have not given us a complete method for determining how difficult a given quantum operation U is—or if, indeed, it is even possible. These questions belong to the fields of *complexity theory* and *computability theory*, respectively. A brief outline of these topics will let us define more precisely what we mean by the "difficulty" or "efficiency" of an algorithm, already alluded to several times above. (More complete introductions to computability and complexity theory can be found in Papadimitriou's two texts [73, 85] and Davis' review [32].)

The primary goal of computability theory is to determine whether a given class of computational problems is solvable using a well-defined algorithm. Development of this theory started with results such as Turing's proof [97] of the unsolvability of Hilbert's Tenth Problem (the *Entscheidungsproblem*[17]) in 1936, and occurred in parallel with the study of various computational models, attempting to formalize the notion of an algorithm, then being developed. Turing's model, the **Turing machine**, consisted of a finite state machine (the **head**) moving along an unbounded linear array of memory registers

common; for a sum of N such terms the overall normalization factor is trivially $\frac{1}{\sqrt{N}}$ and may often be omitted without causing confusion.

[17] Literally "decision problem," referring to the problem of determining the validity of a formula in predicate calculus. Hilbert [57] was interested in algorithmic solutions to Diophantine equations; Gödel showed that arbitrary formulas in predicate calculus could be encoded as Diophantine equations; and Turing showed (once his definition of an *algorithm* was accepted) that such a problem was equivalent to the uncomputable *halting problem*, of determining whether a computation of a particular Turing machine would ever terminate. (See [34, 78] for an elegant, self-contained number-theoretic proof that Hilbert's Tenth Problem is unsolvable.)

(the **tape**) on which it can read and write symbols from a finite alphabet. Turing showed that the *Entscheidungsproblem* is equivalent to the **halting problem**, of determining whether a computation of a particular Turing machine would ever terminate (i.e., reach a particular head state), and that this problem is in general uncomputable by a Turing machine. Another early model, apparently quite different, was the class of *recursive functions* described by Church [29]; other models were proposed later. These models, some invented for ease of theoretical analysis and some invented to model practical computing machines, varied widely in their details; but it was soon noticed that *all of the models which seemed physically realizable agreed on the class of algorithmically-computable functions.* This empirical result, known as the Church-Turing thesis [73, Ch. 5], ascribes a fundamental mathematical importance to the concept of an algorithm, deeper than any of these particular models, which allows them all to agree on the same result.

Computability theory thus tries to answer the qualitative question of whether a particular problem can be solved at all. Complexity theory, on the other hand, attempts to answer the quantitative question of *how hard* a particular problem is to solve (assuming that it has been found to be computable); that is, how much of various limited computational resources must be applied to solve the problem. The computational resources of greatest interest in most applications are TIME, the amount of time required for our algorithm to compute the problem's solution; and SPACE, the memory required by the computer over its course of execution. In some applications other resources such as COMMUNICATION, the amount of information necessary to transmit between distant points, and ENERGY dissipated in the computer are important. In many cases there are tradeoffs between these resources; different solution algorithms will optimize different functions of these resources.

To define these requirements, the costs of each individual operation of the computer must all be defined. This is commonly done by defining a small set of computable *fundamental operations* (often called *gates*, especially in circuit models of computation); the entire computation is then built by composing these fundamental operations. The time complexity of the entire computation, for example, can then be found (for serial computations) by summing over all the fundamental gates in the computation.

Computational resource requirements may be expressed in varying degrees of abstractness. In some cases the exact resource requirements are useful to know; since quantum computers are expected to be difficult to construct, for example, pioneering experimentalists may want to know exactly how many qubits are required to perform a particular calculation and for exactly how long their coherence must be maintained, to know whether a proposed design is sufficient to solve the given problem. However, in many cases we are interested not merely in a single problem but in an entire class of related problems—for example, in finding the factors of a given integer. There may be wide variation in the difficulty among problems of this class ($2^{2^N}+1$ is much harder to factor than 2^{2^N}), and it may be difficult (possibly as difficult as actually solving the problem) to determine exactly how hard a given problem of the class is. Since the usefulness of these resource limits comes from knowing them before we solve the problem, we may be forced to settle for finding upper bounds on the resources required (average-case and worst-case bounds are both of interest: average-case bounds when the problems are proposed by a neutral party, but worst-case bounds if proposed by an adversary). These bounds are expressed in terms of easily-computed functions of the input parameters; the simplest such function, and the most common value used to express these complexity bounds, is simply some measure of the total *length* of the input (most bounds on factoring an integer N, for example, are given in terms of the length $\log N$ of N).

The actual values of these resources required will vary, of course, in obvious and relatively uninteresting ways due to differences in parameters like the computer's clock speed. The resource requirements may be stated in units which try to normalize for these differences; for instance, Time is measured by counting the number of *fundamental operations* of the computer required to solve the problem. Alternatively, it is quite common to express resource bounds using Knuth's "big-O" asymptotic-growth notation $\mathcal{O}(\cdot)$, in which only the scaling of the leading-order term is expressed: For $f, g : \mathbb{N} \to \mathbb{N}$, we say $f(n) = \mathcal{O}(g(n))$ iff there exist $N \in \mathbb{N}$ and $c > 0$ such that $f(n) \leq cg(n)$ for all $n > N$—that is, whenever f grows no more quickly than g. (For example, for an algorithm requiring $\textsc{Time}(n) = \sum_{i=0}^{k} \alpha_i n^i$ with $\alpha_k \neq 0$, we would say $\textsc{Time}(n) = \mathcal{O}(n^k)$.) Knuth also defined the related growth classes Ω, Θ, and o. $f(n) = \Omega(g(n))$ whenever $g(n) = \mathcal{O}(f(n))$ (f

grows at least as quickly as g); $f(n) = \Theta(g(n))$ whenever both $f(n) = \mathcal{O}(g(n))$ and $g(n) = \mathcal{O}(f(n))$ (f and g have the same asymptotic rate of growth); and $f(n) = o(g(n))$ whenever $f(n) = \mathcal{O}(g(n))$ but $f(n) \neq \Theta(g(n))$ (f grows strictly more slowly than g).

A more interesting variation is between different computational models—for example, between a Turing machine, in which the memory cells can only be accessed in a fixed sequence, and a computer with random-access memory. Because different computational models have different sets of fundamental operations, one model may prove more efficient than another at solving a particular problem. A somewhat surprising empirical result, noticed in the course of investigating the many various reasonable[18] computational models proposed, is that they are all *polynomially reducible to Turing machines* [85, p. 36]. That is, given an algorithm for solving a class of problems on some reasonable computational model which uses Resource $f(n)$ for a problem of input size n, a Turing machine may be programmed to *simulate* this computer and solve the problem using Resource at most $P_1(f(P_2(n)))$, where P_1 and P_2 are polynomial functions. This empirical result (sometimes called the "strong Church-Turing thesis" [83, p. 140]) seems to argue that searching for new and more powerful models of computation is not likely to be very fruitful, the efficiency gains over existing models likely at best polynomial. It is common in complexity theory not only to entirely ignore constant factors (as with $\mathcal{O}$-notation) but also to denigrate "merely" polynomial improvements in resource usage. This is somewhat surprising at first, since a polynomial improvement from, say, $\mathcal{O}(n^{1000})$ to $\mathcal{O}(n^3)$ can make a difference between a theoretical curiosity and a practical algorithm. In practice, though, such large exponents are rare. Furthermore, the strong Church-Turing thesis implies that at least some polynomial reductions are plausible consequences of architecture changes, so that differences in the polynomial degree may not be fundamental in the way that $\Theta(2^n)$ and $\Theta(n^2)$ algorithms differ.[19] The algorithms in the class P of polynomial-time algorithms[20] are often considered for these reasons to be practically computable

[18] Some computational models (e.g., the nondeterministic models) are proposed for theoretical reasons and are not expected to be physically realizable; those are not considered "reasonable" for this discussion.

[19] The class of polynomial-time algorithms is said to be **stable**, i.e., invariant, under polynomial-time reductions.

[20] Strictly speaking, P is defined as the class of polynomial-time **decision algorithms**, i.e., algorithms

(and, similarly, algorithms which require superpolynomial time to complete are often considered impractical). A complexity class closely related to P, and often considered to be practically equivalent to the set of practically-computable functions, is BPP, the set of decision problems computable in polynomial time with bounded (i.e., bounded away from $\frac{1}{2}$) probability of error, in the presence of a random-number generator.

There are many other complexity classes of interest besides P. (Algorithms can also be classified, for example, based on how much ancillary memory they require.) One other class of great interest, already mentioned above, is NP, the class of *nondeterministic* polynomial-time algorithms.[21] A nondeterministic class can be described as one in which *guessing correctly* is an allowable operation: so clearly this is not meant to directly describe a realistic computational model. However, given the list of correct guesses a normal computer can find and check the answer in polynomial time, so NP can also be described as the class of problems with polynomial-time *verification* algorithms. The list of guesses is sometimes called a *certificate* for the solution; certificates will be further discussed in Chapter 2 below.

(More formally, we say the Boolean function $f(x) \in \mathsf{NP}$ if there exist functions $g(x)$ and $\hat{f}(x, y)$, with $f(x) \equiv \hat{f}(x, g(x))$, and there exists a polynomial-time algorithm computing $\hat{f}$. Here $g(x)$ represents the set of "guesses"—the certificate, an alleged proof that $f(x)$ is true—for x, and $\hat{f}$ can be thought of as checking the validity of the certificate. Clearly $\mathsf{P} \subseteq \mathsf{NP}$; a major open question in classical complexity theory is the question of whether $\mathsf{P} = \mathsf{NP}$. NP is not a symmetric class; it requires that there exist a single valid proof $g(x)$ whenever $f(x)$ is true, but that there are *no* valid proofs $g(x)$ when $f(x)$ is false. The complementary class coNP is the class of Boolean functions f for which $\neg f \in \mathsf{NP}$.)

One implicit detail common to all of these different "reasonably-physical" computational models (NP is obviously excluded here) discussed so far is that they were all based

computing boolean functions. We say the Boolean function $f(x) \in \mathsf{P}$ if there exists a polynomial-time algorithm for computing $f(x)$.

[21] The class NP appears to contain some very difficult problems. The most difficult of these problems are called the **NP-complete** problems; a problem is NP-complete if any other problem in NP can be efficiently reduced to an instance of this problem class: so solving this one problem is essentially equivalent to solving all problems in NP.

on *classical* physics. Likewise when computers were actually constructed, although their design details differed their operations, whether mechanical or electrical, were understandable using classical physics. As mentioned above, this assumption was not noted explicitly during the first few decades of computer science research when the Church-Turing theses were being noticed. This recognition, that physical theories form a foundation for theories of computation, immediately suggests an important question: *Do different physical theories provide different computers with different computational power?* —That is, do the Church-Turing theses still hold for computers based on new physics?

In this thesis our primary interest, as already mentioned, will be in utilizing nonrelativistic quantum-mechanical effects in computation. For such quantum computers, the class of computable functions remains unchanged, and the (weak form of the) Church-Turing thesis still holds, because quantum-mechanical systems can be simulated on classical computers to any desired level of fidelity (though the Space and Time required may be large). The interesting question is whether the *strong* form of the Church-Turing thesis is true for quantum as well as classical computers. At this time there are some indications that it may *not* be true; Shor's factoring algorithm [91, 92], the Deutsch-Jozsa algorithm [37], and Feynman's early idea of simulating quantum systems [46, 47] all provide better-than-polynomial improvements in time complexity over the best classical algorithms known. (Since the classical algorithms are not proven optimal, however, this does not prove a superpolynomial separation between the computational power of quantum and classical mechanics.[22])

[22] The Deutsch-Jozsa result is, strictly speaking, provably exponentially faster than the optimal classical algorithm (in the worst case), as stated above. However, the Deutsch-Jozsa problem is an oracle problem, presuming the existence of a black-box function f (the function to be tested for constant/balanced behavior); since the oracle may not be strictly part of classical (or quantum-mechanical) theory, these results, "relative to an oracle," do not necessarily relate strictly to physically-based "absolute" complexity classes, though they may provide hints as to the relationship of the corresponding absolute complexity classes. The random-number generator mentioned in our definition of BPP can also be thought of as an oracle, although since quantum mechanics allows the preparation of truly random bits this oracle, at least, can be considered a physical device.

1.2.3 The quantum computer, revisited

In our earlier description of a quantum computer, its allowable operations were described by reference to the axioms for evolution of quantum systems. In a sense this description follows the first approach, of finding limits to the power of a computer. Now we complete the description of the quantum computer as a model for computation, by explaining the assignment of complexity costs to the individual quantum operations.

For simplicity we will use the *quantum circuit model* of computation throughout the work. In this model, all of the qudits (or quantum registers) within a computer are considered mutually local, or, equivalently, the cost of transporting qudits around the system is negligibly small, so that (for example) it is equally easy to perform a given n-qudit operation U_n on any n-qudit subset of the entire system.[23] In the quantum circuit model of computation, as in the classical circuit model, an operation will be considered as composed of a sequence of fundamental gates, each having known resource costs. In the classical case, there are several essentially equivalent *universal* gate sets[24], i.e., sets of fundamental gates which allow construction of a circuit for any computable function, with equivalent complexities (up to the usual uninteresting constant factors). Similarly in the case of quantum circuits one can construct different interesting universal gate sets. Because Hilbert space is continuous, in contrast with the discrete Boolean space of classical computation, a gate set which allows exact construction of a arbitrary quantum operators must be infinite; however, finite gate sets which allow approximate construction with any desired fidelity also exist. (See [83, §4.5] for several such constructions.) An example of such a universal gate set is the set of all one-qubit unitary operators together with a single two-qubit operator, the *controlled-not* gate $\textsc{cnot} \equiv \Lambda(\textsc{not}) \equiv \Lambda(\sigma_x)$ (the operator $\textsc{not}$

[23] This model can be contrasted with various proposed physical models such as Lloyd's lattice model [74], in which (usually) qubits are localized in a periodic lattice of some sort and able to interact, via some local coupling, only with neighboring qubits. In such models the cost of qubit transport may be nontrivial (though they can be bounded by a factor at most linear in the size of the system); such considerations complicate the design and discussions.

[24] Two universal classical gate sets in common use, for example, are $\{\textsc{and}(a,b) \equiv a \wedge b, \textsc{or}(a,b) \equiv a \vee b, \textsc{not}(a) \equiv \bar{a}\}$ and $\{\textsc{nand}(a,b) \equiv \overline{a \wedge b}\}$.

is of course identical to the Pauli operator σ_x) defined by[25]

$$\begin{aligned}
\textsc{cnot}_{[\![\mathrm{A}]\!],\mathrm{B}} \left|0\right\rangle_{\mathrm{A}} \left[\alpha \left|0\right\rangle_{\mathrm{B}} + \beta \left|1\right\rangle_{\mathrm{B}}\right] &= \left|0\right\rangle_{\mathrm{A}} \left[\alpha \left|0\right\rangle_{\mathrm{B}} + \beta \left|1\right\rangle_{\mathrm{B}}\right] \\
\textsc{cnot}_{[\![\mathrm{A}]\!],\mathrm{B}} \left|1\right\rangle_{\mathrm{A}} \left[\alpha \left|0\right\rangle_{\mathrm{B}} + \beta \left|1\right\rangle_{\mathrm{B}}\right] &= \left|1\right\rangle_{\mathrm{A}} \left[\alpha \left|1\right\rangle_{\mathrm{B}} + \beta \left|0\right\rangle_{\mathrm{B}}\right]
\end{aligned}$$

(and, of course, acting linearly on superpositions); that is, CNOT leaves its second qubit (the **target**) unchanged when the first qubit (the **control**) is $|0\rangle$ and flips the second qubit when the first qubit is $|1\rangle$. Here we have introduced Kitaev's notation $\Lambda(U)$ for a controlled-U operator,

$$\Lambda(U)_{[\![\mathrm{A}]\!],\mathrm{B}} \Big[\left|a\right\rangle_{\mathrm{A}} \left|\psi\right\rangle_{\mathrm{B}} \Big] = \left|a\right\rangle_{\mathrm{A}} \Big[{U^a}_{\mathrm{B}} \left|\psi\right\rangle_{\mathrm{B}} \Big] \, . \tag{1.2}$$

Note that this generalizes the original definition; we allow the "control" Hilbert space $\mathcal{H}_{\mathrm{A}}$ to have basis states $a > 1$, in which the unitary operation is performed multiple times.[26]

As mentioned above, we may usually assume that the computer performs only unitary actions over the course of its evolution, preceded by an initial state preparation and followed by a final measurement. We must take care to consider the costs of the state preparation and measurement (if necessary) in computing the resource costs. This may be done by assuming that state preparation and measurement in the computational basis is easy; we take the computer's initial state to be the state $|0\cdots 0\rangle$ (in the computational basis) and require any final measurement to be performed in the computational basis. In analogy with the classical complexity class BPP, the class of algorithms computable on a quantum computer in polynomial time with bounded probability of error is denoted BQP.

The relations between BQP and the various classical complexity classes is of obvious interest. One interesting result—the oracle-relative exponential separation for the Deutsch-Jozsa problem—has already been briefly mentioned above. Bernstein and Vazirani, in [20], investigated BQP more generally, finding classical complexity classes bounding BQP both

[25] We put brackets around the control qubit labels to distinguish them from the targets.

[26] This is not practically very useful unless $U^{d_{\mathrm{A}}}$ is efficiently implementable (or $d_{\mathrm{A}} \equiv \dim \mathcal{H}_{\mathrm{A}}$ is small) but it is a useful notation.

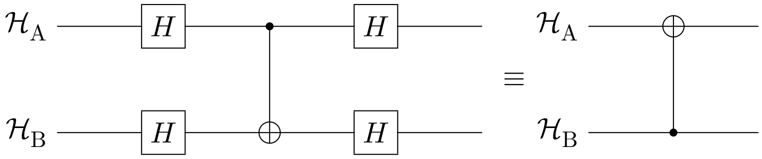

$\mathcal{H}_{\mathrm{A}}$
$\mathcal{H}_{\mathrm{B}}$
H
H
H
H
≡
$\mathcal{H}_{\mathrm{A}}$
$\mathcal{H}_{\mathrm{B}}$

Figure 1.1: An interesting quantum circuit equivalence.

above and below: $\mathsf{BPP} \subseteq \mathsf{BQP} \subseteq \mathsf{P}^{\#\mathsf{P}}$.[27] Other work, such as the paper by Adleman *et al.* [2] considering quantum computers with restricted operation sets, has slightly tightened these bounds but has not sufficed to answer the question of whether BQP contains problems which are classically difficult to solve. Bennett *et al.* [17], Simon [93], and Vazirani [98] have proved several oracle-relative results, giving both upper and lower (relative) bounds. These indicate that BQP, while showing occasional surprising strength, cannot provide superpolynomial reductions for black-box problems: the quantum computer is apparently well suited to attacking some special structure in particular types of problems.

For future use, we close this section with the definition of a unitary operator particularly useful for quantum computation, the *Hadamard* gate (so named because it performs the one-bit Walsh-Hadamard transform on the vector of amplitudes in the computational basis) H:

$$\begin{aligned} H\,|0\rangle &= \tfrac{1}{\sqrt{2}}\big[\,|0\rangle + |1\rangle\big] \\ H\,|1\rangle &= \tfrac{1}{\sqrt{2}}\big[\,|0\rangle - |1\rangle\big]\ . \end{aligned} \tag{1.3}$$

Quantum circuits are often represented pictorially (in much the same way as classical circuits), with qubits flowing along "wires" and boxes and other symbols representing various operations. For example, the circuits describe graphically the operator equivalence

$$H_{\mathrm{A}} H_{\mathrm{B}} \mathrm{CNOT}_{[\![\mathrm{A}]\!],\mathrm{B}} H_{\mathrm{A}} H_{\mathrm{B}} \equiv \mathrm{CNOT}_{[\![\mathrm{B}]\!],\mathrm{A}}\ .$$

[27] #P denotes, essentially, the class of functions *counting* the number of solutions of a function in P; it is apparently a very powerful class. The superscript notation $\mathsf{P}^{\#\mathsf{P}}$ denotes the class of functions computable in polynomial time *given a* #P *oracle*, i.e., a black box capable of efficiently computing the answers to all problems in #P.

The circuit is to be read left-to-right, with the qubits entering from the left as shown by the two wires. The boxed H indicate the Hadamard gates acting on each qubit; the vertically-aligned pairs commute and can be performed in either order, $H_{\mathrm{A}}H_{\mathrm{B}}$ or $H_{\mathrm{B}}H_{\mathrm{A}}$ (or even simultaneously, if parallelization is possible). The symbol in the center represents the CNOT gate, with the filled circle indicating the control qubit and the open circle the target qubit. Thus we have the somewhat surprising result that the sense of the controlled-not gate, with its misleadingly-named "source" and "target" qubits, can be changed by surrounding it with Hadamard gates, implementing changes of basis on the source and target Hilbert spaces. (This equivalence may trivially and inelegantly be proved by writing the operators H_{A}, H_{B}, $\textsc{cnot}_{\mathrm{A,B}}$ explicitly as 4×4 matrices and computing their product. A more elegant proof is possible using the language of stabilizers.)

This introduction to quantum computing and the quantum-circuit model has necessarily been much abbreviated. For more background and more detailed descriptions of various quantum algorithms see Nielsen and Chuang [83, Ch. 4–7] and Preskill [88, Ch. 6].

1.3 In conclusion, an introduction

The following two chapters contain examinations of quantum computation from the dual perspectives outlined above. In Chapter 3 I explore the first approach, showing how physical laws—quantum mechanics plus causality—result in restrictions on nonlocal measurements; some such measurements are disallowed as violating causality. The question of whether all measurements not forbidden by these physical laws can be implemented is considered. In Chapter 2 I explore the second approach, with an emphasis on the classical combinatoric problem of determining whether two graphs are isomorphic. Various quantum approaches to solving this problem are proposed.

Chapter 2

Ruminations on Graph Isomorphism

2.1 Introduction

The development of efficient quantum algorithms to solve interesting problems is one of the major driving forces behind the growth of the field of quantum information theory, just as in the classical case the development of more efficient algorithms is a major source of interest. Sadly, there are as yet few quantum algorithms providing substantial gains over their classical counterparts. In this chapter we approach the subject of quantum computation with the goals of understanding quantum algorithms and developing new ones. Only a few distinctively-quantum algorithms have been discovered so far, giving us a fairly small set of known essentially-quantum tools to use in new algorithms. We begin with a review of some of these primitives, to try to gain hints as to the source of whatever extra computational power quantum evolution can give us.

To guide our discussions we have chosen to consider a particular classical problem of interest, that of determining whether two graphs are isomorphic to one another. This problem is of interest in classical complexity theory because it, like Factoring, is believed to be a relatively easy problem in NP (see, e.g., [66]) which is, however, not known to lie in BPP. We present several potentially-fruitful quantum approaches to a solution of the problem.

2.1.1 The problem

We begin with definitions of a few terms. An **undirected graph** (in this work, when we refer to a **graph** we will be considering an undirected graph) is a set V of **vertices** together with a set E of **edges**, unordered pairs of distinct vertices. (We can think of E as a symmetric, antireflexive binary relation on V.) Two graphs $G = (V, E)$, $G' = (V, E')$ are called **isomorphic** (we write $G \sim G'$ or $G \overset{\phi}{\sim} G'$) if there is a bijection $\phi \in S_{|V|}$, the **isomorphism**, such that $(e_1, e_2) \in E$ iff $(\phi(e_1), \phi(e_2)) \in E'$. Intuitively, two graphs are isomorphic if they represent the same pattern of vertices and edges, or if the drawing of one graph can be topologically distorted (allowing edges to pass through one another, unlike in knot theory) into the drawing of the other. A **topological graph** $\overline{G}$ is an equivalence class of isomorphic graphs, $\overline{G} \equiv \{H : H \sim G\}$. The **automorphism group** $\Phi(G)$ of a graph G is the group of **automorphisms** of G, $\Phi(G) = \{\phi : G \overset{\phi}{\sim} G\}$. A graph is **rigid** if it has trivial automorphism group, $\Phi(G) = e$.

Finally, we define a particular representation of a graph useful for considering algorithms on graphs. The **adjacency matrix** A for a graph $G = (V, E)$ with vertices $V = \{1, \ldots, n\}$ is the $n \times n$ matrix $A = (a_{ij})$ with $a_{ij} = 1$ if $\{i, j\} \in E$ and $a_{ij} = 0$ otherwise.[1] (Thus the adjacency matrix of an undirected graph is a symmetric, traceless $0 - 1$ matrix.)

Graph Isomorphism (GI) is the following decision problem in computer science: Given two graphs G, H over the same set of vertices, is $G \overset{\phi}{\sim} H$? In terms of the adjacency matrices A_G and A_H, we may equivalently ask whether there is a permutation matrix P_ϕ such that $A_G = P_\phi A_H P_\phi^{-1}$. Clearly GI $\in$ NP: the permutation P_ϕ is a polynomial certificate for the isomorphism. Whether GI $\in$ P, and even whether GI $\in$ coNP (i.e., that there exists a succinct proof that two graphs are *not* isomorphic) are open problems.

The isomorphism problem for graphs is well studied in classical computer science as

[1] Another common representation of graphs is as an **edge list** $(e_1, \ldots, e_n)$ (where $E \equiv \{e_1, \ldots, e_n\}$). (This representation is strictly valid only as long as G has no isolated vertices, since V is defined only implicitly as the set of all edge endpoints, the domain of the relation E.) This representation is more efficient for *sparse* graphs, having $o(n^2)$ edges, but for general graphs is equivalent in power to the adjacency-matrix representation.

an interesting problem in its own right—many problems of equivalence of algebraic structures, such as GroupIso, the problem of isomorphism of two abstract groups (explicitly defined by their multiplication tables), can be reduced to GI, essentially by encoding the multiplication table in the structure of the graph; see [81] for a general review of the problem. GI is also interesting as an example of a problem in NP which is neither known to be NP-complete nor known to be in P. In this it is rather similar to the problem of Factoring[2] (given a positive integer N, find its prime factors), although unlike GI, Factoring is known to lie in $\mathsf{NP} \cap \mathsf{coNP}$. (This is because there is a succinct proof that an integer is prime; thus the list of factors multiplying to N, together with proofs that each is prime, provide a succinct certificate for N.) Factoring is known to have a polynomial-time quantum algorithm (Shor's algorithm [92]); whether the same is true for GI is currently an open question.

GI is one of a number of related isomorphism problems, many of which are easily reducible to each other [67]. GA, the problem of determining whether a graph has a non-trivial automorphism, is (polynomial-time) reducible to GI, but it is not known whether GI is reducible to GA. However, the related **counting problems** #GI (find the number of distinct permutations taking G to H) and #GA (find the number of distinct permutations taking G to itself, i.e., the order of the automorphism group of G) are polynomially equivalent to each other and to GI. The precise definition of "graph" is not very important to the complexity of GI: GI for (undirected) graphs, **directed graphs** (where edges are *ordered* pairs of vertices), **multigraphs** (where two vertices may have more than one edge joining them), and **pseudographs** (where **loops**, edges joining a vertex to itself, are allowed) are all polynomially equivalent. (These equivalences can be proved by mapping the vertices and edges of these graph variants onto **gadgets**, clusters of vertices and edges of an undirected graph having the required properties. Another example of such a gadget is shown below, in Figure 2.1.) The Subgraph Isomorphism problem SubGI (given two

[2] Strictly speaking, it is the decision-problem version of Factoring (given $N, k > 1$, does N have a prime factor less than k?) that is in NP. This distinction is not crucial: both the version of Factoring defined here and GI (and many other problems in NP) have **self-computable** solutions: the function problems are polynomial-time reducible to the decision problems, by a sort of binary search on the solution space. An example of such a search for the graph isomorphism problem is given in this article.

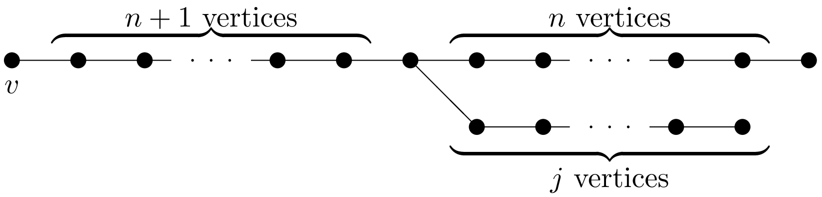


Figure 2.1: A label j ($1 \leq j \leq n$) for the vertex v, in a graph with n vertices.

graphs $G = (V, E)$ and $G' = (V', E')$, determine whether there is an injective mapping $\varphi : V \to V'$ such that if $\{e_1, e_2\} \in E$, $\{\varphi(e_1), \varphi(e_2)\} \in E'$) is apparently harder than GI, though: it is known to be NP-complete.[3] (Recall that a problem X in NP is called NP-complete if any problem in NP has a polynomial-time reduction to X.)

One way of looking at the difficulty in solving GI is to think of it as a difficulty in finding one of an exponentially large number of permutations. Consider the closely related function problem FGI: Given two graphs G, H, return a permutation P such that $A_G = PA_HP^{-1}$ if $G \sim H$ (otherwise return an arbitrary P). Clearly there is a polynomial reduction from GI to FGI: we merely check the returned P to see if it is in fact an isomorphism. There is also a polynomial reduction from FGI to GI: Suppose we have an algorithm that determines whether two graphs G, H are isomorphic. To determine an isomorphism with polynomially many uses of this algorithm, we inductively construct **partially labelled** graphs from G and H and test these graphs for isomorphism. We label a graph by attaching **labels** (gadgets made up of extra vertices and edges) to vertices in such a way that any isomorphism must map a vertex with a label to another vertex with an identical label; when the algorithm is complete, the isomorphism can be read directly from the labelled graphs. It is easy to construct such a gadget. One example is given here as Figure 2.1 (from [67]): The idea is that this set of gadgets is distinguishable from all structures appearing in the original graph (having more vertices than the base graph) and recognizable as the label j; any isomorphism must therefore map a label to an equivalent label.

The algorithm is as follows:

[3] This is easy to see; in particular, reduction of the well-known NP-complete problem HamiltonCycle to SubGI is trivial.

```
BEGIN with two isomorphic graphs $G$, $H$ on $n$ vertices.
    $G_0$ := $G$
    $H_0$ := $H$
    for $j$ := 1 to $n$
        Choose an unlabeled vertex $v$ of $G_{j-1}$ and give it the label $j$.
        for each unlabeled vertex $w$ of $H_{j-1}$ do
            Label $w$ with label $j$.                  [try partial matching $v \mapsto w$]
            if $G_{j-1} \sim H_{j-1}$ then
                break               [exit inner loop, having found that $\phi : v \mapsto w$]
            else
                remove label $j$ from vertex $w$.              [fail this partial matching]
        endif
        $G_j$ := $G_{j-1}$                       [save this partial matching and continue]
        $H_j$ := $H_{j-1}$
    $G'$ := $G_n$
    $H'$ := $H_n$
END; the identical labels of $G'$ and $H'$ explicitly define $\phi : G \overset{\phi}{\sim} H$.
```

This algorithm has $n = |V|$ steps; each step involves at most n queries to the GI subroutine, on a graph with at most $\mathcal{O}(n^2)$ vertices (since each label has $\mathcal{O}(n)$ vertices). Thus FGI is polynomial-time reducible to GI.

2.1.2 Classical algorithms

As mentioned above, GI seems to be hard because the space S_n of potential solutions is so large. It is not so surprising, then, that when this space can be reduced substantially, the problem becomes tractable. Polynomial-time algorithms are known for several types of graphs in which some *graph invariant* allows this search space to be reduced: Graphs with bounded degree [77] (the **degree** of a vertex v is the number of edges incident on v), graphs with bounded eigenvalue degeneracy [10] (the **eigenvalues** of an undirected graph are the (real) eigenvalues of its adjacency matrix), and graphs with bounded genus [48, 80] (the **genus** of a graph is that of the minimum-genus orientable surface on which the graph may be embedded with no edge crossings; for the particular case of planar graphs, a more readable algorithm is in Kučera [68]) all have polynomial-time isomorphism algorithms. (Typically the time bound for such results is of the form $\mathcal{O}(n^{ax+b})$, where x is the bounded quantity and a and b are constants.)

A simple instructive example is the algorithm for TREEGI where the graphs are restricted to be **trees** (connected acyclic graphs). The algorithm is as follows:

```
BEGIN with an unlabeled tree.
    Give each leaf (vertex of degree 1) the label 1.
    while the tree contains unlabeled vertices do
        for each unlabeled vertex $v$ adjacent to a vertex labeled in a
                previous iteration, attach the label $\{\ell_e\}$, the set of all
                labels of labeled vertices adjacent to $v$.
END with a labeled tree; the list of labels, sorted lexicographically, is a
        faithful invariant for the tree.
```

To determine whether two trees are isomorphic, one computes each tree's lexicographically-sorted list of labels and compares them. If the two lists are identical, the trees are isomorphic; if the lists differ, the trees are not isomorphic.[4]

This example illustrates an important concept in isomorphism problems, also mentioned in passing above: the idea of *invariants*. We now formalize this concept, since it will be useful in later discussion. A function $r(x) : X \to R$ (where the domain X is the **presentation space**, on which the isomorphism $\sim$ is defined, and the range R is the **representation space**) is an **invariant**, or a **canonical** representation, if it is constant on equivalence classes of X: i.e., $r(x) = r(y)$ if $x \sim y$. An invariant is **complete**, or **faithful**, if it is constant *only* on equivalence classes of X: i.e., $r(x) = r(y)$ iff $x \sim y$. A faithful invariant (often called a **certificate**) preserves all the information about an object's equivalence class, and so an efficient method of finding faithful invariants allows

[4] This algorithm may be improved somewhat, if all we care about is whether $G \sim H$, by noting that the lexicographically-sorted lists must agree *at each stage of labeling*; so we may compare the partial lists after each iteration of the `while` loop, returning NOT ISOMORPHIC if the lists ever differ. The algorithm is presented as shown to give an example of an invariant for a graph.

a solution to the isomorphism problem. In the example above, the set of labels for the completely-labelled tree is just such a faithful invariant. Unfortunately, no **deterministic** (polynomial-time computable) certificate r is known for a general graph ([31], for example, computes a certificate in exponential, but subfactorial, time). **Succinct** (polynomial-size) certificates *are* known: e.g., $\mathrm{first}\{H : H \sim G\}$, where $\mathrm{first}\, X$ gives the element of X which is lexicographically minimal.[5] Unfortunately, succinct certificates for graphs seem to be difficult to compute in general.

Note that a general isomorphism problem reduces polynomially to the problem of finding succinct certificates, but that these two problems are not necessarily equivalent: it may be easier to determine whether two objects are isomorphic than to find a succinct certificate for an object. It may, however, be more *useful* to find a succinct certificate. For instance, consider the application of finding a match to a previously-classified graph in a database of N graphs. If we only have an algorithm to solve the decision-problem version of GI for a graph, we must apply it on average $\mathcal{O}(N)$ times to the graphs in the database (we cannot sort the database, since we don't have a useful ordering function) to find a match. On the other hand, if we have an algorithm to find a succinct certificate for a graph, we can apply it once to find a certificate $\overline{G}$ for the unknown graph, and then compare (in average time $\mathcal{O}(|\overline{G}| \log N)$) the certificate to the certificates in the database (which is sorted, of course, on the certificate). Thus for large databases, finding succinct certificates is more useful than merely determining whether two graphs are isomorphic.

It is important, in considering the quantum case, to examine carefully what is needed for this method to work. The definition of an invariant needs no modification for the quantum case, but we must decide what we mean by a quantum certificate: Let us call a quantum representation (i.e., a representation by elements of a Hilbert space) **faithful** if distinct objects map to orthogonal states[6], $\langle r(x) \,|\, r(y) \rangle = 0$ if $x \not\sim y$. In the classical case, of course, it is easy to compare the two certificates, and they may be copied at

[5] It is easy to see that in fact all problems in P have deterministic certificates, and all problems in NP have succinct certificates.

[6] This restriction, at least, seems necessary to preserve the idea of a faithful representation; otherwise one cannot perfectly distinguish different certificates. Further restrictions on the allowable classes of basis states may also be useful to consider.

will; so the comparison routine is *perfect* and *information-preserving.* But if we have two *quantum* certificates, neither of these conditions necessarily holds: I know of no efficient ways to implement either a perfect comparison routine or a perfect copier for states chosen from an arbitrary (but known) orthogonal set, so a database of quantum certificates is more difficult to use than a database of classical certificates.[7] Thus, it seems that generic quantum certificates have less power than classical certificates.

2.2 Some primitives for quantum computation

Only a few generic primitive operations seem to differentiate the currently known quantum algorithms from classical ones. (This is not intended as a rigorous statement; it is merely an observation that very few extra generally-interesting tools seem to be available thus far to the quantum computer programmer.) These are the useful primitives which differ from the tools available classically, as I see them:

(i) The simplest obviously quantum primitive is the **creation of superpositions**

$$\tfrac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \ , \tag{2.1}$$

which can be implemented in time $\mathcal{O}(\log N)$. (This primitive can be viewed as a special case of the quantum fast Fourier transform, (iv) below, for input $|0\rangle$. We consider it to be its own primitive, though, both because it seems more fundamental than the QFFT, appearing in more algorithms, and because it is somewhat easier to implement than the general QFFT.)

(ii) Since unitary evolution is necessarily reversible, there is generally no exact analog of the classical primitive of **function evaluation**, $x \mapsto f(x)$. Instead, to ensure

[7] One way to perform these operations, since the orthonormal set is assumed known, is to rotate this set into the computational basis, perform the measurement or copy operation, and then rotate back into the set of interest. However, these rotations may be difficult to perform (because the certificates may be difficult to construct) so this is an unsatisfactory solution. In §2.3.1 we will discuss some simple comparison methods which are independent of the given set but which provide only a probabilistic answer.

reversibility we must in general preserve the input value:

$$|x\rangle_{\mathrm{X}}\,|0\rangle_{\mathrm{Y}} \xrightarrow{U[f]} |x\rangle_{\mathrm{X}}\,|f(x)\rangle_{\mathrm{Y}}\ , \tag{2.2}$$

which can be implemented efficiently whenever f is efficiently computable classically. Because $U(f)$ is necessarily linear, it acts linearly on superpositions; this is the operation often pointed to as “quantum parallelism.” But although this primitive, along with superposition creation ((i) above) is often credited with giving quantum computers their added power, these primitives alone are insufficient; the ability for both constructive and destructive interference between different “computational paths” is also necessary.[8]

(iii) If f is injective, then the operator mapping $|x\rangle$ to $|f(x)\rangle$ is unitary. If we want to save space by reusing qubit registers, we need to erase the input registers containing x after computing the outputs $f(x)$. (If we have a state $\sum_i |x_i\rangle_{\mathrm{X}}\,|f(x_i)\rangle_{\mathrm{Y}}$, we can't merely discard the x register and prepare new qubits; such an operation would cause the coherent superposition to decohere into a mixed state.) The quantum implementation of such an operator seems to require, in general, an algorithm to evaluate f^{-1} as well as f. (f^{-1} is needed to “uncompute” the input $|x\rangle_{\mathrm{X}}$ after the output $|f(x)\rangle_{\mathrm{Y}}$ has been computed.) Thus, a **bijection evaluation** can be implemented in two steps,

$$|x\rangle_{\mathrm{X}}\,|0\rangle_{\mathrm{Y}} \xrightarrow{U[f]} |x\rangle_{\mathrm{X}}\,|f(x)\rangle_{\mathrm{Y}} \xrightarrow{U[f^{-1}]^{-1}} |0\rangle_{\mathrm{X}}\,|f(x)\rangle_{\mathrm{Y}}\ , \tag{2.3}$$

which can be implemented efficiently whenever *both* f and f^{-1} are efficiently computable classically.

(iv) The **quantum (fast) Fourier transform** (QFFT), defined (for the simplest case

[8] Exactly what characteristics of a quantum computer are responsible for its apparent power beyond that of a classical computer is a subject of some debate. A few interesting discussions can be found in the manifesto of the Quantum Computation Collective [30] and in articles by Steane [95] and by Jozsa [59] and Ekert and Jozsa [42], but many other opinions are there for the asking.

of a cyclic group $\mathbb{Z}_n$) by

$$|x\rangle \overset{\text{QFFT}}{\longrightarrow} \frac{1}{\sqrt{N}} \sum_{i=0}^{N-1} e^{2\pi i x y/N} |y\rangle \,, \tag{2.4}$$

can be used to sample the eigenvalues $e^{i\phi_n}$ of an operator U, with probability distribution for an eigenvalue given by the corresponding eigenstate's overlap with the given initial state, if the controlled-U operator

$$\Lambda(U)_{[\![\mathrm{A}]\!],\mathrm{B}} : |k\rangle_{\mathrm{A}} |\psi\rangle_{\mathrm{B}} \mapsto |k\rangle_{\mathrm{A}} (U_{\mathrm{B}})^k |\psi\rangle_{\mathrm{B}}$$

can be implemented.[9] For instance, if we are given an eigenstate $|\phi\rangle$ for U, with $U|\phi\rangle = e^{i\phi}|\phi\rangle$, the unitary operator $\mathrm{Eig}(U)_{\mathrm{A,B}} \equiv \mathrm{QFFT}_{\mathrm{A}} \cdot \Lambda(U)_{[\![\mathrm{A}]\!],\mathrm{B}} \cdot \mathrm{QFFT}_{\mathrm{A}}$

[9] The basic definition of the controlled-U operator, as mentioned previously, has a two-dimensional Hilbert space for the control register (so that U is performed when the control qubit is $|1\rangle$, in the computational basis, and not performed when the control qubit is $|0\rangle$). Here we extend the definition to an N-dimensional control-qubit Hilbert space, where the unitary operation performed is U^k if the control register is in state $|k\rangle$. This can be performed efficiently by, for example, performing the sequence of operations

$$\prod_{i=0}^{a-1} \Lambda(U^{2^i})_{[\![\mathrm{A}_i]\!],\mathrm{B}}$$

where A_i refers to the Hilbert space of bit i of register A.

allows estimation of ϕ:

$$|0\rangle_{\mathrm{A}}\,|\phi\rangle_{\mathrm{B}}$$

$$\mathrm{QFFT}_{\mathrm{A}}\Big\downarrow\text{Perform the QFFT on state } |0\rangle_{\mathrm{A}} \text{ to create the superposition}$$

$$\tfrac{1}{\sqrt{N}}\sum_{k=0}^{N-1}|k\rangle_{\mathrm{A}}\,|\phi\rangle_{\mathrm{B}}$$

$$\Lambda(U)_{\llbracket\mathrm{A}\rrbracket,\mathrm{B}}\Big\downarrow\text{Perform } U^k\text{, where } k \text{ is the value in register A, on register B}$$

$$\tfrac{1}{\sqrt{N}}\sum_{k=0}^{N-1}|k\rangle_{\mathrm{A}}(U^k)_{\mathrm{B}}\,|\phi\rangle_{\mathrm{B}} = \tfrac{1}{\sqrt{N}}\sum_{k=0}^{N-1}e^{ik\phi}\,|k\rangle_{\mathrm{A}}\,|\phi\rangle_{\mathrm{B}}$$

$$\mathrm{QFFT}_{\mathrm{A}}\Big\downarrow\text{Perform a second QFFT on register A}$$

$$\tfrac{1}{N}\sum_{j=0}^{N-1}\left[\sum_{k=0}^{N-1}e^{ik(\phi+\frac{2\pi j}{N})}\right]|j\rangle_{\mathrm{A}}\,|\phi\rangle_{\mathrm{B}}$$

$$\Big\downarrow\text{Measure register A}$$

$$|j\rangle_{\mathrm{A}}\,|\phi\rangle_{\mathrm{B}}\ , \quad \text{with probability}$$

$$\left|\sum_{k=0}^{N-1}e^{ik\theta_j}\right|^2 = \frac{\sin^2\frac{N}{2}\theta_j}{N^2\sin^2\frac{1}{2}\theta_j}\ \text{, where } \theta_j \equiv \phi + \tfrac{2\pi j}{N}.$$

This probability is small unless θ_j is small, i.e., unless $j \approx -\frac{N\phi}{2\pi}$; so sampling j can give a good estimate (to $\mathcal{O}(\frac{1}{N})$) of ϕ. In particular, if $\Lambda(U)_{\llbracket\mathrm{A}\rrbracket,\mathrm{B}}$ can be *efficiently* implemented (i.e., in time polynomial in $n = \lg N$, for an n-qubit register A), then we may estimate ϕ with n bits of precision in time polynomial in n. (If $U^N = \mathbf{1}$, so that the eigenvalues are Nth roots of unity, then the correlation is exact; the bracketed sum is zero except for a single value of j.) Sampling the eigenvalue also has the happy side effect of projecting out (approximately) an eigenstate with that eigenvalue, so we can use the $\mathrm{Eig}(U)$ to sample eigenstates of U.

The QFFT allows, for example, solution of the problem of **period finding**, the determination of the period r of a strictly periodic black-box function f (i.e., one for which $f(n+m) = f(n)$ if and only if $m = kr$, $r \in \mathbb{Z}$), in time $\mathcal{O}\left[(\log r)^2\right]$. (The optimal classical algorithm, for black-box f, requires average time $\mathcal{O}(r^{1/2})$ and worst-case time $\mathcal{O}(r)$.) PeriodFinding is closely related to the above eigenvalue-estimation problem; the operator $\Lambda(U)$ is replaced with operator $U[f]$ having action $U[f]_{[\![\mathrm{X}]\!],\mathrm{Y}}\,|n\rangle_{\mathrm{X}}\,|y\rangle_{\mathrm{Y}} = |n\rangle_{\mathrm{X}}\,|y + f(n) - f(0)\rangle_{\mathrm{Y}}$. Now $U[f]$ acts as the identity on the subspace spanned by $\{|kr\rangle_{\mathrm{X}}\,|y\rangle_{\mathrm{Y}} : k \in \mathbb{Z}\}$, so (using the same argument as above) the phases measured by the QFFT must have the form $\phi_k = 2\pi i \frac{k}{r}$, for some integers k. Large divisors of the denominator r (the desired period) can be found with high probability, assuming the values k are uniformly distributed in $\{0, \ldots, r-1\}$, using continued-fraction expansions of the measured phases $\frac{\phi_k}{2\pi}$. (If k and r have common divisors, the continued-fraction expansion of the phase will actually find the fraction $\frac{k'}{r'}$ *reduced to lowest form*, giving only a divisor of r; sampling the result a few times, together with guessing at small missing factors, suffices to determine r with high probability. A detailed consideration of PeriodFinding, including the distribution of the measured denominators r', is somewhat involved; see Shor's pioneering papers [91, 92] for the full details.[10]) The QFFT is quite fast: it can be implemented in time $\mathcal{O}\left[(\log N)^2\right]$, compared with $\mathcal{O}(N \log N)$ for an implementation of the classical FFT.[11] The reason the QFFT is so fast is that the vector of complex amplitudes is encoded directly in the phases of the elements of the superposition, so that the $\mathcal{O}(N)$ complex additions are performed automatically by the linear rules of quantum mechanics. (The Fourier transform often shows up in describing changes of basis, e.g., between position and momentum bases, so it might not be too surprising

[10] Shor's algorithm uses a probabilistic reduction of IntegerFactoring to PeriodFinding. The overall goal, following a method originated by Fermat and refined by Kraitchik (and shared with many classical factoring algorithms such as the quadratic sieve—see [26] for a review), is to find a nontrivial solution of $x^2 \equiv 1 \pmod{N}$; given such x, $\gcd(x \pm 1, N)$ yields a nontrivial factor of N. This can be probabilistically reduced to the problem of finding the order of an element a of the multiplicative group modulo N, an instance of PeriodFinding: the order r of a is even with reasonable probability, $r = 2s$; so $a^r \equiv (a^s)^2 \equiv 1 \pmod{N}$. The modular-exponentiation operator whose period we want to find, $\textsc{Exp}(a,N)_{[\![\mathrm{X}]\!],\mathrm{Y}} : |x\rangle_{\mathrm{X}}\,|1\rangle_{\mathrm{Y}} \mapsto |x\rangle_{\mathrm{X}}\,|a^x \bmod N\rangle_{\mathrm{Y}}$, is efficiently implementable (see [11] for an explicit $\mathcal{O}\left[(\log N)^3\right]$ implementation). Again with reasonable probability we have $a^s \not\equiv \pm 1 \pmod{N}$, giving a nontrivial factor of N. Preskill [88, §§6.9-6.12] also has a nice explanation of the QFFT and its uses in PeriodFinding and IntegerFactoring.

[11] See, e.g., [83, Chapter 5] for a detailed implementation.

that it has a simple implementation in quantum computation.) Note, however, that the QFFT is less powerful than the classical FFT: after a classical FFT, the sampled values of the transformed function $F(k)$ are all known classically; after a QFFT, the sampled values are represented as state amplitudes, which cannot be measured directly. In addition, for the complexity reduction to be important there must be an efficient—polynomial in $\log N$—method for preparing the initial state; in many classical applications of the FFT, the preparation algorithm is $\mathcal{O}(N)$ and so improvement of the FFT to $o(N)$ will not substantially reduce the overall complexity. The QFFT is in fact a solution to a particular case of a more general problem, the **Abelian stabilizer problem** (given an Abelian group A, a finite set X acted on by A, and $x \in X$, find the **stabilizer** $A_x = \{g \in A : g(x) = x\}$ of x in A, e.g., by giving a basis for A_x), which is also efficiently solvable on a quantum computer [61, 62]. Unfortunately, though, the *nonabelian* stabilizer problem, to which GI easily reduces, has no known polynomial-time algorithm. (An algorithm for GI based on this reduction, developed by Ettinger and Høyer [43], is discussed in §2.3.3. The algorithm does not appear to be polynomial-time; one of the observables used seems to be difficult to implement.)

(v) If a search doesn't seem to be reducible to a period-finding problem, all is not lost. Given an arbitrary function $f : \{0, \ldots, N-1\} \to \{0, 1\}$, the **Grover search** algorithm [55] (the presentation here follows Preskill [88, §6.4] in a formalism originally due to Brassard and Høyer [24]) allows determination of a solution i to $f(i) = 1$, if one exists, in $\mathcal{O}(\sqrt{N})$ evaluations of f, a quadratic improvement over the $\mathcal{O}(N)$ evaluations of f required classically. The Grover search can be thought of as a rotation of a quantum state in a two-dimensional space, spanned by the two vectors

$$\begin{aligned} |\alpha\rangle &= \tfrac{1}{\sqrt{N}} \sum_{i=0}^{N-1} |i\rangle \\ |\omega\rangle &= \tfrac{1}{\sqrt{n}} \sum_{i:f(i)=1} |i\rangle\,, \quad n \equiv \big|\{i : f(i) = 1\}\big|\,. \end{aligned} \tag{2.5}$$

Such a rotation can be implemented as the product of reflections along the axes defined by $|\alpha\rangle\langle\alpha|$ and $|\omega\rangle\langle\omega|$, i.e., the product $-U_\omega U_\alpha$ of the unitary operators $U_\alpha =$

$\mathbf{1} - 2\,|\alpha\rangle\langle\alpha|$ and $-U_\omega = 2\,|\omega\rangle\langle\omega| - \mathbf{1}$ (for details see, e.g., Jozsa [60]). U_α can be constructed if we can *detect* the state $|\alpha\rangle$ (e.g., by a unitary operator U with $U\,|\alpha\rangle\,|0\rangle = |\alpha\rangle\,|1\rangle$ and $U\,|\alpha^\perp\rangle\,|0\rangle = |\alpha^\perp\rangle\,|0\rangle$ for any $\langle\alpha\,|\,\alpha^\perp\rangle = 0$); similarly U_ω can be constructed if we can detect the state $|\omega\rangle$. Thus the Grover search is feasible if we can *reliably detect* (on the two-dimensional Hilbert space spanned by $|\alpha\rangle$ and $|\omega\rangle$) the **initial state** $|\alpha\rangle$ and the desired **final state** $|\omega\rangle$. In particular, these two operations can be efficiently implemented to find a succinct certificate for any problem in NP: using the terminology of §1.2.2, we wish to solve the equation $\hat{f}(x, y) = 1$ for a valid certificate y (if one exists), given a problem instance x.

The operator $-U_\omega U_\alpha$ rotates a state in this two-dimensional space by the angle $2\theta = 2\sin^{-1}|\langle\alpha\,|\,\omega\rangle|$ (acting trivially on the orthogonal complement of this space), so rotating the initial state $|\alpha\rangle$ to the final state $|\omega\rangle$ with the Grover rotation $(-U_\omega U_\alpha)^M$ will take about $M \sim \frac{\pi}{2|\langle\alpha|\omega\rangle|} = \mathcal{O}\big(\,|\langle\alpha\,|\,\omega\rangle|^{-1}\,\big)$ iterations.[12] For the simplest case of finding the unique solution to $f(i) = 1$ presented above, the Grover algorithm with projectors on the states (2.5) takes $\mathcal{O}\big(\,|\langle\alpha\,|\,\omega\rangle|^{-1}\,\big) = \mathcal{O}\big(\frac{1}{\sqrt{N}}\big)$ iterations. The initial state may also be some state other than $|\alpha\rangle$, as analyzed by Brassard *et al.* [25] and by Biron *et al.* [23, 22]; clever choice of initial state, if partial information is known about the solution set, can reduce the time required to find a solution.

There are a number of possible generalizations to the form of Grover's algorithm as presented above. For instance, note that $|\alpha\rangle$ and $|\omega\rangle$ need not have the specific form of (2.5); given any two noncommuting unitary matrices U_α, U_ω, each performing an inversion about a single axis (i.e., each having one eigenvalue -1, the remainder being $+1$), the Grover iteration will perform a rotation in the plane spanned by these two axes, so the Grover algorithm can be used to construct states $|\omega\rangle$ besides the equally-weighted superpositions of (2.5).

[12] Unless the Grover rotation angle 2θ is commensurate with $\frac{\pi}{2}$, M will not be an integer, and no integral number of Grover iterations will take $|\alpha\rangle$ exactly to the desired state $|\omega\rangle$. However, for large N the state after M iterations will have high fidelity with the state $|\omega\rangle$, so measurement will give the desired result with high probability. In addition, one can construct a modified Grover iteration with rotation angle $2\lambda\theta$ for any $\lambda \in [0, 1]$; using one of these iterations, a rotation by any desired angle can be implemented exactly. This technique is discussed in Høyer [58] but is not necessary for this paper.

In fact, U_α and U_ω need not be single-axis inversions. For instance, suppose more generally that $U_\alpha = \mathbf{1} - 2P_\alpha$ and $U_\omega = \mathbf{1} - 2P_\omega$, where P_α and P_ω are **projectors** (idempotent Hermitian operators). In the simple case where either P_α or P_ω is still a one-dimensional projector, the situation is essentially the same as before; for example, if P_α is one-dimensional, we can decompose $P_\omega = \sum_i P_{\omega i}$ as a sum of one-dimensional projectors, all but one of which are orthogonal to P_α. The remaining one-dimensional projector $P_{\omega 0}$ may be substituted for P_ω in the Grover algorithm with essentially no change (a single iteration $-U_\omega U_\alpha$ will effect a reflection along the axes specified by $P_{\omega i}$ with $i \neq 0$; but a *pair* of Grover iterations is unaffected:

$$\begin{aligned}
(-U_\omega U_\alpha)^2 &= \Big(\mathbf{1} - 2\sum_i P_{\omega i}\Big)(\mathbf{1} - 2P_\alpha)\Big(\mathbf{1} - 2\sum_i P_{\omega i}\Big)(\mathbf{1} - 2P_\alpha) \\
&= \Big[\prod_i (\mathbf{1} - 2P_{\omega i})\Big](\mathbf{1} - 2P_\alpha)\Big[\prod_i (\mathbf{1} - 2P_{\omega i})\Big](\mathbf{1} - 2P_\alpha) \\
&= (\mathbf{1} - 2P_{\omega 0})(\mathbf{1} - 2P_\alpha)(\mathbf{1} - 2P_{\omega 0})(\mathbf{1} - 2P_\alpha)
\end{aligned}$$

since the orthogonal projectors commute). The case in which both projectors are multidimensional is more complicated. In this general case it is easy to see that $U_\alpha U_\omega$ still decomposes as a direct sum of rotations in orthogonal two-dimensional subspaces (the eigenvalues of $U_\alpha U_\omega$ other than ±1 appear in complex-conjugate pairs). Generically there will be one such rotation for each dimension of the lower-dimensional projector[13]; the rotation angles will generally be incommensurate, leading to complicated trajectories for generic input states.[14] Another generalization, studied by Brassard *et al.* [25] and by Biron *et al.* [23, 22], is the removal of the requirement that the initial state be a uniform superposition over the solution space.

As a prelude to the discussion of GI, I would like to discuss one other primitive which does not make the list above (because it has not yet proved generically useful in the most interesting quantum algorithms). The notion of a *set*, an unordered collection of elements, is of course a very useful mathematical primitive. In classical computer programming, a

[13] More precisely, in a space of dimension N with $n_\alpha \equiv \operatorname{tr} P_\alpha$ and $n_\omega \equiv \operatorname{tr} P_\omega$, there will generically be $\min\{n_\alpha, n_\omega, N - n_\alpha, N - n_\omega\}$ rotation angles.

[14] I don't know of any situations where this case is useful.

faithful, canonical representation of a set with polynomially many elements is trivial to create: We merely arrange the elements in lexicographic order. This reduces a set to a list, which is a far more natural structure to implement algorithmically. The problem is somewhat more subtle for the quantum computer programmer: The transformation from an unsorted list to a sorted list is not reversible (since different orderings of the same elements map to the same sorted list) and so not unitary, so we cannot expect set construction and manipulation operations to be trivial to implement.

Ignoring for the moment this difficulty, there are two essentially different constructions for implementing a set. In the first method, we use the classical representation as a list (sorted according to an arbitrary but fixed total order), or, in a pretty quantum variant, we represent the set as a superposition over all orderings of the list of elements:

$$|\{x_1, \ldots, x_n\}_1\rangle \equiv \tfrac{1}{\sqrt{n!}} \sum_{\sigma \in S_n} |x_{\sigma_1}\rangle \cdots |x_{\sigma_n}\rangle \,. \tag{2.6}$$

This representation is *always faithful* (that is, distinct objects map to orthogonal states): two sets X, Y with $|X| = |Y| = n$ have $\langle X \,|\, Y\rangle \neq 0$ iff $X = Y$, in which case $\langle X \,|\, Y\rangle = 1$. (The requirement that $|X| = |Y|$ is only necessary so that the inner product $\langle X \,|\, Y\rangle$ is a scalar. We could alternately "pad" the smaller set with values guaranteed not to be in either X or Y to perform the comparison even in the case $|X| \neq |Y|$.) Thus this representation is in fact a certificate for the set.

A second method represents the set as a coherent superposition of its elements:

$$|\{x_1, \ldots, x_n\}_2\rangle \equiv \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} |x_i\rangle \,. \tag{2.7}$$

This method has an advantage over the first, and over the classical representation above: The restriction to polynomial-sized sets is no longer necessary (since we only require enough space to hold a single element). But there is a corresponding disadvantage. This representation is *not always faithful*: two sets X, Y have $\langle X \,|\, Y\rangle = \frac{|X \cap Y|}{\sqrt{|X| \cdot |Y|}}$. (Thus this representation, although an invariant, is not a certificate for general sets. It is, however, a certificate for classes of disjoint sets, though.) Which representation is useful in a particular instance depends on the desired behavior and on the class of sets we wish to represent.[15] Below I will present a faithful certificate for a topological graph which uses both of these set representations.

2.3 Quantum algorithms

The simplest quantum technique to speed up a classical search algorithm is Grover's algorithm. Given a classical algorithm which uses an initial guess to try to find a solution to an instance of GI and which requires $\mathcal{O}(N)$ guesses on average to find a correct solution, Grover's algorithm can reduce the average run-time (measured by the number of times a comparison of the form $H \overset{?}{=} \sigma(G)$ is made) quadratically, from $\mathcal{O}(N)$ to $\mathcal{O}(\sqrt{N})$. For an instance of GI on n vertices, the solution space has size $S_n = n!$; thus a brute-force search of the solution space requires time $\mathcal{O}(n!)$, and Grover's algorithm reduces this time to $\mathcal{O}\big(\sqrt{n!}\big)$. (Of course, for a graph with more structure, e.g., one in which not all vertices have the same degree, the search space can be easily reduced by classical methods before the Grover search is started.) Unfortunately, since the speedup is only quadratic, such a search is not polynomial in n unless the size of the original classical search space is also polynomial.

Other attempts at quantum algorithms for GI make use of extra structure in the problem. For instance, GI reduces (as mentioned above) to the problem of finding a succinct (classical) certificate for a topological graph $\overline{G}$ (given a graph G). We may try to solve GI by considering the analogous quantum solution method: Suppose that we have an algorithm producing a succinct *quantum* certificate $\big|\overline{G}\big\rangle$ given a graph G. Then we may determine whether two graphs G, H are isomorphic by determining whether the two quantum certificates $\big|\overline{G}\big\rangle$, $\big|\overline{H}\big\rangle$ are identical or orthogonal. Two probabilistic algorithms to make this measurement are presented here. (In each case, however, the given certificate states may end up entangled with the ancilla registers, so we cannot guarantee that they are reusable.) Later, in §2.4, we consider the problem of actually constructing such states.

[15] Since the time of writing, this set primitive has found use in Watrous' certificate for the group non-membership problem (discussed in §2.5.1 below).

2.3.1 Solving GI given a succinct quantum certificate

In the first construction, we begin with the state $|0\rangle\left|\overline{G}\right\rangle + |1\rangle\left|\overline{H}\right\rangle$, a superposition of two succinct certificates. (Given a quantum algorithm to construct the certificates $\left|\overline{G}\right\rangle$, such a state can be created by running the algorithm twice, conditioned on the state of the control qubit $|\cdot\rangle_{\mathrm{c}}$.)

$$\frac{1}{\sqrt{2}}\Big(|0\rangle_{\mathrm{c}}\left|\overline{G}\right\rangle_{\mathrm{A}} + |1\rangle_{\mathrm{c}}\left|\overline{H}\right\rangle_{\mathrm{A}}\Big)$$

$$H_1 \quad \Big\downarrow \text{Perform a Hadamard gate on qubit c}$$

$$\frac{1}{2}\Big[|0\rangle_{\mathrm{c}}\Big(\left|\overline{G}\right\rangle_{\mathrm{A}} + \left|\overline{H}\right\rangle_{\mathrm{A}}\Big) + |1\rangle_{\mathrm{c}}\Big(\left|\overline{G}\right\rangle_{\mathrm{A}} - \left|\overline{H}\right\rangle_{\mathrm{A}}\Big)\Big]$$

$$\Big\downarrow \text{Measure qubit c}$$

$$\begin{cases} |0\rangle_{\mathrm{c}}, \text{ with probability 1, if } G \sim H, \text{ so that } \left\langle\overline{G}\,\middle|\,\overline{H}\right\rangle = 1 \\ |0\rangle_{\mathrm{c}} \text{ or } |1\rangle_{\mathrm{c}}, \text{ each with probability } \frac{1}{2}, \text{ if } G \not\sim H, \text{ so that } \left\langle\overline{G}\,\middle|\,\overline{H}\right\rangle = 0 \end{cases}$$

If we repeat the state preparation and measurement n times, then conclude that $G \sim H$ if only $|0\rangle$s were measured and $G \not\sim H$ if at least one $|1\rangle$ was measured, our conclusion will be in error with probability at most 2^{-n} (but will be certain if a $|1\rangle$ is ever measured).[16]

A prettier method [63] uses the controlled swap (a series of Fredkin gates). This time

[16] This construction can be thought of as a special case of the QFFT algorithm for PeriodFinding, §2.2(iv), where the period is 1 if the graphs are isomorphic and 2 if not. Similar $H\Lambda(U)H$ constructions occur in many contexts in quantum information theory.

we may begin with the certificates in a product state $|\overline{G}\rangle|\overline{H}\rangle$.

$$\begin{tikzcd}
{|0\rangle_{\mathrm{c}}|\overline{G}\rangle_{\mathrm{A}}|\overline{H}\rangle_{\mathrm{B}}} \arrow[d, "H_{\mathrm{c}}"', "\text{Perform a Hadamard gate on qubit c}"] \\
{\frac{1}{\sqrt{2}}\left(|0\rangle_{\mathrm{c}}+|1\rangle_{\mathrm{c}}\right)|\overline{G}\rangle_{\mathrm{A}}|\overline{H}\rangle_{\mathrm{B}}} \arrow[d, "\text{Swap registers A, B conditioned on qubit c}"] \\
{\frac{1}{\sqrt{2}}\left(|0\rangle_{\mathrm{c}}|\overline{G}\rangle_{\mathrm{A}}|\overline{H}\rangle_{\mathrm{B}}+|1\rangle_{\mathrm{c}}|\overline{H}\rangle_{\mathrm{A}}|\overline{G}\rangle_{\mathrm{B}}\right)} \arrow[d, "H_{\mathrm{c}}"', "\text{Perform a second Hadamard gate on qubit c}"] \\
{\frac{1}{2}\left[|0\rangle_{\mathrm{c}}\left(|\overline{G}\rangle_{\mathrm{A}}|\overline{H}\rangle_{\mathrm{B}}+|\overline{H}\rangle_{\mathrm{A}}|\overline{G}\rangle_{\mathrm{B}}\right)+|1\rangle_{\mathrm{c}}\left(|\overline{G}\rangle_{\mathrm{A}}|\overline{H}\rangle_{\mathrm{B}}-|\overline{H}\rangle_{\mathrm{A}}|\overline{G}\rangle_{\mathrm{B}}\right)\right]} \arrow[d, "\text{Measure qubit c}"] \\
{\begin{cases} |0\rangle_{\mathrm{c}}\text{, with probability 1, if } G\sim H \text{ (so } |\overline{G}\rangle=|\overline{H}\rangle\text{)} \\ |0\rangle_{\mathrm{c}} \text{ or } |1\rangle_{\mathrm{c}}\text{, each with probability } \frac{1}{2}\text{, if } G\nsim H \text{ (so } \langle\overline{G}\,|\,\overline{H}\rangle=0\text{)} \end{cases}}
\end{tikzcd}$$

This construction has the same failure probability as the first and can be subjected to the same probability amplification process. Note that this second construction (unlike the first) only requires the states $|\overline{G}\rangle$, $|\overline{H}\rangle$ to be equal *up to phase*: it works as long as $|\langle\overline{G}\,|\,\overline{H}\rangle| = 1$.

These constructions both provide only probabilistic answers. It is easy to show that no construction can decide with probability 1 whether a given pair of states chosen from an *unknown* orthonormal set. In fact, for the case of an unknown set of orthonormal states, if we require the algorithm to return $|0\rangle$ with certainty if the states are parallel and allow the state to return $\sqrt{1-p}\,|0\rangle + \sqrt{p}\,|1\rangle$ if the states are orthogonal (as in the two constructions above), then we *must* have $p \leq \frac{1}{2}$: the two algorithms above are optimal. (As mentioned previously, we may be able to do better in this case, because we know the form of the states $|\overline{G}\rangle$. It is not clear whether there is a computationally efficient method for improving the accuracy of this state comparison for these graph certificates.)

In the constructions above, we actually only require a reasonable approximation $|\widetilde{\overline{G}}\rangle$

to $|\overline{G}\rangle$, since the methods above will succeed as long as the two probabilities

$$\begin{aligned} p_{\mathrm{y}} &\equiv P\{|0\rangle \text{ measured} : G \sim H\} \\ p_{\mathrm{n}} &\equiv P\{|0\rangle \text{ measured} : G \not\sim H\} \end{aligned}$$

are polynomially distinguishable, i.e., $|p_{\mathrm{y}} - p_{\mathrm{n}}| \geq \frac{1}{p(n)}$ for some polynomial $p(n)$, so that with polynomially many trials we can resolve the two distributions. For the first construction described above,

$$P\{|0\rangle\} = \tfrac{1}{4}\Big|\,|\widetilde{\overline{G}}\rangle + |\widetilde{\overline{H}}\rangle\Big|^2 = \tfrac{1}{2}\Big[1 + \Re\langle\widetilde{\overline{G}}\,|\,\widetilde{\overline{H}}\rangle\Big] \ ;$$

if we require that $\Re\langle\overline{G}\,|\,\widetilde{\overline{G}}\rangle > 1 - p$ for all G, then some simple algebra gives worst-case bounds on $\langle\overline{G}\,|\,\widetilde{\overline{G}}\rangle$:

$$\begin{aligned} \Re\langle\widetilde{\overline{G}}\,|\,\widetilde{\overline{H}}\rangle &\geq 1 - 4p + 2p^2 \ , \quad G \sim H \\ \Re\langle\widetilde{\overline{G}}\,|\,\widetilde{\overline{H}}\rangle &\leq 2\sqrt{2p - p^2} \ , \qquad G \not\sim H \ ; \end{aligned}$$

and so

$$|p_{\mathrm{y}} - p_{\mathrm{n}}| \geq \tfrac{1}{2} - \sqrt{2p - p^2} - 2p + p^2 \ ,$$

and the distributions can be efficiently resolved as long as $p < 1 - \sqrt[4]{\frac{3}{4}} \approx 0.07$ (with at least a polynomial gap). For the second construction (in which the relative phase of the certificates does not matter),

$$P\{|0\rangle\} = \tfrac{1}{4}\Big|\,|\widetilde{\overline{G}}\rangle|\widetilde{\overline{H}}\rangle + |\widetilde{\overline{H}}\rangle|\widetilde{\overline{G}}\rangle\Big|^2 = \tfrac{1}{2}\left(1 + |\langle\widetilde{\overline{G}}\,|\,\widetilde{\overline{H}}\rangle|^2\right) \ ;$$

if we require that $|\langle\overline{G}\,|\,\widetilde{\overline{G}}\rangle| = 1 - p$, the worst-case bounds on $|\langle\overline{G}\,|\,\widetilde{\overline{G}}\rangle|$ are

$$\begin{aligned} |\langle\widetilde{\overline{G}}\,|\,\widetilde{\overline{H}}\rangle| &\geq 1 - 4p + 2p^2 \ , \quad G \sim H \\ |\langle\widetilde{\overline{G}}\,|\,\widetilde{\overline{H}}\rangle| &\leq 2\sqrt{2p - p^2} \ , \qquad G \not\sim H \ ; \end{aligned}$$

and

$$|p_{\mathrm{y}} - p_{\mathrm{n}}| \geq \tfrac{1}{2} - 8p + 12p^2 - 8p^3 + 2p^4 = 2(1-p)^4 - \tfrac{3}{2} \ ,$$

(as long as $p \leq 1 - \frac{1}{2}\sqrt{2}$), so the distributions can again be resolved as long as $p < 1 - \sqrt[4]{\frac{3}{4}}$

(with at least a polynomial gap). (These are worst-case estimates assuming that all the phases conspire against us; it is likely that far worse fidelities for the states $\left|\overline{G}\right\rangle$ will suffice for distinguishability in reality.)

2.3.2 Examples of succinct certificates

We now give an explicit example of a succinct certificate $\left|\overline{G}\right\rangle$: Given a graph G, represented by an $n \times n$ adjacency matrix A_G, let[17]

$$\left|\overline{G}\right\rangle = \sum_{\sigma \in S_n} \left| P_\sigma A_G P_\sigma^{-1} \right\rangle . \tag{2.8}$$

It is clear that this state is a faithful representation of the topological graph $\overline{G}$. A similar construction works for other representations of the graph; the representation of the topological graph is made canonical by *summing over all possible representations of the topological graph.* For example, using the edge-list representation $\big((v_1, w_1), \ldots, (v_e, w_e)\big)$ for a graph $G = (V, E)$ with no isolated vertices (so that $V = \{v_1, \ldots, v_e, w_1, \ldots, w_e\}$ and $E = \{(v_1, w_1), \ldots, (v_e, w_e)\}$),

$$\left|\overline{G}\right\rangle = \sum_{\sigma \in S_n} \sum_{\pi \in S_e} \left|(\sigma_{v_{\pi_1}}, \sigma_{w_{\pi_1}}), \ldots, (\sigma_{v_{\pi_e}}, \sigma_{w_{\pi_e}})\right\rangle , \tag{2.9}$$

summing both over all permutations of vertex labels and over all orderings of the edges in the list. Note that this construction uses *both* representations of sets discussed above: The set of edges of the labelled graph is represented using the first method $|\{\cdots\}_1\rangle$, while the set of all graphs in the topological graph's equivalence class is represented using the second method $|\{\cdots\}_2\rangle$. We can use the second form here because the set of topological graphs is a partition of the set of graphs; so $\overline{G} \cap \overline{H} \neq \varnothing$ iff $\overline{G} = \overline{H}$.

Unfortunately, such representations seem to be difficult to construct, even approximately. Several methods for constructing $\overline{G}$ (none of which seem to be efficient) are considered in §2.4 below.

The above arguments show that there is a probabilistic polynomial-time reduction

[17] We omit the normalization factors where this will not cause confusion.

from GI to the problem of constructing the states $\left|\overline{G}\right\rangle$. In fact, it is easy to see that a solution to GI allows construction of such states, so the reduction also goes the other way: GI is *equivalent* to the problem of constructing the states $\left|\overline{G}\right\rangle$. As discussed above, a solution to FGI is polynomial-time reducible to a solution to GI, so we may assume WLOG that we have an algorithm for FGI. Suppose the algorithm can be written as a unitary operator $U[\mathrm{FGI}]_{\mathrm{A,B,P,x}}$ with action

$$U[\mathrm{FGI}]_{\mathrm{A,B,P,x}} \left|G\right\rangle_{\mathrm{A}} \left|H\right\rangle_{\mathrm{B}} \left|0\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} = \begin{cases} \left|G\right\rangle_{\mathrm{A}} \left|H\right\rangle_{\mathrm{B}} \left|\sigma\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} & \text{if } G \sim H \text{ (where } G \overset{\sigma}{\sim} H\text{)} \\ \left|G\right\rangle_{\mathrm{A}} \left|H\right\rangle_{\mathrm{B}} \left|0\right\rangle_{\mathrm{P}} \left|1\right\rangle_{\mathrm{x}} & \text{if } G \not\sim H \ . \end{cases}$$

(We assume that G is rigid, so that the isomorphism σ is well-defined. This assumption is not necessary but makes the construction easier to describe.)

We can use this operator to construct the state $\left|\overline{G}\right\rangle$ for a given G. We begin with the state $\left|G\right\rangle$:

$$\begin{tikzcd}
\left|G\right\rangle_{\mathrm{A}} \left|0\right\rangle_{\mathrm{B}} \left|0\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} \arrow[d, "\begin{array}{l}\text{Create a superposition of all permutations in } S_n \text{, where } n \text{ is} \\ \text{the number of vertices in } G \text{ (an easy construction)}\end{array}"] \\
\left|G\right\rangle_{\mathrm{A}} \left|0\right\rangle_{\mathrm{B}} \left[\sum_{\sigma\in S_n} \left|\sigma\right\rangle_{\mathrm{P}}\right] \left|0\right\rangle_{\mathrm{x}} \arrow[d, "\begin{array}{l}\text{Act on } G \text{ with } \sigma \text{, putting the result } \sigma(G) \text{ in the second} \\ \text{register}\end{array}"] \\
\left|G\right\rangle_{\mathrm{A}} \left[\sum_{\sigma\in S_n} \left|\sigma(G)\right\rangle_{\mathrm{B}} \left|\sigma\right\rangle_{\mathrm{P}}\right] \left|0\right\rangle_{\mathrm{x}} \arrow[d, "U[\textsc{fgi}]^{-1}"', "\begin{array}{l}\text{Note that for a rigid graph } G\text{,} \\ \quad U[\mathrm{FGI}] \left|G\right\rangle_{\mathrm{A}} \left|\sigma(G)\right\rangle_{\mathrm{B}} \left|0\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} = \left|G\right\rangle_{\mathrm{A}} \left|\sigma(G)\right\rangle_{\mathrm{B}} \left|\sigma\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} \ .\end{array}"] \\
\left|G\right\rangle_{\mathrm{A}} \left[\sum_{\sigma\in S_n} \left|\sigma(G)\right\rangle_{\mathrm{B}}\right] \left|0\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} = \left|G\right\rangle_{\mathrm{A}} \left|\overline{G}\right\rangle_{\mathrm{B}} \left|0\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}} \ .
\end{tikzcd}$$

(Since this is a tensor-product state, we may discard the states $\left|G\right\rangle_{\mathrm{A}} \left|0\right\rangle_{\mathrm{P}} \left|0\right\rangle_{\mathrm{x}}$ without affecting state of $\left|\overline{G}\right\rangle_{\mathrm{B}}$.) Thus a unitary algorithm implementing GI is polynomial-time equivalent to creation of the succinct certificates $\left|\overline{G}\right\rangle$. As is not too surprising, algorithms that *approximately* solve FGI can be used to approximate $\left|\overline{G}\right\rangle$: Suppose we have a unitary operator $\widetilde{U}[\mathrm{FGI}]_{\mathrm{A,B,P,x}}$ satisfying, for all graphs G, H,

$$\begin{aligned}
\left|1 - \left\langle {}_{\mathrm{A}}\langle G|\, {}_{\mathrm{B}}\langle \sigma(G)|\, {}_{\mathrm{P}}\langle \sigma|\, {}_{\mathrm{x}}\langle 0|\, \widetilde{U}[\mathrm{FGI}]_{\mathrm{A,B,P,x}}\, |G\rangle_{\mathrm{A}}\, |\sigma(G)\rangle_{\mathrm{B}}\, |0\rangle_{\mathrm{P}}\, |0\rangle_{\mathrm{x}} \right\rangle_{\sigma}\right| &< \epsilon \\
{}_{\mathrm{A}}\langle G|\, {}_{\mathrm{B}}\langle H|\, {}_{\mathrm{P}}\langle 0|\, {}_{\mathrm{x}}\langle 1|\, \widetilde{U}[\mathrm{FGI}]_{\mathrm{A,B,P,x}}\, |G\rangle_{\mathrm{A}}\, |H\rangle_{\mathrm{B}}\, |0\rangle_{\mathrm{P}}\, |0\rangle_{\mathrm{x}} &= 1\ , \quad G \not\sim H\ .
\end{aligned}$$

Because the isomorphism σ is a certificate for FGI, we can easily check that a purported isomorphism is actually an isomorphism; this is why we can require that the second condition hold exactly, when $G \not\sim H$, even for an approximation $\widetilde{U}[\mathrm{FGI}]$. A randomized FGI algorithm, however, may only find the isomorphism σ, if one exists, with some finite probability; because GI is not known to lie in coNP, there is no known certificate allowing this result to be checked. The first requirement is that, averaged over all isomorphisms σ (and over all internal random draws used by the algorithm), the algorithm fails to find σ only ϵ of the time. (In these cases the algorithm can be assumed to transform $|G\rangle_{\mathrm{A}}\, |\sigma(G)\rangle_{\mathrm{B}}\, |0\rangle_{\mathrm{P}}\, |0\rangle_{\mathrm{x}}$ to $|G\rangle_{\mathrm{A}}\, |\sigma(G)\rangle_{\mathrm{B}}\, |0\rangle_{\mathrm{P}}\, |1\rangle_{\mathrm{x}}$; any other spurious results could be easily checked and rejected.) Then[18]

$$\begin{aligned}
&\left| {}_{\mathrm{A}}\langle G|\, {}_{\mathrm{B}}\left\langle \overline{G}\right|\, {}_{\mathrm{P}}\langle 0|\, {}_{\mathrm{x}}\langle 0| \cdot \widetilde{U}[\mathrm{FGI}]_{\mathrm{A,B,P,x}}{}^{-1} \cdot \tfrac{1}{\sqrt{n!}} \sum_{\sigma \in S_n} |G\rangle_{\mathrm{A}}\, |\sigma(G)\rangle_{\mathrm{B}}\, |\sigma\rangle_{\mathrm{P}}\, |0\rangle_{\mathrm{x}} \right| \\
&= \tfrac{1}{n!} \left| \sum_{\rho,\sigma \in S_n} \langle G, \rho(G), 0, 0| \cdot \widetilde{U}[\mathrm{FGI}]^{-1} \cdot |G, \sigma(G), \sigma, 0\rangle \right| \\
&= \tfrac{1}{n!} \left| \sum_{\rho,\sigma \in S_n} \langle G, \sigma(G), \sigma, 0|\ \widetilde{U}[\mathrm{FGI}]\ |G, \rho(G), 0, 0\rangle \right| \\
&= \left| \tfrac{1}{n!} \sum_{\sigma \in S_n} \langle G|\, \langle \sigma(G)|\, \langle \sigma|\, \langle 0|\ \widetilde{U}[\mathrm{FGI}]\ |G\rangle\, |\sigma(G)\rangle\, |0\rangle\, |0\rangle \right| \\
&= \left| \left\langle \langle G|\, \langle \sigma(G)|\, \langle \sigma|\, \langle 0|\ \widetilde{U}[\mathrm{FGI}]\ |G\rangle\, |\sigma(G)\rangle\, |0\rangle\, |0\rangle \right\rangle_{\sigma} \right| \\
&\geq\ 1 - \epsilon\ ,
\end{aligned}$$

[18] We explicitly include the normalization factors here. We concatenate the registers, $|\alpha\rangle\, |\beta\rangle \equiv |\alpha, \beta\rangle$ so that there aren't so many $|\cdot\rangle$s floating around.

so we can use the operator $\widetilde{U}[\mathrm{FGI}]^{-1}$ (on the state $\sum_{\sigma} |G\rangle |\sigma(G)\rangle |\sigma\rangle |0\rangle$, which, as previously shown, is easily constructed) to create a good approximation $|\widetilde{\overline{G}}\rangle$ to $|\overline{G}\rangle$.

This is an interesting difference from the classical case: Although GI is of course reducible to the problem of finding a succinct certificate for a topological graph $\overline{G}$ given the graph G, classically it is *not* known whether a reduction in the other direction is possible. With a quantum computer, however, these two problems become equivalent in complexity. [I have not seen this result explicitly noted before, although others [3] have independently constructed succinct certificates $|\overline{G}\rangle$ in the same manner as described below.] Unfortunately, as noted above, these certificates do not seem as powerful as classical certificates; it seems to be difficult to do operations on the basis states $\{|\overline{G}\rangle\}$.

This equivalence is also potentially interesting from the other point of view described in the Introduction: we have shown that (a class of) quantum algorithms for GI are equivalent in power to construction of a class of quantum states. If we can find lower bounds on the difficulty of constructing such states, we have immediate lower bounds on the quantum complexity, and hence on the *classical* complexity, of GI. (I know of no such bounds, however.)

2.3.3 An observable solving GI

The above state-construction method defines states which are apparently difficult to construct but uses operators which are easy to implement. It is possible to transfer the difficult operations from the state construction to the operator implementation. This allows a conceptual simplification: in the state-construction method, construction of a *class* of states $\{|\overline{G}\rangle\}$, one state for each distinct topological graph, is required; Ettinger and Høyer [43] define a single observable, essentially by the direct sum of the operators over this entire class. (It is not clear whether there is any real gain in this, though, since there seems to be no reason to believe that their observable is any easier to implement than the state construction.)

Before we describe the actual construction, a little terminology is in order: Consider the graph G, the **disjoint union** of two connected graphs G_1, G_2 each on n vertices (where the two graphs are placed “side by side”, sharing no vertices). The automorphism group

$\Phi \equiv \Phi(G)$, a subgroup of the wreath product group[19] $\Gamma \equiv S_n \wr S_2$, contains an **involutive swap**, an element of the form $s(\phi) \equiv (\phi^{-1}, \phi, (1\ 2))$ (i.e., a bijection between the vertices of G_1 and those of G_2), if and only if $G_1 \sim G_2$ (in which case ϕ is the isomorphism, $G_1 \overset{\phi}{\sim} G_2$). For convenience let us define $\Gamma_0 \equiv S_n \wr \{e\}$ ($\simeq S_n \times S_n$) to be the index-2 normal subgroup of Γ (the subgroup $\Gamma_0 \lhd \Gamma$ which does not swap the vertices of the G_i).

Now if $G_1 \overset{\phi}{\sim} G_2$, then Φ is of the form $\Phi = \Phi_0 \cup \Phi_0 s(\phi)$, where $\Phi_0 \equiv \Phi \cap \Gamma_0$ is the subgroup of Φ which does not swap the G_i; so the cosets of Φ all have the form $\gamma\Phi = \Psi_0 \cup \Psi_0 s(\phi)$ (where, similarly, none of the elements of Ψ_0 swap the G_i). On the other hand, if $G_1 \not\sim G_2$, then $\Phi = \Phi_0 \subset \Gamma_0$, and the elements of a coset $\gamma\Phi$ either all swap the G_i or all leave the G_i fixed. The essential point, and the main result used in the construction, is that if $G_1 \sim G_2$, elements of $\gamma\Phi$ appear in pairs, γ and $\gamma s(\phi)$; but if $G_1 \not\sim G_2$, only one of these can appear in $\gamma\Phi$.

Let us represent a coset $\gamma\Phi$ using the set representation $|\{\cdots\}_2\rangle$ of (2.7), in the $2(n!)^2$-dimensional Hilbert space $\mathcal{H} \equiv \text{span}\{|\gamma\rangle : \gamma \in \Gamma\}$:

$$|\gamma\Phi\rangle = \tfrac{1}{\sqrt{|\Phi|}} \sum_{\xi \in \Phi} |\gamma\xi\rangle\,; \tag{2.10}$$

note that distinct coset states (of the same group Φ) are orthogonal.[20] Now if $G_1 \overset{\phi}{\sim} G_2$,

[19] The **wreath product** group $S_n \wr S_2 \cong (S_n \times S_n) \rtimes S_2$ is defined by the multiplication

$$(\alpha_1, \alpha_2, a)(\beta_1, \beta_2, b) \equiv (\alpha_1\beta_{a(1)}, \alpha_2\beta_{a(2)}, ab)$$

and its action on G is induced by $(\alpha_1, \alpha_2, a)v_i^{(j)} \equiv v_{i'}^{(a(j))}$, $i' \equiv \alpha_{a(j)}(i)$. The two S_n subgroups should be thought of as acting on the vertices of G_1 and G_2, and the S_2 subgroup as swapping the vertices of G_1 with those of G_2 using a particular fixed bijection. (Notice that the action on G is read right-to-left, as conventional for composition of permutations.)

[20] Although finding Φ is not easy—FGI, and hence GI, reduce to finding Φ—random coset states $|\gamma\Phi\rangle$ *can* be efficiently constructed, as follows: Begin with a superposition of all elements $\gamma \in S_n \wr S_2$, together with a representation of the graph G. Act on G with the γ, creating the entangled state

$$\sum_{\gamma\in\Gamma} |\gamma\rangle|G\rangle \mapsto \sum_{\gamma\in\Gamma} |\gamma\rangle|\gamma(G)\rangle\,.$$

Now measure the second register; a measurement result of $|\gamma_0(G)\rangle$ creates a superposition (in the first then $|\gamma\Phi\rangle$ is a superposition of states of the form $|k(\phi,\gamma)\rangle \equiv \frac{1}{\sqrt{2}}\Big[\,|\gamma\rangle + |\gamma s(\phi)\rangle\Big]$. Thus $|\gamma\Phi\rangle$ lies in the $(n!)^2$-dimensional Hilbert subspace $\mathcal{H}_1(\phi) \equiv \mathrm{span}\{|k(\phi,\gamma)\rangle : \gamma \in \Gamma\}$, and

$$\langle\gamma\Phi|\, P_{\mathcal{H}_1(\phi)}\, |\gamma\Phi\rangle = 1 \quad \text{if } G_1 \overset{\phi}{\sim} G_2 \ . \tag{2.11}$$

The states $|k(\phi,\gamma)\rangle$ are, for fixed ϕ, orthonormal (since $s(\phi)^2 = 1$). Hence we may write

$$P_{\mathcal{H}_1(\phi)} = \sum_{\gamma\in\Gamma_0} P_{k(\phi,\gamma)} = \sum_{\gamma\in\Gamma_0} |k(\phi,\gamma)\rangle\langle k(\phi,\gamma)| \ . \tag{2.12}$$

Let ϕ be an arbitrary element of S_n. If $G_1 \not\sim G_2$, then for each term $|\gamma\xi\rangle$, $\xi \in \Phi$, in the superposition $|\gamma\Phi\rangle$, exactly one projector $P_{k(\phi,\zeta)}$ of this sum has nonzero projection $P_{k(\phi,\zeta)}\, |\gamma\xi\rangle$: namely, $\zeta = \gamma\xi$ if $\gamma \in \Gamma_0$, and $\zeta = \gamma\xi s(\phi^{-1})$ if $\gamma \notin \Gamma_0$. Since $\langle\gamma\xi|\, P_{k(\phi,\zeta)}\, |\gamma\xi\rangle = \frac{1}{2}$, this implies

$$\langle\gamma\Phi|\, P_{\mathcal{H}_1(\phi)}\, |\gamma\Phi\rangle = \tfrac{1}{2} \quad \text{if } G_1 \not\sim G_2 \ . \tag{2.13}$$

Hence the binary measurement observable $\{P_{\mathcal{H}_1(\phi)}, \mathbf{1} - P_{\mathcal{H}_1(\phi)}\}$ (a two-valued von Neumann measurement) acting on a random coset state $|\gamma\Phi\rangle$ can distinguish (with bounded error probability) whether the graphs G_i are isomorphic.

Unfortunately, we don't know ϕ (this is the isomorphism we're trying to find), so we can't expect to implement $P_{\mathcal{H}_1(\phi)}$. We can solve this problem by considering a projector onto a larger Hilbert space $\mathcal{H}_1 = \mathrm{span}\{\mathcal{H}_1(\varphi) : \varphi \in S_n\}$. Of course we still have $\langle\gamma\Phi|\, P_{\mathcal{H}_1}\, |\gamma\Phi\rangle = 1$ if $G_1 \sim G_2$. It is difficult to find an exact value when $G_1 \not\sim G_2$, but we can set a trivial bound: Clearly $P_{\mathcal{H}_1} \leq \sum_{\varphi\in S_n} P_{\mathcal{H}_1(\varphi)}$. For each term an argument like the one above applies, $\langle\gamma\Phi|\, P_{\mathcal{H}_1(\varphi)}\, |\gamma\Phi\rangle = \frac{1}{2}$, and there are $n!$ terms in the sum; so $\langle\gamma\Phi|\, P_{\mathcal{H}_1}\, |\gamma\Phi\rangle \leq \frac{n!}{2}$.

This bound is unfortunately *too* trivial; we have solved one problem but created another. But this problem can also be solved, to make the bound interesting. The trick now is to consider tensor products of the coset states, $|\vec{\gamma}\Phi\rangle \equiv |\gamma_1\Phi\rangle \cdots |\gamma_m\Phi\rangle$, and of the subspaces $\mathcal{H}_m(\phi) \equiv \mathcal{H}_1(\phi)^{\otimes m}$ (note that, as before, $|\vec{\gamma}\Phi\rangle \in \mathcal{H}_m(\phi)$ if $G_1 \overset{\phi}{\sim} G_2$). Again we

register) of all $\gamma \in S_n \wr S_2$ with the same action on G. Thus we create the state $|\{\gamma : \gamma(G) = \gamma_0(G)\}_2\rangle = |\gamma_0\Phi\rangle$ with probability $\frac{|\gamma_0\Phi|}{|\Gamma|} = \frac{|\Phi|}{2n!^2}$, a uniform distribution over all cosets.

consider the projector onto the span of all these subspaces, $\mathcal{H}_m \equiv \mathrm{span}\{\mathcal{H}_m(\varphi) : \varphi \in S_n\}$. By the arguments above, $\langle \vec{\gamma}\Phi | \, P_{\mathcal{H}_m(\phi)} \, | \vec{\gamma}\Phi \rangle = 1$ if $G_1 \overset{\phi}{\sim} G_2$, but $\langle \vec{\gamma}\Phi | \, P_{\mathcal{H}_m(\phi)} \, | \vec{\gamma}\Phi \rangle = \frac{1}{2^m}$ if $G_1 \not\sim G_2$. The difference between the expectation values of $P_{\mathcal{H}_m}$ has thus been amplified:

$$\begin{aligned} \langle \vec{\gamma}\Phi | \, P_{\mathcal{H}_m} \, | \vec{\gamma}\Phi \rangle &= 1 \qquad \text{if } G_1 \sim G_2 \\ \langle \vec{\gamma}\Phi | \, P_{\mathcal{H}_m} \, | \vec{\gamma}\Phi \rangle &\leq \tfrac{n!}{2^m} \quad \text{if } G_1 \not\sim G_2 \; . \end{aligned} \tag{2.14}$$

For sufficiently large $m \gtrapprox \log_2 n!$, this separation is measurable. (The point is that the dimension of $\mathcal{H}_m$, the Hilbert space spanned by the coset states of all groups Φ containing an involutive swap, grows more slowly with m than the dimension of $\mathcal{H}^{\otimes m}$, the Hilbert space spanned by all coset states of all subgroups of Γ:

$$\frac{\dim \mathcal{H}_m}{\dim \mathcal{H}^{\otimes m}} \leq \frac{(n!)^{2m+1}}{2^m (n!)^{2m}} = \frac{n!}{2^m} \; . \tag{2.15}$$

So we can make m large enough that random coset states $|\vec{\gamma}\Phi\rangle$ have small overlap with the subspace $\mathcal{H}_m$ unless Φ contains an involutive swap, i.e., $G_1 \sim G_2$.) However, there is no obvious reason to believe that the desired measurement observable $\{P_{\mathcal{H}_m}, \mathbf{1} - P_{\mathcal{H}_m}\}$ can be efficiently implemented.

2.4 Trying to construct $\left|\overline{G}\right\rangle$

As mentioned above, I know of no efficient method of constructing the states $\left|\overline{G}\right\rangle$. Here I outline a few ideas and try to understand why they don't work.

2.4.1 Structured Grover search

First, we can use the Grover search to construct the state. Recall that implementation of Grover's algorithm requires that we have *efficient detectors* for the initial and final state. Here, detection of the initial state $|G\rangle$ is easy, since $|G\rangle$ is a known computational basis state. We can also detect the final state $\left|\overline{G}\right\rangle$, by detecting the $+1$ eigenvalues of a complete set of generators of the group Γ being summed over[21], since the state $\left|\overline{G}\right\rangle = \sum_{\gamma \in \Gamma} |\gamma(G)\rangle$

[21] This actually detects not the state $\left|\overline{G}\right\rangle$ but the space spanned by all such states; so we are using Grover's algorithm with $P_\alpha = |G\rangle\langle G|$ and $P_\omega = \sum_{\overline{G}} \left|\overline{G}\right\rangle\left\langle\overline{G}\right|$. P_α is one-dimensional, so (as discussed in

is a simultaneous $+1$ eigenstate of the operators U_σ ($\sigma \in \Gamma$) which implement the group action $U_\sigma \,|x\rangle = |\sigma(x)\rangle$. For convenience in the following discussion, let us assume that G is rigid, and that we are using the adjacency-matrix representation (2.8) for the graph; then $\Gamma = S_n$. Thus if we begin with the system in a state $|G\rangle$, application of Grover's algorithm will rotate the system to the state $\big|\overline{G}\big\rangle$ in $\mathcal{O}\big(|\langle G \,|\, \overline{G}\rangle|^{-1}\big) \sim \sqrt{n!}$ operations.

Of course, this is not a polynomial-time algorithm. We can try to remedy this by using a **structured Grover search** [28, 45, 56], in which we use extra information about the structure of the search space to perform a multi-stage search. For example, suppose that in a search problem we are able to efficiently and reliably detect (in the sense explained in §2.2(v)) not only the initial state $|\alpha\rangle \equiv |\phi_0\rangle$ and the final state $|\omega\rangle \equiv |\phi_n\rangle$, but also **intermediate** states $|\phi_1\rangle, \ldots, |\phi_{n-1}\rangle$, where the inner products of consecutive terms $\langle \phi_k \,|\, \phi_{k+1}\rangle$ are large relative to $\langle \alpha \,|\, \omega\rangle$. Then we can construct the desired state $|\omega\rangle$ in n stages of Grover rotation; stage i ($i \in \{1, \ldots, n\}$), from $|\phi_{i-1}\rangle$ to $|\phi_i\rangle$, requires $\mathcal{O}\left(|\langle \phi_{i-1} \,|\, \phi_i\rangle|^{-1}\right)$ applications of the operator $\Big[\mathbf{1} - 2\,|\phi_{i-1}\rangle\langle\phi_{i-1}|\,\Big]\Big[\mathbf{1} - 2\,|\phi_i\rangle\langle\phi_i|\,\Big]$, for a total time of $\mathcal{O}\left(\sum_{i=0}^{n-1} |\langle \phi_i \,|\, \phi_{i+1}\rangle|^{-1}\right)$.

In the optimal (but unrealistic) case, for example, the presence of a single efficiently-detectable intermediate state $|\phi\rangle \propto |\alpha\rangle + |\omega\rangle$ (bisecting the angle between $|\alpha\rangle$ and $|\omega\rangle$, so that $|\langle \alpha \,|\, \phi\rangle| = |\langle \phi \,|\, \omega\rangle| \approx \frac{1}{\sqrt{2}}$) would reduce the time required from $\mathcal{O}(|\langle \alpha \,|\, \omega\rangle|^{-1})$ to $\mathcal{O}(|\langle \alpha \,|\, \phi\rangle|^{-1} + |\langle \phi \,|\, \omega\rangle|^{-1}) = \mathcal{O}(1)$. In more realistic cases (examples of which are given in the references listed above) the intermediate states are not in the plane spanned by $|\alpha\rangle$ and $|\omega\rangle$ but still have larger transition rates than for the unstructured search problem. The problem structure, for instance, might allow rejection of some classes of solutions as impossible (in the examples above, these classes represent partial solutions in a partially-prunable search tree), giving rise to a descending chain $X \equiv X_0 \supset X_1 \supset \cdots \supset X_n \equiv Z$ of candidate solution sets, beginning with the given domain X and ending with the set Z of actual solutions. In this case we may define $|\phi_k\rangle \equiv \sum_{x \in X_k} |x\rangle$; the structured Grover search may take much less time than the unstructured search if the sets X_i **efficiently resolve** X (i.e., if all the ratios $\frac{|X_i|}{|X_{i+1}|}$ are about equal).

§2.2(v)) this performs precisely as if P_ω were the one-dimensional projector $\big|\overline{G}\big\rangle\big\langle\overline{G}\big|$ (the one projector in the subspace P_ω which does not commute with P_α).

Suppose in our case, for instance, that we have a set of generators for the group of interest, $\Gamma = \langle \gamma_1, \ldots, \gamma_n \rangle$. Let us set $\Gamma_i = \langle \gamma_1, \ldots, \gamma_i \rangle$ (with $\Gamma_0 \equiv \{1\}$ and $\Gamma_n \equiv \Gamma$). If we can efficiently detect the intermediate states $|\overline{G}_i\rangle = \sum_{\gamma \in \Gamma_i} |\gamma(G)\rangle$, then we can use the structured Grover search to construct the desired state $|\overline{G}\rangle$; the search will take time $\mathcal{O}\left(\sum_{i=0}^{n-1} |\Gamma_i : \Gamma_{i+1}|\right)$. It is possible to choose the generators γ_i such that this time is polynomial in n (for instance, using the generators of (2.23) or of (2.25) below). The remaining question is then whether each intermediate state $|\overline{G}_i\rangle$, a sum over the subgroup Γ_i, can be efficiently detected, so that each may be efficiently constructed in turn from the previous one.

One tempting idea is to use, instead of projectors onto the states $|\overline{G}_i\rangle$, projectors P_H onto the states transforming as the trivial representation of some subgroup H of Γ: that is, P_H projects onto the simultaneous $+1$ eigenstate of all $\eta \in H$. P_H is the projection onto the span, over *all graphs* G, of all states of the form $\sum_{\eta \in H} |\eta(G)\rangle$ (so, for example, $P_{\{e\}} \equiv \mathbf{1}$, and $P_\Gamma \equiv P_{S_n} \equiv \sum_{\overline{G}} |\overline{G}\rangle\langle\overline{G}|$). Such a projector is easy to implement, since it is merely the projection onto the simultaneous $+1$ eigenstate of all $\eta \in H$ (where η acts by permutation of the vertices of G). Unfortunately, use of these easily-implemented projectors alone does not suffice to construct the desired state: Suppose that only such projectors are used, so that at each stage the Grover iteration is of the form $(\mathbf{1} - 2P_{H_{i+1}})(\mathbf{1} - 2P_{H_i})$. We show below that this iteration preserves the subspaces corresponding to the irreducible representations of the group $\Gamma = S_n$.

First let us define, for all $\gamma \in \Gamma$, the unitary operator $V(\gamma)$ acting by the group action on the graph states: $V(\gamma)\,|G\rangle \equiv |\gamma G\rangle$. The projector P_H onto a subgroup $H \subset \Gamma$ can be written as a linear combination of $V(\gamma)$:

$$P_H = \tfrac{1}{|H|} \sum_{\eta \in H} V(\eta)\ . \tag{2.16}$$

(Clearly P_H acting on a graph state prepares a state which is invariant under action by $V(\eta)$, for all $\eta \in H$; it is also easy to verify that $(P_H)^2 = P_H$, so P_H is a projector.)

Now recall the orthonormality and completeness relations from classical representation theory [96],

$$
\begin{aligned}
\frac{n_\mu}{n_\Gamma}\sum_{\gamma\in\Gamma} D^\dagger_\mu(\gamma)^k{}_i D^\nu(\gamma)^j{}_l &= \delta^\nu_\mu \delta^j_i \delta^k_l && (2.17)\\
\sum_{\mu ij}\frac{n_\mu}{n_\Gamma} D^\mu(\gamma)^j{}_i D^\dagger_\mu(\gamma')^i{}_j &= \delta_{\gamma,\gamma'} && (2.18)
\end{aligned}
$$

where $D^\mu(\gamma)$ is the representation matrix for group element γ in the n_μ-dimensional unitary irreducible representation μ of Γ and $n_\Gamma \equiv |\Gamma|$, and $D^\dagger_\mu(\gamma)^j{}_i \equiv [D^\mu(\gamma)^i{}_j]^*$. We can use these relations to define an alternate basis for the space spanned by $\{|\gamma G\rangle : \gamma \in \Gamma\}$:

$$
\left|\phi_G(\mu)^i{}_j\right\rangle \equiv \sqrt{\frac{n_\mu}{n_\Gamma}} \sum_{\gamma\in\Gamma} D^\mu(\gamma)^i{}_j \left|\gamma G\right\rangle . \qquad (2.19)
$$

The orthonormality and completeness of these basis states follows immediately from the orthonormality (2.17) and completeness (2.18) relations above.

The action of $V(\gamma)$ on $\left|\phi_G(\mu)^i{}_j\right\rangle$ preserves the representation μ, only mixing the states *within* each irreducible representation:

$$
\begin{aligned}
V(\gamma)\left|\phi_G(\mu)^i{}_j\right\rangle &= \sqrt{\frac{n_\mu}{n_\Gamma}}\sum_{\eta\in\Gamma} D^\mu(\eta)^i{}_j \left|\gamma\eta G\right\rangle = \sqrt{\frac{n_\mu}{n_\Gamma}}\sum_{\eta\in\Gamma} \left[D^\mu(\gamma^{-1})D^\mu(\gamma\eta)\right]^i{}_j \left|\gamma\eta G\right\rangle \\
&= \sqrt{\frac{n_\mu}{n_\Gamma}}\sum_{\eta\in\Gamma}\sum_k D^\mu(\gamma^{-1})^i{}_k D^\mu(\gamma\eta)^k{}_j \left|\gamma\eta G\right\rangle \\
&= \sum_k D^\mu(\gamma^{-1})^i{}_k \left|\phi_G(\mu)^k{}_j\right\rangle . \qquad (2.20)
\end{aligned}
$$

So the action of the unitary operator $\mathbf{1} - 2P_H$ (one of the two reflections composing a Grover iteration), in the representation basis, is

$$
\begin{aligned}
\left(\mathbf{1} - 2P_H\right)\left|\phi_G(\mu)^i{}_j\right\rangle &= \left|\phi_G(\mu)^i{}_j\right\rangle - \tfrac{2}{|H|}\sum_{\eta\in H} V(\eta)\left|\phi_G(\mu)^i{}_j\right\rangle \\
&= \left|\phi_G(\mu)^i{}_j\right\rangle - \tfrac{2}{|H|}\sum_k \left[\sum_{\eta\in H} D^\mu(\eta^{-1})^i{}_k\right]\left|\phi_G(\mu)^k{}_j\right\rangle , \qquad (2.21)
\end{aligned}
$$

acting only within the representation μ. Any product of such unitary operators will thus preserve the amplitude of the projection onto a representation subspace P_μ (the space spanned by all $\left|\phi_G(\mu)^i{}_j\right\rangle$). The desired state $\left|\overline{G}\right\rangle$ is just the state $|\phi_G(1)\rangle$ of the trivial representation 1 (defined by $D^1(\gamma) \equiv 1$), so the amplitude in this state is preserved throughout the structured Grover search. (This amplitude, if we begin in the state $|G\rangle$, is just $|\langle G \,|\, \overline{G}\rangle| = \frac{1}{\sqrt{|\Gamma|}} = \frac{1}{\sqrt{n!}}$.)

Thus such projectors (or any projectors which can be written as linear combinations of the $V(\gamma)$) do not suffice for implementing the Grover search. What we would like is a more powerful detector which projects onto the space $\left|\overline{G}_i\right\rangle\left\langle\overline{G}_i\right|$ (for the given graph G) instead of onto the span of all such states, so that the Grover iteration is $\left[\mathbf{1} - 2\left|\overline{G}_{i+1}\right\rangle\left\langle\overline{G}_{i+1}\right|\right]\left[\mathbf{1} - 2\left|\overline{G}_i\right\rangle\left\langle\overline{G}_i\right|\right]$. If we could efficiently implement all of these projectors, then each Grover iteration in the structured search could be implemented efficiently. (In fact, $\left[\mathbf{1} - 2P_{H_{i+1}}\right]\left[\mathbf{1} - 2\left|\overline{G}_i\right\rangle\left\langle\overline{G}_i\right|\right]$ would have exactly the same action, as discussed in §2.2(v).)

Unfortunately, these projectors seem to be about as difficult to implement (for $i \sim n$) as a classical solution to GI: We can decompose $|\overline{G}_i\rangle\langle\overline{G}_i| = P_{H_i} P_{\Gamma_i G}$, where $P_{\Gamma_i G}$ projects onto the space of graphs isomorphic to G under the group Γ_i (i.e., onto the space spanned by $|H\rangle$ with $G \overset{\phi}{\sim} H$ and $\phi \in \Gamma_i$). But implementing this projector then requires being able to solve GI for the groups Γ_i; for large i, $|\Gamma_i| \sim n!$, and this is still believed to be a hard problem.

2.4.2 Stabilizer formalism

An **abelian stabilizer code** for a quantum system is a subspace $\bar{\mathcal{H}}$ of a Hilbert space $\mathcal{H}$ characterized by its **stabilizer** subgroup S (the subgroup of the Pauli group[22] $\mathcal{P}^{\otimes n}$ leaving states in $\bar{\mathcal{H}}$ invariant). The **normalizer** subgroup $N(S)$, another subgroup of $\mathcal{P}^{\otimes n}$, is the group of operators in $\mathcal{P}^{\otimes n}$ commuting with all $s \in S$, and thus mapping $\bar{\mathcal{H}}$ onto itself; elements in the normalizer subgroup represent "encoded operations" on the logical subspace $\bar{\mathcal{H}}$ of a stabilizer code. S is an abelian subgroup of the Pauli group $\mathcal{P}^{\otimes n}$, which in turn is a discrete nonabelian subgroup of $\mathcal{U}(n)$. The motivation for requiring

[22] The n-qubit **Pauli group** $\mathcal{P}_2^{\otimes n}$ is the group generated by the single-qubit Pauli matrices and -1; that is, $\mathcal{P}_2^{(i)} \equiv \mathcal{P}^{(i)} \equiv \{\pm 1^{(i)}, \pm\sigma_x^{(i)}, \pm\sigma_y^{(i)}, \pm\sigma_z^{(i)}\} = \langle\sigma_x^{(i)}, \sigma_z^{(i)}\rangle$. Elements of the qubit Pauli group are thus both Hermitian and unitary.

S to be an abelian group is that this allows simultaneous diagonalization of all stabilizer operators; hence there is a simultaneous eigenbasis (for $\mathcal{H}$) of all stabilizer operators. $\bar{\mathcal{H}}$ is just the subspace of $\mathcal{H}$ with all eigenvalues $+1$.

The theory of abelian stabilizer codes has proved useful in the development of quantum error-correcting codes (see [52, 90, 64, 65, 94] for several examples) and is well developed (Preskill [88, Ch. 7] and Nielsen and Chuang [83, Ch. 10] have two introductory treatments). Various useful operations, including **stabilizer measurements** (projecting onto the code subspace and its coset spaces), **error-basis operations** (operations in $\mathcal{P}^{\otimes n}$ permuting among these subspaces), and **encoded operations** (operations in $N(S)$ acting in interesting ways within the code subspace), can be easily implemented.

The techniques used in quantum error correction can also be used in state preparation. Construction of a standard state $|\bar{0}\rangle$ (the $\bar{0}$ representing an encoded, "logical" qubit, possibly constructed as a superposition of several physical qubits) in an abelian stabilizer code S can be thought of as an error-correction operation applied to an abelian stabilizer $S' \rhd S$ having a one-dimensional code subspace (the stabilizer S' being generated by the elements of the original stabilizer S together with all the logical $\sigma_{\bar{z}}^{(i)}$ operators of $N(S)$): For any basis $\{M_i\}$ of S' there is an error operator $E_i \in \mathcal{P}^{\otimes n}$ such that $\{E_i, M_i\} = 0$ and $[E_i, M_j] = 0$ for $i \neq j$. (There are actually many such operators; for each stabilizer generator M_i there is a coset $E_i N(S)$ of the normalizer which anticommutes with M_i and commutes with the other generators. Typically the error operator used is the one which represents the most likely error, according to some assumed error model.) The error-correction technique for such an abelian stabilizer is to apply a measurement corresponding to each of the (Hermitian) M_i in turn, and then apply the (unitary) operator E_i iff the measurement result was -1. It is easy to see that because the stabilizer is abelian, the result is a state in the simultaneous $+1$ eigenspace of all stabilizer operators, i.e., a state in the stabilizer subspace.

For an example of why we might want to extend this theory to nonabelian groups, consider the state $|\overline{G}\rangle$ defined above. All states $|\overline{G}\rangle$ (for graphs G on v vertices) are simultaneous $+1$ eigenstates of all permutation operators Λ_σ defined (for the adjacency-matrix representation of $\left|\overline{G}\right\rangle$) by its action on the computational basis:

$$\Lambda_\sigma \left|A_G\right\rangle = \left|P_\sigma A_G P_\sigma^{-1}\right\rangle . \tag{2.22}$$

The Λ_σ clearly do not commute (the group $\{\Lambda_\sigma\}$ is isomorphic to S_v), but on the subspace symmetrized as above they all have trivial action. Of course, even in the abelian case there are operators in $\mathcal{U}(n)$ which do not commute with S but which do have trivial action on $\bar{\mathcal{H}}$; these are just the unitary operators on $\bar{\mathcal{H}}$ which are not in the (discrete) normalizer subgroup. But the subspace spanned by the $\left|\overline{G}\right\rangle$ does not seem to have a convenient representation as an abelian stabilizer code. And although the stabilizer operators are easy to implement (they are just permutations) the normalizer operators seem to be difficult to implement, and the error basis operators are hard to even define.

Suppose we want to modify this technique to allow for projection into the simultaneous $+1$ eigenspace of a nonabelian stabilizer group. (For concreteness, and because it is the group of interest, let us assume that the nonabelian group is just S_n, the symmetric group.) The important facts used above were the existence of two sets of operators: the stabilizer generators M_i (which all commute) and the error operators E_i (which commute with all of the M_j except that $E_i M_i = -M_i E_i$). Of course, in the nonabelian case our stabilizer generators M_i do not all commute, and so operation with all the M_i in turn (even if we can find E_i satisfying the condition above) will not generally project us into the stabilizer subspace; but perhaps it will get us closer to it, and repeated action of all of these operators will get us close enough to use the above algorithms.

It is not clear that all sets of generators for S_n will give the same results. We mention here three different generating sets, each having some nice properties (stated without proof). The first generating set contains exactly one k-cycle for each $2 \leq k \leq n$:

$$\{c_2 \equiv (1\ 2), c_3 \equiv (1\ 2\ 3), \ldots, c_n \equiv (1\ \cdots\ n)\}\ . \tag{2.23}$$

These generators are potentially interesting because of the following identity on the algebra over S_n:

$$\sum_{\sigma \in S_n} \sigma = \prod_{k=2}^{n} \sum_{i=0}^{k-1} (c_k)^i \; . \tag{2.24}$$

Thus each element of S_n can be uniquely expressed as a product of the c_k (in increasing order of k) to some power.

The second generating set contains only 2-cycles:

$$\{(1\ 2), (2\ 3), \ldots, (n{-}1\ \ n)\} \; . \tag{2.25}$$

Since 2-cycles square to the identity, operators implementing 2-cycles have eigenvalues ±1 and so are both Hermitian and unitary.

Finally, the third generating set contains only two generators:

$$\{(1\ 2), (1\ 2\ \cdots\ n)\} \; . \tag{2.26}$$

This is a subset of the first set mentioned; it may be useful for explicit calculations.

2.5 Related problems

2.5.1 Group non-membership

The algorithms discussed above for GI have the obvious shortcoming that they have no obvious, efficient implementation; they use an "oracle," performing an operation that seems to be difficult to implement on a quantum computer, for one of the computational steps. In particular, we have no algorithm for creating the state $\left|\overline{G}\right\rangle$ (2.8) in polynomial time given a graph G, or for measuring the observable (2.14) given a pair of graphs.

One way of weakening the oracle condition is to consider an *untrustworthy* oracle. Specifically, we consider the standard complexity class MA introduced by Babai (defined, e.g., in [67]) of problems which Arthur can solve if he is allowed to consult a powerful oracle, Merlin, who can solve arbitrarily difficult problems but who may try to trick poor Arthur by giving him a false proof. Arthur must therefore verify that Merlin's utterances (his **certificate** for the problem) are correct, as well as using the certificate to solve the problem. The algorithms for GI above do not obviously fall into this class: Arthur does not know how to completely verify the correctness of Merlin's certificate $|\overline{G}\rangle|\overline{H}\rangle$ for the graph isomorphism problem. (He can check that the states are correctly symmetrized, and he can also measure some graph invariants; but since no complete, efficient, deterministic set of graph invariants is known, he cannot be certain that Merlin is not tricking him.) Of course, if the graphs are isomorphic, Arthur can compute an isomorphism between them and use this to check that Merlin was truthful; the problem is that Arthur cannot easily check that Merlin was truthful when he finds that the graphs are *not* isomorphic. Classically the same problem exists: Succinct, verifiable classical certificates are known for GI but not for GNI (the problem of proving that two graphs are not isomorphic).[23]

Watrous [101] has independently used many of the ideas above to find a *verifiable* certificate for GNM (Group Non-Membership), the problem of proving that a particular element g of some finite group G is not in the subgroup $H = \langle h_1, \ldots, h_k \rangle$ with given generators. As with the quantum certificate for GI given above, this certificate may be difficult to construct; the major difference from the problem above is that for GNM, Watrous has found a way to verify that the given certificate is valid. Thus GNM is in QMA, the class of problems solvable in polynomial time on a quantum computer given a powerful but untrustworthy oracle. (The problem GM (Group Membership) is, like GI, in MA; if $g \in H$, Arthur need only ask for a list of generators of H whose product is g. Merlin can't trick Arthur, because Arthur can easily multiply the generators together to verify that their product is g.)

Watrous' quantum certificate for GNM (to be provided by Merlin) is the state

$$|H\rangle = \tfrac{1}{\sqrt{|H|}} \sum_{h \in H} |h\rangle \,, \tag{2.27}$$

the representation $|\{\cdots\}_2\rangle$ of the set of elements of H. Given such a certificate, one can

[23] This is just a restatement of the fact that GI is in NP but is not known to be in coNP.

easily check whether $g \in H$:

$$\begin{tikzcd}
\lvert 0\rangle_{\mathrm{c}} \lvert H\rangle_{\mathrm{A}} \lvert g\rangle_{\mathrm{B}} \arrow[d, "H_{\mathrm{c}}"', "\text{Perform a Hadamard gate on qubit c}"] \\
\frac{1}{\sqrt{2}}\left(\lvert 0\rangle_{\mathrm{c}} + \lvert 1\rangle_{\mathrm{c}}\right) \lvert H\rangle_{\mathrm{A}} \lvert g\rangle_{\mathrm{B}} \arrow[d, "\substack{\text{Perform a left multiplication of register A by } h \text{ (here stored} \\ \text{in register B), controlled by qubit c}}"] \\
\frac{1}{\sqrt{2}}\left(\lvert 0\rangle_{\mathrm{c}} \lvert H\rangle_{\mathrm{A}} + \lvert 1\rangle_{\mathrm{c}} \lvert gH\rangle_{\mathrm{A}}\right) \lvert g\rangle_{\mathrm{B}} \arrow[d, "H_{\mathrm{c}}"', "\text{Perform a second Hadamard gate on qubit c}"] \\
\frac{1}{2}\left[\lvert 0\rangle_{\mathrm{c}}\left(\lvert H\rangle_{\mathrm{A}} + \lvert gH\rangle_{\mathrm{A}}\right) + \lvert 1\rangle_{\mathrm{c}}\left(\lvert H\rangle_{\mathrm{A}} - \lvert gH\rangle_{\mathrm{A}}\right)\right] \lvert g\rangle_{\mathrm{B}} \arrow[d, "\text{Measure qubit c}"] \\
\begin{cases} \lvert 0\rangle_{\mathrm{c}}, \text{ with probability 1, if } g \in H \text{ so that } gH = H \\ \lvert 0\rangle_{\mathrm{c}} \text{ or } \lvert 1\rangle_{\mathrm{c}}, \text{ each with probability } \frac{1}{2}, \text{ if } g \notin H \text{ so that } H \cap gH = \varnothing\ . \end{cases}
\end{tikzcd}$$

(Note that this works because distinct cosets gH of a group H are disjoint.) This part of the algorithm is quite similar to the certificate solution for GI above. The difference is that prior to doing this measurement procedure, we can *verify* the validity of the certificate: Any $h \in H$ must satisfy $hH = H$, so the above procedure, run with h instead of g, should always return $\lvert 0\rangle$ if the certificate is invariant under left multiplication by h. This doesn't mean that the certificate has to be $\lvert H\rangle$, of course; it could, for instance, be any $\lvert G\rangle$, where G is any finite group containing H. But *any certificate* $\lvert\psi\rangle$ which is invariant under left multiplication by any element $h \in H$ suffices to prove that $g \notin H$. (If the algorithm always returns $\lvert 0\rangle$, Arthur cannot guarantee that $g \in H$, since Merlin may be giving Arthur a state such as $\lvert G\rangle$ with $H < G$; but if it returns $\lvert 0\rangle$ and $\lvert 1\rangle$, each with probability $\frac{1}{2}$, then Arthur can be certain, having verified the certificate, that in fact $g \notin H$.)

So to verify the certificate $\lvert\psi\rangle$ we should check that it is invariant under left multiplication by all $h \in H$. This can be done explicitly, in polynomial time, by generating a superposition of all elements of H and using it in place of register B in the algorithm above. A theorem of Babai [9] shows that elements in a finite group (defined by its generators) can be efficiently generated with a random distribution which is very close[24] to uniform. This allows generation of a state which is very close to the uniform (incoherent) superposition over elements of H,

$$B\,|0\rangle\,|0\rangle \equiv \sum_{h\in H} \sqrt{p_h}\,|h\rangle\,|s_h\rangle \approx \frac{1}{\sqrt{|H|}} \sum_{h\in H} |h\rangle\,|s_h\rangle\ ,$$

where $|s_h\rangle$ represents the random numbers and scratch registers used to compute h. (This theorem does *not* immediately allow generation of a state $|\widetilde{H}\rangle$ close to $|H\rangle$, because it may be difficult (or even impossible) to invert the map from the random numbers to the group elements; hence each element will be correlated with some other state information that cannot efficiently be erased—recall the discussion about bijection evaluation in §2.2(iii).)

Watrous actually uses a slightly cleaner technique than the one above to verify the certificate, dispensing with the control register: Suppose Merlin supplies the certificate $|\psi\rangle = \sum_{x\in G} c_x\,|x\rangle$. (We assume that Arthur can easily verify that the state is a superposition only of elements of G, so this is the most general certificate possible.) For clarity, we assume that the Babai construction (with operator B) gives an exactly uniform superposition over elements of H; it is straightforward to work out the inequalities for the actual state.

[24] The algorithm is polynomial in $\log|G|$ and $\log\epsilon$, where we require that the probability of generating $h \in H$ satisfies $\frac{1}{|H|} - \epsilon < p_h < \frac{1}{|H|} + \epsilon$.

$$|\psi\rangle_{\mathrm{X}} |0\rangle_{\mathrm{H}} |0\rangle_{\mathrm{S}} = \sum_{x\in G} c_x |x\rangle_{\mathrm{X}} |0\rangle_{\mathrm{H}} |0\rangle_{\mathrm{S}}$$

$$B_{\mathrm{H,S}} \Big\downarrow$$

Generate random elements of H in register H, approximately uniformly (using register S for scratch space), using Babai's algorithm B

$$\sum_{x\in G} c_x |x\rangle_{\mathrm{X}} \frac{1}{\sqrt{|H|}} \sum_{h\in H} |h\rangle_{\mathrm{H}} |s_h\rangle_{\mathrm{S}}$$

$$\Big\downarrow$$

Perform a left multiplication of register X by register H

$$\frac{1}{\sqrt{|H|}} \sum_{x\in G} \sum_{h\in H} c_x |hx\rangle_{\mathrm{X}} |h\rangle_{\mathrm{H}} |s_h\rangle_{\mathrm{S}}$$

$$(B^{-1})_{\mathrm{H,S}} \Big\downarrow$$

Uncompute the random superposition of elements (if unentangled with register X)

$$\frac{1}{\sqrt{|H|}} \sum_{x\in G} \sum_{h\in H} c_x |hx\rangle_{\mathrm{X}} (B^{-1})_{\mathrm{H,S}} |h\rangle_{\mathrm{H}} |s_h\rangle_{\mathrm{S}}$$

$$\Big\downarrow$$

Measure registers H and S; accept the certificate as a valid certificate for H iff the measurement result is $|0\rangle_{\mathrm{H}} |0\rangle_{\mathrm{S}}$

$$\begin{aligned} &\frac{C}{\sqrt{|H|}} \sum_{x\in G} \sum_{h\in H} c_x |hx\rangle_{\mathrm{X}}\, {}_{\mathrm{H}}\langle 0|\, {}_{\mathrm{S}}\langle 0|\, (B^{-1})_{\mathrm{H,S}} |h\rangle_{\mathrm{H}} |s_h\rangle_{\mathrm{S}} \\ = \; &\frac{C}{|H|} \sum_{x\in G} \sum_{h\in H} c_x |hx\rangle_{\mathrm{X}} \end{aligned}$$

for an accepting measurement, where C is a normalization constant. Note that if this verification procedure accepts the certificate, then the certificate is placed in a uniform superposition[25] over cosets of H; but this is exactly the condition required of the certificate, if Arthur wants to determine whether $g \in H$.

This technique immediately gives Watrous verifiable, succinct certificates for several related properties on finite groups. However, verifiable certificates for graph-isomorphism problems do not seem to follow from his result.

[25] The superposition is actually only approximately uniform since B only approximates a uniform superposition.

Chapter 3

Causal Restrictions on Nonlocal Measurement Superoperators

3.1 Introduction

The previous chapter contained several examples of operations on quantum systems which can be used to solve interesting problems. But in our attempt to discover quantum algorithms we may find that implementing a particular quantum operation seems difficult or impossible. In such cases we may suspect that in fact some general principle is forbidding implementation (or at least efficient implementation) of such an operation. We may ask more generally, *Given a quantum system, what operations on the system are physically realistic?* The basic formalism of quantum mechanics allows, in principle, the measurement of any observable. But this does not provide a complete answer to our question, because the answer may depend on other properties of the system—for example, the presence of certain conservation laws may forbid the measurement of certain observables.

One limit on the sorts of operations we allow can arise as a consequence of symmetries of the system. The WAY theorem [7, 103] (discussed in Peres [86, pp. 421–422]), for example, states that it is impossible to implement exactly the von Neumann measurement of any observable which does not commute with an additive conserved dynamical variable.[1]

[1] For example, since the x component of angular momentum J_x is an additive conserved quantity which does not commute with the z component J_z, it is impossible to *exactly* measure J_x for a system using a von Neumann measurement.

The idea is that such an operator has broken the symmetry implied by the conservation law, and so an interaction Hamiltonian effecting such a measurement also must violate the conservation law; only operators respecting this symmetry are measurable.

The proof is simple: Suppose (following [51]) that M is the observable (on $\mathcal{H}_{\mathrm{S}}$) to be measured, with orthonormal eigenstates $\{|m,\alpha\rangle\}$ satisfying $M\,|m,\alpha\rangle = m\,|m,\alpha\rangle$. (The index α accounts for any degeneracy in the eigenvalues of M.) Let U be the measurement unitary operator on $\mathcal{H}_{\mathrm{S}} \otimes \mathcal{H}_{\mathrm{A}}$, producing the desired ideal correlations between the system $\mathcal{H}_{\mathrm{S}}$ and the apparatus $\mathcal{H}_{\mathrm{A}}$ while preserving the eigenstates of M: $U\,|m,\alpha\rangle_{\mathrm{S}}\,|0\rangle_{\mathrm{A}} = |m,\alpha\rangle_{\mathrm{S}}\,|m,\alpha\rangle_{\mathrm{A}}$ (where $|0\rangle_{\mathrm{A}}$ is the initial state, and ${}_{\mathrm{A}}\langle n,\beta\,|\,m,\alpha\rangle_{\mathrm{A}} \propto \delta_{mn}$; we do not require orthogonality of apparatus states corresponding to degenerate eigenstates of M). Now suppose that Π is an additive conserved quantity, so that all allowable operations preserve the value of $\Pi = \Pi_{\mathrm{S}} \otimes \mathbf{1}_{\mathrm{A}} + \mathbf{1}_{\mathrm{S}} \otimes \Pi_{\mathrm{A}}$. Then in particular the interaction unitary U must preserve Π, $U^{\dagger}\Pi U = \Pi$, and

$$\begin{aligned}
{}_{\mathrm{S}}\langle n,\beta|\,[\Pi_{\mathrm{S}},M]\,|m,\alpha\rangle_{\mathrm{S}} &= {}_{\mathrm{S}}\langle n,\beta|\,{}_{\mathrm{A}}\langle 0|\,[\Pi,M]\,|m,\alpha\rangle_{\mathrm{S}}\,|0\rangle_{\mathrm{A}} \\
&= (m-n)\,{}_{\mathrm{S}}\langle n,\beta|\,{}_{\mathrm{A}}\langle 0|\,\Pi\,|m,\alpha\rangle_{\mathrm{S}}\,|0\rangle_{\mathrm{A}} \\
&= (m-n)\,{}_{\mathrm{S}}\langle n,\beta|\,{}_{\mathrm{A}}\langle 0|\,U^{\dagger}\Pi U\,|m,\alpha\rangle_{\mathrm{S}}\,|0\rangle_{\mathrm{A}} \\
&= (m-n)\,{}_{\mathrm{S}}\langle n,\beta|\,{}_{\mathrm{A}}\langle n,\beta|\,\Pi\,|m,\alpha\rangle_{\mathrm{S}}\,|m,\alpha\rangle_{\mathrm{A}} \\
&= (m-n)\Big[{}_{\mathrm{S}}\langle n,\beta|\,\Pi_{\mathrm{S}}\,|m,\alpha\rangle_{\mathrm{S}}\,{}_{\mathrm{A}}\langle n,\beta\,|\,m,\alpha\rangle_{\mathrm{A}} \\
&\quad + {}_{\mathrm{S}}\langle n,\beta\,|\,m,\alpha\rangle_{\mathrm{S}}\,{}_{\mathrm{A}}\langle n,\beta|\,\Pi_{\mathrm{A}}\,|m,\alpha\rangle_{\mathrm{A}}\Big] \\
&= 0
\end{aligned}$$

for all m,n,α,β; hence we must have $[\Pi,M]=0$.[2]

Another simple example, demonstrating the subtlety required for a description of the set of observables, is given by Nielsen [82]: There exist Hermitian operators which, if measurable, could be used to compute functions which are not classically computable. For example, the halting function $h:\mathbb{N}\to\{0,1\}$, defined to be 1 if the Turing machine

[2] Approximate measurements of observables M which do not commute with Π are still possible; however, lower bounds can be placed on the amount of *distortion* of the eigenstates caused by the measurement operator U, as shown, e.g., in [7, 51]. That is, U cannot leave eigenstates of M unchanged: $\big[\mathbf{1} - |m\alpha\rangle_{\mathrm{S}}\,{}_{\mathrm{S}}\langle m\alpha|\,\big]U\,|m\alpha\rangle_{\mathrm{S}}\,|0\rangle_{\mathrm{A}} \neq 0$.

described by the integer n halts on input n, and 0 if not, was proved by Turing [97] to be uncomputable. But we may define the "halting observable" $\hat{h}$, for a system with orthogonal basis states $\{|n\rangle : n \in \mathbb{N}\}$, as $\hat{h} = \sum_{n=0}^{\infty} h(n)\, |n\rangle\langle n|$.[3] If such an observable could be measured, then a measurement of $h(n)$ could be effected, for arbitrary $n \in \mathbb{N}$, merely by preparing the system in the state $|n\rangle$ and measuring $\hat{h}$. (One can, similarly, find examples of unitary operators which allow by their implementation the determination of uncomputable quantities.) While such an operator is an observable, and is therefore theoretically measurable by the formal principles of quantum mechanics, it is hard to believe that such a thing could actually exist—at any rate, $\hat{h}$ could not be implemented on a quantum computer as a product of operators each acting on small numbers of qubits, as the evolution of such a system is classically computable, while the evolution of a system under $\hat{h}$ is not computable.

The requirement in special relativity that an operation preserve causality provides another source of restrictions on the class of allowed operations; these are the restrictions we will primarily consider in this chapter. In some situations the naïve definition of an operator allows violations of causality (e.g., superluminal communication) and hence is not physically realistic; so further restrictions must be placed on the class of physical operators. Such restrictions depend, of course, on the spacetime regions on which the operators are considered to act. For example, one can formally write down a superoperator $\$$ on a bipartite Hilbert space $\mathcal{H}_{\mathrm{AB}} \equiv \mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ which allows communication from one half of the Hilbert space ($\mathcal{H}_{\mathrm{A}}$) to the other ($\mathcal{H}_{\mathrm{B}}$): i.e., when a state in $\mathcal{H}_{\mathrm{A}}$ is acted on by one of two different local operations on $\mathcal{H}_{\mathrm{A}}$, and then by the superoperator $\$$, the results are distinguishable by examining only $\mathcal{H}_{\mathrm{B}}$. The requirements of causality imply that such an operator is physical only if the region $\mathcal{S}_{\mathrm{A}}$ (in which we allow Alice to interact with the Hilbert space $\mathcal{H}_{\mathrm{A}}$) intersects the causal past of the region $\mathcal{S}_{\mathrm{B}}$. One of the purposes of this paper is to consider the problem of characterizing the operations on a quantum

[3] This construction is somewhat reminiscent of Chaitin's constant

$$\chi = \sum_{n=0}^{\infty} h(n) 2^{-n}$$

in classical computability theory.

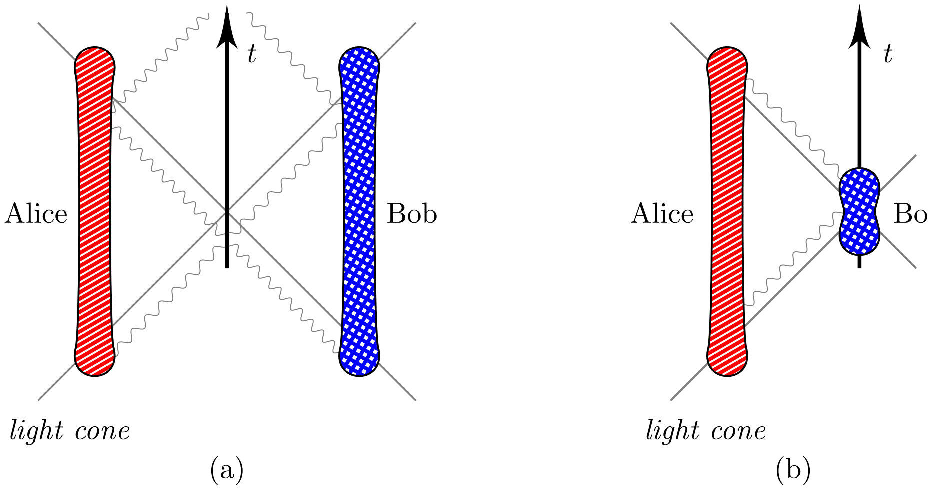


Figure 3.1: Two causal geometries with different characteristics, with information transmission denoted by wavy lightlike lines. In (b) there is sufficient time for a roundtrip transmission; in (a) there is not.

system which do not allow violations of causality. (This question has also been raised by Aharonov, Albert, Popescu, and Vaidman [4, 5, 6, 87], who were largely concerned with single-state nondemolition measurements. In this paper we consider a different class of quantum operations.

It is also interesting to consider the related question of which superoperators Alice and Bob are *actually able* to implement, using local operations, in such cases. In particular, we may ask whether, for a given "causal geometry," any operator whose generation is not forbidden by causality considerations can actually be implemented. (For some such geometries, as we shall see, the answer is *No.*)

Two examples of separated regions of spacetime are shown in Figure 3.1. In Figure 3.1(a), there is time for two one-way messages (which could consist of either classical or quantum communication) to be sent: Alice can send a message from $\mathcal{S}_{\mathrm{A}}$ to Bob in $\mathcal{S}_{\mathrm{B}}$, and Bob can send a message to Alice; each receives the other's message after sending his own. In Figure 3.1(b), there is time for one *round-trip* message to be sent: Alice can send a message to Bob, who can then send a return message to Alice *after* receiving

Alice's message. The situation in (b) is, then, at least as powerful as that in (a):[4] Bob of (b) can, after all, ignore Alice's message until after sending his own, if he wishes. If we allow Alice and Bob to share a quantum channel, then in fact the situation in (b) allows performance of any unitary operation on $\mathcal{H}_{\text{AB}}$: Alice merely sends her system $\mathcal{H}_{\text{A}}$ to Bob, who performs the desired operation and sends her system back to her.[5] It seems unlikely that all operators allowed by the geometry of (b) must also be allowed by the geometry of (a), though; we may gain something by allowing Bob to send Alice information based not only on his own local operations but also on her earlier communication.

3.1.1 The problem

In this paper we consider the simple geometries that result from assuming that $\mathcal{S}_{\text{A}}$ and $\mathcal{S}_{\text{B}}$ are small, widely separated regions of Minkowski spacetime (i.e., with a Euclidean separation much larger than their Euclidean diameters). Figure 3.2 shows the two possible causal structures given this assumption. In Figure 3.2(a), $\mathcal{S}_{\text{A}}$ and $\mathcal{S}_{\text{B}}$ can have no causal contact, since neither intersects the other's future light cone (they have spacelike separation); for this geometry, an operator on $\mathcal{H}_{\text{AB}}$ must not allow signals to be sent either from Alice to Bob or from Bob to Alice. In Figure 3.2(b), $\mathcal{S}_{\text{B}}$ intersects the causal future of $\mathcal{S}_{\text{A}}$, but not vice versa (they have timelike separation); for this geometry, causality allows signals to be sent from Alice to Bob, but not from Bob to Alice.

[4] Here we ignore "complexity-based" constraints: that is, although Figure 3.1(b) seems to imply that Bob has very little time to perform any computations after receiving a signal from Alice but before sending a reply, we will assume that he has sufficient time to perform any desired operation on his system.

[5] If we allow only a *classical* channel between Alice and Bob, then not all operations can be performed even with arbitrarily many rounds of communication between Alice and Bob. For example, Alice and Bob cannot generate shared quantum entanglement using only local operations and classical communication, if they begin with a tensor-product state (and no shared entanglement). Even "separable" superoperators (those for which the Kraus operators can all be written as tensor products, as defined in [89] and discussed in [18, 99]) cannot all be simulated using local operations and arbitrarily many rounds of classical communication; this is proved for a specific superoperator in [18]. Below we exhibit another class of superoperators for which the classical channel provides no aid in simulation; see Theorem 3.

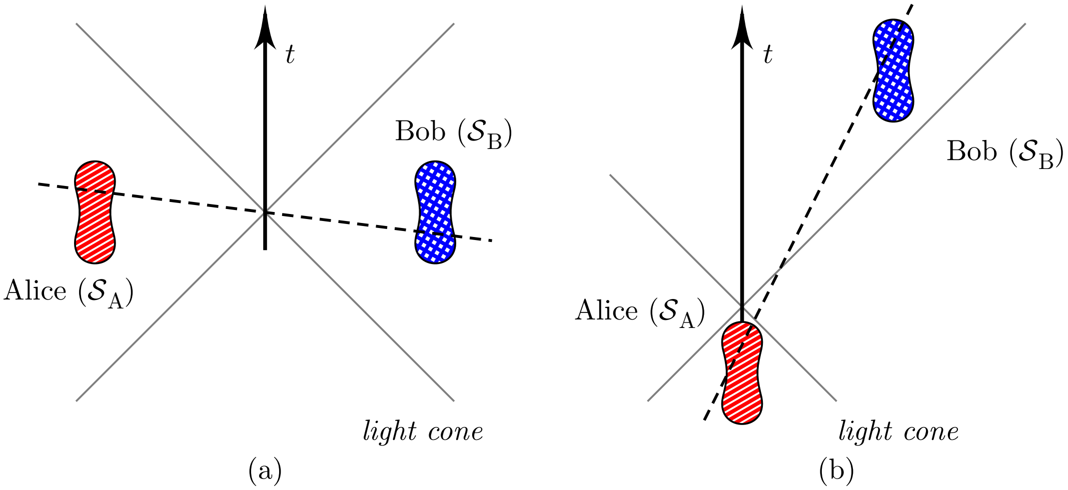

t
Bob ($\mathcal{S}_\mathrm{B}$)
Alice ($\mathcal{S}_\mathrm{A}$)
light cone
(a)
t
Bob ($\mathcal{S}_\mathrm{B}$)
Alice ($\mathcal{S}_\mathrm{A}$)
light cone
(b)

Figure 3.2: The causal geometries considered in this paper. In (a), $\mathcal{S}_\mathrm{A}$ and $\mathcal{S}_\mathrm{B}$ are separated by a spacelike interval. In (b), the spaces have timelike separation. The dashed lines represent the regions on which the superoperator is to act.

Causality

If an operator allows no communication in either direction, from Alice to Bob or from Bob to Alice (as must be the case, for any physical operator, in the situation shown in Figure 3.2(a)), we will say that the operator is **causal**. If it allows no communication from Bob to Alice (as in Figure 3.2(b)), we say it is **1-causal** or **semicausal** (more precisely, 1-causal *from Bob to Alice*); thus a causal operator is 1-causal "in both directions." Operators which can be written as tensor products of local operators clearly satisfy this limitation, but there exist causal operators which are not of this form—for example, the superoperator projecting onto the Bell basis is also a causal operator (a consequence of Theorem 1 below).

Nonlocal operators which allow superluminal signalling if implemented on a spacelike section of spacetime (e.g., as suggested in Figure 3.2(a)) are easily found, however. For example, consider the superoperator which causes decoherence onto the basis $\{|\Phi^+\rangle, |\Phi^-\rangle, |01\rangle, |10\rangle\}$ (where Alice and Bob each have one qubit of the two-qubit system). Suppose Alice and Bob begin with their system in the state $|00\rangle$. If the superoperator acts on this state, Alice's density matrix is then $\sigma(\mathbf{1}) = \frac{1}{2}\mathbf{1}$ (the entire system is an incoherent superposition of the two Bell states $|\Phi^{\pm}\rangle$). But suppose that before the superoperator acts, Bob applies the unitary operator σ_x to his qubit, changing the state to $|01\rangle$. If the superoperator acts on this state instead, Alice's density matrix is $\sigma(\sigma_x) = |0\rangle\langle 0|$, a pure state.[6]

These two results are different and hence distinguishable: Bob can send a message to Alice. If Alice and Bob are far apart and the superoperator is viewed as acting on a spacelike slice, this allows Bob to send a superluminal signal to Alice; such superoperators must therefore be viewed as nonphysical. (Such superoperators may be considered physical if they are sufficiently "smeared out," acting on a timescale at least as long as (c^{-1} times) the spacelike separation between Alice and Bob, so that no superluminal signalling can occur.)

Let $\$$ be a superoperator acting on the Hilbert space $\mathcal{H}_{\rm AB} \equiv \mathcal{H}_{\rm A} \otimes \mathcal{H}_{\rm B}$, where $\dim \mathcal{H}_{\rm A} = \dim \mathcal{H}_{\rm B} = d$. We wish to find conditions on $\$$ for which Bob can send a superluminal signal to Alice. Alice and Bob may, of course, have ancillary systems $\mathcal{H}_{\rm R}$ and $\mathcal{H}_{\rm S}$ (which may initially be entangled or classically correlated, since such correlations may have been created at some time in the past), on which $\$$ acts trivially.[7] We can state the semicausality condition simply, as follows:

Semicausality condition *Signalling from Bob to Alice is possible utilizing superoperator $\$$, and $\$$ is therefore not 1-causal, iff there exist an initial pure state*[8] *$|\psi\rangle_{\rm ARBS}$ and a superoperator $\mathcal{B}$ on $\mathcal{H}_{\rm BS}$ such that*

$$\mathrm{tr}_{\rm BS}\left[\$_{\rm AB}\left(\,|\psi\rangle\langle\psi|\,\right)\right] \neq \mathrm{tr}_{\rm BS}\left[\$_{\rm AB}((\mathbf{1}_{\rm AR}\otimes\mathcal{B}_{\rm BS})(|\psi\rangle\langle\psi|))\right] , \tag{3.1}$$

so that the two density matrices for Alice's system $\mathcal{H}_{\rm AR}$ are different and hence distinguishable (by local actions of Alice).

[6] We use the notation $\sigma(\mathcal{B})$ throughout to denote the final value of Alice's density matrix, after Bob has applied the local operator $\mathcal{B}$ to his system, and after the global superoperator $\$$ has acted.

[7] It turns out that adding ancillary systems doesn't change the class of causal measurements, at least in the special case discussed in Theorem 1.

[8] The requirement that the initial state be pure is no restriction: If signalling is possible with an initial density matrix $\rho_{\rm AB}$, it must also be possible for one of the pure states in any decomposition of $\rho_{\rm AB}$.

Localizability

A closely related condition on a superoperator is that it be **localizable**: implementable using only local operations (performable by Alice alone or by Bob alone) and previously-shared resources. Alice and Bob, using their local quantum computers in the geometry of Figure 3.2(a) and possibly using pre-shared entanglement, can effect an operator on $\mathcal{H}_{\mathrm{AB}}$ iff it is localizable.[9] If Bob's Hilbert space intersects the causal future of Alice's, as in Figure 3.2(b), it is also reasonable to allow one-way communication from Alice to Bob; we call these superoperators **1-localizable** (or **semilocalizable**).[10] (Of course, if Alice just sends her state to Bob then Bob can perform any superoperator he wants, but Alice ends up without her quantum system; we require that at the end of the operation $\mathcal{H}_{\mathrm{A}}$ remains with Alice and $\mathcal{H}_{\mathrm{B}}$ with Bob.) Restricting the communications to a classical channel does not tighten this class, as long as we allow Alice and Bob to share Bell pairs as part of their pre-shared resources: Alice may use the Bell pairs and classical channel to teleport quantum information to Bob. Since we are primarily interested in placing bounds on what is possible, in this paper we will mostly confine our attention to the broad case of quantum channels.

(We note in passing that all superoperators are "2-localizable" (in the sense of Figure 3.1(b)) when a round-trip quantum channel, or its simulation via teleportation, is available. However, as shown by Lo and Popescu [76], two-way *classical* communication is no more powerful than one-way classical communication in terms of entanglement manipulation. Bennett *et al.* [18] also demonstrated a measurement operator on a bipartite state which cannot be implemented using only local operations and classical communication.)

[9] We distinguish *localizable* and *bilocal* superoperators: A localizable superoperator on $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$ is one which can be simulated as a bilocal operation $\$_{\mathrm{AR}} \otimes \$_{\mathrm{BS}}$ on a Hilbert space $\mathcal{H}_{\mathrm{AR}} \otimes \mathcal{H}_{\mathrm{BS}}$: the original Hilbert space augmented with a possibly-entangled ancilla system $\mathcal{H}_{\mathrm{R}} \otimes \mathcal{H}_{\mathrm{S}}$.

[10] In order to keep the notation from becoming too cumbersome, the term "1-localizable" and its earlier analogue "1-causal" leave the Hilbert space labels implicit. Throughout this paper I have tried to maintain a convention consistent with that suggested by Figure 3.2(b) and the example above: That is, a 1-causal superoperator $\$_{\mathrm{AB}}$ does not allow Bob to send a signal to Alice (though it may allow *Alice* to send a signal to *Bob*); a 1-localizable superoperator $\$_{\mathrm{AB}}$ is one that can be implemented with use of a one-way quantum channel from Alice to Bob, with no communication from Bob to Alice.

We can immediately see the following relations on these four classes of superoperators:

$$\begin{array}{ccc} \{\text{causal superoperators}\} & \subsetneqq & \{\text{1-causal superoperators}\} \\ \cup\!\| & & \cup\!\| \;\text{[11]} \\ \{\text{localizable superoperators}\} & \subsetneqq & \{\text{1-localizable superoperators}\} \end{array}$$

The remaining two questions about the inclusion relations (the questions about whether the two vertical inclusions are strict) can be stated as two conjectures:

Conjecture 1 *The set of localizable superoperators is exactly the set of causal superoperators.*

Conjecture 2 (DiVincenzo [39]) *The set of 1-localizable superoperators is exactly the set of 1-causal superoperators.*

DiVincenzo's conjecture holds for the special case of *complete projective superoperators*, as proved and discussed at length in §3.2 below; the general case was not proved here. After publication of these results (in [12]), the general case was proved by Eggeling, Schlingemann, and Werner [40]; thus DiVincenzo's conjecture is true.

However, as will be shown below, the first conjecture is false. This indicates that some further physical constraint, besides causality, is required to tighten the set of operations which are not forbidden until it agrees with the set of operations which we can actually do. Preskill has shown [12] that the CHSH and Cirel'son inequalities can be used to further constrain the class of operations allowed by physical principles, for spacelike-separated systems. It is unclear whether restrictions of this type suffice to reduce the class of "allowed" superoperators to exactly the set of localizable superoperators.

[11] That is (using our convention), the class of superoperators which can be implemented with a one-way quantum channel from Alice to Bob is contained in the class of superoperators which do not allow signalling from Bob to Alice.

3.2 Complete projective superoperators

Let us restrict our attention to a special case: Suppose $\$$ is a **complete von Neumann measurement superoperator** on $\mathcal{H}_{\rm AB}$,[12]

$$\$(\rho) \equiv \sum_a E_a \rho E_a \ , \tag{3.2}$$

where the $\{E_a \equiv |\phi_a\rangle\langle\phi_a|\}$ are a complete set of one-dimensional projectors on $\mathcal{H}_{\rm AB}$ (so $E_a^\dagger = E_a$, $E_a E_b = \delta_{ab} E_b$, $\operatorname{tr} E_a = 1$, and $\sum_a E_a = \mathbf{1}_{\rm AB}$). Let us define $\sigma_a \equiv \operatorname{tr}_{\rm B} E_a$, Alice's density matrices for the basis states $|\phi_a\rangle$.

3.2.1 Causality and 1-causality

For this special case, we can give a complete, intuitive characterization of the measurements which allow signalling:

Theorem 1 *A complete von Neumann measurement superoperator $\$_{\rm AB}$ with (orthogonal, one-dimensional) projectors $\{E_a \equiv |\phi_a\rangle\langle\phi_a|\}$, having partial traces $\sigma_a \equiv \operatorname{tr}_{\rm B} E_a$, is 1-causal (i.e., cannot be used for signalling from Bob to Alice) iff for all σ_a and σ_b, either $\sigma_a = \sigma_b$ or $\sigma_a \sigma_b = 0$.*

We will say that Bob can cause a **transition** $a{\to}b$ (i.e., from state $|\phi_a\rangle$ to state $|\phi_b\rangle$) if, with the system $\mathcal{H}_{\rm AB}$ initially in the pure state $|\phi_a\rangle$, Bob can perform some local superoperator $\mathcal{B}_{\rm BS}$ such that after $\$$ acts (and Alice and Bob throw away their ancilla systems), there is a nonzero probability for the system to be in the final state $|\phi_b\rangle$, i.e.,

$$P_{a\to b} \equiv \operatorname{tr}_{\rm RS}\left[\langle\phi_b|\$_{\rm AB}\Big(\mathcal{B}_{\rm BS}\big(\,|\phi_a\rangle_{\rm AB}\,{}_{\rm AB}\langle\phi_a| \otimes \rho_{\rm RS}\big)\Big)\,|\phi_b\rangle\right] \neq 0 \ , \tag{3.3}$$

where $\rho_{\rm RS}$ is the state of the ancilla system shared between Alice and Bob. Intuitively, the meaning of the theorem is clear: Bob can only induce transitions between subspaces which have nontrivial overlap on Alice's system (hence the requirement that $\sigma_a\sigma_b \neq 0$).

[12] We think of the measurement being implemented "by the environment"; the measurement result is not told to Alice or Bob. That is, this superoperator causes decoherence, or "dephasing," onto the basis $\{|\phi_a\rangle\}$.

But Bob can use such a transition to send a signal to Alice iff the transition is between two subspaces that Alice can distinguish (hence the requirement that $\sigma_a \neq \sigma_b$).

The proof of this theorem divides naturally into two parts, which we state as lemmas. First we find the conditions under which Bob can induce a transition $a \to b$ (namely, precisely when Alice's density matrices σ_a and σ_b have nontrivial overlap):

Lemma 1 *Bob can induce a transition $a \to b$ iff $\sigma_a \sigma_b \neq 0$. Furthermore, when such a transition is allowed, he can always induce the transition with a local unitary operator U acting only on $\mathcal{H}_{\mathrm{B}}$—no ancillary system $\mathcal{H}_{\mathrm{RS}}$ is needed.*

Proof of Lemma 1:

Write the Schmidt decompositions of $\$$'s eigenstates $|\phi_a\rangle$ as

$$|\phi_a\rangle_{\mathrm{AB}} \equiv \sum_i \sqrt{\lambda_{ai}}\, |a,i\rangle_{\mathrm{A}}\, |a,i\rangle_{\mathrm{B}} \qquad \sigma_a = \sum_i \lambda_{ai}\, |a,i\rangle_{\mathrm{A}}\, {}_{\mathrm{A}}\langle a,i| \ ,$$

where all $\lambda_{ai} > 0$ (the sum implicitly runs only over the nonzero terms in the Schmidt decomposition) and ${}_{\mathrm{A}}\langle a,i\,|\,a,j\rangle_{\mathrm{A}} = {}_{\mathrm{B}}\langle a,i\,|\,a,j\rangle_{\mathrm{B}} = \delta_{ij}$. So

$$\sigma_a \sigma_b = \sum_{i,j} \lambda_{ai} \lambda_{bj}\, {}_{\mathrm{A}}\langle a,i\,|\,b,j\rangle_{\mathrm{A}}\, |a,i\rangle_{\mathrm{A}}\, {}_{\mathrm{A}}\langle b,j|$$

and $\sigma_a \sigma_b = 0$ iff all ${}_{\mathrm{A}}\langle a,i\,|\,b,j\rangle_{\mathrm{A}} = 0$ (i.e., the two subspaces are orthogonal). But local action by Bob (on $\mathcal{H}_{\mathrm{BS}}$) cannot change Alice's density matrix. In particular, if the subspaces of σ_a and σ_b are orthogonal, then for *any* local operator X_{BS} (not necessarily Hermitian or trace-preserving) on Bob's system,

$$\begin{aligned} {}_{\mathrm{AB}}\langle\phi_a|\, X_{\mathrm{BS}}\, |\phi_b\rangle_{\mathrm{AB}} &= \sum_{ij} \lambda_{ai} \lambda_{bj}\, {}_{\mathrm{A}}\langle a,i\,|\,b,j\rangle_{\mathrm{AB}}\langle a,i|\, X_{\mathrm{BS}}\, |b,j\rangle_{\mathrm{B}} \\ &= 0\ . \end{aligned}$$

Let Bob's superoperator $\mathcal{B}_{\mathrm{BS}}(\rho) \equiv \sum_c B_c \rho B_c{}^{\dagger}$. $\$$ will project onto the subspace E_b with probability

$$P_{a \to b} = \mathrm{tr}_{\mathrm{RS}}\, {}_{\mathrm{AB}}\langle\phi_b|\, \mathcal{B}_{\mathrm{BS}} \left(|\phi_a\rangle_{\mathrm{AB}}\, {}_{\mathrm{AB}}\langle\phi_a| \otimes \rho_{\mathrm{RS}}\right) |\phi_b\rangle_{\mathrm{AB}}\ ;$$

then if $\sigma_a\sigma_b = 0$,

$$\begin{aligned} P_{a\to b} &= \mathrm{tr}_{\mathrm{RS}} \sum_{cijkl} \sqrt{\lambda_{bi}\lambda_{aj}\lambda_{ak}\lambda_{bl}}\, {}_{\mathrm{A}}\langle b,i\,|\,a,j\rangle_{\mathrm{AA}}\langle a,k\,|\,b,l\rangle_{\mathrm{A}} \\ &\qquad {}_{\mathrm{B}}\langle b,i|\, B_c \left(|a,j\rangle_{\mathrm{BB}}\langle a,k| \otimes \rho_{\mathrm{RS}}\right) B_c{}^{\dagger}\,|b,l\rangle_{\mathrm{B}} \\ &= 0\ . \end{aligned}$$

On the other hand, if $\sigma_a\sigma_b \neq 0$ so that (say) ${}_{\mathrm{A}}\langle a,1\,|\,b,1\rangle_{\mathrm{A}} \neq 0$, then we can just define $\mathcal{B}$ to be a unitary operator U on $\mathcal{H}_{\mathrm{B}}$ which rotates the basis $\left\{\,|a,i\rangle_{\mathrm{B}}\right\}$ into the basis $\left\{\,|b,j\rangle_{\mathrm{B}}\right\}$, with appropriate phases so that the sum is necessarily positive, then $P_{a\to b} > 0$, and Bob may induce the transition $a\to b$ with nonzero probability. (We may assume without loss of generality that ${}_{\mathrm{A}}\langle a,i\,|\,b,i\rangle_{\mathrm{A}} \geq 0$ by moving all phases to the definition of the $|a,i\rangle_{\mathrm{B}}$, a freedom of the Schmidt decomposition; for this choice of basis we then define $U\,|a,i\rangle_{\mathrm{B}} = |b,i\rangle_{\mathrm{B}}$. The sum now consists only of nonnegative terms, at least one of which is positive.) ■

Next we show that a transition $a\to b$ allows signalling from Bob to Alice precisely if Alice's density matrices σ_a and σ_b differ.

Lemma 2 *Signalling from Bob to Alice via the nonlocal superoperator* $\$_{\mathrm{AB}}$ *is possible iff Bob can induce a transition* $a\to b$ *with* $\sigma_a \neq \sigma_b$. *Furthermore, if such a signalling transition can be induced, it can be induced via a local unitary operator* U *acting only on* $\mathcal{H}_{\mathrm{B}}$*; again, no ancillary system is needed.*

Proof of Lemma 2:

First we show that if Bob can only cause transitions $a\to b$ for which $\sigma_a = \sigma_b$, then Alice's final density matrix is unaffected by any local operation performed by Bob. Since the $|\phi_a\rangle$ define an orthonormal basis for $\mathcal{H}_{\mathrm{AB}}$, we may write the initial state as $|\psi\rangle_{\mathrm{ARBS}} \equiv \sum_a \alpha_a\,|\phi_a\rangle_{\mathrm{AB}}\,|\chi_a\rangle_{\mathrm{RS}}$ (where the $|\chi_a\rangle$ need not be orthogonal). Suppose Bob performs the superoperator $\mathcal{B}_{\mathrm{BS}}$ (with Kraus operators $\{B_d\}$, acting on his Hilbert space $\mathcal{H}_{\mathrm{BS}}$) prior to the action of the environment's measurement superoperator $\$$. Alice's final density matrix, as a function of Bob's operation $\mathcal{B}_{\mathrm{BS}}$, is then

$$\begin{aligned}
\sigma(\mathcal{B}) &\equiv \mathrm{tr}_{\mathrm{BS}}\$\Big[\mathcal{B}_{\mathrm{BS}}\big(\,|\psi\rangle\langle\psi|\,\big)\Big] \\
&= \mathrm{tr}_{\mathrm{BS}}\$\Big[\mathcal{B}_{\mathrm{BS}}\Big(\sum_{ab}\alpha_a\alpha_b^*\,|\phi_a\rangle_{\mathrm{AB}}{}_{\mathrm{AB}}\langle\phi_b|\otimes|\chi_a\rangle_{\mathrm{RS}}{}_{\mathrm{RS}}\langle\chi_a|\,\Big)\Big] \\
&= \mathrm{tr}_{\mathrm{BS}}\left[\sum_{abcd}\alpha_a\alpha_b^*(E_c)_{\mathrm{AB}}(B_d)_{\mathrm{BS}}\,|\phi_a\rangle_{\mathrm{AB}}\,|\chi_a\rangle_{\mathrm{RS}}{}_{\mathrm{AB}}\langle\phi_b|\,{}_{\mathrm{RS}}\langle\chi_b|\,(B_d^\dagger)_{\mathrm{BS}}(E_c)_{\mathrm{AB}}\right] \\
&= \mathrm{tr}_{\mathrm{S}}\left[\sum_{abcd}\alpha_a\alpha_b^*\Big[\langle\phi_c|\,B_d\,|\phi_a\rangle\,|\chi_a\rangle\langle\phi_b|\,\langle\chi_b|\,B_d^\dagger\,|\phi_c\rangle\Big]\,\sigma_c\right]\ .
\end{aligned}$$

(in the final line we have dropped the subspace labels for compactness). But each term in this sum has factors $\langle\phi_c|\,B_d\,|\phi_a\rangle$ and $\langle\phi_c|\,B_d\,|\phi_b\rangle^*$, which can only be nonzero if the transition probabilities $P_{a\to c}$ and $P_{b\to c}$ (respectively) are nonzero; hence a term in the sum is nonzero only if both transitions $a\to c$ and $b\to c$ are allowed. Then by assumption, a nonzero term in the sum has $\sigma_a = \sigma_b = \sigma_c$, and

$$\sigma(\mathcal{B}) = \mathrm{tr}_{\mathrm{S}}\left[\sum_{abcd}\alpha_a\alpha_b^*\Big[\langle\phi_c|\,B_d\,|\phi_a\rangle\,|\chi_a\rangle\langle\phi_b|\,\langle\chi_b|\,{B_d}^\dagger\,|\phi_c\rangle\Big]\,\sigma_b\right]$$

Now we may use, successively, the identities $\sum_c|\phi_c\rangle_{\mathrm{AB}}{}_{\mathrm{AB}}\langle\phi_c| = \mathbf{1}$, $\sum_d {B_d}^\dagger B_d = \mathbf{1}$, and ${}_{\mathrm{AB}}\langle\phi_b\,|\,\phi_a\rangle_{\mathrm{AB}} = \delta_{ab}$ to reduce this sum[13] to

$$\sigma(\mathcal{B}) = \sum_a|\alpha_a|^2(\sigma_a)_{\mathrm{B}}\otimes\mathrm{tr}_{\mathrm{S}}\left[|\chi_a\rangle_{\mathrm{RS}}{}_{\mathrm{RS}}\langle\chi_a|\right]\ ,$$

which is independent of $\mathcal{B}$. So Alice can learn nothing by any measurement on $\mathcal{H}_{\mathrm{BS}}$ about whether Bob performed $\mathcal{B}$, and no signalling from Bob to Alice is possible.

Suppose, on the other hand, that Bob can cause a transition $a\to b$ with $\sigma_a \neq \sigma_b$. We show that signalling from Bob to Alice is possible in this case, by showing that there is a particular initial state $|\phi_0\rangle$ from which he can, by a local unitary operation on $\mathcal{H}_{\mathrm{B}}$ alone, affect Alice's final density matrix and hence send Alice a signal.

Suppose Alice and Bob begin with their system in the state $|\phi_a\rangle$. After Bob

[13] It is awkward to write these transformations in bra-ket notation; the presence of the ancilla system $\mathcal{H}_{\mathrm{RS}}$ prevents the two conjugate factors from being true scalars, though the ancilla's presence does not prevent us from commuting these identities through. Writing the equation in tensor notation may help convince you if you don't believe it.

performs a unitary U and the superoperator $\$$ acts, Alice's density matrix is

$$\begin{aligned}\sigma(U) &= \mathrm{tr}_{\mathrm{B}}\left[\sum_b E_b\Big(\mathbf{1}_{\mathrm{A}}\otimes U_{\mathrm{B}}\,|\phi_a\rangle\langle\phi_a|\,\mathbf{1}_{\mathrm{A}}\otimes(U^\dagger)_{\mathrm{B}}\Big)E_b\right]\\ &= \sum_b \Big[\langle\phi_b|\,\mathbf{1}_{\mathrm{A}}\otimes U_{\mathrm{B}}\,|\phi_a\rangle\langle\phi_a|\,\mathbf{1}_{\mathrm{A}}\otimes(U^\dagger)_{\mathrm{B}}\,|\phi_b\rangle\Big]\,\sigma_b\\ &= \sum_b \big|\langle\phi_b|\,\mathbf{1}_{\mathrm{A}}\otimes U_{\mathrm{B}}\,|\phi_a\rangle\big|^2\,\sigma_b\ . \qquad (3.4)\end{aligned}$$

(In particular, if Bob performs the identity operation, then Alice's final density matrix is just $\sigma(\mathbf{1}) = \sigma_a$.) So $\sigma(U)$ is a convex sum of elements of the **reachable set** $S_a \equiv \{\sigma_b : \sigma_b\sigma_a \neq 0\}$ of density matrices which have nonzero transition amplitudes from σ_a. We know by Lemma 1 that all such transitions are possible (i.e., for any $\sigma \in S_a$ there is some U for which σ's coefficient is nonzero).

Suppose that there is a subspace E_a for which σ_a is **extremal** in S_a, i.e., for which σ_a cannot be expressed as a convex sum of elements in $S_a \setminus \{\sigma_a\}$; then σ_a can be written, as a convex combination of elements of S_a, *only* as the trivial sum $\sigma_a = \sigma_a$. But if S_a contains at least two distinct elements ($|S_a| > 1$; we will call such an S_a **nontrivial**), then by Lemma 1 there is a local unitary operator U on $\mathcal{H}_{\mathrm{A}}$ inducing a transition to another $\sigma_b \in S_a$ (with $\sigma_b \neq \sigma_a$). For such a U, Alice's final density matrix $\sigma(U)$ is a convex combination (3.4) of elements of S_a whose coefficient of σ_b is nonzero. But $\sigma(\mathbf{1}) = \sigma_a$ cannot be written in this form; hence $\sigma(U) \neq \sigma(\mathbf{1})$, and signalling is possible.

Now we must show that such an extremal σ_a exists. Note that if $\sigma_a \in S_b$, then since

$$0 \neq \sigma_b\sigma_a = (\sigma_a^\dagger\sigma_b^\dagger)^\dagger = (\sigma_a\sigma_b)^\dagger\ ,$$

$\sigma_b \in S_a$ as well; that is, if Bob can induce the transition $a \rightarrow b$, he can also induce the transition $b \rightarrow a$. So if S_a is nontrivial and $\sigma_b \in S_a$, then S_b is nontrivial as well. Now consider among all nontrivial reachable sets the set S_0 having maximal $\mathrm{tr}\,\sigma_0^2$ (i.e., $\mathrm{tr}\,\sigma_0^2 \geq \mathrm{tr}\,\sigma_a^2$ for all a with nontrivial S_a). For all $\sigma_a \in S_0$, S_a is nontrivial (since $\sigma_0, \sigma_a \in S_a$), so necessarily $\mathrm{tr}\,\sigma_0^2 \geq \mathrm{tr}\,\sigma_a^2$ for all $\sigma_a \in S_0$.

Suppose that σ_0 is not extremal in S_0. Then σ_0 is a nontrivial convex combination of elements of S_0,

$$\sigma_0 = \sum_{\sigma_a \in S_0} p_a \sigma_a$$

(where $p_a \geq 0$ and $\sum_a p_a = 1$), containing at least two nonzero terms. Hence

$$\operatorname{tr} \sigma_0^2 < \sum_{\sigma_a \in S_0} p_k \operatorname{tr} \sigma_a^2 \leq \max_{\sigma_a \in S_0} \operatorname{tr} \sigma_a^2 = \operatorname{tr} \sigma_0^2 \ ,$$

a contradiction. (The first inequality follows from Corollary IX.4.4 in Bhatia [21], using the trace norm $|\!|\!| \cdot |\!|\!| \equiv \operatorname{tr} | \cdot |$.[14]) Hence such a σ_0 is in fact extremal in S_0.

Now if Alice and Bob begin in the state $|\phi_0\rangle$, and Bob performs either the identity $\mathbf{1}$ or a unitary U inducing a (nontrivial) transition $0 \to a$, where $\sigma_a \in S_0 \setminus \{\sigma_0\}$ (i.e., U satisfies $\langle \phi_a | \, \mathbf{1}_{\rm A} \otimes U_{\rm B} \, | \phi_0 \rangle \neq 0$; U may be chosen, e.g., via Lemma 1), Alice's two possible density matrices have $\operatorname{tr} \sigma(\mathbf{1})^2 > \operatorname{tr} \sigma(U)^2$ and so $\sigma(\mathbf{1}) \neq \sigma(U)$, and signalling from Bob to Alice is possible. ■

Now we may prove Theorem 1 as a straightforward application of the two lemmas above.

Proof of Theorem 1:

By Lemma 2, a superoperator $\$$ is 1-causal iff all allowed transitions $a \to b$ have $\sigma_a = \sigma_b$. But by Lemma 1, a transition $a \to b$ is allowed unless $\sigma_a \sigma_b = 0$. Thus $\$$ is 1-causal iff we have either $\sigma_a \sigma_b = 0$ (transition forbidden) or $\sigma_a = \sigma_b$ (transition doesn't allow signalling). This completes the proof of Theorem 1. ■

We can massage these conditions on the σ_a into a more interesting form. Let P_a project onto the d_a-dimensional support subspace of σ_a, and define $d_{\rm A} \equiv \dim \mathcal{H}_{\rm A}$, $d_{\rm B} \equiv \dim \mathcal{H}_{\rm B}$. Then we find

$$\begin{aligned} d_{\rm B} \mathbf{1} &= \operatorname{tr}_{\rm B} \sum_b E_b &&= \sum_b \sigma_b \\ P_a d_{\rm B} &= P_a \sum_b \sigma_b &&= \sum_{b : \sigma_b = \sigma_a} \sigma_b &&= k \sigma_a \end{aligned}$$

[14] It is also straightforward to prove this directly, by showing that

$$\operatorname{tr}(pX + qY)^2 \leq p \operatorname{tr} X^2 + q \operatorname{tr} Y^2$$

(where $p, q > 0$ and $p + q = 1$) with equality iff $X = Y$, and generalizing with induction.

(where $k \equiv \big|\{b : \sigma_b = \sigma_a\}\big|$), since the σ_b which are not equal to σ_a are orthogonal to σ_a. Taking the trace of both sides, and using $\operatorname{tr}\sigma_b = 1$, we find that there are $k = d_a d_\mathrm{B}$ projectors E_b with $\sigma_b = \sigma_a$ and thus that $\sigma_a = \frac{1}{d_a} P_a$: the $|\phi_a\rangle$ are d_a-dimensional Bell states, and the σ_a are completely mixed states on their support subspaces.

(In general, we say that a bipartite state is a d-dimensional **Bell state** if it has Schmidt decomposition $\frac{1}{\sqrt{d}} \sum_{i=1}^{d} |i\rangle_\mathrm{A} |i\rangle_\mathrm{B}$, where the $|i\rangle_\mathrm{A}$ and $|i\rangle_\mathrm{B}$ are orthonormal sets of vectors on $\mathcal{H}_\mathrm{A}$ and $\mathcal{H}_\mathrm{B}$. A **Bell basis** for a $d \times d$-dimensional Hilbert space is an orthonormal basis all of whose elements are d-dimensional—that is, full-rank—Bell states. The famous **standard Bell basis**, for example, is the basis $\{|\Phi^\pm\rangle, |\Psi^\pm\rangle\}$, with

$$\left|\Phi^\pm\right\rangle \equiv \tfrac{1}{\sqrt{2}}\Big[\,|00\rangle \pm |11\rangle\Big] \qquad \left|\Psi^\pm\right\rangle \equiv \tfrac{1}{\sqrt{2}}\Big[\,|01\rangle \pm |10\rangle\Big]\ ,$$

on a 2 × 2-dimensional Hilbert space.)

A causal superoperator is 1-causal in both directions, so such a superoperator induces partitions $\{P_a\}$ and $\{Q_b\}$ of both Alice's and Bob's Hilbert spaces $\mathcal{H}_\mathrm{A}$ and $\mathcal{H}_\mathrm{B}$. Suppose that not all the P_a and Q_b have the same dimension d; then there must exist particular P_a and Q_b with $m \equiv \operatorname{tr} P_a \neq \operatorname{tr} Q_b \equiv n$. But now, since the $|\phi_i\rangle$ form a basis for $\mathcal{H}_\mathrm{AB}$, there must be mn states in $\{|\phi_i\rangle\}$ with support only on $P_a \times Q_b$. Such a state $|\phi\rangle$ has (as we saw in the discussion above)

$$\operatorname{tr}_\mathrm{B} |\phi\rangle\langle\phi| = \tfrac{1}{m} P_a \quad \text{and} \quad \operatorname{tr}_\mathrm{A} |\phi\rangle\langle\phi| = \tfrac{1}{n} Q_b\ ;$$

but these density matrices have different Schmidt ranks, which is not possible. So all the P_a and Q_b must have the same dimension d (which, therefore, must divide both d_A and d_B), and we have:

Corollary 1 *A causal, complete von Neumann measurement superoperator on $\mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$ partitions the space into $d \times d$-dimensional subspaces, each of which contains d^2 projectors onto rank-d Bell states.*

That is, on each $d \times d$ subspace, the superoperator acts by decoherence onto some Bell basis for that subspace.

3.2.2 Localizability and 1-localizability

Thus we have a complete and convenient way to characterize those complete von Neumann measurement superoperators which are causal or 1-causal. What can we say about the localizable and 1-localizable superoperators?

Tensor-product measurement superoperators (of the form $\mathcal{A}_{\rm A} \otimes \mathcal{B}_{\rm B}$, where both $\mathcal{A}_{\rm A}$ and $\mathcal{B}_{\rm B}$ are measurement superoperators) are clearly localizable.[15] Less obvious is that projections onto some Bell bases are also localizable. The following superoperator equivalence demonstrates how this is done for the standard Bell basis $\{|\Phi^{\pm}\rangle, |\Psi^{\pm}\rangle\}$ in 2×2 dimensions; the extension to $d \times d$ dimensions is analogous.[16]

$$\begin{aligned} \$_{\rm Bell}(\rho) &= E_{\Phi^+}\rho E_{\Phi^+} + E_{\Phi^-}\rho E_{\Phi^-} + E_{\Psi^+}\rho E_{\Psi^+} + E_{\Psi^-}\rho E_{\Psi^-} \\ &= \tfrac{1}{4}[\rho + (\sigma_x \otimes \sigma_x)\rho(\sigma_x \otimes \sigma_x) + (\sigma_y \otimes \sigma_y)\rho(\sigma_y \otimes \sigma_y) + (\sigma_z \otimes \sigma_z)\rho(\sigma_z \otimes \sigma_z)] \ . \end{aligned} \tag{3.5}$$

Thus to implement this superoperator, Alice and Bob can use a pre-shared random number or pre-shared entanglement to decide which of $\mathbf{1}$, σ_x, σ_y, σ_z (each with probability $\frac{1}{4}$) to act with. (Such a construction is not always possible; below we demonstrate that not all Bell-basis measurement superoperators are localizable.) Note that this Bell basis "measurement superoperator" differs from a *measurement* in the Bell basis; such a measurement of course cannot be performed locally, since it is an entangling measurement.

This superoperator, causing decoherence in a Bell basis, can also be viewed as a special

[15] Although the "measurement" (decoherence) of the causal superoperator is considered to be performed by the environment, so that neither Alice nor Bob necessarily has any information about the particular final state of the system (i.e., which projective result occurred), we will often speak loosely of simulating such a measurement by actions which could result in Alice or Bob gaining some information about the measurement's result. In such a situation, rather than *measure* the system, Alice and Bob should correlate it with an ancillary quantum system (postponing all measurements until the end of the computation, in the usual way); at the end of the operation, instead of measuring the ancilla, it should just be discarded. Similarly, if (as in the next example) a superoperator has a realization by bilocal Kraus operators, Alice and Bob could gain information about which Kraus operator was implemented by reading their random numbers; these should also be encoded in an ancilla and discarded after use. (Of course, in either case the ensemble average of the resulting density matrix, over many such simulations, will agree with $\$(\rho)$.)

[16] This is closely related to the "twirling" operation of [19, 18]. I am grateful to David DiVincenzo for pointing this out.

case of a superoperator which projects onto the stabilizer subspaces (eigenspaces) of an arbitrary qudit stabilizer code: For a stabilizer subgroup $\mathcal{S} = \{S_j : j = 1, \ldots, n\}$, an abelian subgroup of the Pauli group,[17] the superoperator

$$\$_{\mathcal{S}}(\rho) = \tfrac{1}{n} \sum_j S_j \rho S_j^\dagger \tag{3.6}$$

is localizable (with the obvious extension of the definition to multipartite systems), since elements of the Pauli group are tensor products of local operators; the technique is the same as above. (This technique can be applied more generally; any superoperator which can be written as a random local unitary superoperator,

$$\$(\rho) = \sum_i p_i U_i \rho U_i^\dagger \ , \quad U_i \equiv (V_i)_{\mathrm{A}} \otimes (W_i)_{\mathrm{B}} \ , \quad \sum_i p_i = 1 \ ,$$

is localizable. I do not know of an easy way of determining in general whether a given superoperator has such a decomposition, though.)

A simple calculation shows that

Lemma 3 $\$_{\mathcal{S}}$*, the superoperator implemented by applying a random unitary operator chosen uniformly from the stabilizer group* $\mathcal{S} < \mathcal{P}_d^{\otimes n}$ *(as in (3.6)), induces decoherence into the stabilizer subspaces of* $\mathcal{S}$*.*

Proof of Lemma 3:

Let $\{|\phi_i\rangle\}$ be a simultaneous eigenbasis of all $S_j \in \mathcal{S}$, with $S_j |\phi_a\rangle = v_{aj} |\phi_a\rangle$ ($v_{aj} \in \{1, \omega, \ldots, \omega^{d-1}\}$), and define the **syndromes**[18] $\boldsymbol{v}_a \equiv (v_{aj})$. Because $\mathcal{S}$ is a

[17] We define the **Pauli group** $\mathcal{P}_d$ in $U(d)$ as the group generated by the generalizations X and Z of the Pauli matrices σ_x and σ_z, defined by their actions on the computational basis: $X|a\rangle \equiv |(a+1) \bmod d\rangle$ and $Z|a\rangle \equiv \omega^a |a\rangle$, where ω is a primitive dth root of unity. $\mathcal{P}_d$ has order d^3; as a projective group (ignoring the $U(1)$ phase factors) it has order d^2. On a Hilbert space of n qudits, we define the Pauli group as $\mathcal{P}_d^{\otimes n}$. (This group is commonly used in analyses of quantum error-correcting codes on d-dimensional Hilbert spaces. There is a large body of work on this subject; see, e.g., [53, 64, 90] for more details.) There is also extensive literature describing the theory of stabilizer codes; for example, see Gottesman's readable papers on binary [52] and nonbinary [53] stabilizer codes.

[18] The syndromes v_{aj} for a set of generators of $\mathcal{S}$ uniquely define the syndromes for the remaining group elements, since $\mathcal{S}$ is abelian; in quantum error correction it is common to consider the syndromes over a generating set. But for the proof it is useful to consider the syndromes of all group elements.

group, if $\boldsymbol{v}_a \neq \boldsymbol{v}_b$ (so that $|\phi_a\rangle$ and $|\phi_b\rangle$ are in different stabilizer subspaces) then $\boldsymbol{v}_a \cdot \boldsymbol{v}_b \equiv \sum_c v^*_{ac} v_{bc} = 0$: Suppose $\boldsymbol{v}_a \neq \boldsymbol{v}_b$ so that for some k, $v_{ak} \neq v_{bk}$. Then

$$\begin{aligned}
\boldsymbol{v}_a \cdot \boldsymbol{v}_b \, |\phi_b\rangle\langle\phi_a| &= \sum_c S_c \, |\phi_b\rangle\langle\phi_a| \, S_c^\dagger \quad = \quad \sum_c (S_c S_k) \, |\phi_b\rangle\langle\phi_a| \, (S_c S_k)^\dagger \\
& \qquad \text{(just a reordering of the sum, since } \mathcal{S} \text{ is a group)} \\
&= \sum_c S_c S_k \, |\phi_b\rangle\langle\phi_a| \, S_k^\dagger S_c^\dagger \quad = \quad v^*_{ak} v_{bk} \boldsymbol{v}_a \cdot \boldsymbol{v}_b \, |\phi_b\rangle\langle\phi_a| \\
0 &= (1 - v^*_{ak} v_{bk}) \boldsymbol{v}_a \cdot \boldsymbol{v}_b \, |\phi_b\rangle\langle\phi_a| \\
0 &= \boldsymbol{v}_a \cdot \boldsymbol{v}_b \, .
\end{aligned}$$

So an initial pure state $\sum_a \alpha_a |\phi_a\rangle$ decoheres to

$$\begin{aligned}
\$\Big(\sum_{ab} \alpha^*_a \alpha_b \, |\phi_b\rangle\langle\phi_a| \Big) &= \tfrac{1}{n} \sum_{jab} \alpha^*_a \alpha_b S_j \, |\phi_b\rangle\langle\phi_a| \, S_j^\dagger \\
&= \sum_{ab} \alpha^*_a \alpha_b \Big(\tfrac{1}{n} \sum_j v^*_{aj} v_{bj} \Big) \, |\phi_b\rangle\langle\phi_a| \\
&= \sum_{ab} \alpha^*_a \alpha_b \frac{\boldsymbol{v}_a \cdot \boldsymbol{v}_b}{n} \, |\phi_b\rangle\langle\phi_a| \ :
\end{aligned}$$

the $|\phi_b\rangle\langle\phi_a|$ matrix element vanishes if $|\phi_a\rangle$ and $|\phi_b\rangle$ belong to different stabilizer subspaces and is unaffected if they belong to the same stabilizer subspace. This is precisely the operation of decoherence into the set of stabilizer subspaces. ■

Measurement of stabilizer operators (e.g., the syndrome measurement for a stabilizer code) is a 1-localizable operation, using the usual method of measuring a stabilizer operator with one ancilla qudit, as long as the stabilizer generators are localizable. Alice does an operation between her qudit and the ancilla and then passes the ancilla to Bob, who does an operation on his qudit and the ancilla and then measures the ancilla qubit, allowing Bob (but not, in general, Alice) to know the syndrome. We will discuss this further below.

3.2.3 Causal, nonlocalizable superoperators

However, not all causal measurement superoperators are localizable. Counterexamples exist for $d_\mathrm{A} \times d_\mathrm{B} = 4 \times 4$. In order to demonstrate this, the following theorem will be useful:

Theorem 2 *Let $\$$ be a localizable superoperator on a bipartite Hilbert space $\mathcal{H} \equiv \mathcal{H}_\mathrm{A} \otimes \mathcal{H}_\mathrm{B}$,* *and suppose that $|\psi\rangle$, $A \otimes \mathbf{1}\,|\psi\rangle$, and $\mathbf{1} \otimes B\,|\psi\rangle$ are all eigenstates of $\$$ (where A and B are unitary operators on $\mathcal{H}_{\mathrm{A}}$, $\mathcal{H}_{\mathrm{B}}$ respectively). Then $A \otimes B\,|\psi\rangle$ is also an eigenstate of $\$$.*

(We say that a state $|\psi\rangle \in \mathcal{H}$ is an **eigenstate** of superoperator $\$$ iff $|\psi\rangle\langle\psi|$ is an eigenstate (necessarily having eigenvalue 1) of $\$$. We will also say that an eigenstate of an operator (or superoperator) is **stabilized** by that operator.) This result makes sense intuitively; it means that local action by Alice, $A \otimes \mathbf{1}$, and local action by Bob, $\mathbf{1} \otimes B$, can't affect each other.

The superoperators of interest to us here, derived from localizable protocols, can be taken to have the form

$$\$(\rho) = \mathrm{tr}_{\mathrm{RS}} \$' \left(\rho_{\mathrm{AB}} \otimes |\chi\rangle_{\mathrm{RS}\,\mathrm{RS}}\langle\chi|\right)$$

for some localizable superoperator $\$' = \mathcal{A}_{\mathrm{AR}} \otimes \mathcal{B}_{\mathrm{BS}}$ and an ancilla state $|\chi\rangle_{\mathrm{RS}}$. An easy lemma characterizing the eigenstates of $\$'$ will be useful in further discussion of localizable superoperators.

Lemma 4 *A state $|\psi\rangle$ is an eigenstate of a superoperator $\$$ of the form*

$$\$(\rho) = \mathrm{tr}_{\mathrm{R}} (\$')_{\mathrm{AR}} \big[\, |\psi\rangle_{\mathrm{A}\,\mathrm{A}}\langle\psi| \otimes |\chi\rangle_{\mathrm{R}\,\mathrm{R}}\langle\chi| \,\big]$$

iff for each Kraus operator M_i' of $\$'$ we have

$$M_i'\,|\psi\rangle_{\mathrm{A}}\,|\chi\rangle_{\mathrm{R}} = |\psi\rangle_{\mathrm{A}}\,|\chi_i\rangle_{\mathrm{R}}$$

for some (not necessarily normalized) state $|\chi_i\rangle$.

Proof of Lemma 4:

Suppose $|\psi\rangle$ is an eigenstate of such a superoperator. Then

$$\begin{aligned} |\psi\rangle\langle\psi| \;&=\; \$(|\psi\rangle\langle\psi|) \;=\; \mathrm{tr}_{\mathrm{R}} (\$')_{\mathrm{AR}} \big[\, |\psi\rangle_{\mathrm{A}\,\mathrm{A}}\langle\psi| \otimes |\chi\rangle_{\mathrm{R}\,\mathrm{R}}\langle\chi| \,\big] \\ &=\; \mathrm{tr}_{\mathrm{R}} \sum_i M_i'\,|\psi\rangle_{\mathrm{A}}\,|\chi\rangle_{\mathrm{RA}}\langle\psi|\,{}_{\mathrm{R}}\langle\chi|\,(M_i')^\dagger \;. \end{aligned}$$

Define $P_\psi \equiv |\psi\rangle\langle\psi|$ and $P_\psi^\perp \equiv \mathbf{1} - P_\psi$. Then

$$\begin{aligned}
0 &= \langle\psi|\, P_\psi^\perp \,|\psi\rangle \quad = \quad \mathrm{tr}\, P_\psi^\perp \,|\psi\rangle\langle\psi| \quad = \quad \mathrm{tr}\, P_\psi^\perp \$(|\psi\rangle\langle\psi|) \\
&= \sum_i {}_{\mathrm{A}}\langle\psi|\, {}_{\mathrm{R}}\langle\chi|\, (M_i')^\dagger P_\psi^\perp M_i' \,|\psi\rangle_{\mathrm{A}} \,|\chi\rangle_{\mathrm{R}} \\
&= \sum_i \left\| P_\psi^\perp M_i' \,|\psi\rangle_{\mathrm{A}} \,|\chi\rangle_{\mathrm{R}} \right\|^2 .
\end{aligned}$$

Hence all

$$\big(P_\psi^\perp\big)_{\mathrm{A}} \big(M_i'\big)_{\mathrm{AR}} \,|\psi\rangle_{\mathrm{A}} \,|\chi\rangle_{\mathrm{R}} = 0,$$

so $M_i' \,|\psi\rangle_{\mathrm{A}} \,|\chi\rangle_{\mathrm{R}}$ has no component orthogonal to $|\psi\rangle$, and

$$M_i' \,|\psi\rangle_{\mathrm{A}} \,|\chi\rangle_{\mathrm{R}} = |\psi\rangle_{\mathrm{A}} \,|\chi_i\rangle_{\mathrm{R}}$$

for some (possibly unnormalized) state $|\chi_i\rangle$. (The converse is obvious.) ■

In particular, we can take $\mathcal{H}_{\mathrm{R}}$ to be one-dimensional to see that $|\psi\rangle$ must be an eigenstate of each Kraus operator M_i of $\$$.

With this result in hand we can prove Theorem 2:

Proof of Theorem 2:

Our first step is to write formally what we mean by a "localizable superoperator." We will account for entanglement (or classical randomness) previously shared between Alice and Bob by adjoining to the system a fixed pure state[19] $|\chi\rangle_{\mathrm{RS}}$ in an ancillary Hilbert space $\mathcal{H}_{\mathrm{RS}} \equiv \mathcal{H}_{\mathrm{R}} \otimes \mathcal{H}_{\mathrm{S}}$ shared between them; Alice can operate only on $\mathcal{H}_{\mathrm{AR}}$ and Bob can operate only on $\mathcal{H}_{\mathrm{BS}}$. The action of $\$$ on $\mathcal{H}_{\mathrm{AB}}$ can be taken to consist of the action of a bilocal superoperator $\mathcal{A}_{\mathrm{AR}} \otimes \mathcal{B}_{\mathrm{BS}}$ on the augmented bipartite system $\mathcal{H}_{\mathrm{AR}} \otimes \mathcal{H}_{\mathrm{BS}}$, following which the ancillary system $\mathcal{H}_{\mathrm{RS}}$ is discarded:

$$\begin{aligned}
\$(\rho) &= \mathrm{tr}_{\mathrm{RS}} \left[\left(\mathcal{A}_{\mathrm{AR}} \otimes \mathcal{B}_{\mathrm{BS}} \right) \left(\rho_{\mathrm{AB}} \otimes |\chi\rangle_{\mathrm{RS}\,\mathrm{RS}}\langle\chi| \right) \right] \\
&= \mathrm{tr}_{\mathrm{R'S'}} \left[\left(U_{\mathrm{AR'}} \otimes V_{\mathrm{BS'}} \right) \left(\rho_{\mathrm{AB}} \otimes |\chi'\rangle_{\mathrm{R'S'}\,\mathrm{R'S'}}\langle\chi'| \right) \right] . \qquad (3.7)
\end{aligned}$$

[19] There is no loss of generality in restricting the ancilla to a pure state; it is easy to see that if a state $|\psi\rangle$ is an eigenstate when a mixed-state ancilla ρ_{RS} is used, then it must also be an eigenstate for each pure state in any decomposition of ρ_{RS}.

To get this second form, in which U and V are local *unitary* (super)operators and $|\chi'\rangle \equiv |\chi\rangle|0\cdots 0\rangle$, we use the unitary representation [88, §3.2] of an arbitrary superoperator (extending $\mathcal{H}_{\mathrm{R}}$ and $\mathcal{H}_{\mathrm{S}}$ as necessary by adding the ancilla states $|0\cdots 0\rangle$ required for the unitary simulations of the superoperators $\mathcal{A}$ and $\mathcal{B}$). Henceforth we will drop the primes on the ancillary system labels, and consider $\$$ to consist of a bilocal unitary superoperator $U_{\mathrm{AR}} \otimes V_{\mathrm{BS}}$ followed by a trace over the ancilla space $\mathcal{H}_{\mathrm{RS}}$.

Now the unitary superoperator $U_{\mathrm{AR}} \otimes V_{\mathrm{BS}}$ has a single Kraus operator, $U_{\mathrm{AR}} \otimes V_{\mathrm{BS}}$. So by Lemma 4, a state $|\phi_i\rangle \in \mathcal{H}_{\mathrm{AB}}$ is an eigenstate of $\$$ if and only if

$$U_{\mathrm{AR}} V_{\mathrm{BS}} |\phi_i\rangle_{\mathrm{AB}} |\chi\rangle_{\mathrm{RS}} \quad = \quad |\phi_i\rangle_{\mathrm{AB}} |\chi_i\rangle_{\mathrm{RS}} \tag{3.8}$$

for some $|\chi_i\rangle_{\mathrm{RS}} \in \mathcal{H}_{\mathrm{RS}}$.[20]

Now suppose that $|\psi\rangle$, $A \otimes \mathbf{1}\,|\psi\rangle$, and $\mathbf{1} \otimes B\,|\psi\rangle$ are all eigenstates of $\$$, so that

$$\begin{aligned}
U_{\mathrm{AR}} V_{\mathrm{BS}} \quad |\psi\rangle_{\mathrm{AB}} \; |\chi\rangle_{\mathrm{RS}} &= \quad |\psi\rangle_{\mathrm{AB}} \; |\chi_{\mathbf{1}}\rangle_{\mathrm{RS}} \\
U_{\mathrm{AR}} V_{\mathrm{BS}} \Big[A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}}\Big] |\chi\rangle_{\mathrm{RS}} &= \Big[A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}}\Big] |\chi_A\rangle_{\mathrm{RS}} \\
U_{\mathrm{AR}} V_{\mathrm{BS}} \Big[B_{\mathrm{B}} |\psi\rangle_{\mathrm{AB}}\Big] |\chi\rangle_{\mathrm{RS}} &= \Big[B_{\mathrm{B}} |\psi\rangle_{\mathrm{AB}}\Big] |\chi_B\rangle_{\mathrm{RS}} \,.
\end{aligned}$$

The commutator Δ_{AR} of $U_{\mathrm{AR}} V_{\mathrm{BS}}$ and A_{A} has a particularly simple form when acting on $A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}$:

$$\begin{aligned}
\Delta_{\mathrm{AR}} \Big[A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}\Big] &= \Big(U_{\mathrm{AR}} A_{\mathrm{A}} U^{\dagger}{}_{\mathrm{AR}} A^{\dagger}{}_{\mathrm{A}}\Big) A_{\mathrm{A}} U_{\mathrm{AR}} V_{\mathrm{BS}} |\psi\rangle_{\mathrm{AB}} |\chi\rangle_{\mathrm{RS}} \\
&= U_{\mathrm{AR}} V_{\mathrm{BS}} A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi\rangle_{\mathrm{RS}} \\
&= A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_A\rangle_{\mathrm{RS}}
\end{aligned}$$

so Δ_{AR} acts on $A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}$ merely by changing the state of the *ancilla* system,[21] from $|\chi_{\mathbf{1}}\rangle$ to $|\chi_A\rangle$. Thus we can write

$$A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_A\rangle_{\mathrm{RS}} = \Delta_{\mathrm{AR}} \Big[A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}\Big] = (\mathbf{1}_{\mathrm{A}} \otimes R_{\mathrm{R}}) A_{\mathrm{A}} |\psi\rangle_{\mathrm{AB}} |\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}$$

[20] Although the proof of this theorem does not require a characterization of the $|\chi_i\rangle$, it is worth noting that if $\$$ is a complete projective measurement superoperator with (orthogonal) eigenstates $|\phi_i\rangle$, then necessarily $\langle\chi_i\,|\,\chi_j\rangle = \delta_{ij}$.

[21] Note that this is *not* an operator equation; it does not necessarily follow that $\Delta_{\mathrm{AR}} \equiv \mathbf{1}_{\mathrm{A}} \otimes R_{\mathrm{R}}$.

for some unitary R_{R} on $\mathcal{H}_{\mathrm{R}}$. (The content of this statement is that $|\chi_A\rangle_{\mathrm{RS}} = R_{\mathrm{R}}\,|\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}$ for some unitary operator R_{R} acting *trivially on* $\mathcal{H}_{\mathrm{S}}$, and nontrivially only on $\mathcal{H}_{\mathrm{R}}$.) Thus

$$\begin{aligned} U_{\mathrm{AR}}V_{\mathrm{BS}}A_{\mathrm{A}}\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} &= A_{\mathrm{A}}R_{\mathrm{R}}U_{\mathrm{AR}}V_{\mathrm{BS}}\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} \\ U_{\mathrm{AR}}A_{\mathrm{A}}\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} &= A_{\mathrm{A}}R_{\mathrm{R}}U_{\mathrm{AR}}\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}}\,. \end{aligned}$$

Similarly we can commute $U_{\mathrm{AR}}V_{\mathrm{BS}}$ through B_{B} at the cost of a unitary operator S_{S} acting only on $\mathcal{H}_{\mathrm{S}}$:

$$V_{\mathrm{BS}}B_{\mathrm{B}}\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} = B_{\mathrm{B}}S_{\mathrm{S}}V_{\mathrm{BS}}\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}}\,.$$

So

$$\begin{aligned} &U_{\mathrm{AR}}V_{\mathrm{BS}}\Big[A_{\mathrm{A}}B_{\mathrm{B}}\,|\psi\rangle_{\mathrm{AB}}\Big]\,|\chi\rangle_{\mathrm{RS}} \\ &= (V_{\mathrm{BS}}B_{\mathrm{B}})(U_{\mathrm{AR}}A_{\mathrm{A}})\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} \\ &= (V_{\mathrm{BS}}B_{\mathrm{B}})(A_{\mathrm{A}}R_{\mathrm{R}}U_{\mathrm{AR}})\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} \\ &= (A_{\mathrm{A}}R_{\mathrm{R}}U_{\mathrm{AR}})(V_{\mathrm{BS}}B_{\mathrm{B}})\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} \\ &= (A_{\mathrm{A}}R_{\mathrm{R}}U_{\mathrm{AR}})(B_{\mathrm{B}}S_{\mathrm{S}}V_{\mathrm{BS}})\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} \\ &= (A_{\mathrm{A}}B_{\mathrm{B}})(R_{\mathrm{R}}S_{\mathrm{S}})(U_{\mathrm{AR}}V_{\mathrm{BS}})\,|\psi\rangle_{\mathrm{AB}}\,|\chi\rangle_{\mathrm{RS}} \\ &= (A_{\mathrm{A}}B_{\mathrm{B}})(R_{\mathrm{R}}S_{\mathrm{S}})\,|\psi\rangle_{\mathrm{AB}}\,|\chi_{\mathbf{1}}\rangle_{\mathrm{RS}} \\ &= \Big[A_{\mathrm{A}}B_{\mathrm{B}}\,|\psi\rangle_{\mathrm{AB}}\Big]\Big[R_{\mathrm{R}}S_{\mathrm{S}}\,|\chi_{\mathbf{1}}\rangle_{\mathrm{RS}}\Big]\;; \end{aligned}$$

thus $A \otimes B\,|\psi\rangle$ also has the form (3.8) and is also an eigenstate of $\$$. ■

Example 1

With this result in hand, we discuss several cases of causal measurements which are not localizable. First, a simple example will demonstrate the sort of difficulties encountered in trying to bilocally simulate a causal operator. Consider a causal measurement superoperator $\$$ in 4×4 dimensions, with its 16 eigenstates consisting of four "Bell subbases," one rotated relative to the rest by a local unitary operator $\Upsilon_{\mathrm{B}} \in U(2)$ (acting nontrivially only on the subspace spanned by $|2\rangle_{\rm B}$ and $|3\rangle_{\rm B}$):

$$\begin{array}{cc} & \text{Alice } (\mathcal{H}_{\rm A}\ ) \\ \text{Bob } (\mathcal{H}_{\rm B}) & \begin{array}{|c|c|}\hline |\phi^{\pm}_{00}\rangle, \qquad |\psi^{\pm}_{00}\rangle & |\phi^{\pm}_{20}\rangle, \qquad |\psi^{\pm}_{20}\rangle \\ \hline |\phi^{\pm}_{02}\rangle, \qquad |\psi^{\pm}_{02}\rangle & \Upsilon_{\rm B}\,|\phi^{\pm}_{22}\rangle, \Upsilon_{\rm B}\,|\psi^{\pm}_{22}\rangle \\ \hline \end{array} \end{array} \tag{3.9}$$

where we define $|\phi^{\pm}_{ij}\rangle \equiv \frac{1}{\sqrt{2}}(|i,j\rangle \pm |i+1,j+1\rangle)$ and $|\psi^{\pm}_{ij}\rangle \equiv \frac{1}{\sqrt{2}}(|i,j+1\rangle \pm |i+1,j\rangle)$. (The boxes indicate the division of $\mathcal{H}_{\rm AB}$ into its four 2×2-dimensional Bell subbases.) Theorem 1 immediately tells us that such a superoperator is causal.

But suppose that $|\phi^{+}_{00}\rangle$, $|\phi^{+}_{20}\rangle = (X^2)_{\rm A} \otimes \mathbf{1}_{\rm B}\,|\phi^{+}_{00}\rangle$, and $|\phi^{+}_{02}\rangle = \mathbf{1}_{\rm A} \otimes (X^2)_{\rm B}\,|\phi^{+}_{00}\rangle$ are all eigenstates of a localizable superoperator $\$'$ (where $X\,|i\rangle = |(i+1) \bmod 4\rangle$, an element of $\mathcal{P}_4$). Then Theorem 2 tells us that $|\phi^{+}_{22}\rangle = (X^2)_{\rm A} \otimes (X^2)_{\rm B}\,|\phi^{+}_{00}\rangle$ must also be an eigenstate of $\$'$; thus $\$$ is not localizable unless $\Upsilon_{\rm B}$ merely permutes the Bell subbasis $\{|\phi^{\pm}_{22}\rangle, |\psi^{\pm}_{22}\rangle\}$.

Corollary 2 *Not all causal operators are localizable.*

Intuitively, if a causal measurement consists of several disjoint $d \times d$-dimensional subspaces of Bell-basis projectors, the projectors on each subspace may use Bell states in a different basis: each subspace has complete freedom in the choice of Schmidt basis. For generic choice of Schmidt bases, Alice and Bob cannot locally determine what the measurement basis should be, so they are unable to implement the measurement superoperator bilocally, even with shared entanglement.

$\$$ is 1-localizable, however: Alice can make a partial measurement, with projectors $P_{01} = |0\rangle\langle 0| + |1\rangle\langle 1|$, $P_{23} = |2\rangle\langle 2| + |3\rangle\langle 3|$ (allowing her to determine which column the state is in) and send the measurement result to Bob; then Bob, after making the analogous partial measurement to determine which row the state is in, knows what the Bell measurement basis should be. Then Alice and Bob can apply the superoperator (3.5) on the appropriate subspace, as determined by their measurement results; if Bob determines that the state was in the $P_{23} \otimes P_{23}$ subspace, he must rotate his basis by $(\Upsilon^{\dagger})_{\rm B}$ first. (This construction is further generalized below.) Implementation of this superoperator requires only a *classical* channel from Alice to Bob; they do not need to use any nonlocal quantum resources.

Bell bases

In fact, not even all superoperators projecting onto a *single* Bell basis are localizable. As noted above, the superoperator projecting onto the **standard Bell basis** in $d \times d$ dimensions, consisting of states $|\phi_{ab}\rangle \equiv X^a Z^b |\Phi_d^+\rangle$, where $|\Phi_d^+\rangle \equiv \frac{1}{\sqrt{d}} \sum_{i=0}^{d-1} |i\rangle |i\rangle$, is localizable. However, in some dimensions there are other Bell bases inequivalent to the standard basis under local unitary transformations; by mixing these inequivalent Bell bases we can find a new Bell basis for which the measurement superoperator is not localizable. Some examples in 4×4 dimensions are discussed below.[22]

It will be convenient in further discussion of Bell bases to encode the basis information as a set of operators, so that we need not explicitly work with the states $|\phi_i\rangle$. Let us characterize a Bell basis $\{|\phi_i\rangle\}$ in a $d \times d$-dimensional Hilbert space $\mathcal{H}_{\mathrm{AB}}$ by the set $\mathcal{W} = \{W_i\}$ of d^2 unitary operators on $\mathcal{H}_{\mathrm{A}}$ with $|\phi_i\rangle \equiv W_i \otimes \mathbf{1} |\Phi_d^+\rangle$. (Thus for the standard Bell basis, we may take $\mathcal{W} = \bigl\{X^a Z^b : a, b \in \{0, \ldots, d-1\}\bigr\}$.) The $|\phi_i\rangle$ are orthonormal, $\langle\phi_i \,|\, \phi_j\rangle = \delta_{ij}$, iff the W_i are orthonormal with respect to the Hilbert-Schmidt inner product

$$\operatorname{tr} W_i^\dagger W_j = d\, \delta_{ij} \ . \tag{3.10}$$

Since we are interested here not strictly in Bell bases, but in *projections* onto these bases, we will not concern ourselves with the $U(1)$ phase of the W_i. (That is, we are interested in $U(1)$ cosets of elements of $U(d)$: elements of the **projective** group $U(d)\big/U(1) \equiv \bigl\{\{e^{i\phi}U : \phi \in [0, 2\pi)\} : U \in U(d)\bigr\}$.) It will also be convenient to locally define the computational bases in such a way that $\mathbf{1} \in \mathcal{W}$. We will call such a Bell basis **normal**.

The following well-known result will be useful in working further with Bell states:

Lemma 5 *Let X and Y be arbitrary linear operators ($X, Y \in GL(d, \mathbb{C})$). Then*

$$X_{\mathrm{A}} \otimes \mathbf{1}_{\mathrm{B}} |\Phi_d^+\rangle_{\mathrm{AB}} = \mathbf{1}_{\mathrm{A}} \otimes Y_{\mathrm{B}} |\Phi_d^+\rangle_{\mathrm{AB}}$$

if and only if

$$Y = X^{\mathrm{T}} \ ,$$

[22] I don't know whether there are examples in 3×3 dimensions.

where the transpose (a basis-dependent operation) is taken relative to the computational bases (the Schmidt bases of $|\Phi_d^+\rangle_{\rm AB}$).

Proof of Lemma 5:

We have

$$\begin{aligned} X_{\rm A}\otimes \mathbf{1}_{\rm B}\,|\Phi_d^+\rangle_{\rm AB} &= \sum_j X_{\rm A}\,|j\rangle_{\rm A}\,|j\rangle_{\rm B} &= \sum_{ij}\left[\langle i|\,X\,|j\rangle\right]|i\rangle_{\rm A}\,|j\rangle_{\rm B} \\ \mathbf{1}_{\rm A}\otimes Y_{\rm B}\,|\Phi_d^+\rangle_{\rm AB} &= \sum_i |i\rangle_{\rm A} Y_{\rm B}\,|i\rangle_{\rm B} &= \sum_{ij}\left[\langle j|\,Y\,|i\rangle\right]|i\rangle_{\rm A}\,|j\rangle_{\rm B}\ , \end{aligned}$$

which are clearly equal if and only if $\langle i|\,X\,|j\rangle = \langle j|\,Y\,|i\rangle$: that is, if $Y = X^{\rm T}$. ■

Now we can derive a useful operator form of the stabilizer equation

$$S\,|\phi\rangle = e^{i\chi}\,|\phi\rangle\ ,$$

where $|\phi\rangle$ is a full-rank Bell state with $|\phi\rangle \equiv W\otimes \mathbf{1}\,|\Phi_d^+\rangle$, so that

$$SW\,|\Phi_d^+\rangle = e^{i\chi}W\,|\Phi_d^+\rangle\ .$$

Of particular interest to us will be the case where S is a bilocal unitary operator $S_{\rm AB} = U_{\rm A}\otimes V_{\rm B}$ on $\mathcal{H}_{\rm A}\otimes\mathcal{H}_{\rm B}$ and $W\in U(d)$ is a local unitary operator on $\mathcal{H}_{\rm A}$. In this case we have

$$UW\otimes V\,|\Phi_d^+\rangle = e^{i\chi}W\otimes\mathbf{1}\,|\Phi_d^+\rangle\ ,$$

which (via Lemma 5) implies

$$V^{\rm T} = e^{i\chi}W^\dagger U^\dagger W\ .$$

If S also stabilizes $|\Phi_d^+\rangle$ itself, $S\,|\Phi_d^+\rangle = e^{i\chi_0}\,|\Phi_d^+\rangle$ (e.g., if S stabilizes a normal Bell basis), then we also have

$$V^{\rm T} = e^{i\chi_0}U^\dagger\ .$$

We may take $\chi_0 = 0$ WLOG by a local change of basis; this merely redefines the phases of all the eigenstates of S, and these phases affect the projectors trivially. Thus we may take $V^{\rm T} = U^\dagger$, and hence $S = U\otimes U^*$ (we call this the **normal** form for S), and arrive at our desired stabilizer criterion

$$UW = e^{i\chi}WU\ . \tag{3.11}$$

We will say that such a W is **stabilized** by S (and by U).

We make note now of the simple facts (which we will use later) that

Lemma 6 *If a bilocal stabilizer (in its normal form, $S = U \otimes U^*$) stabilizes both of the local unitary operators W and W', with eigenvalues $e^{i\chi}$ and $e^{i\chi'}$, then S stabilizes their product WW', with eigenvalue $e^{i(\chi+\chi')}$; and the operator inverse W^{-1}, with eigenvalue $e^{-i\chi}$. Furthermore, if $e^{i\chi} = e^{i\chi'}$, then any operator of the form $\alpha W + \alpha' W'$ is also stabilized by S, also with eigenvalue $e^{i\chi}$.*

Proof of Lemma 6:

(Trivial.) ■

Now suppose $\mathbf{1}, U, V \in \mathcal{W}$, and suppose that superoperator $\$$ induces decoherence in the basis $\mathcal{W}$; then the states $|\Phi^+\rangle$, $U \otimes \mathbf{1} |\Phi^+\rangle$, and $V \otimes \mathbf{1} |\Phi^+\rangle = \mathbf{1} \otimes V^{\mathrm{T}} |\Phi^+\rangle$ are eigenstates of $\$$. If $\$$ is localizable, then Theorem 2 applies, and $(U \otimes \mathbf{1})(\mathbf{1} \otimes V^{\mathrm{T}}) |\Phi^+\rangle = U \otimes V^{\mathrm{T}} |\Phi^+\rangle = UV \otimes \mathbf{1} |\Phi^+\rangle$ is also an eigenstate of the superoperator. This means that if decoherence onto the normal basis $\mathcal{W} \ni U, V$ can be simulated by a localizable superoperator, then for some phase ϕ, $e^{i\phi} UV \in \mathcal{W}$—that is, the set $U(1)\mathcal{W} \equiv \{e^{i\phi} W : \phi \in [0, 2\pi), W \in \mathcal{W}\}$ is closed under operator composition as long as $\mathbf{1} \in \mathcal{W}$. (This, along with the continuity of $\$$, implies that $U(1)\mathcal{W}$ forms a group.)

Note that local rotations R and local basis transformations B on Alice's Hilbert space transform the W_i by $W_i \mapsto R W_i B^\dagger$; so local operations can ensure that in fact $\mathbf{1} \in \mathcal{W}$, and Alice can choose a basis (or make a local unitary rotation) so that $U(1)\mathcal{W}$ is closed. (In general, $\mathcal{W}$ forms a coset of a projective subgroup of $U(d)$.) The condition that the basis of d^2 operators $\{W_i\}$ is projectively closed (i.e., up to phase) is equivalent to Knill's condition in [64] that an error basis be **nice**. We are primarily interested here in the projectors $|\phi_i\rangle\langle\phi_i|$ and not the states $|\phi_i\rangle$, so in fact the phases of the W_i are not relevant. We may (partially) define this phase by the condition $\det W_i = 1$ and restrict ourselves (WLOG) to Knill's **very nice** bases.

Example 2

Now consider the superoperator $\$$ in 4×4 dimensions projecting onto the (normal) Bell basis

$$\begin{aligned}\mathcal{W} \;=\; \{ \;& \mathbf{1}\;, \quad Z\;, \quad Z^2, \quad Z^3, \\ & X\;,\; X\;Z\;,\; X\;Z^2,\; X\;Z^3, \\ & X^2,\; X^2Z\;,\; X^2Z^2,\; X^2Z^3, \\ & X^3,\; X^3Z_1,\; X^3Z_2,\; X^3Z_1Z_2 \;\;\}\;,\end{aligned} \tag{3.12}$$

where $X\left|a\right\rangle \equiv \left|(a+1) \bmod 4\right\rangle$, $Z\left|a\right\rangle \equiv i^a\left|a\right\rangle$, $Z_1\left|a\right\rangle \equiv (-1)^a\left|a\right\rangle$ (so $Z_1 \equiv Z^2$), and $Z_2\left|a\right\rangle \equiv (-1)^{\lfloor a/2 \rfloor}\left|a\right\rangle$. (That is, X and Z are elements of the four-dimensional Pauli group $\mathcal{P}_4$, and Z_1 and Z_2 are elements of two independent two-dimensional Pauli groups $\mathcal{P}_2$. This basis, loosely speaking, is a mixture of the basis $\mathcal{P}_4$ and the basis $\mathcal{P}_2 \times \mathcal{P}_2$.) It is easy to verify that this set gives an orthogonal basis—that is, that the $W_i \in \mathcal{W}$ are orthonormal in the Hilbert-Schmidt norm (3.10). However, $\mathcal{W}$ is not projectively closed under operator composition:[23] The product $X \cdot X^2Z = X^3Z$, for example, is not proportional to any element of $\mathcal{W}$. Hence $\$$ is not localizable.

This superoperator is, however, 1-localizable (when we allow *quantum* communication from Alice to Bob). Note that the basis states are all eigenstates of the stabilizer subgroup with generators $Z \otimes Z^{-1}$ and $X^2 \otimes X^2$: the operators W_i all satisfy the stabilization condition (3.11), both with $U = Z$ and with $U = X^2$. (The states in a standard Bell basis are all eigenstates of the stabilizer subgroup generated by $Z \otimes Z^{-1}$ and $X \otimes X^{-1}$. X and Z_2 do not satisfy (3.11), but X^2 and Z_2 do.) Alice and Bob can measure these stabilizer generators 1-localizably.[24] For example, to measure $Z \otimes Z^{-1}$, Alice and Bob can perform the unitary operation $\Lambda(X^{-1})_{[\![\mathrm{B}]\!],\mathrm{R}}\Lambda(X)_{[\![\mathrm{A}]\!],\mathrm{R}}$ (where $\mathcal{H}_\mathrm{R}$ is an ancilla system which Alice sends to Bob after performing the $\Lambda(X)_{[\![\mathrm{A}]\!],\mathrm{R}}$ operation), after which Bob can perform $\Lambda(X^{-1})_{[\![\mathrm{B}]\!],\mathrm{R}}$ and measure $\mathcal{H}_\mathrm{R}$ in the computational basis to determine the value of $Z \otimes Z^{-1}$. Measurement of these stabilizer generators, a partial "syndrome measurement," effects a

[23] Thus more generally, normal bases which are not projectively closed (in the terminology of Knill [64], bases which are not nice) are not localizable.

[24] More precisely, Alice and Bob can perform a 1-localizable operation, after which Bob knows the measurement result. The theory of stabilizer codes provides efficient algorithms for performing 1-local measurements of stabilizer operators.

projection into one of eight two-dimensional subspaces, partitioning the set $\mathcal{W}$ above into eight two-element subsets, tabulated here with the stabilizer generators' eigenvalues as shown:

$$
\begin{array}{rr|cc}
 & & \multicolumn{2}{c}{X^2 \otimes X^2 :} \\
 & & +1 & -1 \\ \cline{2-4}
 & +1 & \{\mathbf{1}, Z^2\} & \{Z, Z^3\} \\
Z \otimes Z^{-1} : & +i & \{X, XZ^2\} & \{XZ, XZ^3\} \\
 & -1 & \{X^2, X^2Z^2\} & \{X^2Z, X^2Z^3\} \\
 & -i & \{X^3, X^3Z^2\} & \boxed{\{X^3Z_2, X^3Z^2Z_2\}}
\end{array} \tag{3.13}
$$

(the boxed subspace is the one causing all the problems). Now the remaining stabilizer generator, which will remove the degeneracies of these eight subspaces and thus complete the projection, depends on the subspace—if it didn't, the operator would be localizable. It can be written as $X \otimes X$ for seven of the eight subspaces; for the remaining (boxed) subspace, with syndrome $(X^2 \otimes X^2, Z \otimes Z^{-1}) = (-1, -i)$, the generator can be written as $X \otimes Z_2XZ_2$. Now Alice and Bob can perform the (localizable) projection onto the subspaces of this stabilizer, using the construction of (3.6) above: Alice and Bob will use a shared random number c, and Alice will *always* perform X^c; Bob, who does know the result of the original measurement, can decide whether to perform $Z_2X^cZ_2$ (for the $(-1, -i)$ subspace) or X^c (for all others).[25] It's important here that *Alice's* stabilizer generator X is the same for each of the subspaces, since she doesn't know the results of the projective measurements of the first two stabilizer generators and hence doesn't know what subspace the state is in.

For the superoperator in the first example, we exhibited a 1-localizable simulation which used only *classical* resources. But any 1-localizable simulation of the superoperator of Example 2 requires an "effective quantum channel" (that is, either a quantum channel or a simulation of one using teleportation; thus, nonlocal quantum resources are required). We can prove this using the following result:

Theorem 3 *Suppose $\$$ is a 1-localizable superoperator on $\mathcal{H}_{\mathrm{AB}}$ which can be implemented*

[25] Alternatively, they can perform a 1-localizable measurement of this stabilizer generator using the same technique that they used to measure $X^2 \otimes X^2$ and $Z \otimes Z^\dagger$, allowing Bob to measure the eigenvalue of this generator as well.

when the only nonlocal resources are classical (classical channel and classical shared random bits), and suppose that $\$$ *has an eigenstate* $|\phi\rangle$ *of Schmidt rank* $d_{\mathrm{A}} \equiv \dim \mathcal{H}_{\mathrm{A}}$*. Then* $\$$ *is localizable; the classical channel adds no power.*

So the set of 1-localizable superoperators is strictly larger than the set of superoperators which can be simulated when the only nonlocal resources available (i.e., the one-way channel and the previously shared state) are classical resources. (The reason this argument doesn't apply to Example 1 is that it requires that the eigenstate $|\phi\rangle$ be a *full-rank* Bell state. The eigenstates (3.9) for Example 1 are Bell states, but not full-rank Bell states: they have rank 2, not 4.)

Proof of Theorem 3:

Suppose we have only a one-way classical channel available (and no shared entanglement!). Now the most general operation that Alice and Bob can implement can be written as a generalized measurement superoperator (POVM) by Alice, with the measurement result sent to Bob, who performs a superoperator $\mathcal{B}_i$ conditioned on Alice's measurement result.[26] We have left one possibility out here. The definition of a POVM does not specify the operator's action on the Hilbert space being measured (here $\mathcal{H}_{\mathrm{A}}$), and there are many ways that a POVM could be implemented as the action of a superoperator on $\mathcal{H}_{\mathrm{AM}}$ (where $\mathcal{H}_{\mathrm{M}}$ is the space used to hold the result of the POVM, initially in the state $|0\rangle_{\mathrm{M}}$). Clearly, however, these all differ only by local actions of Alice, applied after her superoperator implementation of the POVM, and so these will not affect whether the operator is 1-localizable.

We consider Alice's POVM, with generalized projectors $F_i = A_i^\dagger A_i$, to act with operator elements A_i, and we define the Kraus operators B_{ij} of Bob's conditional superoperator $\mathcal{B}_i$. Thus the most general operation satisfying these hypotheses

[26] With a one-way classical channel, Alice can generate random bits as part of the POVM and send them to Bob as part of her measurement result, so we need not explicitly include shared random bits in the protocol; this is why $\mathcal{H}_{\mathrm{R}}$ does not appear here.

can be taken to have the form

$$\begin{aligned}\$(\rho) &= \sum_{ij}\big[(A_i)_{\rm A}(B_{ij})_{\rm B}\big]\rho_{\rm AB}\big[(A_i)_{\rm A}(B_{ij})_{\rm B}\big]^\dagger \\ &= \mathrm{tr}_{\rm S}\Big[\sum_i \big[(A_i)_{\rm A}(V_i)_{\rm BS}\big]\big(\rho_{\rm AB}\otimes|0\rangle_{\rm S\,S}\langle 0|\,\big)\big[(A_i)_{\rm A}(V_i)_{\rm BS}\big]^\dagger\Big]\ ,\end{aligned}$$

where $\sum_j B_{ij}^\dagger B_{ij} = \mathbf{1}$ for all i, $\sum_i A_i^\dagger A_i = \mathbf{1}$, and V_i is a unitary implementation on $\mathcal{H}_{\rm BS}$ for superoperator $\mathcal{B}_i$ (implicitly extending $\mathcal{H}_{\rm S}$ as we did in the proof of Theorem 2). An eigenstate $|\phi\rangle$ of $\$$, by Lemma 4, is an eigenstate of each Kraus operator $(A_i)_{\rm A}(V_i)_{\rm BS}$ of $\$$:

$$(A_i)_{\rm A}(V_i)_{\rm BS}\,|\phi\rangle_{\rm AB}\,|0\rangle_{\rm S} = \alpha_i\,|\phi\rangle_{\rm AB}\,|\chi_i\rangle_{\rm S}\ .$$

Now we can put restrictions on the allowed POVM operators A_i that Alice can use: Suppose one of the eigenstates $|\phi\rangle$ of $\$$ is a rank-$d_{\rm A}$ Bell state.[27] Then

$$\begin{aligned}A_i A_i^\dagger &= A_i\mathbf{1}A_i^\dagger \\ &= d_{\rm A}(A_i)_{\rm A}\,\mathrm{tr}_{\rm BS}\Big[(V_i^\dagger)_{\rm BS}(V_i)_{\rm BS}\big(\,|\phi\rangle_{\rm AB\,AB}\langle\phi|\,|0\rangle_{\rm S\,S}\langle 0|\,\big)\Big](A_i^\dagger)_{\rm A} \\ &= d_{\rm A}\,\mathrm{tr}_{\rm BS}\Big[\big[(A_i)_{\rm A}(V_i)_{\rm BS}\big]\big(\,|\phi\rangle_{\rm AB\,AB}\langle\phi|\,|0\rangle_{\rm S\,S}\langle 0|\,\big)\big[(A_i)_{\rm A}(V_i)_{\rm BS}\big]^\dagger\Big] \\ &= d_{\rm A}\,\mathrm{tr}_{\rm BS}\Big[|\alpha_i|^2\,|\phi\rangle_{\rm AB\,AB}\langle\phi|\otimes|\chi_i\rangle_{\rm S\,S}\langle\chi_i|\,\Big] \\ &= |\alpha_i|^2\mathbf{1}\end{aligned}$$

and hence, for a finite-dimensional Hilbert space $\mathcal{H}_{\rm A}$, we may conclude that $U_i \equiv \frac{1}{\alpha_i}A_i$ is unitary.[28] Thus $\$$ has the form

$$\$(\rho) = \mathrm{tr}_{\rm S}\Big[\sum_i |\alpha_i|^2\big[(U_i)_{\rm A}(V_i)_{\rm BS}\big]\big(\rho_{\rm AB}\otimes|0\rangle_{\rm S\,S}\langle 0|\,\big)\big[(U_i)_{\rm A}(V_i)_{\rm BS}\big]^\dagger\Big]\ .$$

But in this form, we see that $\$$ can be simulated using random bilocal unitary operations: with probability $|\alpha_i|^2$, Alice applies U_i and Bob applies V_i. Such a superoperator is *localizable*, not just 1-localizable (cf. the discussion following (3.6)).

[27] By Theorem 1, of course, since $\$$ is 1-causal, this implies that all of the eigenstates are rank-$d_{\rm A}$ Bell states.

[28] If $\mathcal{H}_{\rm A}$ is infinite-dimensional, we can't immediately conclude that $\frac{1}{\alpha_i}A_i$ is unitary; we can say only that it is an isometry. I have not investigated this case further.

Thus Theorem 2 applies, and a one-way classical channel doesn't allow implementation of any nonlocalizable Bell-basis measurement superoperators. ■

This example illustrates a technique for simulating 1-localizable projective superoperators which is more general than the stabilizer-decoherence technique of (3.6). The idea is to *recursively* decompose a space into stabilizer subspaces. With a one-way quantum channel, decoherence into stabilizer subspaces can be implemented directly, by ancilla-based measurements of each of the stabilizer generators $S_i \equiv A_i \otimes B_i$ in turn, instead of by random bilocal unitary operators as in (3.6). Bob learns the eigenvalue s_i of each stabilizer generator S_i (the "syndrome") and so implements the projection onto the subspaces, as desired. But Bob can do something more general than this: Since, before the measurement of the ith operator, Bob already knows the previous measurement results $\boldsymbol{s}_i \equiv (s_1, \ldots, s_{i-1})$, Bob can choose his unitary operator based on these measurement results; the ith operator measured by Alice and Bob can thus have the form $S_{\boldsymbol{s}_i} \equiv A_i \otimes B_{\boldsymbol{s}_i}$. (Alice's operators A_i are still not allowed to depend on the measurement record $\boldsymbol{s}_i$, because she cannot in general know the result of the syndrome measurement; only Bob knows $\boldsymbol{s}_i$.)

Instead of n bilocal stabilizer generators forming a commutative subgroup of $\mathsf{U}(d)$, the stabilizers on different subspaces need not be the same. More precisely, consider a **tree** of bilocal unitary stabilizer operators $S_{i,\boldsymbol{s}_i} = A_i \otimes B_{i,\boldsymbol{s}_i}$, where $\boldsymbol{s}_i \equiv (s_1, \ldots, s_{i-1})$ indicates the record of measurements made so far. (That is, s_i is the result measured for the operator $S_{i,\boldsymbol{s}_i}$ if the previous measurement results are $\boldsymbol{s}_i$. Bob's operators $B_{i,\boldsymbol{s}_i}$ are allowed to depend on $\boldsymbol{s}_i$, because he knows the measurement record; Alice's are not, because she cannot in general know the result of the syndrome measurement.)

For the above construction to work, the stabilizer operators $S_{i,\boldsymbol{s}_i}$ should have the following characteristics: Each route from the root of the tree to a leaf should define a stabilizer code (i.e., each operator in such a path should commute with all the other operators along the path). The stabilizer subspace defined by a partial set of measurements $\boldsymbol{s}_k \equiv (s_1, \ldots, s_k)$ (the measurement results of stabilizers $S_1, \ldots, S_{k(s_1,\ldots,s_{k-1})}$) is precisely the span of some subset $V_{\boldsymbol{s}_k}$ of the set of eigenstates $V = \{|\phi_i\rangle\}$. Furthermore, the stabilizer $S_{(k+1)\boldsymbol{s}_k}$ partitions $V_{\boldsymbol{s}_k}$ by its measurement result. The stabilizer subspaces at the lowest (leaf) levels have dimension at most 1.

Example 3

But for this "recursive decomposition" into stabilizer subspaces, we still need to be able to find at least one nontrivial bilocal stabilizer element, in order that the "syndrome tree" actually branches. It turns out that we can't always do this; there are causal, complete von Neumann measurement superoperators whose measurement bases cannot be partitioned by any bilocal stabilizer. To show this we prove a result on the necessary structure of eigenstates of a bilocal stabilizer operator.

Lemma 7 *Suppose a unitary operator $U \in \mathsf{U}(d)$ stabilizes the linearly-independent unitary operators $P, P' \in \mathsf{U}(d)$ (in the sense of (3.11)), with eigenvalues $e^{i\chi}$ and $e^{i\chi'}$ respectively. Then either P and P' are orthogonal (i.e., $\operatorname{tr} P^\dagger P' = 0$), or they have the same eigenvalue ($e^{i\chi} = e^{i\chi'}$) and span a two-dimensional eigenspace of U.*

That is, if $|\phi\rangle \equiv (P \otimes \mathbf{1})\,|\Phi_d^+\rangle$ and $|\phi'\rangle \equiv (P' \otimes \mathbf{1})\,|\Phi_d^+\rangle$ are both stabilized by $S = U \otimes U^*$, then either $\langle \phi \,|\, \phi' \rangle = 0$, or $|\phi\rangle$ and $|\phi'\rangle$ have the same eigenvalue and span a two-dimensional eigenspace.

Proof of Lemma 7:

The proof is straightforward. We have

$$\begin{aligned} UP &= e^{i\chi} P U \\ UP' &= e^{i\chi'} P' U \end{aligned}$$

by assumption. Now we have

$$\begin{aligned} \operatorname{tr}(P^\dagger P') &= \operatorname{tr}(P^\dagger U^\dagger U P') \quad = \quad \operatorname{tr}\left[(UP)^\dagger (UP')\right] \\ &= e^{i(\chi'-\chi)} \operatorname{tr}\left[(PU)^\dagger (P'U)\right] \quad = \quad e^{i(\chi'-\chi)} \operatorname{tr}(U^\dagger P^\dagger P' U) \\ &= e^{i(\chi'-\chi)} \operatorname{tr}(P^\dagger P' U U^\dagger) \quad = \quad e^{i(\chi'-\chi)} \operatorname{tr}(P^\dagger P') \\ 0 &= \left[e^{i(\chi'-\chi)} - 1\right] \operatorname{tr}(P^\dagger P') \end{aligned}$$

from which we have either $e^{i\chi} = e^{i\chi'}$ (P and P' have the same eigenvalue) or $\operatorname{tr} P^\dagger P' = 0$ (P and P' are orthogonal in the Hilbert-Schmidt inner product), as stated. ■

Now we can use this result to prove

Theorem 4 *There exist Bell bases for which the only bilocal stabilizer operators are proportional to the identity.*

Proof of Theorem 4:

To prove this, we want to find a Bell basis whose elements W_i are mixed in such a way that Lemma 7 puts strong constraints on the form of any local stabilizer. Now an arbitrary Bell basis $\mathcal{W} = \{W_i\}$ on a $d \times d$-dimensional Hilbert space can be written as a unitary transformation from the standard Bell basis $\mathcal{P}_d = \{P_i\}$,

$$W_i = \sum_j U_{ij} P_j \ , \quad U \in \mathsf{U}(d^2) \ . \tag{3.14}$$

This is just because we can view the $\{W_i\}$ and the $\{P_i\}$ as two orthonormal bases (with respect to the Hilbert-Schmidt inner product) for the vector space $\mathsf{GL}(d, \mathbb{C})$ of $d \times d$ linear operators, and any two orthonormal bases on the same space are related by a unitary transformation. Unfortunately, analysis of Bell bases is complicated by the fact that not all $U \in \mathsf{U}(d^2)$ specify Bell bases: The W_i trivially continue to satisfy the Hilbert-Schmidt orthogonality condition (3.10), but they are not generally unitary. For the two-qubit Pauli group $\mathcal{P} \equiv \mathcal{P}_2^{\otimes 2}$ (and in fact for any tensor product $\mathcal{P}_2^{\otimes n}$ of qubit Pauli groups), though, there is a simple class of unitary transformations which preserves the unitarity of the elements, based on the fact that for any $P \in \mathcal{P}_2^{\otimes n}$, $P^2 = \pm \mathbf{1}$.

In particular, suppose $P, Q \in \mathcal{P} \equiv \mathcal{P}_2^{\otimes 2}$. We can modify this standard basis by replacing the elements P and Q by the two unitary operators

$$\begin{array}{rcl} P\cos\theta & + & iQ\sin\theta \\ iP\sin\theta & + & Q\cos\theta \end{array} \tag{3.15}$$

if $P^\dagger Q - Q^\dagger P = 0$ (i.e., $(P^\dagger Q)^2 = +\mathbf{1}$); and by

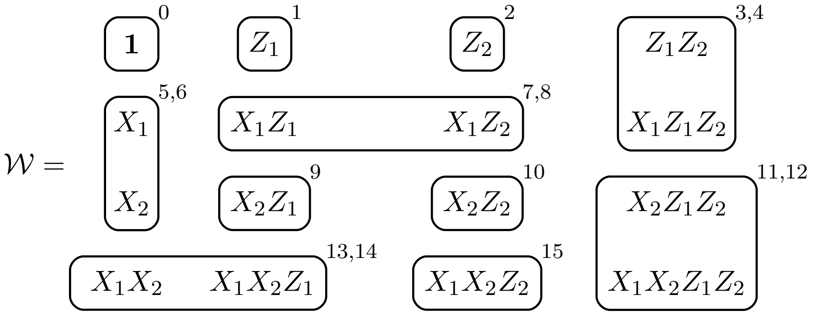
$$\begin{array}{rcl} P\cos\theta & + & Q\sin\theta \\ -P\sin\theta & + & Q\cos\theta \end{array} \tag{3.16}$$

if $P^\dagger Q + Q^\dagger P = 0$ (i.e., $(P^\dagger Q)^2 = -\mathbf{1}$). Here θ, in (3.15) and (3.16), is an arbitrary free parameter.[29] This clearly gives us a lot of freedom in our choice of

[29] Each of these transformations is a unitary transformation of the P_i, in the sense of (3.14); so the

bases: We can partition $\mathcal{P}$ into pairs in any way we like and mix the operators in each two-element set by arbitrary and unrelated angles θ while maintaining the orthogonality and unitarity of the elements.

Using these results we demonstrate a choice of basis for which the only bilocal stabilizer elements are trivial:



(3.17)

In this graphical representation of the new Bell basis, ovals are drawn around pairs of elements which are to be mixed according to the appropriate mixing equations (3.15), (3.16) above; thus, for instance, $W_3 = Z_1Z_2\cos\theta + iX_1Z_1Z_2\sin\theta$ and $W_4 = iZ_1Z_2\sin\theta + X_1Z_1Z_2\cos\theta$, since $\big((Z_1Z_2)^\dagger(X_1Z_1Z_2)\big)^2 = +\mathbf{1}$. The mixing angles are arbitrary (except that of course they should not be multiples of $\frac{\pi}{2}$). (Ovals enclosing a single element indicate, of course, that that element is not modified: e.g., $W_1 = Z_1$.)

Now we can follow a twisted path through this Bell basis to show that any bilocal stabilizer S for the basis $\mathcal{W}$ acts only as a global phase. First, since $W_0 = \mathbf{1} \in \mathcal{W}$ (this Bell basis is normal), Lemma 5 tells us that any bilocal stabilizer has the form $S = e^{i\chi_0}U \otimes U^*$. As usual we will assume WLOG that $\chi_0 = 0$. Now the stabilization criterion (3.11) applies, and $UW_i = e^{i\chi_i}W_iU$ for all the $W_i \in \mathcal{W}$.

Since both $W_1 = Z_1$ and $W_2 = Z_2$ are stabilized by U, so is Z_1Z_2 (by Lemma 6).

orthogonality of these operators, with each other and with the rest of the P_i, is immediate in both cases. The unitarity of these operators is what depends on these "commutation relations" of P and Q. For higher-order Pauli groups, the same sort of relation will probably hold, but the linear combinations need to have more elements in order to satisfy the unitarity constraints.

But now since $W_3 = \alpha_3 Z_1 Z_2 + \beta_3 X_1 Z_1 Z_2$ is also stabilized by U (where α_3 and β_3 are the nonzero mixing coefficients, here given by (3.15)), Lemma 7 implies that *both* $Z_1 Z_2$ and $X_1 Z_1 Z_2$ are stabilized by U, *with the same eigenvalue.* Applying Lemma 6 twice, we find that $X_1 = (X_1 Z_1 Z_2)(Z_1 Z_2)^{-1}$ is also stabilized by U, with eigenvalue 1.

Now since both X_1 and $W_5 = \alpha_5 X_1 + \beta_5 X_2$ are stabilized by U, Lemma 7 shows that X_1 and X_2 are both stabilized by U, with the same eigenvalue. Hence X_2 also has eigenvalue 1. Next, considering $W_{13} = \alpha_{13} X_1 X_2 + \beta_{13} X_1 X_2 Z_1$ and using an argument analogous to that with W_3 above, we find that Z_1 is also stabilized by U with eigenvalue 1. Finally, since $X_1 Z_1$ is stabilized by U with eigenvalue 1, the eigenstate $W_7 = \alpha_7 X_1 Z_1 + \beta_7 X_1 Z_2$ shows us that $X_1 Z_2$ also has eigenvalue 1, and hence Z_2 has eigenvalue 1 as well.

So X_1, X_2, Z_1, and Z_2 all are stabilized by U with eigenvalue 1, and hence every element in $\mathcal{P}$ is stabilized by U with eigenvalue 1. Since $\mathcal{P}$ forms a basis for $GL(4, \mathbb{C})$, all linear operators are stabilized by U with eigenvalue 1. Thus by Schur's lemma U must be only a phase, and the stabilizer $S = U \otimes U^* = \mathbf{1}$ is trivial. ■

For this Bell basis, then, we can't use the "recursive stabilizer" idea, because we can't even find a single nontrivial stabilizer to partition the eigenstates.

3.2.4 DiVincenzo's conjecture

The above examples of causal operators have made a progression away from the set of localizable operators; each has been more complicated to simulate even 1-localizably. What of DiVincenzo's conjecture, then, that not only all causal operators but all *1-causal* operators are 1-localizable?

For the restricted class of operators discussed here (the complete von Neumann measurement superoperators), the conjecture holds true. Even the superoperator of Example 3, for which there are no nontrivial bilocal stabilizer operators, can be simulated 1-localizably.

Theorem 5 *All 1-causal complete von Neumann measurement superoperators (allowing no communication from Bob to Alice) can be simulated 1-localizably (with a one-way quantum channel from Alice to Bob).*

Since of course all 1-localizable superoperators are 1-causal, this means that these two classes of superoperators are equivalent.

To prove this, we will display a protocol for 1-localizably simulating any 1-causal superoperator of the form (3.2). The idea behind the protocol is that since all full-rank Bell states are *locally* (not just bilocally) interconvertible, Alice can send to Bob her half of the original system, $\mathcal{H}_{\text{A}}$, along with $\mathcal{H}_{\text{S}}$, containing half of a newly-minted Bell state $|\Phi_d^+\rangle_{\text{RS}}$; then Bob, having all of $\mathcal{H}_{\text{AB}}$, can perform the measurement onto the basis of eigenstates of $\$$ and locally (via some unitary operator on $\mathcal{H}_{\text{S}}$) rotate the new Bell state into the correct basis state, to match the result of his projective measurement. For the more general case of Theorem 1, in which Alice's Hilbert space is partitioned into disjoint subspaces spanned by Bell subbases, Alice first does a partial measurement to determine which subspace the system is in (and hence which Bell subbasis is needed); she then sends to Bob her partially-projected system and an appropriate Bell state.

The complete proof is little more than a rewriting of this argument:

Proof of Theorem 5:

Let $\$$ be a 1-causal complete von Neumann measurement superoperator, allowing no signal to be sent from Bob to Alice. We wish to demonstrate a 1-localizable protocol (allowing quantum communication from Alice to Bob) by which $\$$ can be simulated.

By Theorem 1, we know that the eigenstates[30] $|\phi_{ai}\rangle$ are rank-d_a Bell states on orthogonal subspaces $E_a \times \mathcal{H}_{\text{B}}$; that is, $\text{tr}_{\text{B}} |\phi_{ai}\rangle\langle\phi_{ai}| = \frac{1}{d_a} E_a$, where $E_a E_b = \delta_{ab} E_b$ and $\sum_a E_a = \mathbf{1}$. Since the Bell state $|\phi_{ai}\rangle$ has support entirely on the subspace $E_a \times \mathbf{1}$, we may write (where $E_{ai} \equiv |\phi_{ai}\rangle\langle\phi_{ai}|$) $E_{ai} E_a = E_{ai} = E_a E_{ai}$.

[30] Here we have given the eigenstates two indices. The first index a explicitly labels the subspace of $\mathcal{H}_{\text{A}}$ on which the projector has its support; the second runs over all of the states with support on this subspace.

Now by Lemma 5, we can write[31]

$$|\phi_{ai}\rangle_{\mathrm{A}}{}^{\mathrm{B}} = \mathbf{1} \otimes W_{ai}\, |\Phi^a\rangle_{\mathrm{A}}{}^{\mathrm{B}} \,,$$

where $|\Phi^a\rangle_{\mathrm{A}}{}^{\mathrm{B}}$ is a particular fixed Bell state on $E_a \times \mathcal{H}^{\mathrm{B}}$ (say,

$$|\Phi^a\rangle_{\mathrm{A}}{}^{\mathrm{B}} \equiv \tfrac{1}{\sqrt{d_a}} \sum_{i=0}^{d_a-1} |a,i\rangle_{\mathrm{A}}\, |a,i\rangle^{\mathrm{B}} \,,$$

where $\{|a,i\rangle_{\mathrm{A}}\}$ is an orthonormal basis for E_a and $\{|a,i\rangle^{\mathrm{B}}\}$ is a set of d_a orthonormal states in $\mathcal{H}^{\mathrm{B}}$).

The protocol is as follows:

$$\rho_{\mathrm{A}}{}^{\mathrm{B}}$$

$\downarrow$ Alice performs the partial (local projective) measurement defined by the projectors $\{E_a\}$, storing the (classical) measurement result in a classical ancilla system $\mathcal{H}_{\mathrm{M}}$.

$$\sum_a (E_a)_{\mathrm{A}} \rho_{\mathrm{A}}{}^{\mathrm{B}} (E_a)_{\mathrm{A}} \otimes |a\rangle_{\mathrm{M}}{}_{\mathrm{M}}\langle a|$$

$\downarrow$ Alice uses her measurement result to create the state $|\Phi^a\rangle_{\mathrm{RS}}$ on an additional bipartite Hilbert space $\mathcal{H}_{\mathrm{RS}}$ (a copy of the original Hilbert space $\mathcal{H}_{\mathrm{A}}{}^{\mathrm{B}}$, so that $\dim \mathcal{H}_{\mathrm{R}} = \dim \mathcal{H}_{\mathrm{A}}$ and $\dim \mathcal{H}_{\mathrm{S}} = \dim \mathcal{H}^{\mathrm{B}}$), and swaps the states of systems A and R.

$$\sum_a \left[(E_a)_{\mathrm{R}} \rho_{\mathrm{R}}{}^{\mathrm{B}} (E_a)_{\mathrm{R}} \right] \otimes |a\rangle_{\mathrm{M}}{}_{\mathrm{M}}\langle a| \otimes |\Phi^a\rangle_{\mathrm{AS}}{}_{\mathrm{AS}}\langle \Phi^a|$$

$\downarrow$ Alice sends her measurement result (system $\mathcal{H}_{\mathrm{M}}$), her half of the partially-measured original system (now in $\mathcal{H}_{\mathrm{R}}$), and Bob's half of the new-minted Bell state (in $\mathcal{H}_{\mathrm{S}}$), to Bob.

$$\sum_a (E_a)^{\mathrm{R}} \rho^{\mathrm{RB}} (E_a)^{\mathrm{R}} \otimes |a\rangle^{\mathrm{M}}{}^{\mathrm{M}}\langle a| \otimes |\Phi^a\rangle_{\mathrm{A}}{}^{\mathrm{S}}{}_{\mathrm{A}}{}^{\mathrm{S}}\langle \Phi^a|$$

[31] Throughout the description of this protocol, the indices for systems which Alice holds will be subscripted, while the indices for systems Bob holds will be superscripted.

$$\sum_a (E_a)^{\rm R}\rho^{\rm RB}(E_a)^{\rm R}\otimes|a\rangle^{\rm M}\,{}^{\rm M}\langle a|\otimes|\Phi^a\rangle_{\rm A}^{\;\rm S}\,{}_{\rm A}^{\rm S}\langle\Phi^a|$$

$\Big\downarrow$ Bob performs the *complete* von Neumann measurement $\{(E_{bi})^{\rm RB}\}$, storing the measurement results (b,i) in a second ancilla system $\mathcal{H}^{\rm N}$.

$$\begin{aligned}&\sum_{abi}\left(E_{bi}{}^{\rm RB}E_a{}^{\rm R}\rho^{\rm RB}E_a{}^{\rm R}E_{bi}{}^{\rm RB}\right)\otimes|a\rangle^{\rm M}\,{}^{\rm M}\langle a|\otimes|b,i\rangle^{\rm N}\,{}^{\rm N}\langle b,i|\otimes|\Phi^a\rangle_{\rm A}^{\;\rm S}\,{}_{\rm A}^{\rm S}\langle\Phi^a|\\ &\quad=\sum_{ai}\left(E_{ai}{}^{\rm RB}\rho^{\rm RB}E_{ai}{}^{\rm RB}\right)\otimes|a\rangle^{\rm M}\,{}^{\rm M}\langle a|\otimes|a,i\rangle^{\rm N}\,{}^{\rm N}\langle a,i|\otimes|\Phi^a\rangle_{\rm A}^{\;\rm S}\,{}_{\rm A}^{\rm S}\langle\Phi^a|\end{aligned}$$

(since $E_{bi}{}^{\rm RB}E_a{}^{\rm R}=\delta_{ab}E_{bi}{}^{\rm RB}$)

$\Big\downarrow$ Bob performs the local unitary rotation W_{ai} on $\mathcal{H}^{\rm S}$ (conditioned on measurement result $|a,i\rangle^{\rm N}$), and swaps $\mathcal{H}^{\rm B}$ and $\mathcal{H}^{\rm S}$.

$$\begin{aligned}&\sum_{ai}(W_{ai})^{\rm B}\,|\Phi^a\rangle_{\rm A}^{\;\rm B}\,{}_{\rm A}^{\rm B}\langle\Phi^a|\,(W_{ai}^\dagger)^{\rm B}\otimes(E_{ai}{}^{\rm RS}\rho^{\rm RS}E_{ai}{}^{\rm RS})\otimes|a\rangle^{\rm M}\,{}^{\rm M}\langle a|\otimes|a,i\rangle^{\rm N}\,{}^{\rm N}\langle a,i|\\ &\quad=\sum_{ai}|\phi_{ai}\rangle_{\rm A}^{\;\rm B}\,{}_{\rm A}^{\rm B}\langle\phi_{ai}|\otimes(E_{ai}{}^{\rm RS}\rho^{\rm RS}E_{ai}{}^{\rm RS})\otimes|a\rangle^{\rm M}\,{}^{\rm M}\langle a|\otimes|a,i\rangle^{\rm N}\,{}^{\rm N}\langle a,i|\end{aligned}$$

$\Big\downarrow$ Bob discards the ancilla systems $\mathcal{H}^{\rm MNRS}$.

$$\begin{aligned}&\sum_{ai}\mathrm{tr}\left(E_{ai}\rho\right)\langle a\,|\,a\rangle\langle a,i\,|\,a,i\rangle\,|\phi_{ai}\rangle_{\rm A}^{\;\rm B}\,{}_{\rm A}^{\rm B}\langle\phi_{ai}|\\ &\quad=\sum_{ai}{}_{\rm A}^{\rm B}\langle\phi_{ai}|\,\rho\,|\phi_{ai}\rangle_{\rm A}^{\;\rm B}\;\;|\phi_{ai}\rangle_{\rm A}^{\;\rm B}\,{}_{\rm A}^{\rm B}\langle\phi_{ai}|\\ &\quad=\sum_{ai}(E_{ai})_{\rm A}^{\;\rm B}\rho(E_{ai})_{\rm A}^{\;\rm B}\end{aligned}$$

which is precisely the action of $\$$.

The ancilla systems $\mathcal{H}_{\rm M}$ and $\mathcal{H}_{\rm N}$ are not actually necessary for the protocol, but they make the argument a little easier to follow. ■

This is clearly a much more powerful protocol than the earlier classical-only protocols discussed; the simplicity of the protocol may make DiVincenzo's conjecture more plausible. The equivalence of local unitary operations on the two halves of a bipartite Bell state means that Bob can perform operations that it initially seems must be performed by Alice.

Bell bases in 2×2 dimensions

As a final aside to this section, we note that

Lemma 8 *In 2×2 dimensions, all (complete) Bell-basis measurement superoperators are localizable.*

Proof of Lemma 8:

As before, let us consider a set $\mathcal{W} \equiv \{W_i\}$ of elements of $\mathsf{SU}(2)$ such that the eigenstates of the superoperator $\$$ are $|\phi_i\rangle \equiv W_i \otimes \mathbf{1}\,|\Phi^+\rangle$. Now any element $U \in \mathsf{SU}(2)$ can be written (uniquely) as $U = \hat{v}_U \cdot \vec{P}$, where $\vec{P} \equiv (P_\beta) \equiv (P_0, \boldsymbol{P}) \equiv (\mathbf{1}, i\sigma_x, i\sigma_y, i\sigma_z)$ is the vector of $\mathsf{SU}(2)$ Pauli matrices and $\hat{v}_U$ is a unit vector in $\mathbb{R}^4$. Matrices $U, V \in \mathsf{SU}(2)$ can be seen to be orthogonal in the Hilbert-Schmidt norm ($\operatorname{tr} U^\dagger V = 0$) iff the corresponding vectors are orthogonal in Euclidean $\mathbb{R}^4$, $\hat{v}_U \cdot \hat{v}_V = 0$. Hence any Bell basis W_α in 2×2 dimensions can be derived by a (real) orthogonal transformation $M = (M_{\alpha\beta}) \in \mathsf{SO}(4)$ from the Pauli basis $\{P_\beta\}$:

$$W_\alpha = \sum_\beta M_{\alpha\beta} P_\beta \ ,$$

and any such M defines a Bell basis.

Let us characterize the set $\mathcal{W}$. First, let us use our basis freedom $\mathcal{W} \mapsto \mathcal{W}U$ to ensure that $\mathbf{1} \equiv W_0 \in \mathcal{W}$. Now we can (by permuting the rows of M if necessary) consider $\mathbf{1}$ to be preserved by the orthogonal transformation M. Hence M is of the form $M = 1 \oplus M'$, where $M' \in \mathsf{SO}(3)$, and the three W_a are of the form $W_a = \boldsymbol{m}_a \cdot \boldsymbol{P}$ (where the $\boldsymbol{m}_a$ are the three orthonormal row vectors of M').

The set $\mathcal{W}$ now can be seen to be isomorphic to the projective group $\{P_\beta\}$ (to form a group we must take $\{\pm P_\beta\} \sim Q_8$, the quaternionic group): We have (for $a, b \in \{1,2,3\}$, $a \neq b$) $W_a^\dagger = -W_a$, $W_a W_a = -\mathbf{1}$, and

$$\begin{aligned}
W_b W_a &= \sum_{ij} m_{bi} m_{aj} P_i P_j &&= -\sum_{ij} m_{bi} m_{aj} (\delta_{ij}\mathbf{1} + \varepsilon_{ijk} P_k) \\
&= -(\boldsymbol{m}_a \cdot \boldsymbol{m}_b + \boldsymbol{m}_b \times \boldsymbol{m}_a \cdot \boldsymbol{P}) &&= \boldsymbol{m}_a \times \boldsymbol{m}_b \cdot \boldsymbol{P} \\
&= \sum_c \varepsilon_{abc} \boldsymbol{m}_c \cdot \boldsymbol{P} &&= \sum_c \varepsilon_{abc} W_c \ .
\end{aligned}$$

We see that the case of 2×2 dimensions is special because there are only three nontrivial basis states, any two of them determining the third up to a sign.[32] This is precisely the group formed by the $\{P_\beta\}$; hence we can use the method of (3.5),

$$\$(\rho) = \tfrac{1}{4}\sum_\alpha (W_\alpha \otimes W_\alpha^*)\rho(W_\alpha \otimes W_\alpha^*)^\dagger \ ,$$

to implement the decoherence superoperator for this Bell basis. ■

3.3 Unitary superoperators

Another special case, that of unitary superoperators $\$(\rho) = U\rho U^\dagger$, turns out to be trivial; every causal unitary superoperator is in fact a tensor product of local unitary superoperators (and so trivially localizable):

Theorem 6 (Nielsen [12]) *A unitary operator U_{AB} is causal iff it is a tensor product operator, $U_{\mathrm{AB}} = U_{\mathrm{A}} \otimes U_{\mathrm{B}}$.*

Proof of Theorem 6:

The "if" direction is trivial. Now suppose that U_{AB} is not a tensor product operator on $\mathcal{H}_{\mathrm{A}} \otimes \mathcal{H}_{\mathrm{B}}$. Then the Schmidt decomposition for U_{AB},

$$\begin{array}{ll} U_{\mathrm{AB}} \equiv \sum_i \zeta_i A_i \otimes B_i \ , & \mathrm{tr}(A_i^\dagger A_j) = d_{\mathrm{A}}\delta_{ij} \ , \quad d_{\mathrm{A}} \equiv \dim \mathcal{H}_{\mathrm{A}} \\ (\text{all } \zeta_i > 0) & \mathrm{tr}(B_i^\dagger B_j) = d_{\mathrm{B}}\delta_{ij} \ , \quad d_{\mathrm{B}} \equiv \dim \mathcal{H}_{\mathrm{B}} \ , \end{array}$$

has at least two terms.

We show that Bob may send a signal to Alice. Let Alice introduce an ancilla system $\mathcal{H}_{\mathrm{R}}$ (of dimension d_{A}), and let the initial state be

$$|\Psi\rangle_{\mathrm{ARB}} \equiv \frac{1}{\sqrt{d_{\mathrm{A}}}} \sum_{i=1}^{d_{\mathrm{A}}} |i\rangle_{\mathrm{A}}\, |i\rangle_{\mathrm{R}}\, |\psi\rangle_{\mathrm{B}} \equiv |\Phi^+_{d_{\mathrm{A}}}\rangle_{\mathrm{AR}}\, |\psi\rangle_{\mathrm{B}} \ .$$

After action of U_{AB} on $|\Psi\rangle_{\mathrm{ARB}}$, Alice's density matrix (on $\mathcal{H}_{\mathrm{AR}}$) is

$$\begin{aligned} \sigma(|\psi\rangle) &\equiv \mathrm{tr}_{\mathrm{B}}\left[U_{\mathrm{AB}}\, |\Psi\rangle_{\mathrm{ARB}\ \mathrm{ARB}}\langle\Psi|\, (U^\dagger)_{\mathrm{AB}}\right] \\ &= \sum_{ab} \zeta_a \zeta_b A_a\, |\Phi^+_{d_{\mathrm{A}}}\rangle_{\mathrm{AR}\ \mathrm{AR}}\langle\Phi^+_{d_{\mathrm{A}}}|\, A_b^\dagger\ {}_{\mathrm{B}}\langle\psi|\, B_b^\dagger B_a\, |\psi\rangle_{\mathrm{B}} \ . \end{aligned}$$

[32] In the proof we have assumed WLOG that $(\boldsymbol{m}_1, \boldsymbol{m}_2, \boldsymbol{m}_3)$ form a right-handed coordinate system (otherwise we could just rename $W_2 \leftrightarrow W_3$ to get one).

The $\{A_a \,|\Phi^+_{d_\mathrm{A}}\rangle_\mathrm{AR}\}$ are orthonormal states:

$$\begin{aligned}{}_\mathrm{AR}\langle\Phi^+_{d_\mathrm{A}}|\, A_b^\dagger A_a \,|\Phi^+_{d_\mathrm{A}}\rangle_\mathrm{AR} &= \frac{1}{d_\mathrm{A}}\sum_{ij} {}_\mathrm{A}\langle j|\, A_b^\dagger A_a \,|i\rangle_\mathrm{AR}\langle j\,|\,i\rangle_\mathrm{R} \\ &= \frac{1}{d_\mathrm{A}}\sum_{i} {}_\mathrm{A}\langle i|\, A_b^\dagger A_a \,|i\rangle_\mathrm{A} = \frac{1}{d_\mathrm{A}}\,\mathrm{tr}\, A_b^\dagger A_a = \delta_{ab}\ .\end{aligned}$$

Case I: There is some B_i not unitary.

Then B_i has two distinct singular values, and $B_i^\dagger B_i$ has two distinct eigenvalues, λ_0 and λ_1. Let $|\psi_0\rangle_\mathrm{B}$ and $|\psi_1\rangle_\mathrm{B}$ be eigenvectors of $B_i^\dagger B_i$ with these two distinct eigenvalues.

Case II: All B_i are unitary.

$\mathrm{tr}(B_i^\dagger B_j) = 0$ for $i \neq j$. Then since $B_i^\dagger B_j$ is unitary it must have two distinct eigenvalues, λ_0 and λ_1. Let $|\psi_0\rangle_\mathrm{B}$ and $|\psi_1\rangle_\mathrm{B}$ be eigenvectors of $B_i^\dagger B_j$ with these two distinct eigenvalues.

Bob can prepare either $|\psi_0\rangle_\mathrm{B}$ or $|\psi_1\rangle_\mathrm{B}$ (clearly he can locally rotate between these two states), and this affects Alice's density matrix $\sigma(|\psi_n\rangle)$, and in particular the matrix element between $A_i\,|\Phi^+_{d_\mathrm{A}}\rangle_\mathrm{AR}$ and $A_j\,|\Phi^+_{d_\mathrm{A}}\rangle_\mathrm{AR}$:

$$\begin{aligned}{}_\mathrm{AR}\langle\Phi^+_{d_\mathrm{A}}|\, A_i^\dagger\sigma(|\psi_n\rangle)A_j \,|\Phi^+_{d_\mathrm{A}}\rangle_\mathrm{AR} &= \zeta_i\zeta_j\,{}_\mathrm{B}\langle\psi_n|\, B_j^\dagger B_i \,|\psi_n\rangle_\mathrm{B} \\ &= \zeta_i\zeta_j\lambda_n\end{aligned}$$

distinct for $n = 0, 1$. ■

3.4 More general superoperators

It remains to consider the case of more general superoperators. Let us consider only incomplete von Neumann measurement superoperators (i.e., those in which the Kraus operators, while still projectors, need not be one-dimensional). (Note that since a POVM may be implemented as a von Neumann measurement on an augmented Hilbert space, we may immediately say that if the POVM allows signalling, so must *any* implementation of the POVM as a von Neumann measurement. In this sense, allowing POVMs does not add any surprises.)

As an example of the new subtleties for this case, consider the three-outcome incomplete Bell measurement superoperator with

$$\begin{aligned} E_1 &= \left|\Phi^+\right\rangle\left\langle\Phi^+\right| \\ E_2 &= \left|\Psi^+\right\rangle\left\langle\Psi^+\right| + \left|\Psi^-\right\rangle\left\langle\Psi^-\right| \\ E_3 &= \left|\Phi^-\right\rangle\left\langle\Phi^-\right| \ ; \end{aligned}$$

$\mathrm{tr}_{\mathrm{A}}\, E_a \propto \mathbf{1}$ for all E_a, but signalling is possible (e.g., with initial state $|\psi\rangle = |00\rangle$ and Bob's local operator either $\mathbf{1}$ or $U = \sigma_x$), even though signalling was not possible in the case of the complete Bell measurement superoperator (which was in fact localizable—recall (3.5)). The difference is that this superoperator has another completion which (from Theorem 1) can be used for signalling:

$$\begin{aligned} E_{2,1} &= |01\rangle\langle 01| \\ E_{2,2} &= |10\rangle\langle 10| \ . \end{aligned}$$

It is true, though, that an incomplete measurement can be used for signalling only if it has a completion which can be used for signalling (and so satisfies the hypotheses of Theorem 1):

Theorem 7 *A von Neumann measurement operator $\$$ with (orthogonal) projectors E_a can be used for signalling from Bob to Alice only if there exists a completion $\widehat{\$}$ of $\$$ (i.e., a superoperator derived from a set of one-dimensional orthogonal projectors $\widehat{E}_{ai}$ with $E_a \equiv \sum_i \widehat{E}_{ai}$) which allows signalling from Bob to Alice, with the same initial state $|\psi\rangle_{\mathrm{AB}}$ and optional unitary operator U_{B}.*

Proof of Theorem 7:

Fix an initial state $|\psi\rangle_{\mathrm{AB}}$ and optional unitary operator U_{B}, from a signalling protocol for $\$$. Let us define $|0_a\rangle$, $|1_a\rangle$, the projections of the two possible initial states onto the measurement eigenspaces: $\sqrt{\alpha_{a0}}\,|0_a\rangle \equiv E_a\,|\psi\rangle$ and $\sqrt{\alpha_{a1}}\,|1_a\rangle \equiv E_a(\mathbf{1} \otimes U)\,|\psi\rangle$, where $|0_a\rangle$ and $|1_a\rangle$ are normalized states—but not necessarily orthogonal—and the $\alpha_{ai} \geq 0$. Further, let $\sigma_{a0} = \mathrm{tr}_{\mathrm{B}}\,|0_a\rangle\langle 0_a|$ and $\sigma_{a1} = \mathrm{tr}_{\mathrm{B}}\,|1_a\rangle\langle 1_a|$ be Alice's density matrices corresponding to these states. Hence Alice's density matrix is either $\sigma(\mathbf{1}) = \sum_a \alpha_{a0}\sigma_{a0}$ or $\sigma(U) = \sum_a \alpha_{a1}\sigma_{a1}$; these differ (so that Alice can distinguish them, and this protocol allows signalling) iff

$$\mathbf{0} \neq \sigma(\mathbf{1}) - \sigma(U) = \sum_a \delta\sigma_a \equiv \sum_a (\alpha_{a0}\sigma_{a0} - \alpha_{a1}\sigma_{a1}) \ . \tag{3.18}$$

If the subspace $E_a^{||} \equiv \Big\langle E_a \left|\psi\right\rangle, E_a U \left|\psi\right\rangle \Big\rangle$ is one-dimensional (i.e., $\alpha_{a0} = 0$, $\alpha_{a1} = 0$, or $|\langle 0_a \,|\, 1_a\rangle| = 1$), then we can trivially complete E_a to a set of one-dimensional projectors while leaving the superoperator's action on the states $\left|\psi\right\rangle$ and $\mathbf{1} \otimes U \left|\psi\right\rangle$ unchanged (we just take $\widehat{E}_{a1}$ to be the projector onto this one-dimensional space, and complete this basis in E_a). We call such an E_a a **degenerate** projector (though clearly this "degeneracy" depends also on the chosen protocol ($\left|\psi\right\rangle$, U)).

Suppose now that $E_a^{||}$ is two-dimensional (E_a a **nondegenerate** projector). Again, we may complete the orthogonal complement (in E_a) of $E_a^{||}$ however we please, without affecting the superoperator's action. So without loss of generality we may assume that the projectors of the given incomplete von Neumann measurement operator are all either one-dimensional, or two-dimensional and nondegenerate (again, relative to the chosen signalling protocol).

Let E be a nondegenerate two-dimensional projector; as before let $\sqrt{\alpha_0}\left|0\right\rangle \equiv E\left|\psi\right\rangle$ and $\sqrt{\alpha_1}\left|1\right\rangle \equiv EU\left|\psi\right\rangle$. We will show that either we can complete E to one-dimensional projectors $E \equiv \widehat{E}_a + \widehat{E}_b$ in such a way that the completion has no effect on Bob's density matrix, or we can complete E to one-dimensional projectors in two ways with *different effects* on Bob's ability to distinguish the states $\left|\psi\right\rangle$ and $U\left|\psi\right\rangle$. Then if all of the E_a are of the first type, we can complete all of them without affecting Bob's density matrices at all, and signalling is clearly still possible. If at least one of the E_a is of the second type, we can complete all of the other projectors however we want; then at least one of the two completions of E_a must allow Bob to distinguish the resulting density matrices $\widehat{\sigma}(\mathbf{1})$, $\widehat{\sigma}(U)$.

For convenience, we define two particular complete von Neumann measurements on this space: the *randomizing* measurement $E \equiv \widehat{E}_+ + \widehat{E}_-$ (providing no

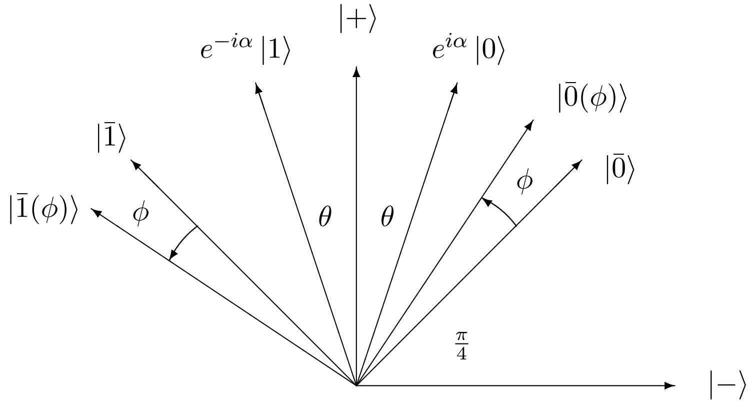


Figure 3.3: States used in the completions of E.

information distinguishing $\lvert 0_a\rangle$, and $\lvert 1_a\rangle$), where $\widehat{E}_\pm \equiv \lvert\pm\rangle\langle\pm\rvert$ and

$$\lvert \pm_a\rangle \equiv \frac{e^{i\alpha}\lvert 0_a\rangle \pm e^{-i\alpha}\lvert 1_a\rangle}{\left\| e^{i\alpha}\lvert 0_a\rangle \pm e^{-i\alpha}\lvert 1_a\rangle\right\|} \ , \quad \langle 0_a \,\vert\, 1_a\rangle \equiv e^{2i\alpha}\cos 2\theta \ ; \tag{3.19}$$

and the *optimal von Neumann* measurement $E \equiv \widehat{E}_{\bar{0}} + \widehat{E}_{\bar{1}}$ ($\widehat{E}_{\bar{a}} \equiv \lvert\bar{a}\rangle\langle\bar{a}\rvert$), where

$$\begin{array}{rcl} \lvert\bar{0}\rangle & \equiv & \frac{1}{\sqrt{2}}\big(\lvert+\rangle + \lvert-\rangle\big) \\ \lvert\bar{1}\rangle & \equiv & \frac{1}{\sqrt{2}}\big(\lvert+\rangle - \lvert-\rangle\big) \ . \end{array} \tag{3.20}$$

The phases have been chosen so that the states $e^{i\alpha}\lvert 0\rangle$, $e^{-i\alpha}\lvert 1\rangle$, $\lvert+\rangle$, $\lvert-\rangle$, $\lvert\bar{0}\rangle$, and $\lvert\bar{1}\rangle$ all lie in a *real* two-dimensional vector space (shown in Figure 3.3). Thus we can define a one-parameter family of unitary operators $V(\phi)$ on $\mathcal{H}_{\mathrm{AB}}$, acting by rotation within E and as the identity on $\mathbf{1} - E$. (Explicitly, take $V(\phi)$ to have eigenvectors $\lvert\bar{0}\rangle \pm i\,\lvert\bar{1}\rangle$ with eigenvalues $e^{\mp i\phi}$, and eigenvalue 1 on $\mathbf{1} - E$.)

We can now use $V(\phi)$ to define a family of complete measurements on the subspace E: We take $E \equiv \widehat{E}_{\bar{0}(\phi)} + \widehat{E}_{\bar{1}(\phi)}$, where

$$E_{\bar{a}(\phi)} \quad = \quad V(\phi) E_{\bar{a}} V(\phi)^\dagger \quad = \quad \lvert\bar{a}(\phi)\rangle\langle\bar{a}(\phi)\rvert \ , \tag{3.21}$$

the measurement whose projectors are rotated by ϕ from the projectors $\widehat{E}_{\bar{0}}$ and $\widehat{E}_{\bar{1}}$, as shown in Figure 3.3.

Consider, for this completed measurement superoperator, the difference $\widehat{\delta\sigma}(\phi)$

in Alice's density matrices, depending on whether Bob performed U:

$$\begin{aligned}
\widehat{\delta\sigma}(\phi) &\equiv \alpha_0\widehat{\sigma}_0(\phi) - \alpha_1\widehat{\sigma}_1(\phi) \\
&= \alpha_0 \operatorname{tr}_{\mathrm{B}}\Big[\cos^2\left(\tfrac{\pi}{4}-\theta-\phi\right)|\bar{0}(\phi)\rangle\langle\bar{0}(\phi)| + \sin^2\left(\tfrac{\pi}{4}-\theta-\phi\right)|\bar{1}(\phi)\rangle\langle\bar{1}(\phi)|\Big] \\
&\quad - \alpha_1 \operatorname{tr}_{\mathrm{B}}\Big[\cos^2\left(\tfrac{\pi}{4}+\theta-\phi\right)|\bar{0}(\phi)\rangle\langle\bar{0}(\phi)| + \sin^2\left(\tfrac{\pi}{4}+\theta-\phi\right)|\bar{1}(\phi)\rangle\langle\bar{1}(\phi)|\Big] \\
&= \tfrac{\alpha_0-\alpha_1}{2}\operatorname{tr}_{\mathrm{B}} E \\
&\qquad + \left[\tfrac{\alpha_0-\alpha_1}{2}\sin 2\phi\cos 2\theta + \tfrac{\alpha_0+\alpha_1}{2}\cos 2\phi \sin 2\theta\right] \\
&\qquad \cdot \left[\cos 2\phi \operatorname{tr}_{\mathrm{B}}\left(|\bar{0}\rangle\langle\bar{0}| - |\bar{1}\rangle\langle\bar{1}|\right) + \sin 2\phi \operatorname{tr}_{\mathrm{B}}\left(|+\rangle\langle+| - |-\rangle\langle-|\right)\right] \;.
\end{aligned}$$

If $\widehat{\delta\sigma}(\phi)$ is a nonconstant function of ϕ, then we can certainly choose *some* ϕ_a so that the sum $\widehat{\sigma}(\mathbf{1}) - \widehat{\sigma}(U) = \sum_a \widehat{\delta\sigma}_a(\phi_a)$ is nonzero, and Alice can determine whether Bob performed U.

Thus it remains only to consider the case in which all $\widehat{\delta\sigma}_a(\phi_a)$ are constant. The above equation gives three conditions for which $\widehat{\delta\sigma}(\phi)$ is constant:

I: $\cos 2\theta = 0$ and $\alpha_0 + \alpha_1 = 0$.

This case is immediately excluded since by assumption (E being nondegenerate) $\alpha_0 > 0$ and $\alpha_1 > 0$.

II: $\sin 2\theta = 0$ and $\alpha_0 - \alpha_1 = 0$.

This case is also excluded, since E is degenerate if $2\theta = n\pi$ ($n \in \mathbb{Z}$).

III: $\operatorname{tr}_{\mathrm{B}}\left[\,|\bar{0}\rangle\langle\bar{0}| - |\bar{1}\rangle\langle\bar{1}|\,\right] = \mathbf{0}$ and $\operatorname{tr}_{\mathrm{B}}\left[\,|+\rangle\langle+| - |-\rangle\langle-|\,\right] = \mathbf{0}$.

Only this final case remains; so suppose that

$$\mathbf{0} = \operatorname{tr}_{\mathrm{B}}\left[|+\rangle\langle+| - |-\rangle\langle-|\right] = \operatorname{tr}_{\mathrm{B}}\left[|\bar{0}\rangle\langle\bar{1}| + |\bar{1}\rangle\langle\bar{0}|\right]$$

and so Alice's density matrix, associated with a vector $|\bar{0}(\phi)\rangle$ in this subspace, is

$$\begin{aligned}
&\operatorname{tr}_{\mathrm{B}}\left[V(\phi)\,|\bar{0}\rangle\langle\bar{0}|\,V(\phi)^\dagger\right] \\
&\quad = \operatorname{tr}_{\mathrm{B}}\left[\cos^2\phi\,|\bar{0}\rangle\langle\bar{0}| + \sin^2\phi\,|\bar{1}\rangle\langle\bar{1}| + \cos\phi\sin\phi\left(|\bar{0}\rangle\langle\bar{1}| + |\bar{1}\rangle\langle\bar{0}|\right)\right] \\
&\quad = \operatorname{tr}_{\mathrm{B}}|\bar{0}\rangle\langle\bar{0}| \quad \equiv \quad \tau \;,
\end{aligned}$$

independent of ϕ. So for the incomplete measurement superoperator (with projector E), one of the terms in the Kraus expansion is

$$\begin{aligned}
\mathrm{tr}_{\mathrm{B}}\, E\,|\psi\rangle\langle\psi|\,E &\equiv \alpha_0\sigma_0 \equiv \alpha_0\,\mathrm{tr}_{\mathrm{B}}\,|0\rangle\langle 0| = \alpha_0\tau \\
\mathrm{tr}_{\mathrm{B}}\, EU\,|\psi\rangle\langle\psi|\,U^\dagger E &\equiv \alpha_1\sigma_1 \equiv \alpha_1\,\mathrm{tr}_{\mathrm{B}}\,|1\rangle\langle 1| = \alpha_1\tau \ ,
\end{aligned}$$

(depending on whether or not Bob performs U). Meanwhile for the completed measurement superoperator (with projectors $\widehat{E}_{\bar{0}(\phi)}$, $\widehat{E}_{\bar{1}(\phi)}$), this term in the Kraus expansion is replaced by the two terms

$$\begin{aligned}
&\mathrm{tr}_{\mathrm{B}}\left[\widehat{E}_{\bar{0}(\phi)}\,|\psi\rangle\langle\psi|\,\widehat{E}_{\bar{0}(\phi)} + \widehat{E}_{\bar{1}(\phi)}\,|\psi\rangle\langle\psi|\,\widehat{E}_{\bar{1}(\phi)}\right] \\
&\quad = \ \mathrm{tr}_{\mathrm{B}}\left[\widehat{E}_{\bar{0}(\phi)}\big(E\,|\psi\rangle\langle\psi|\,E\big)\widehat{E}_{\bar{0}(\phi)} + \widehat{E}_{\bar{1}(\phi)}\big(E\,|\psi\rangle\langle\psi|\,E\big)\widehat{E}_{\bar{1}(\phi)}\right] \\
&\quad = \ \alpha_0\,\mathrm{tr}_{\mathrm{B}}\left[\widehat{E}_{\bar{0}(\phi)}\,|0\rangle\langle 0|\,\widehat{E}_{\bar{0}(\phi)} + \widehat{E}_{\bar{1}(\phi)}\,|0\rangle\langle 0|\,\widehat{E}_{\bar{1}(\phi)}\right] \\
&\quad = \ \alpha_0\Big[\cos^2\left(\tfrac{\pi}{4}-\theta-\phi\right)\mathrm{tr}_{\mathrm{A}}\,|\bar{0}(\phi)\rangle\langle\bar{0}(\phi)| + \sin^2\left(\tfrac{\pi}{4}-\theta-\phi\right)\mathrm{tr}_{\mathrm{A}}\,|\bar{1}(\phi)\rangle\langle\bar{1}(\phi)|\,\Big] \\
&\quad = \ \alpha_0\tau \ , \quad \text{and similarly} \\
&\mathrm{tr}_{\mathrm{A}}\left[\widehat{E}_{\bar{0}(\phi)}U\,|\psi\rangle\langle\psi|\,U^\dagger\widehat{E}_{\bar{0}(\phi)} + \widehat{E}_{\bar{1}(\phi)}U\,|\psi\rangle\langle\psi|\,U^\dagger\widehat{E}_{\bar{1}(\phi)}\right] \\
&\quad = \ \alpha_1\tau \ .
\end{aligned}$$

So these terms in the sum are unchanged by any such completion of E.

Thus if $\widehat{\delta\sigma}_a(\phi_a)$ is constant for all (nondegenerate) E_a, we can complete the E_a using (say) the optimal von Neumann measurement without changing the values of $\sigma(\mathbf{1})$ or $\sigma(U)$; so the distinguishability of these states is preserved. ■

In this sense, there are no real surprises in considering incomplete measurements; such a measurement only allows signalling if it is part of a complete measurement scheme which does. But this condition is less satisfactory than the condition for the complete case: it is computationally more difficult to determine, and it is only a necessary condition.

3.5 Related work

In this paper we have concentrated mainly on proving results using the basic physical principle of causality—information cannot be transmitted outside of the forward light cone—for the restricted class of complete projective superoperators. For this case we were able to find interesting and intuitive results; this class of superoperators also provided enough generality to demonstrate the nonequivalence of causal and localizable superoperators (unlike the class of unitary superoperators, §3.3).

Several papers by Aharonov, Albert, Popescu, and Vaidman have studied the issues of causality and localizability for nonlocal measurements, with primary interest in *nondemolition measurements of a single given (multipartite, pure) state* $|\psi\rangle$: that is, for superoperators satisfying

$$\begin{aligned} \$\big(\, |\psi\rangle_{\mathrm{AB}\,\mathrm{AB}}\langle\psi|\,\big) &= |\psi\rangle_{\mathrm{AB}\,\mathrm{AB}}\langle\psi| \otimes |1\rangle_{\mathrm{RS}\,\mathrm{RS}}\langle 1| \\ \$\big(\, |\psi^{\perp}\rangle_{\mathrm{AB}\,\mathrm{AB}}\langle\psi^{\perp}|\,\big) &= |\tilde{\psi}^{\perp}\rangle_{\mathrm{AB}\,\mathrm{AB}}\langle\tilde{\psi}^{\perp}| \otimes |0\rangle_{\mathrm{RS}\,\mathrm{RS}}\langle 0| \ , \quad \text{for any } {}_{\mathrm{RS}}\langle\psi\,|\,\psi^{\perp}\rangle_{\mathrm{RS}} = 0 \ . \end{aligned}$$

Note that the superoperator is only required to be nondemolition for the distinguished state $|\psi\rangle$; we place no special restrictions on the states $|\tilde{\psi}^{\perp}\rangle$. The "measurement" is presumed to be nonlocal in time, as it is in this work, with prior state preparation extending arbitrarily far back in time and the readout of the measurement result extending arbitrarily far forward in time; only the quantum interaction between the system to be observed and the measurement apparatus is required to be constrained in time. The "measurement result," in system RS, may be inaccessible to local measurements by either Alice or Bob. Aharonov and Albert [5] investigated such measurements for the particular case of a pair of spin-$\frac{1}{2}$ systems; in [4] they considered measurements of the nonlocal observable $x_{\mathrm{B}} - x_{\mathrm{A}}$ (where x_{A} is the position operator on system A). Aharonov, Albert, and Vaidman [6] generalized the case of [5] to arbitrary finite-dimensional Hilbert spaces, obtaining a result for single-state nondemolition measurements analogous to Corollary 1. Popescu and Vaidman [87] consider more general measurement superoperators, again for the case of a pair of spin-$\frac{1}{2}$ systems, and prove Corollary 1 for this case.

The inequivalence of causal and localizable operations (Corollary 2) is intriguing. Such a class of forbidden operations hints at further physical principles or conservation laws, beyond causality, restricting the class of operations we may perform. Another class of restrictions is provided by Bell-type inequalities, which place general restrictions on expectation values of random variables arising from classical probability distributions or from quantum-mechanical amplitude distributions. Some such results are presented in [12, §VI], using the (classical) CHSH and (quantum-mechanical) Cirel'son inequalities to constrain the fidelity with which a particular causal operator can be localizably implemented.

The constructions of nonlocalizable superoperators in this paper, especially the construction (3.17) of Example 3, are reminiscent of the construction of the "domino basis" [18] of Bennett *et al.* Their prime consideration is the implementation of *measurements* on a bipartite Hilbert space, using only local operations and classical communication with no shared entanglement, rather than our "measurement superoperators" implemented using arbitrary prior entanglement. Nevertheless, the multipartite states considered in [18] are excellent examples of the sorts of states which we wish to consider in understanding the distinctions between causality and localizability. (Note, however, that their domino basis is not a semicausal measurement basis.)

The restriction to the class of complete measurement operators means that our proof of DiVincenzo's conjecture on the equivalence of 1-causal and 1-localizable operators (Theorem 5) also has restricted applicability. The general form of DiVincenzo's Conjecture (conjecture 2) has now been proved by Eggeling, Schlingemann, and Werner [40]. They also remark on some limitations of the model of quantum mechanics used here: for example, the tensor-product structure we have postulated breaks down at short distances.

Nielsen [12] has derived a formula for determining whether a general bipartite superoperator is causal: $\$_{\mathrm{AB}}$ is causal iff the mutual information, between systems $\mathcal{H}_{\mathrm{R}}$ and $\mathcal{H}_{\mathrm{S}}$, of the state

$$[\$_{\mathrm{AB}} \otimes \mathbf{1}_{\mathrm{RS}}]\left[\,|\Phi^+\rangle_{\mathrm{AR}\,\mathrm{AR}}\langle\Phi^+| \otimes |\Phi^+\rangle_{\mathrm{BS}\,\mathrm{BS}}\langle\Phi^+|\,\right]$$

is zero. This formula is computationally tractable, an advantage over the causality condition described here. It is not clear whether this approach provides similarly simple methods for determining (for example) localizability, however.

3.6 Discussion

As noted in [18], there are several different operations that can reasonably be called measurement superoperators in the bipartite case considered here. In the most powerful form of measurement, both Alice and Bob have local knowledge of the measurement results after the superoperator acts:

$$\$(\rho) \equiv \sum_a (E_a \rho E_a) \otimes |i\rangle_{\mathrm{R}\,\mathrm{R}}\langle i| \otimes |i\rangle_{\mathrm{S}\,\mathrm{S}}\langle i| \ , \tag{3.22}$$

where R and S are measurement records created by the superoperator and held by (respectively) Alice and Bob. A more general, and less powerful, form of measurement causes the measurement record to be split nonlocally, in a form which may be inaccessible to Alice or Bob alone but which is readable when Alice and Bob bring their measurement records together:

$$\$(\rho) \equiv \sum_a (E_a \rho E_a) \otimes |i\rangle_{\mathrm{RS}\,\mathrm{RS}}\langle i| \ . \tag{3.23}$$

The "von Neumann measurement superoperators" considered here (3.2) are a third form of measurement, in which neither Alice nor Bob have any information about the measurement results. However, it is clear that if Alice and Bob may implement such a superoperator then they may implement it by unitary operations, using the unitary implementation of an arbitrary superoperator, while saving their ancilla states rather than discarding them (as we have been assuming in taking the partial trace $\mathrm{tr}_{\mathrm{RS}}[\cdots]$). Thus in fact they may implement it in such a way as to preserve the measurement results, as in (3.23), and this form of measurement is equivalent to the form we have considered. The first form (3.22) is, however, strictly more powerful than this—and rather unrealistic in the situation considered here: even the superoperator projecting on the computational basis $\{|i\rangle_{\mathrm{A}}\,|j\rangle_{\mathrm{B}}\}$ allows signalling between Alice and Bob, since Alice can locally modify i while Bob can locally modify j, and both Alice and Bob learn the pair (i,j) as a result of the measurement. Because it is a more powerful form of measurement, of course, the superoperators which we have ruled out on grounds of causality immediately rule out the corresponding measurements of the first form (3.22).

Another way of weakening the measurement is to consider measurements which need not leave the measurement subspaces undisturbed; these have been considered by Groisman and Reznik [54], for example. For this destructive measurement procedure to be theoretically interesting, however, it must be restricted somewhat; otherwise *arbitrary* measurements can easily be performed in this formalism by the "exchange measurement" in which Alice and Bob merely exchange their subsystems with their ancilla systems, transferring all information about the state into the ancillas (and placing the original system $\mathcal{H}_{\mathrm{AB}}$ in a fixed state which is independent of the original state). When these ancillas are brought together, an arbitrary measurement can trivially be performed. (This type of measurement can be ruled out, for example, by requiring that the measurement records be classical: Alice and Bob must measure their ancilla states, in some basis, at the close of their measurement protocol.)

The related questions in the previous section have given some indications of opportunities for related research. In particular, the relation between causal and localizable operators (i.e., the gap between operations allowed by special relativity and those allowed by quantum mechanics) would be interesting to explore further. We have shown that not all causal operators are localizable, and thus that the rules of quantum mechanics provide stricter requirements on the local evolution of a state than those provided by causality. The Bell-type inequalities provide one such source of stricter requirements; it would be interesting to know whether this type of additional constraint is sufficient to close the gap between the causal and the localizable superoperators, or whether some other physical limitations are required.

The questions asked here have some of the same flavor as questions asked in the field of quantum communication complexity, where we are concerned with the shared resources required to solve a particular problem. Here we have considered primarily the question of whether the quantum operation in question *could be* implemented. Another extension of this work would be to consider the corresponding complexity bounds, i.e., to ask more quantitatively *how much* of each given resource is required to implement an allowed operation.

Bibliography

[1] ACM. *12th Annual ACM Symposium on Theory of Computing*, 1980.

[2] Leonard M. Adleman, Jonathan Demarrais, and Ming-Deh A. Huang. Quantum computability. *SIAM Journal on Computing*, 26(5):1524–1540, 1997.

[3] Dorit Aharonov, 1997. Private communication.

[4] Yakir Aharonov and David Z. Albert. States and observables in relativistic quantum field theories. *Physical Review D*, 21(12):3316–3324, 1980.

[5] Yakir Aharonov and David Z. Albert. Can we make sense out of the measurement process in relativistic quantum mechanics? *Physical Review D*, 24(2):359–370, 1981.

[6] Yakir Aharonov, David Z. Albert, and Lev Vaidman. Measurement process in relativistic quantum theory. *Physical Review D*, 34(6):1805–1813, 1986.

[7] Huzihiro Araki and Mutsuo M. Yanase. Measurement of quantum mechanical operators. *Physical Review*, 120(2):622–626, 1960.

[8] William Aspray and Arthur Burks, editors. *Papers of John von Neumann on Computing and Computer Theory*, volume 12 of *Charles Babbage Institute reprint series for the History of Computing*. The MIT Press, 1987.

[9] László Babai. Local expansion of vertex-transitive graphs and random generation in finite groups. In *23rd Annual ACM Symposium on Theory of Computing*, pages 164–174. ACM, 1991.

[10] László Babai, D. Yu. Grigoryev, and David M. Mount. Isomorphism of graphs with bounded eigenvalue multiplicity. In *14th Annual ACM Symposium on Theory of Computing*, pages 310–324. ACM, 1982.

[11] David Beckman, Amalavoyal N. Chari, Srikrishna Devabhaktuni, and John Preskill. Efficient networks for quantum factoring. *Physical Review A*, 54:1034–1063, 1996. Available from the LANL preprint archive [71]: `quant-ph/9602016`.

[12] David Beckman, Daniel Gottesman, M. A. Nielsen, and John Preskill. Causal and localizable quantum operations. *Physical Review A*, 64:052309, 2001. Available from the LANL preprint archive [71]: `quant-ph/0102043`.

[13] Jacob D. Bekenstein. Energy cost of information transfer. *Physical Review Letters*, 46(10):623–626, 1981.

[14] Jacob D. Bekenstein. Entropy content and information flow in systems with limited energy. *Physical Review D*, 30(6):1669–1679, 1984.

[15] C. H. Bennett. Logical reversibility of computation. *IBM Journal of Research and Development*, 17:525–532, 1973.

[16] Charles H. Bennett. The thermodynamics of computation—a review. *International Journal of Theoretical Physics*, 21(12):905–940, 1982.

[17] Charles H. Bennett, Ethan Bernstein, Gilles Brassard, and Umesh Vazirani. Strengths and weaknesses of quantum computing. *SIAM Journal on Computing*, 26(5):1510–1523, 1997. Available from the LANL preprint archive [71]: `quant-ph/9701001`.

[18] Charles H. Bennett, David P. DiVincenzo, Christopher A. Fuchs, Tal Mor, Eric Rains, Peter W. Shor, John A. Smolin, and William K. Wootters. Quantum nonlocality without entanglement. *Physical Review A*, 59(2):1070–1091, 1999. Available from the LANL preprint archive [71]: `quant-ph/9804053`.

[19] Charles H. Bennett, David P. DiVincenzo, John A. Smolin, and William K. Wootters. Mixed state entanglement and quantum error correction. *Physical Re-*

view A, 54(5):3824–3851, 1996. Available from the LANL preprint archive [71]: `quant-ph/9604024`.

[20] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997.

[21] Rajendra Bhatia. *Matrix Analysis.* Springer, 1997.

[22] Eli Biham, Ofer Biham, David Biron, Markus Grassl, and Daniel A. Lidar. Grover's quantum search algorithm for an arbitrary initial amplitude distribution. 1999. Available from the LANL preprint archive [71]: `quant-ph/9807027`.

[23] David Biron, Ofer Biham, Eli Biham, Markus Grassl, and Daniel A. Lidar. Generalized Grover search algorithm for arbitrary initial amplitude distribution. In Williams [104], pages 140–147. Available from the LANL preprint archive [71]: `quant-ph/9801066`.

[24] Gilles Brassard and Peter Høyer. An exact quantum polynomial-time algorithm for Simon's problem. In *Fifth Israeli Symposium on Theory of Computing and Systems*, pages 12–23. IEEE, IEEE Computer Society Press, 1997. Available from the LANL preprint archive [71]: `quant-ph/9704027`.

[25] Gilles Brassard, Peter Høyer, Michele Mosca, and Alain Tapp. Quantum amplitude amplification and estimation. 2000. Available from the LANL preprint archive [71]: `quant-ph/0005055`.

[26] David M. Bressoud. *Factorization and Primality Testing.* Springer, 1989.

[27] Arthur Burks, editor. *Theory of Self-Reproducing Automata.* University of Illinois Press, 1966.

[28] Nicolas J. Cerf, Lov K. Grover, and Colin P. Williams. Nested quantum search and NP-complete problems. *Physical Review A*, 61:032303, 2000. Available from the LANL preprint archive [71]: `quant-ph/9806078`.

[29] Alonzo Church. An unsolvable problem of elementary number theory. *American Journal of Mathematics*, 58(2):345–363, 1936. Included in [33].

[30] The Quantum Computation Collective. What makes quantum computers powerful? 1997. Available from John Preskill's Quantum Computation home page as `http://www.theory.caltech.edu/people/preskill/ph229/manifesto5.ps`.

[31] Derek Corneil and Mark Goldberg. A non-factorial algorithm for canonical numbering of a graph. *Journal of Algorithms*, 5:345–362, 1984.

[32] Martin Davis. *Computability and Unsolvability*. McGraw-Hill, 1958. Reprinted by Dover, 1982.

[33] Martin Davis, editor. *The Undecidable: Basic Papers on Undecidable Propositions,Unsolvable Problems and Computable Functions*. Raven Press, Hewlett, N.Y., 1965.

[34] Martin Davis. Hilbert's tenth problem is unsolvable. *American Mathematical Monthly*, 80:233–269, 1973. Explanation of result of Ju. V. Matijasevič.

[35] David Deutsch. Quantum theory, the Church-Turing principle, and the universal quantum computer. *Proceedings of the Royal Society of London A*, 400(1818):97–117, 1985.

[36] David Deutsch. Quantum computational networks. *Proceedings of the Royal Society of London A*, 425(1868):73–90, 1989.

[37] David Deutsch and Richard Jozsa. Rapid solution of problems by quantum computation. *Proceedings of the Royal Society of London A*, 439(1907):553–558, 1992.

[38] D. Dieks. Communication by EPR devices. *Physics Letters A*, 92(6):271–272, 1982.

[39] David P. DiVincenzo, 2000. Private communication.

[40] T. Eggeling, D. Schlingemann, and R. F. Werner. Semicausal operations are semilocalizable. 2001. Available from the LANL preprint archive [71]: `quant-ph/0104027`.

[41] A. Ekert, R. Jozsa, and R. Penrose, editors. *Proceedings of Royal Society Discussion Meeting "Quantum Computation: Theory and Experiment"*, volume 356(1743). The

Royal Society of London, Philosophical Transactions of the Royal Society (London) A, 1998.

[42] Artur Ekert and Richard Jozsa. Quantum algorithms: entanglement-enhanced information processing. In Ekert et al. [41]. Available from the LANL preprint archive [71]: `quant-ph/9803072`.

[43] Mark Ettinger and Peter Høyer. A quantum observable for the graph isomorphism problem. 1999. Available from the LANL preprint archive [71]: `quant-ph/9901029`.

[44] Edward Farhi, Jeffrey Goldstone, Sam Gutmann, and Michael Sipser. A limit on the speed of quantum computation in determining parity. *Physical Review Letters*, 81(24):5442–5444, 1998. Available from the LANL preprint archive [71]: `quant-ph/9802045`.

[45] Edward Farhi and Sam Gutmann. Quantum mechanical square root speedup in a structured search problem. 1997. Available from the LANL preprint archive [71]: `quant-ph/9711035`.

[46] Richard P. Feynman. Simulating physics with computers. *International Journal of Theoretical Physics*, 21(6/7):467–488, 1982.

[47] Richard P. Feynman. Quantum mechanical computers. *Foundations of Physics*, 16(6):507–531, 1986.

[48] I. S. Filotti and Jack N. Mayer. A polynomial-time algorithm for determining the isomorphism of graphs of fixed genus. In *12th Annual ACM Symposium on Theory of Computing* [1], pages 236–243.

[49] Edward Fredkin and Tommaso Toffoli. Conservative logic. *International Journal of Theoretical Physics*, 21(3/4):219–253, 1982.

[50] Michael R. Garey and David S. Johnson. *Computers and Intractability: A Guide to the Theory of* NP*-Completeness*. Freeman, 1979.

[51] G. C. Ghirardi, F. Miglietta, A. Rimini, and T. Weber. Limitations on quantum measurements (i; ii). *Physical Review D*, 24(2):347–352; 353–358, 1981.

[52] Daniel Gottesman. Theory of fault-tolerant quantum computation. *Physical Review A*, 57(1):127–137, 1998. Available from the LANL preprint archive [71]: quant-ph/9702029.

[53] Daniel Gottesman. Fault-tolerant quantum computation with higher-dimensional systems. In Williams [104], pages 302–313. Available from the LANL preprint archive [71]: quant-ph/9802007.

[54] Berry Groisman and Benni Reznik. Measurements of semi-local and non-maximally entangled states. *Physical Review A*, 66:022110, 2002. Available from the LANL preprint archive [71]: quant-ph/0111012.

[55] Lov K. Grover. A fast quantum mechanical algorithm for database search. In *28th Annual ACM Symposium on Theory of Computing*, pages 212–219. ACM, 1996. Available from the LANL preprint archive [71]: quant-ph/9605043.

[56] Lov K. Grover. Quantum search on structured problems. In Williams [104], pages 126–139. Available from the LANL preprint archive [71]: quant-ph/9802035.

[57] David Hilbert. Mathematical problems. *Bulletin of the American Mathematical Society*, 8:437–479, 1901–1902.

[58] Peter Høyer. On arbitrary phases in quantum amplitude amplification. 2000. Available from the LANL preprint archive [71]: quant-ph/0006031.

[59] Richard Jozsa. Quantum effects in algorithms. In Williams [104], pages 103–112. Available from the LANL preprint archive [71]: quant-ph/9805086.

[60] Richard Jozsa. Searching in Grover's algorithm. 1999. Available from the LANL preprint archive [71]: quant-ph/9901021.

[61] A. Yu. Kitaev. Quantum measurements and the abelian stabilizer problem. 1995. Available from the LANL preprint archive [71]: quant-ph/9511026.

[62] A. Yu. Kitaev. Quantum computations: algorithms and error correction. *Russian Mathematical Surveys*, 52(6):1191–1249, 1997.

[63] A. Yu. Kitaev, 1998. Private communication.

[64] E. Knill. Non-binary unitary error bases and quantum codes. 1996. Available from the LANL preprint archive [71]: `quant-ph/9608048`.

[65] Emanuel Knill and Raymond Laflamme. A theory of quanum error-correcting codes. *Physical Review A*, 55(2):900–911, 1997. Available from the LANL preprint archive [71]: `quant-ph/9604034`.

[66] Johannes Köbler, Uwe Schöning, and Jacobo Torán. Graph isomorphism is low for PP. *Journal of Computational Complexity*, 2(4):301–330, 1992. Available at `http://citeseer.nj.nec.com/obler92graph.html`.

[67] Johannes Köbler, Uwe Schöning, and Jacobo Torán. *The Graph Isomorphism Problem: Its Structural Complexity*. Birkhäuser, 1993.

[68] Luděk Kučera. *Combinatorial Algorithms*. Adam Hilger, 1990.

[69] R. Landauer. Irreversibility and heat generation in the computing process. *IBM Journal of Research and Development*, 5:183–191, 1961.

[70] R. Landauer. Uncertainty principle and minimal energy dissipation in the computer. *International Journal of Theoretical Physics*, 21(3/4):283–297, 1982.

[71] LANL arXiv: The Los Alamos National Laboratory e-print archive. At `http://xxx.lanl.gov/`.

[72] A. K. Lenstra, H. W. Lenstra, Jr., M. S. Manasse, and J. M. Pollard. The number field sieve. In *22rd Annual ACM Symposium on Theory of Computing*, pages 564–572. ACM, 1990. Available at `http://citeseer.ist.psu.edu/lenstra90number.html`.

[73] Harry R. Lewis and Christos H. Papadimitriou. *Elements of the Theory of Computation*. Prentice Hall, 1981.

[74] Seth Lloyd. A potentially realizable quantum computer. *Science*, 261:1569–1571, 1993.

[75] Seth Lloyd. Ultimate physical limits to computation. *Nature*, 406:1047–1054, 2000. Available from the LANL preprint archive [71]: `quant-ph/9908043`.

[76] Hoi-Kwong Lo and Sandu Popescu. Concentrating entanglement by local actions—beyond mean values. 1997. Available from the LANL preprint archive [71]: `quant-ph/9707038`.

[77] Eugene M. Luks. Isomorphism of graphs of bounded valence can be tested in polynomial time. *Journal of Computer and System Sciences*, 25:42–65, 1982.

[78] Ju. V. Matijasevič. Enumerable sets are Diophantine. *Soviet Mathematics—Doklady*, 11(2):354–358, 1970. Translation from the original Russian, in *Doklady Akademii Nauk SSSR*, 191(2):279–282, 1970.

[79] James D. Meindl, Qiang Chen, and Jeffrey A. Davis. Limits on silicon nanoelectronics for terascale integration. *Science*, 293:2044–2049, 2001.

[80] Gary Miller. Isomorphism testing for graphs of bounded genus. In *12th Annual ACM Symposium on Theory of Computing* [1], pages 225–235.

[81] Gary L. Miller. Graph isomorphism, general remarks. *Journal of Computer and System Sciences*, 18:128–142, 1979.

[82] M. A. Nielsen. Computable functions, quantum measurements, and quantum dynamics. *Physical Review Letters*, 79(15):2915–2918, 1997. Available from the LANL preprint archive [71]: `quant-ph/9706006`.

[83] Michael A. Nielsen and Isaac L. Chuang. *Quantum Computation and Quantum Information.* Cambridge University Press, 2000.

[84] Yuri Ozhigov. Quantum computer can not speed up iterated applications of a black box. In Williams [104], pages 152–159. Available from the LANL preprint archive [71]: `quant-ph/9712051`.

[85] Christos H. Papadimitriou. *Computational Complexity.* Addison Wesley, 1994.

[86] Asher Peres. *Quantum Theory: Concepts and Methods.* Kluwer Academic, 1995.

[87] Sandu Popescu and Lev Vaidman. Causality constraints on nonlocal quantum measurements. *Physical Review A*, 49(6):4331–4338, 1994.

[88] John Preskill. Quantum information and computation. 1997–1998. Lecture notes for Caltech Physics 229.

[89] Eric M. Rains. Entanglement purification via separable superoperators. 1997. Available from the LANL preprint archive [71]: `quant-ph/9707002`.

[90] Eric M. Rains. Nonbinary quantum codes. 1997. Available from the LANL preprint archive [71]: `quant-ph/9703048`.

[91] Peter W. Shor. Algorithms for quantum computation: Discrete logarithms and factoring. In *35th Annual Symposium on Foundations of Computer Science*, pages 124–134. IEEE, IEEE Computer Society Press, 1994. Available from the LANL preprint archive [71]: `quant-ph/9508027`.

[92] Peter W. Shor. Polynomial-time algorithms for prime factorization and discrete logarithms on a quantum computer. *SIAM Journal on Computing*, 26:1484–1509, 1997. Expanded version of [91]; Available from the LANL preprint archive [71]: `quant-ph/9508027`.

[93] Daniel R. Simon. On the power of quantum computation. *SIAM Journal on Computing*, 26(5):1474–1483, 1997.

[94] A. M. Steane. Introduction to quantum error correction. In Ekert et al. [41], pages 1739–1758.

[95] A. M. Steane. A quantum computer only needs one universe. In *Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics*, chapter 8, pages 469–478. Elsevier Science, 2003. Available from the LANL preprint archive [71]: `quant-ph/0003084`.

[96] Wu-Ki Tung. *Group Theory in Physics*. World Scientific, 1985.

[97] A. M. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings, London Mathematical Society, series 2*, 42:230–265, 1936. Corrections and clarifications in 43:544–546. Included in [33].

[98] Umesh Vazirani. On the power of quantum computation. In Ekert et al. [41], pages 1759–1768.

[99] V. Vedral and M. B. Plenio. Entanglement measures and purification procedures. *Physical Review A*, 57(3):1619–1633, 1998. Available from the LANL preprint archive [71]: `quant-ph/9707035`.

[100] John von Neumann. Computers and information. Collected in [8, Part IV, Ch. 11 (pp.433–490)] and in [27], 1949. Lectures at the University of Illinois.

[101] John Watrous. Succinct quantum proofs for properties of finite groups. In *41st Annual Symposium on Foundations of Computer Science*, pages 537–546. IEEE, IEEE Computer Society Press, 2000. Available from the LANL preprint archive [71]: `cs.CC/0009002`.

[102] Steven Weinberg. *The Quantum Theory of Fields; Volume I: Foundations.* Cambridge University Press, 1995.

[103] E. P. Wigner. Die Messung quantenmechanischer Operatoren. *Zeitschrift für Physik*, 133:101–108, 1952. Especially useful if you can read German.

[104] C. P. Williams, editor. *Quantum Computing and Quantum Communications: First NASA International Conference, QCQC'98*, volume 1509. Springer (Lecture Notes in Computer Science), 1999.

[105] W. K. Wootters and W. H. Zurek. A single quantum cannot be cloned. *Nature*, 299:802–803, 1982.