Nonlocality and communication complexity

Buhrman, H.; Cleve, R.; Massar, S.; de Wolf, R.

DOI
10.1103/RevModPhys.82.665
Publication date
2010
Document Version
Final published version
Published in
Reviews of Modern Physics

REVIEWS OF MODERN PHYSICS, VOLUME 82, JANUARY–MARCH 2010

Nonlocality and communication complexity

Harry Buhrman

Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands

Richard Cleve

Institute for Quantum Computing and School of Computer Science, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada N2L 3G1 and Perimeter Institute for Theoretical Physics, 31 Caroline Street North, Waterloo, Ontario, Canada N2L 2Y5

Serge Massar

Laboratoire d'Information Quantique, CP 225, Université Libre de Bruxelles, Boulevard du Triomphe, B-1050 Bruxelles, Belgium

Ronald de Wolf

Centrum Wiskunde & Informatica, Science Park 123, 1098 XG Amsterdam, The Netherlands

(Published 11 March 2010)

Quantum information processing is the emerging field that defines and realizes computing devices that make use of quantum mechanical principles such as the superposition principle, entanglement, and interference. Until recently the common notion of computing was based on classical mechanics and did not take into account all the possibilities that physically realizable computing devices offer in principle. The field gained momentum after Shor developed an efficient algorithm for factoring numbers, demonstrating the potential computing powers that quantum computing devices can unleash. In this review the information counterpart of computing is studied. It was realized early on by Holevo that quantum bits, the quantum mechanical counterpart of classical bits, cannot be used for efficient transformation of information in the sense that arbitrary k-bit messages cannot be compressed into messages of $k-1$ qubits. The abstract form of the distributed computing setting is called *communication complexity*. It studies the amount of information, in terms of bits or in our case qubits, that two spatially separated computing devices need to exchange in order to perform some computational task. Surprisingly, quantum mechanics can be used to obtain dramatic advantages for such tasks. The area of quantum communication complexity is reviewed and it is shown how it connects the foundational physics questions regarding nonlocality with those of communication complexity studied in theoretical computer science. The first examples exhibiting the advantage of the use of qubits in distributed information-processing tasks were based on nonlocality tests. However, by now the field has produced strong and interesting quantum protocols and algorithms of its own that demonstrate that entanglement, although it cannot be used to replace communication, can be used to reduce the communication exponentially. In turn, these new advances yield a new outlook on the foundations of physics and could even yield new proposals for experiments that test the foundations of physics.

DOI: 10.1103/RevModPhys.82.665 PACS number(s): 03.65.Ta, 03.65.Ud, 03.67.Ac, 03.67.Hk

CONTENTS

I. Introduction 666
A. Background 666
B. Communication complexity 667
C. Quantum nonlocality 667
D. Unity of quantum communication complexity and quantum nonlocality 668
E. Resources 668
F. Basic scenarios 669
G. Mappings between communication complexity and nonlocality 670
H. Summary of the review 670
II. Simple Nonlocality Examples 670
A. GHZ and Mermin 671
B. CHSH 672
C. Tsirelson's upper bound for CHSH 673
D. Magic square game 673
III. Communication Complexity 675
A. The setting 675
B. The quantum question 676
C. The first examples 676
D. Distributed Deutsch-Jozsa 676
E. The intersection problem 677
F. Raz's problem 678
G. The hidden matching problem 678
H. Inner product 679
IV. Nonlocality and Communication Complexity 679

A. Converting communication complexity to nonlocality 679
1. Mapping one-way quantum communication complexity to nonlocality 679
B. Nonlocal version of the distributed Deutsch-Jozsa problem 680
C. Nonlocal version of the hidden matching problem 680
V. Quantum Fingerprinting and the Simultaneous Message Passing Model 681
A. Quantum fingerprints 681
B. Classical protocol for equality 681
C. Quantum protocol for equality 681
D. Subsequent work in the SMP model 682
VI. Other Aspects of Quantum Nonlocality 682
A. Nonlocal boxes 682
B. Bell inequalities and Tsirelson bounds 684
C. Classical simulation of quantum correlations and quantum communication 685
1. When no communication is needed 685
2. One-way quantum communication 685
a. Simulation of one-way communication of a single qubit and subsequent projective measurement 685
b. Simulation of one-way communication of n qubits 686
3. Entanglement simulation 686
a. Simulation of projective measurements on a single ebit 686
b. Simulation of entangled states of dimension 2^n 686
4. Exact classical simulations 686
VII. Implementations 687
A. Inefficient detectors 687
1. The detection loophole 687
2. Communication complexity and the detection loophole 687
3. Asymmetric detection loophole 688
B. Present and future experiments 688
1. Experimental quantum nonlocality 688
C. Future nonlocality experiments 689
1. Experimental communication complexity 690
VIII. Conclusion 690
A. Open questions 690
B. What have we learned from quantum communication complexity? 691
Acknowledgments 691
Appendix A: Nayak's Proof of a Consequence of Holevo's Bound 692
Appendix B: Rectangles and the Lower Bound for Distributed Deutsch-Jozsa 692
1. Rectangles 692
2. Randomized protocols 693
3. Discrepancy of the inner product function 693
4. The lower bound for the distributed Deutsch-Jozsa problem 693
Appendix C: Razborov's Lower Bound for the Quantum Communication Complexity of Intersection 694
1. The Kremer-Razborov-Yao lemma and its consequences 694
2. Translation from protocols to polynomials 695
3. The quantum lower bound for intersection 695
Appendix D: Asymmetric Detection Efficiency 696
References 696

I. INTRODUCTION

A. Background

During the last decades of the 20th century it was realized that information processing at the quantum level could offer tremendous advantages over conventional "classical" information processing. Quantum information admits extremely efficient algorithms, such as Shor's factoring algorithm (Shor, 1997), and qualitatively superior cryptographic protocols, such as the BB84 key distribution protocol (Bennett and Brassard, 1984). Many other works contributed to put this field on solid foundations. Quantum error-correcting codes and fault-tolerant quantum computation showed that these ideas could in principle be realized experimentally. These codes, combined with Holevo's theorem, Schumacher compression, and entanglement distillation (which are analogs of Shannon's noiseless coding theorem), gave us the foundations of an information theory pertaining to quantum systems in terms of quantum bits, or *qubits*, and entanglement that is measured (in the bipartite case) in entanglement bits, or *ebits*. These discoveries generated much excitement. By now quantum information has become a well-established field, and there are many reviews and textbooks to which we refer the reader for background information [see, for example, Nielsen and Chuang (2000)].

In view of the advantages that quantum information offers for computation and cryptography, it is natural to enquire whether quantum information is also a superior medium for efficient communication. In this article we review progress on this specific question and its relation to the problem of quantum nonlocality, which has fascinated physicists for decades.

On the face of it, there are important reasons for doubting that quantum information provides such a communication efficiency advantage. Many years before the "quantum information" discipline took hold on a large scale, Holevo (1973) proved an important theorem about the classical information capacity of quantum channels. Holevo's theorem—as it is now called—states that, for any classical message, the cost of transmitting it from one party (Alice) to another party (Bob) in terms of quantum bits (qubits) is the same as the cost of transmitting it in terms of classical bits. If the task requires k bits on average, then it also requires k qubits on average. The latter consequence of Holevo's theorem can be proven quite simply using a different approach (Nayak, 1999), and this proof is reproduced in Appendix A. Thus one would naively expect that quantum information cannot provide a communication efficiency advantage. This intuition turns out to be wrong. Tremendous communication savings are possible with the use of quantum information, as explained in the next section.

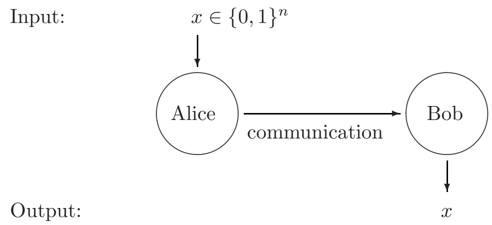

FIG. 1. The basic communication scenario: Alice receives an n-bit string x as input and sends one message to Bob, who must output x. For this task, a quantum message is no more efficient than a classical message.

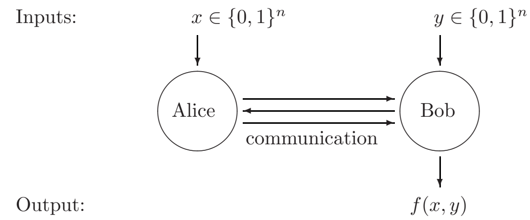

FIG. 2. The basic communication complexity scenario: Alice and Bob receive n-bit strings x and y, respectively, as input and their goal to compute some function of these values $f(x,y)$ as Bob's output. There are tasks of this form where communication in terms of quantum messages is much more efficient than communication in terms of classical messages. The number of qubits can be exponentially smaller than the number of bits. Note that in this framework we do not take into account the time and other resources that Alice and Bob spend locally (although in practice it turns out that their local computations are almost always efficient).

B. Communication complexity

To understand why quantum information can provide a communication advantage without contradicting Holevo's theorem, it is necessary to consider more precisely the various scenarios that can be associated with "communication." The simplest scenario, corresponding to the case covered by Holevo's theorem, is shown in Fig. 1.

There are two parties that we refer to as Alice and Bob. Alice has an n-bit string x that she would like to convey to Bob by sending one message. Here it is indeed true, by Holevo's theorem (Holevo, 1973), that quantum messages are no more efficient than classical messages. Alice must send n qubits to accomplish this specific task.

A variant of the communication scenario is where Bob's goal is not to determine Alice's data x but to determine some information that is a *function* of x in a way that may depend on other data y that resides with Bob (while y is unknown to Alice). Such a scenario could occur when Alice and Bob each begins with n-bit strings, x and y, respectively (Alice knows x but not y and Bob knows y but not x), and the goal is for Bob to determine the value of some function $f(x,y)$ (where f is known to both parties). An example where such a scenario could arise is where Alice and Bob are interested in scheduling an appointment. Alice's schedule could be represented by x and Bob's by y: if there are n time slots, then we can set the ith bit of x to 1 if Alice is available in time slot i and similarly for y. How much communication is required for Bob to find a time when they are both available (i.e., an i such that $x_i = y_i = 1$)? We shall see that, for this communication scenario, quantum information enables Alice and Bob to accomplish the task with less (asymptotically less in the number of time slots) qubit communication than would be required by any protocol that is restricted to classical bit communication.

This kind of scenario, shown in Fig. 2 (for general functions or relations f on $\{0,1\}^n \times \{0,1\}^n$), is known as *communication complexity*.

It has been extensively studied in the classical case. Indeed, whereas the trivial solution to this problem is for Alice to send Bob her input x and for Bob to compute $f(x,y)$, it is often possible for Bob to compute f with much less than n bits of classical communication. These savings in classical communication are interesting from both a practical and a conceptual point of view. Section III outlines several of the key results in the area, and we refer the reader to Hromkovič (1997) and Kushilevitz and Nisan, (1997) for further information.

When Alice and Bob can communicate qubits, further reductions in the amount of communication are possible, sometimes even exponential reductions. This situation is clearly worthy of study. It is one of the main subjects covered by the present review, and we will see many examples later.

C. Quantum nonlocality

Long before the work on quantum communication complexity mentioned, physicists investigating the foundations of quantum mechanics studied the scenario where local measurements are carried out on two entangled particles. Such entangled states can (at least in principle) be easily produced by having the particles interact together for some time and then sending the particles away to far-off locations. Local measurements are then carried out on the particles. This scenario was first studied by Einstein, Podolsky, and Rosen (1935) (EPR) and immediately afterwards by Schrödinger (1935, 1936) (who coined the word *entanglement*). In these works it was realized that the results of the local measurements would exhibit interesting correlations. For instance, for some pairs of the measurements, the results may always be the same; for other pairs of measurements, the results may always be opposite, etc.

Nevertheless, one can easily show—this follows immediately from the structure of quantum mechanics—that the parties carrying out the measurements cannot use the entangled particles to communicate to each other. More precisely, if two physically separated parties Alice and Bob initially possess entangled particles and then Alice is given an arbitrary bit x, there is no way for Alice to manipulate her particles in order to convey any information about x to Bob when he performs measurements on his particles.

Given that these correlations cannot be used for communication, one would naively expect that if a (quantum or classical) model can reproduce these correlations, then it is not necessary for that model to use communication. This is indeed the case in the quantum scenario where, having established the entanglement through some interaction in the past, no communication is needed at the time of the measurement. But if one wants to reproduce these correlations in a purely classical model, then classical communication between the parties is required at the moment of the measurements. This situation is even more surprising if the particles are widely separated from each other and the measurements take place during a very short time interval, so short that the two measurement events are spacelike separated. In this case the communication would have to occur faster than the speed of light.

This feature of quantum mechanics was discovered by Bell (1965) and is now known as "quantum nonlocality." It has been the subject of much further theoretical and experimental studies since. Indeed it is one of the most surprising and counterintuitive features of quantum mechanics. Bell's theorem shows that Einstein's program of trying to rationalize quantum mechanics by reducing it to classical mechanics is futile and doomed to failure, as it cannot be done without giving up another cornerstone of 20th century physics (discovered by Einstein himself), namely, the fact that information cannot travel faster than the speed of light. More recently, another reason why such a reduction is doomed emerged through the study of quantum information. Namely, we expect any such classical description of quantum mechanics to be exponentially inefficient, i.e., to use *exponentially* more resources than the quantum theory. We discuss quantum nonlocality extensively in the present review, focusing on its connection to communication complexity.

D. Unity of quantum communication complexity and quantum nonlocality

The reason why in this review we deal with quantum communication complexity and quantum nonlocality together is that these two topics are intimately related. Indeed they can be formulated in a unified way, and furthermore many questions can be mapped from one topic to the other. In fact, during the past dozen years an intense cross-fertilization has occurred between these two fields, which has considerably enriched both of them.

To see the unity between the two subjects, recall that in both cases the parties Alice and Bob are given some inputs x and y. In one case these inputs correspond to the arguments of the function that must be computed. In the other case these inputs correspond to a description of the measurements that must be carried out on the particles (the "measurement settings"). In both cases Alice and Bob must provide an output a and b. In communication complexity we require that $b=f(x,y)$ and a is irrelevant; in nonlocality we are interested in the correlations between a,b and x,y (for instance, we request that $a=b$ when x and y have certain values and that $a \neq b$ when x and y have some other values). We can unify these descriptions by saying that the aim in both cases is to produce a joint probability distribution

$$P(a,b|x,y)$$

of the outputs given the inputs, such that $P(a,b|x,y)$ has certain desirable properties.

E. Resources

In both communication and nonlocality, the basic questions one wants to answer are what is the minimum amount of resources necessary to reproduce the distribution $P(a,b|x,y)$ and how does this amount change when one changes the model, i.e., when one changes the type of resource that can be used. There are in fact many different types of resources that can be compared, and we now review them. We return to them in more detail in the body of the review.

- *Quantum communication*. The parties are allowed to send each other quantum states. One quantifies the amount of communication by the number of qubits sent.
- *Classical communication*. The parties are allowed to send each other classical communication. One quantifies the amount of communication by the number of bits sent.
- *Entanglement*. The parties share entangled states. One quantifies the amount of entanglement by the number of qubits that the state locally consists of. For example, we frequently use maximally entangled states of two qubits, called ebits [also known as EPR pairs after Einstein *et al.* (1935)], $(1/\sqrt{2})(|0\rangle|0\rangle + |1\rangle|1\rangle)$, or entangled states that can be obtained from this with local operations.
- *Shared randomness*. The parties have randomness, i.e., they are allowed to toss coins. In the case of *shared* randomness, the parties both share the same string of coins. This could, for instance, be implemented by having the parties toss the coins beforehand, at some earlier time when they are together, and then use the coins later when they need to solve the communication complexity problem.
- *Local randomness*. The parties have randomness, i.e., they are allowed to toss coins. In the case of *local* randomness the coins are tossed locally, and the string of outcomes of the coins for Alice is independent of the string of outcomes of the coins for Bob.

The rationale for measuring classical information in terms of bits is Shannon's noiseless coding theorem (Shannon, 1948), which states that, asymptotically, the information produced by a stochastic source can be encoded in a number of bits equal to the entropy of the source. This is paralleled in the quantum case by Schumacher compression (Schumacher, 1995), which states that, asymptotically, the information produced by a stochastic quantum source can be encoded into a number of qubits equal to the von Neumann entropy of the source. It is paralleled in the case of entanglement, by entanglement distillations, namely, the fact that pure two-party entangled states can, asymptotically in the number of copies of the state, be converted into the number of ebits equal to the von Neumann entropy of the reduced density matrix of each party (Bennett, Bernstein, *et al.*, 1996). In the context of communication complexity, however, we are not dealing with the asymptotic limit of large amounts of communication or large amounts of entanglement. Thus whereas in most cases we keep the basic concepts of bits, qubits, and ebits, it could be relevant in specific cases to consider variants on these resources, such as trits, nonmaximally entangled states, etc.

The above resources have been ordered (more or less) from the strongest to the weakest. Indeed most of these resources imply the ones below them. For instance, one can send classical information using qubits, one can use quantum communication to distribute entanglement, one can measure the entangled particles to produce shared randomness, etc. The only case where the ordering is not so clear is between classical communication and entanglement. Indeed if two parties share an entangled state, they cannot use it to communicate (as discussed above). But, on the other hand (as discussed below), sharing n ebits may allow one to save an exponentially large (in n) amount of bits in some communication scenarios (whereas in all other cases, n uses of one resource allows one to implement n uses of the resources below it).

There are also a number of nontrivial ways in which these resources can be substituted one for the other. *Quantum teleportation* allows one to substitute 1 ebit and 2 bits of classical communication for 1 qubit of quantum communication (Bennett *et al.*, 1993). *Dense coding* shows that sharing 1 ebit and then communicating 1 qubit allow one to communicate 2 bits (Bennett and Wiesner, 1992). Newman's theorem states that in the context of communication complexity, having shared randomness can save only a small amount of communication compared to having local randomness (Newman, 1991).

In addition, we will at some points in this review consider other additional (more specialized or more exotic) resources. For instance, one can consider the following:

- *One-way classical or quantum communication*. Alice is allowed to communicate to Bob, but Bob is not allowed to communicate back to Alice.
- *Simultaneous message passing (SMP) model*. In this model there is a third party, called the referee, and messages are only allowed from Alice to the referee and from Bob to the referee. It is the referee who has to compute the value of the function $f(x,y)$.
- *Multipartite entanglement*. Sometimes one is interested in nonlocality or communication complexity between more than two parties. Contrary to bipartite entanglement where it is sufficient to consider ebits, there are many kinds of multiparticle entanglement [such as Greenberger-Horne-Zeilinger (GHZ) states, W states, etc.] which could be useful for solving different communication problems.
- *Nonlocal (NL) or Popescu-Rohrlich (PR) boxes*. This exotic resource is intermediate between an ebit and a bit. Indeed, it is a resource which does not enable the parties to communicate (in the same way that entanglement does not allow communication). But to be produced physically it requires a bit of communication between the parties at the moment it is used (contrary to entanglement, which once established requires no more communication). Its study provides a deeper understanding of the power and limitations of quantum entanglement in communication complexity.

F. Basic scenarios

The basic question asked in communication complexity and quantum nonlocality is to understand how much of these resources are required in different situations.

Thus classical communication complexity (Kushilevitz and Nisan, 1997) is basically concerned with understanding how much classical communication is required to compute the value of a function $f(x,y)$, possibly using (shared or local) randomness.

In quantum communication complexity the parties are trying to compute the value of f but may now use quantum resources. In the quantum communication model, introduced by Yao (1993), they can communicate qubits, and in the entanglement model, introduced by Cleve and Buhrman (1997), the parties share entangled particles and are allowed to communicate classical bits. When one extends the quantum communication model of Yao such that the parties also share entangled particles, quantum teleportation shows that these two models are essentially equivalent: one qubit in the first model can be replaced by 2 bits and one ebit in the entanglement, and conversely 1 bit can be simulated by one qubit. It is, however, a challenging open problem whether the quantum communication model, *without* shared entanglement, is essentially equivalent to the entanglement model.

Nonlocality, although at first sight a very different topic, is also concerned with comparing resources. Indeed the basic question in this area is to compare the correlations that can be obtained if the parties share entanglement and carry out local measurements on their particles but are not allowed any communication and the correlations that can be obtained if the parties have shared randomness but are not allowed any communication [this is known as a local hidden variable (LHV) model].

Bell's theorem states that these two scenarios are not equivalent: shared randomness alone is not sufficient to reproduce the quantum correlations.

G. Mappings between communication complexity and nonlocality

Quantum communication complexity, classical communication complexity, and nonlocality can be put in a unified framework in which similar kinds of resources are compared. In addition, in some cases there exist mappings between quantum communication complexity scenarios and nonlocality scenarios.

The most simple such mapping occurs in the entanglement model if the parties can solve the communication complexity problem more efficiently using entanglement than without entanglement and if this can be done by measuring their entangled particles before they communicate to each other. Then it immediately follows that the correlations obtained by measuring their entangled particles (but without communicating) cannot be realized in a local hidden variable model.

Conversely it is possible to map any nonlocality experiment to a communication complexity problem in the entanglement model. This was the approach used in the original paper (Cleve and Buhrman, 1997). It mapped the nonlocal correlations that arise in the GHZ paradox to a communication complexity problem. This approach has since been generalized (Brukner *et al.*, 2004), although in the resulting communication complexity problem the function $f(x,y)$ is only computed successfully by the parties with nonzero probability.

Another mapping can occur in the quantum communication model when *one-way* quantum communication from Alice to Bob is more efficient than classical communication. Then it is often possible to construct from the communication complexity problem a nontrivial nonlocality scenario. This approach has yielded some interesting nonlocality scenarios, which we will describe in detail below.

H. Summary of the review

In this review we present some of the main results obtained so far in the field of quantum communication complexity. We start by introducing quantum nonlocality in Sec. II, focusing on its relation with communication complexity. We present simple examples such as the GHZ paradox, the Clauser-Horne-Shimony-Holt (CHSH) example, and the magic square game but rephrasing them in the language of data processing. Next we present quantum communication complexity in Sec. III, illustrating it with examples such as the distributed Deutsch-Jozsa (DJ) problem, the intersection problem, Raz's problem, and the hidden matching problem. In Sec. IV we unite these two approaches, showing how some of the examples from quantum communication complexity can be used to derive new nonlocality games. In Sec. V we discuss another model of communication complexity, the simultaneous message passing model, and show how classical communication, entanglement, and quantum communication can be traded one for the other in this model. In Sec. VI we discuss several additional aspects of quantum nonlocality such as nonlocal boxes, Tsirelson bounds, and simulation of quantum correlations using classical resources. Finally we consider in Sec. VII experimental issues, in particular the detection loophole, and present the outlook for future experiments. We conclude by discussing some open questions in the field. The interested reader can also consult the earlier review (Brassard, 2003) which covers some of the material presented here.

II. SIMPLE NONLOCALITY EXAMPLES

The idea of nonlocality was originally concerned with the possibility that quantum mechanics is actually a *classical theory* that depends on "hidden variables" whose values might be discovered in the future as part of some successor theory to quantum mechanics. Bell (1965) proposed a hypothetical experiment for ruling out such classical theories under the assumptions that measurements of quantum systems can occur at different points in space-time and information cannot be transmitted faster than the speed of light.

Another way of interpreting Bell's experiment is as a method for two (or more) cooperating distributed parties to compute some sort of input-output relation, where each party receives input data and must produce output data consistent with the relation. In Bell's experiment, there is such a task that cannot be accomplished in a setting where the information-processing resources are all classical. In contrast, the task can be accomplished if the parties share prior entanglement.

Since Bell's seminal work, the concept of quantum nonlocality has been extensively studied by physicists, philosophers, and more recently by computer scientists. Some of the important early advances have been the CHSH inequality (Clauser *et al.*, 1969) which allows Bell's surprising predictions to be tested even in the presence of noise, and the GHZ-Mermin scenario (Greenberger *et al.*, 1989; Mermin, 1990b) which was the first "pseudotelepathy" game. More recently there has been a more or less systematic enumeration of Bell inequalities for a small number of settings and/or outcomes [see, e.g., Collins *et al.* (2002) and Collins and Gisin (2004)], the study of the statistical power of nonlocality tests (van Dam *et al.*, 2005), an understanding of the limits to quantum nonlocality (Tsirelson-type bounds) (Cirel'son, 1980) as compared to the larger world of correlations obeying only the no-signaling conditions (e.g., nonlocal boxes), investigations of the power of nonlocality in cryptographic settings (Barrett, Hardy, and Kent, 2005), etc.

Here we review various nonlocality scenarios, casting them in the language of data processing. The interested reader wishing to complement this overview could consult two recent reviews, written more from physics (Werner and Wolf, 2001) and computer science (Brassard *et al.*, 2005) perspectives.

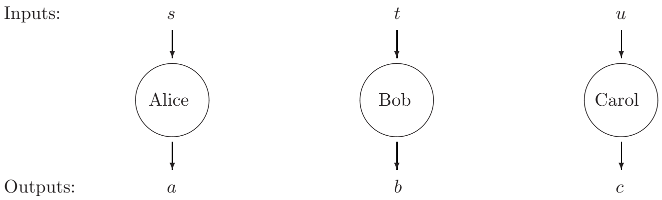

FIG. 3. The general form of a nonlocality scenario involving three parties: Alice, Bob, and Carol receive inputs s, t, and u, respectively, and are required to produce outputs a, b, and c, respectively, satisfying certain conditions. Once the inputs are received, no communication is permitted between the parties. For the specific GHZ scenario, it is possible to accomplish the task if the parties are in possession of a tripartite entangled state. Without the prior entanglement, it is impossible to accomplish the task.

A. GHZ and Mermin

The following scenario essentially underlies those of Greenberger *et al.* (1989) and Mermin (1990b) but is cast in the language of data processing. The basic structure is shown in Fig. 3.

Three physically separated parties—we call them Alice, Bob, and Carol—receive *input* bits s, t, and u, respectively, which are arbitrary subject to the condition that $s \oplus t \oplus u = 0$ ($\oplus$ denotes *exclusive* OR, which is the sum of its arguments in modulo 2 arithmetic). Once they receive their input data, they are forbidden from having any communication between them. Their goal is to produce *output* bits a, b, and c, respectively, such that

$$a \oplus b \oplus c = \begin{cases} 0 & \text{if } stu = 000 \\ 1 & \text{if } stu \in \{011,101,110\}. \end{cases} \tag{1}$$

Note that the task that the three parties are trying to accomplish is the computation of a relation, where there are three input bits (stu) and three output bits (abc). The task is nontrivial in light of the fact that the input bits are distributed among the parties so that each party is given the value of only one of them; the output bits are also distributed.

The first observation is that with classical resources there must be communication among the three parties to succeed. To see why this is so, first consider deterministic strategies (later we analyze the case of probabilistic strategies, where the parties behave stochastically, i.e., they can flip coins). Since Alice cannot receive any information from Bob or Carol, her output bit a can depend only on the value of her input bit s. Let a_0 (a_1) be Alice's output when her input bit is 0 (1). Similarly, let b_0, b_1 and c_0, c_1 be Bob and Carol's outputs for their respective input values. Note that the 6 bits $a_0, a_1, b_0, b_1, c_0, c_1$ completely characterize any deterministic strategy of Alice, Bob, and Carol. The conditions of the problem translate into the equations

$$\begin{aligned} a_0 \oplus b_0 \oplus c_0 &= 0, \\ a_0 \oplus b_1 \oplus c_1 &= 1, \\ a_1 \oplus b_0 \oplus c_1 &= 1, \\ a_1 \oplus b_1 \oplus c_0 &= 1. \end{aligned} \tag{2}$$

It is impossible to satisfy all four equations simultaneously. This is because summing the four equations modulo 2, yields $0=1$ (recall that $1+1=0 \bmod 2$). Therefore, for any strategy, there exists an input configuration $stu \in \{000,011,101,110\}$ for which it fails. Note, however, that for any three out of the four equations from Eqs. (2) there is a strategy that satisfies these three equations perfectly.

To see why probabilistic strategies (that use local or shared randomness) cannot succeed either, note that any such strategy can be modeled as a deterministic strategy where Alice, Bob, and Carol have access to a random variable r (for example, r could be the outcomes of a sequence of uniformly distributed random bits). This r is sometimes referred to as a "local hidden variable." It is assumed that the input bits (stu) are uncorrelated with r. The intuitive way of thinking about this scenario is that the three parties get together before the game starts, randomly select r, and then each party secretly keeps a copy of this information. An example of a probabilistic strategy is for $r \in \{0,1\}^2$ to be two uniformly random bits that specify which three of the four equations in Eqs. (2) are satisfied. This probabilistic strategy succeeds with probability 3/4. We next show that this success probability is optimal.

Suppose that the input data s,t,u are uniformly distributed over $\{000,011,101,110\}$. Then the success probability that any randomized protocol achieves is

$$\sum_r q_r \frac{1}{4} \sum_{s,t,u} P(s,t,u,r), \tag{3}$$

where q_r is the probability that the parties flip r, and $P(s,t,u,r)=1$ if the deterministic protocol corresponding to r is correct on input stu and $P(s,t,u,r)=0$ otherwise. Clearly this is bounded from above by

$$\max_r \frac{1}{4} \sum_{s,t,u} P(s,t,u,r), \tag{4}$$

which by the above discussion is at most 3/4.

Now consider the same problem but where Alice, Bob, and Carol have an additional resource: each is supplied with a qubit, where the state of the combined three-qubit system is[1]

$$\tfrac{1}{2}|000\rangle - \tfrac{1}{2}|011\rangle - \tfrac{1}{2}|101\rangle - \tfrac{1}{2}|110\rangle. \tag{5}$$

The parties are allowed to apply unitary transformations and perform measurements on their individual qubits, but communication between the parties is still forbid-

[1]This is an entangled state that is equivalent to the so-called GHZ state $(1/\sqrt{2})|000\rangle + (1/\sqrt{2})|111\rangle$ (under local unitary operations).

den. It turns out that now the parties can produce a, b, c satisfying Eq. (1). This is achieved by the procedure that follows.

The procedure for Alice is to measure her qubit in the computational basis (consisting of $|0\rangle$ and $|1\rangle$) if her input bit s is 0, and to measure her qubit in the Hadamard basis [consisting of $H|0\rangle=(1/\sqrt{2})(|0\rangle+|1\rangle)$ and $H|1\rangle=(1/\sqrt{2})(|0\rangle-|1\rangle)$] if her input bit is 1. In either case, she sets her output bit a to the outcome of her measurement. The procedures for Bob and Carol are similar to that of Alice, but with Bob's bits being s and b and Carol's bits being u and c.

To see why the described procedure always produces output bits abc satisfying Eq. (1), consider the various cases of the input possibilities stu. In the case where $stu=000$, the state is measured in the computational basis, so clearly the outcomes are from $\{000,011,101,110\}$ and hence satisfy $a\oplus b\oplus c=0$. The case where $stu=011$ can be analyzed by assuming that a Hadamard transform is applied to the last two qubits of the state prior to a measurement in the computational basis. Since

$$
\begin{aligned}
&(I\otimes H\otimes H)(\tfrac{1}{2}|000\rangle-\tfrac{1}{2}|011\rangle-\tfrac{1}{2}|101\rangle-\tfrac{1}{2}|110\rangle)\\
&\quad=(I\otimes H\otimes H)[\tfrac{1}{2}|0\rangle(|00\rangle-|11\rangle)-\tfrac{1}{2}|1\rangle(|01\rangle+|10\rangle)]\\
&\quad=\tfrac{1}{2}|0\rangle(|01\rangle+|10\rangle)-\tfrac{1}{2}|1\rangle(|00\rangle-|11\rangle)\\
&\quad=\tfrac{1}{2}|001\rangle+\tfrac{1}{2}|010\rangle-\tfrac{1}{2}|100\rangle+\tfrac{1}{2}|111\rangle, \qquad (6)
\end{aligned}
$$

$a\oplus b\oplus c=1$, as required, in this case. The remaining cases where $stu=101$ and 110 are similar by the symmetry of the entangled state and protocol.

We have shown that the entangled state enables the three parties to correlate their output bits with their inputs bits in a manner that is impossible to achieve with classical resources unless there is communication among the parties. It should be noted that, in accomplishing this task using the entangled state, no actual communication occurs among the parties. In particular, the output bits a, b, and c individually contain no information about stu; they are uniformly distributed in all cases. It is only the trivariate correlations among a, b, and c that are related to the input data stu.

B. CHSH

The following scenario essentially underlies that of (Clauser *et al.*, 1969) but is cast in the language of data processing. The basic structure is shown in Fig. 4.

Alice and Bob receive input bits s and t, respectively, and, after this, they are forbidden from communicating with each other. Their goal is to produce output bits a and b, respectively, such that

$$a\oplus b=s\wedge t, \qquad (7)$$

("$\wedge$" is the logical AND, which is 1 if all its arguments are 1 and which is 0 otherwise) or, failing that, to satisfy this condition with as high a probability as possible. To analyze the situation in terms of classical information, first again consider the case of deterministic strategies. For these, Alice's output bit depends solely on her input bit s and similarly for Bob. Let a_0, a_1 be the two possibilities for Alice and b_0, b_1 be the two possibilities for Bob. These four bits completely characterize any deterministic strategy. Condition (7) translates into the equations

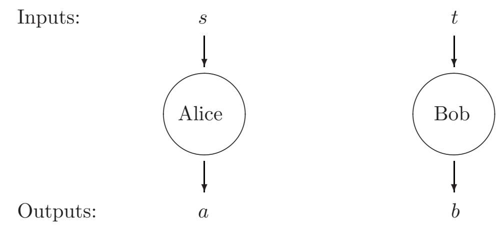

FIG. 4. The nonlocality scenario involving two parties: Alice and Bob receive inputs s and t, respectively, and are required to produce outputs a and b, respectively, satisfying certain conditions. Once the inputs are received, no communication is permitted between the parties. For the specific CHSH scenario, it is possible to accomplish the task with probability $\cos^2(\pi/8)=0.853\ldots$ if the parties are in possession of an ebit. Without the prior entanglement, the highest possible success probability is 3/4.

$$
\begin{aligned}
&a_0\oplus b_0=0,\\
&a_0\oplus b_1=0,\\
&a_1\oplus b_0=0,\\
&a_1\oplus b_1=1.
\end{aligned} \qquad (8)
$$

It is impossible to satisfy all four equations simultaneously (since summing them modulo 2 yields 0=1). Therefore it is impossible to satisfy condition (7) absolutely. Using a probabilistic strategy, Alice and Bob can satisfy condition (7) with probability 3/4. For such a strategy, we allow Alice and Bob have access to shared random variables, whose distribution is independent of that of the inputs s and t. Note that any three of the four equations of Eqs. (8) can be simultaneously satisfied. The probabilistic classical strategy works as follows. Alice and Bob have uniformly distributed random bits that are used to specify which of the four equations of Eqs. (8) is violated and then play the strategy that satisfies the other three perfectly. It is easy to see that (a) for any input st, the resulting outputs satisfy condition (7) with probability 3/4; and (b) this is optimal in that no probabilistic strategy can attain a success probability greater than 3/4.

Now consider the same problem but where Alice and Bob are each supplied with a qubit where the state of the two-qubit system is initialized to

$$\frac{1}{\sqrt{2}}(|00\rangle-|11\rangle). \qquad (9)$$

It turns out that now the parties can produce data that satisfies condition (7) with probability $\cos^2(\pi/8)=0.853\ldots$, which is higher than what is possible in the classical case. This is achieved with the following procedures. Denote the unitary operation that rotates the qubit by angle θ by

$$R(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

(where we have written it out in the computational basis). Alice applies one of two rotations on her qubit, depending on her input bit s: if $s=0$ the rotation is $R(-\pi/16)$; if $s=1$ the rotation is $R(3\pi/16)$. Then Alice measures her qubit in the computational basis and sets her output bit a to the result. Bob's procedure is the same, depending on his input bit t. It is straightforward to calculate that if Alice rotates by θ_1 and Bob rotates by θ_2, then the entangled state becomes

$$\frac{1}{\sqrt{2}}[\cos(\theta_1+\theta_2)(|00\rangle - |11\rangle) + \sin(\theta_1+\theta_2)(|01\rangle + |10\rangle)]. \tag{10}$$

After the measurements, the probability that $a \oplus b = 0$ is $\cos^2(\theta_1+\theta_2)$. It is now a straightforward exercise to verify that condition (7) is satisfied with probability $\cos^2(\pi/8)$ for all four input possibilities.

C. Tsirelson's upper bound for CHSH

Although the protocol in the previous section using entanglement has a higher success probability $[\cos^2(\pi/8)=0.853\ldots]$ than any classical protocol (3/4), it still does not succeed with probability 1. This raises the question of whether there is a different strategy using entanglement that always succeeds or, failing that, whose success probability exceeds $\cos^2(\pi/8)$. Tsirelson (Cirel'son, 1980) first showed that the above quantum protocol is optimal in that it is impossible to exceed success probability $\cos^2(\pi/8)$ regardless of the strategy—including any amount of prior entanglement—the parties start with. What follows is a simple proof of this result.

Consider an arbitrary bipartite entangled state $|\psi\rangle_{AB}$. An arbitrary strategy for Alice that uses this entangled state can be represented by two observables[2] A_0 and A_1, each with eigenvalues in $\{+1,-1\}$. When Alice's input bit is 0, she obtains her output bit by applying the projective measurement corresponding to the eigenspaces of A_0 to the component of $|\psi\rangle_{AB}$ in her possession. The $+1$ eigenspace of A_0 corresponds to output bit 0, while the -1 eigenspace corresponds to the output bit 1. When her input bit is 1, she applies the measurement corresponding to A_1. Similarly, an arbitrary strategy for Bob can be represented by two observables B_0 and B_1.

At this point, the reader might object that $|\psi\rangle_{AB}$, A_0, A_1, B_0, and B_1 do not capture every possible strategy of Alice and Bob since they need not be limited to applying projective measurements. Although nonprojective measurements may be used, such measurements can always be simulated by projective measurements in a larger Hilbert space. Thus, no generality has been lost because any strategy can be converted to the above form.

Since the observables have eigenvalues in $\{+1,-1\}$ rather than $\{0,1\}$, it is more convenient here to think of Alice and Bob's output bits as $a'=(-1)^a$ and $b'=(-1)^b$, respectively. Then the protocol succeeds on input st if and only if $(-1)^{s\wedge t}a'b'=1$.

If s and t are randomly chosen according to the uniform distribution, then the expected value of $(-1)^{s\wedge t}a'b'$ is

$$\langle\psi|_{AB}(\tfrac{1}{4}A_0\otimes B_0 + \tfrac{1}{4}A_0\otimes B_1 + \tfrac{1}{4}A_1\otimes B_0 - \tfrac{1}{4}A_1\otimes B_1)|\psi\rangle_{AB} \tag{11}$$

and is therefore upper bounded by the largest eigenvalue of

$$M = \tfrac{1}{4}A_0\otimes B_0 + \tfrac{1}{4}A_0\otimes B_1 + \tfrac{1}{4}A_1\otimes B_0 - \tfrac{1}{4}A_1\otimes B_1. \tag{12}$$

It is straightforward to calculate that

$$M^2 = \tfrac{1}{4}I - \tfrac{1}{16}(A_0A_1)\otimes(B_0B_1) + \tfrac{1}{16}(A_0A_1)\otimes(B_1B_0) + \tfrac{1}{16}(A_1A_0)\otimes(B_0B_1) - \tfrac{1}{16}(A_1A_0)\otimes(B_1B_0), \tag{13}$$

from which we can impose an upper bound on the maximum eigenvalue of M^2 by the sum of the maximum eigenvalue in each term, obtaining $\frac{1}{4}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}+\frac{1}{16}=\frac{1}{2}$. It follows that the largest eigenvalue of M itself is at most $1/\sqrt{2}$, which therefore imposes an upper bound on the expected value of $(-1)^{s\wedge t}a'b'$. This translates into an upper bound of $(1+1/\sqrt{2})/2=\cos^2(\pi/8)$ for the success probability of the actual protocol (where Alice and Bob output bits a and b). This completes the proof of Tsirelson's upper bound for CHSH.

D. Magic square game

In one respect the GHZ example is more striking than the CHSH example: in the former case, the protocol with entanglement always succeeds, while in the latter case the protocol with entanglement merely succeeds with higher probability. However, the GHZ example involves three parties, whereas the CHSH example only involves two. Is there a *two-party* scenario where the quantum protocol always succeeds, whereas the best classical success probability is bounded below 1? The answer is affirmative [see, for instance, Cabello (2001a, 2001b, 2005)]. A particularly elegant example is the following game, which has been referred to as the *magic square game* (Aravind, 2004).

To define this game, consider the problem of labeling the entries of a 3×3 matrix with bits so that the parity of

[2]An observable is a Hermitian operator. One associates to an observable a projective measurement, with one projector for each of the eigenspaces of the observable.

each row is even, whereas the parity of each column is odd. It is not hard to see that this is impossible.[3] The two matrices

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}$$

each satisfy five out of the six constraints. For the first matrix, all rows have even parity, but only the first two columns have odd parity. For the second matrix, the first two rows have even parity and all columns have odd parity.

Bearing the above in mind, consider the game where Alice receives $s \in \{1,2,3\}$ as input (specifying the number of a row) and Bob receives $t \in \{1,2,3\}$ as input (specifying the number of a column). Their goal is to each produce three-bit outputs, $a_1a_2a_3$ for Alice and $b_1b_2b_3$ for Bob, with these properties: (1) they satisfy the row or column parity constraints, namely, $a_1 \oplus a_2 \oplus a_3 = 0$ and $b_1 \oplus b_2 \oplus b_3 = 1$ and (2) they are consistent where the row intersects the column, namely, $a_t = b_s$.

As usual, Alice and Bob are forbidden from communicating once the game starts, so Alice does not know what t is and Bob does not know what s is. We observe that, classically, the best success probability possible is $8/9$, whereas there is a quantum strategy that always succeeds.

An example of a strategy that attains success probability $8/9$ (when the input st is uniformly distributed) is where Alice plays according to the rows of the first matrix above and Bob plays according the columns of the second matrix above. This succeeds in all cases except where $s=t=3$. To see why this is optimal, note that for any other classical strategy it is possible to represent it as two matrices, as above, but with different entries. Alice plays according to the rows of the first matrix and Bob plays according to the columns of the second matrix. We can assume that the rows of Alice's matrix all have even parity; if she outputs a row with odd parity then they immediately lose regardless of Bob's output. Similarly, we can assume that all columns of Bob's matrix have odd parity.[4] Considering such a pair of matrices, the players lose at each entry where they differ. There must be such an entry since otherwise it would be possible to have all rows even and all columns odd with one matrix. Thus, when the input st is chosen uniformly from $\{1,2,3\} \times \{1,2,3\}$, the success probability is at most $8/9$.

[3]As before, we can express a valid solution in terms of equations, in this case six of them (where arithmetic is modulo 2): $m_{11}+m_{12}+m_{13}=0$, $m_{21}+m_{22}+m_{23}=0$, $m_{31}+m_{32}+m_{33}=0$, $m_{11}+m_{21}+m_{31}=1$, $m_{12}+m_{22}+m_{32}=1$, and $m_{13}+m_{23}+m_{33}=1$. Adding these equations modulo 2 yields $0=1$.

[4]In fact, the game can be simplified so that Alice and Bob each output just 2 bits since the parity constraint determines the third bit.

The quantum strategy for this game is based on the following observation due to Mermin (1990a, 1993). Let I, X, Y, and Z denote the 2×2 Pauli matrices:

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix},$$
$$Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and } Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}. \tag{14}$$

Each is an observable with eigenvalues in $\{+1,-1\}$. Consider the following table of two-qubit observables that are each a tensor product of two Pauli matrices:

$X \otimes X$	$Y \otimes Z$	$Z \otimes Y$
$Y \otimes Y$	$Z \otimes X$	$X \otimes Z$
$Z \otimes Z$	$X \otimes Y$	$Y \otimes X$

For our present purposes, the noteworthy property is that the observables along each row commute and their product is $I \otimes I$, and the observables along each column commute and their product is $-I \otimes I$. This implies that, for any two-qubit state, performing the three measurements along any row results in three $\{+1,-1\}$-valued bits whose product is $+1$. Also, performing the three measurements along any column results in three $\{+1,-1\}$-valued bits whose product is -1. This can be seen more easily when one simultaneously diagonalizes the three commuting observables. They will have 1 and -1 eigenvalues on the diagonal. Each consecutive observable will project the state onto a possible refinement of the current eigenspace the state lies in. This will yield that the product of the outcomes of the three observables will be 1 when the observables belong to a row of the matrix because the product of the row observables is $I \otimes I$, and -1 when they belong to a column since the product of the observables for each column is $-I \otimes I$.

We can now describe the quantum protocol. It uses two pairs of entangled qubits, each of which is in initial state $(1/\sqrt{2})(|01\rangle - |10\rangle)$. Alice, on input s, applies three two-qubit measurements corresponding to the observables in row s of the above table. For each measurement, if the result is $+1$, she outputs 0 and if the result is -1, she outputs 1. Similarly, Bob, on input t, applies the measurements corresponding to the observables in column t and converts the outcomes into bits in the same manner.

We have already established that Alice and Bob's output bits satisfy the required parity constraints. It remains to show that Alice and Bob's output bits that correspond to where the row meets the column are the same. For that measurement, Alice and Bob are measuring with respect to the same observable in the above table. Because all observables in each row and in each column commute, we may assume that the place where they intersect is the first observable applied. Those bits are obtained by Alice and Bob each measuring $\frac{1}{2}(|01\rangle - |10\rangle)(|01\rangle - |10\rangle)$ with respect to the observable in entry (s,t) of the table. To show that their measurements will agree for all cases of st, we consider the individual Pauli measurements on the individual entangled pairs of the form $(1/\sqrt{2})(|01\rangle - |10\rangle)$. Let a' and b' denote the outcomes of the first measurement (in terms of bits) and a'' and b'' denote the outcomes of the second. Since the measurement associated with the tensor product of two observables is operationally equivalent to measuring each individual observable and taking the product of the results, we have that $a_t = a' \oplus a''$ and $b_s = b' \oplus b''$. It is straightforward to verify that if the same measurement from $\{X, Y, Z\}$ is applied to each qubit of $(1/\sqrt{2})(|01\rangle - |10\rangle)$ then the outcomes will be distinct. Therefore, $a' \oplus b' = 1$ and $a'' \oplus b'' = 1$, from which it follows that

$$a_t \oplus b_s = (a' \oplus a'') \oplus (b' \oplus b'') = (a' \oplus b') \oplus (a'' \oplus b'') = 1 \oplus 1 = 0, \tag{15}$$

so $a_t = b_s$. This completes the analysis of the magic square game.

III. COMMUNICATION COMPLEXITY

In the last section we considered scenarios without communication. Here we extend the nonlocality setting to one where the parties (Alice and Bob) are allowed to send information to each other in the form of bits or qubits. They can still have shared randomness and may share an entangled quantum state. We are now interested in the *minimum number* of bits or qubits that are needed in order to compute a function that depends on the inputs of all parties.

The ability to send information to each other departs from the setting of nonlocality. We will see that entanglement can be used to reduce (for certain functions) the communication drastically compared to when the parties share just classical resources. Accordingly, while entanglement cannot be used for signaling, it can be used to significantly reduce the communication needed for certain tasks. Later we shall see how some of the ideas and protocols developed in the setting of communication complexity can be used to formulate new nonlocality games.

Communication complexity has been extensively studied in the area of theoretical computer science and has deep connections with seemingly unrelated areas, such as very large scale integrated design, circuit lower bounds, lower bounds on branching programs, size of data structures, and bounds on the length of logical proof systems, to name just a few. We refer the reader to Hromkovič (1997) and Kushilevitz and Nisan (1997) for more details.

A. The setting

First we sketch the setting for classical communication complexity. Alice and Bob want to compute some function $f: \mathcal{D} \to \{0,1\}$, where $\mathcal{D} \subseteq X \times Y$. If the domain $\mathcal{D}$ equals $X \times Y$ then f is called a *total* function, otherwise it is a *promise* function. Alice receives input $x \in X$, Bob receives input $y \in Y$, with $(x, y) \in \mathcal{D}$. A typical situation, shown in Fig. 2, is where $X = Y = \{0,1\}^n$, so both Alice and Bob receive an n-bit input string. As the value $f(x, y)$ will generally depend on both x and y, some communication between Alice and Bob is required in order for them to be able to compute $f(x, y)$. We are interested in the minimal amount of communication they need.

A communication protocol is a distributed algorithm where first Alice does some individual computation and then sends a message (of one or more bits) to Bob, then Bob does some computation and sends a message to Alice, etc. Each message is called a *round*. After one or more rounds the protocol terminates and outputs some value, which must be known to both players. The *cost* of a protocol is the total number of bits communicated on the worst-case input. A deterministic protocol for f always has to output the right value $f(x, y)$ for all $(x, y) \in \mathcal{D}$. In a bounded-error protocol, Alice and Bob may flip coins and the protocol has to output the right value $f(x, y)$ with probability $\geq 2/3$ for all $(x, y) \in \mathcal{D}$. We could either allow Alice and Bob to toss coins individually (local randomness or "private coin") or jointly (shared randomness or "public coin"). The latter is analogous to the local hidden variables in nonlocality games. A public coin can simulate a private coin and is potentially more powerful. However, Newman's theorem (Newman, 1991) says that having a public coin can save at most $O(\log_2 n)$ bits of communication, compared to a protocol with a private coin.

Some often studied functions are the following:

- *Equality:* $\mathrm{EQ}(x, y) = 1$ if $x = y$ and $\mathrm{EQ}(x, y) = 0$ otherwise.
- *Inner product:* $\mathrm{IP}(x, y) = \sum_{i=1}^{n} x_i y_i (\mathrm{mod}\, 2)$ (for $x, y \in \{0,1\}^n$, x_i is the ith bit of x).
- *Intersection:* $\mathrm{INT}(x, y) = 1$ if there is an i where $x_i = y_i = 1$ and $\mathrm{INT}(x, y) = 0$ otherwise [viewing x as corresponding to the set $\{i : x_i = 1\}$ and similarly for y, $\mathrm{INT}(x, y)$ says whether the sets x and y intersect]. A variant of this problem asks to actually find an i where $x_i = y_i = 1$ or to output that none such i exists.

We first consider the equality problem, which will recur throughout the text. The goal for Alice is to determine whether her n-bit input is the same as Bob's or not. It is not hard to show that in the deterministic case n bits of communication are needed (see Appendix B.1 for a proof), so Bob might as well send his string to Alice after which Alice announces the answer to Bob with one more bit.

To illustrate the power of randomness, we give a simple yet efficient bounded-error protocol for the equality problem. Alice and Bob jointly toss a random string $r \in \{0,1\}^n$. Alice sends the bit $a = x \cdot r$ to Bob (where "$\cdot$" is inner product mod 2). Bob computes $b = y \cdot r$ and compares this with a. If $x = y$ then $a = b$, but if $x \neq y$ then $a \neq b$ with probability 1/2. Repeating this a few times, Alice and Bob can decide equality with small error using $O(n)$ public coin flips and a constant amount of communication.

This protocol uses public coins, but note that Newman's theorem implies that there exists an $O(\log_2 n)$-bit protocol that uses a private coin. We now explicitly describe such a protocol. Alice views her n bits as the coefficients of a polynomial p_x over some finite field $\mathbb{F}$ of about $3n$ elements:[5] $p_x(t)=\sum_{i=1}^{n} x_i t^{i-1}$. She picks a random element $a \in \mathbb{F}$ and sends Bob the pair $a, p_x(a)$, which she can do using $2\log_2(3n)$ bits. Bob computes $p_y(a)$ and outputs 1 if $p_x(a)=p_y(a)$ and outputs 0 otherwise. Clearly, if $x=y$ then Bob always outputs the correct answer 1. However, if $x \neq y$ then the polynomial $p_x(t) - p_y(t)$ is a polynomial in t of degree at most $n-1$ that is not identically equal to 0. Such a polynomial can be 0 on at most $n-1$ elements of $\mathbb{F}$. Hence with probability at least $2/3$, the field element a that Alice chose satisfies $p_x(a) \neq p_y(a)$ and Bob will give the correct output 0 also in this case.

B. The quantum question

Now what happens if we give Alice and Bob a quantum computer and allow them to send each other qubits and/or to make use of ebits that they share at the start of the protocol?

Formally speaking, we can model a quantum protocol as follows. The total state consists of three parts: Alice's private space, the channel, and Bob's private space. The starting state is $|x\rangle|0\rangle|y\rangle$: Alice gets x, the channel is initialized to 0, and Bob gets y. Now Alice applies a unitary transformation to her space and the channel. This corresponds to her private computation as well as to putting a message on the channel (the length of this message is the number of channel qubits affected by Alice's operation). Then Bob applies a unitary transformation to his space and the channel, etc. At the end of the protocol Alice or Bob makes a measurement to determine the output of the protocol. This model was introduced by Yao (1993).

In the second model, introduced by Cleve and Buhrman (1997), Alice and Bob share an unlimited number of ebits at the start of the protocol, but now they communicate via a *classical* channel: the channel has to be in a classical state throughout the protocol. We only count the communication, not the number of ebits used. Protocols of this kind can simulate protocols of the first kind with only a factor of 2 overhead: using teleportation, the parties can send each other a qubit using an ebit and two classical bits of communication. Hence the qubit protocols described below also immediately yield protocols that work with entanglement and a classical channel. Note that an ebit can simulate a public coin toss: if Alice and Bob each measures their half of the pair of qubits, they get the same random bit.

The third variant combines the strengths of the other two: here Alice and Bob start out with an unlimited number of ebits and they are allowed to communicate qubits. This third kind of communication complexity is in fact equivalent to the second, up to a factor of 2, again by teleportation.

Before continuing to study this model, we first have to face an important question mentioned in the Introduction: Is there anything to be gained here? At first sight, the following argument seems to rule out any significant gain. Suppose that in the classical world k bits have to be communicated in order to compute f. Since Holevo's theorem says that k qubits cannot contain more information than k classical bits, it seems that the quantum communication complexity should be roughly k qubits as well (maybe $k/2$ to account for superdense coding but not less). Surprisingly, this argument is false, and quantum communication can sometimes be much less than classical communication complexity. The information-theoretic argument via Holevo's theorem fails because Alice and Bob do not need to communicate the information in the k bits of the classical protocol; they are only interested in the value $f(x,y)$, which is just 1 bit. Below we survey some of the main examples that have so far been found of differences between quantum and classical communication complexities.

C. The first examples

Quantum communication complexity was introduced by Yao (1993) and studied by Kremer (1995), but neither showed any advantages of quantum over classical communication. Cleve and Buhrman (1997) introduced the variant with classical communication and prior entanglement and exhibited the first quantum protocol provably better than any classical protocol. It uses quantum entanglement to save 1 bit of classical communication and involves three parties. This gap was extended by Buhrman *et al.* (2001) and for arbitrary k parties by Buhrman *et al.* (1999).

D. Distributed Deutsch-Jozsa

The first impressively large gaps between quantum and classical communication complexities were exhibited by Buhrman *et al.* (1998). Their protocols are distributed versions of known quantum query algorithms such as the Deutsch-Jozsa (1992) and Grover (1996) algorithms.

We start with the first one. It is actually explained most easily in a direct way, without reference to the Deutsch-Jozsa algorithm (although that is where the idea came from). The problem deals with a promise version of the equality problem. Suppose the n-bit inputs x and y are restricted to the following case.

DJ promise. Either $x=y$ or x and y differ in exactly $n/2$ positions.

Note that this promise only makes sense if n is an even number, otherwise $n/2$ would not be integer. In fact, it will be convenient to assume n a power of 2. Here

[5] For those not familiar with finite fields, it suffices to choose a prime number $p \approx 3n$ and do all additions and multiplications modulo this p.

is a simple quantum protocol to solve this promise version of equality using only $\log_2 n$ qubits:

(1) Alice sends Bob the $\log_2 n$ qubit state $(1/\sqrt{n})\Sigma_{i=1}^{n}(-1)^{x_i}|i\rangle$, which she can prepare unitarily from x and $\log_2 n$ $|0\rangle$ qubits.

(2) Bob applies the unitary map $|i\rangle \mapsto (-1)^{y_i}|i\rangle$ to the state, applies a Hadamard transform to each qubit (for this it is convenient to view i as a $\log_2 n$-bit string), and measures the resulting $\log_2 n$-qubit state.

(3) Bob outputs 1 if the measurement gave $|0^{\log_2 n}\rangle$ and outputs 0 otherwise.

It is clear that this protocol only communicates $\log_2 n$ qubits, but why does it work? Note that the state that Bob measures is

$$H^{\otimes \log_2 n}\left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(-1)^{x_i+y_i}|i\rangle\right)$$
$$= \frac{1}{n}\sum_{i=1}^{n}(-1)^{x_i+y_i}\sum_{j\in\{0,1\}^{\log_2 n}}(-1)^{i\cdot j}|j\rangle.$$

This superposition looks rather unwieldy, but consider the amplitude of the $|0^{\log_2 n}\rangle$ basis state. It is $(1/n)\Sigma_{i=1}^{n}(-1)^{x_i+y_i}$, which is 1 if $x=y$ and 0 otherwise because the promise now guarantees that x and y differ in exactly $n/2$ of the bits. Hence Bob will always give the correct answer.

What about efficient classical protocols (without entanglement) for this problem? Proving lower bounds on communication complexity often requires a very technical combinatorial analysis. Buhrman *et al.* used a deep combinatorial result of Frankl and Rödl (1987) to prove that every classical errorless protocol for this problem needs to send at least $0.007n$ bits. We give the details in Appendix B.4.

This $\log_2 n$-qubit vs $0.007n$-bit example was the first exponentially large separation of quantum and classical communication complexities. Notice, however, that the difference disappears if we move to the bounded-error setting, allowing the protocol to have some small error probability. We can use the randomized protocol for equality discussed above or even simpler: Alice can just send a few (i,x_i) pairs to Bob, who then compares the x_i's with his y_i's. If $x=y$, he will not see a difference, but if x and y differ in $n/2$ positions, then Bob will probably detect this. Hence $O(\log_2 n)$ classical bits of communication suffice in the bounded-error setting, in sharp contrast to the errorless setting.

E. The intersection problem

Now consider the intersection function, which is 1 if $x_i=y_i=1$ for at least one i. Note that this is a decision problem of the appointment-scheduling problem mentioned in the Introduction. Buhrman *et al.* (1998) also presented an efficient quantum protocol for this. Their protocol is based on Grover's famous quantum search algorithm (Grover, 1996), which we sketch here.

Suppose there is some n-bit string z and we would like to find an index i such that $z_i=1$. We cannot "look" at z directly, but we can apply the following unitary map:

$$O_z : |i\rangle \mapsto (-1)^{z_i}|i\rangle.$$

Grover's algorithm starts in a uniform superposition $(1/\sqrt{n})\Sigma_{i=1}^{n}|i\rangle$ and then repeatedly applies the following unitary *Grover iterate* to the state:

$$G = H^{\otimes \log_2 n}O_0 H^{\otimes \log_2 n}O_z,$$

where $H^{\otimes \log_2 n}$ is the $\log_2 n$-qubit Hadamard transform and O_0 is the unitary that puts a "$-$" in front of the all-0 state. Suppose there are exactly t solutions: t indices i where $z_i=1$. We will not give the analysis here [see, for instance, Brassard *et al.* (2002)], but one can show that after about $(\pi/4)\sqrt{n/t}$ Grover iterations, most of the amplitude of the state sits on such solutions. Measuring the state will now with high probability give us a solution. Of course we may not know t in advance, but there is a way to find a solution with high probability using $O(\sqrt{n})$ Grover iterates even in that case.

Now what about the intersection problem? Note that we just want to find a solution for the string $z=x\wedge y$, which is the bitwise AND of x and y since $z_i=1$ whenever both $x_i=1$ and $y_i=1$. The idea is now to let Alice run Grover's algorithm to search for such a solution. Clearly, she can prepare the uniform starting state herself. She can also apply H and O_0 herself. The only thing she needs Bob's help with is in implementing O_z. This they do as follows. Whenever Alice wants to apply O_z to a state

$$|\phi\rangle = \sum_{i=1}^{n}\alpha_i|i\rangle,$$

she tags on her x_i in an extra qubit and sends Bob the state

$$\sum_{i=1}^{n}\alpha_i|i\rangle|x_i\rangle.$$

Bob applies the unitary map

$$|i\rangle|x_i\rangle \mapsto (-1)^{x_i\wedge y_i}|i\rangle|x_i\rangle$$

and sends back the result. Alice sets the last qubit back to $|0\rangle$ (which she can do unitarily because she has x), and now she has the state $O_z|\phi\rangle$! Thus we can simulate O_z using two messages of $\log_2(n)+1$ qubits each. Thus Alice and Bob can run Grover's algorithm to find an intersection, using $O(\sqrt{n})$ messages of $O(\log_2 n)$ qubits each, for total communication of $O(\sqrt{n}\log_2 n)$ qubits. Later Aaronson and Ambainis (2003) gave a more complicated protocol that uses $O(\sqrt{n})$ qubits of communication.

What about lower bounds? It is a well-known result of classical communication complexity that classical bounded-error protocols for the intersection problem need about n bits of communication (Kalyanasundaram and Schnitger, 1992; Razborov, 1992). Thus we have a quadratic quantum-classical separation for this problem. Could the separation be even bigger than quadratic? This question was open for quite a few years after the results of Buhrman *et al.* (1998) appeared until finally Razborov (2003) showed that any bounded-error quantum protocol for intersection needs to communicate about $\sqrt{n}$ qubits. His proof is beautiful but deep and complicated. We sketch it in Appendix C.

F. Raz's problem

Notice the contrast between the examples of the last two sections. For the distributed Deutsch-Jozsa problem we get an exponential quantum-classical separation, but the separation only holds if we require the classical protocol to be errorless. On the other hand, the gap for the disjointness function is only quadratic, but it holds even if we allow classical protocols to have some error probability.

Raz (1999) exhibited a function where the quantum-classical separation has both features: the quantum protocol is exponentially better than the classical protocol even if the latter is allowed some error probability. Consider the following promise problem **P**:

Alice receives a unit vector $v \in \mathbb{R}^m$ and a decomposition of the corresponding space in two orthogonal subspaces $H^{(0)}$ and $H^{(1)}$.

Bob receives an $m \times m$ unitary transformation U.

Promise: Uv is "close" either to $H^{(0)}$ or to $H^{(1)}$ (more precisely, letting P be the projector on subspace H, a vector v is close to H if $\|Pv\|^2 \geqslant 2/3$).

Question: Which of the two?

As stated, this is a problem with continuous input, but it can be discretized in a natural way by approximating each real number by $O(\log_2 m)$ bits. Alice and Bob's input is now $n = O(m^2 \log_2 m)$ bits long. There is a simple yet efficient two-round quantum protocol for this problem: Alice views v as a $\log_2 m$-qubit vector and sends this to Bob. Bob applies U and sends back the result. Alice then measures in which subspace $H^{(i)}$ the vector Uv lies and outputs the resulting i. This takes only $2 \log_2 m = O(\log_2 n)$ qubits of communication.

The efficiency of this protocol comes from the fact that an m-dimensional unit vector can be "compressed" or "represented" as a $\log_2 m$-qubit state. Similar compression is not possible with classical bits, which suggests that any classical protocol for **P** will have to send the vector v more or less literally and hence will require a lot of communication. This turns out to be true but the proof [given by Raz (1999)] is surprisingly hard. It shows that any bounded-error protocol for **P** needs to send at least about $n^{1/4}/\log_2 n$ bits.

G. The hidden matching problem

Consider the following hidden matching (HM) promise problem from Bar-Yossef *et al.* (2004) for even integer n: Alice receives a string $x \in \{0,1\}^n$. Bob receives a perfect matching M on $\{1,\ldots,n\}$ [i.e., a partition into $n/2$ disjoint pairs $M = \{(i_1,j_1),\ldots,(i_{n/2},j_{n/2})\}$]. Question: Output a triple $(i,j,x_i \oplus x_j)$ for some $(i,j) \in M$? This communication problem is not a function but a relation: for each input-pair x,M there are $n/2$ different correct answers instead of only one: $(i,j,x_i \oplus y_i)$ is correct for each $(i,j) \in M$. We consider one-way protocols here, where Alice sends one message to Bob and then Bob should produce a triple $(i,j,x_i \oplus x_j)$.

We now describe a quantum protocol where Alice sends only $O(\log_2 n)$ qubits and Bob gives one of the correct answers with probability 1 (Bar-Yossef *et al.*, 2004). Alice sends Bob the following $\log_2 n$-qubit message:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (-1)^{x_i} |i\rangle.$$

Bob views M as an orthogonal decomposition of the space $\mathbb{C}^n$ into $n/2$ two-dimensional subspaces. For instance, the projector for the subspace corresponding to $(i,j) \in M$ would be $P_{ij} = |i\rangle\langle i| + |j\rangle\langle j|$. Bob applies this measurement on the state he received and obtains the label of some random $(i,j) \in M$ as well as the projected state

$$\frac{1}{\sqrt{2}}[(-1)^{x_i}|i\rangle + (-1)^{x_j}|j\rangle].$$

An appropriate measurement on this state will give Bob the bit $x_i \oplus x_j$ with certainty, and he can output the correct answer $(i,j,x_i \oplus x_j)$.

What about classical protocols? First note that the HM problem can be solved by a short classical message from Bob to Alice: Bob sends Alice a pair $(i,j) \in M$ using $2 \log_2 n$ bits, which allows Alice to compute $x_i \oplus x_j$. But the situation is radically different if we consider classical one-way communication from Alice to Bob only. Indeed, one can show that if Alice sends Bob pairs (i,x_i) for $O(\sqrt{n})$ randomly chosen i's, then Bob probably received both points from at least one pair in M.[6] This allows him to output a correct answer. On the other hand, Bar-Yossef *et al.* (2004) proved that any classical protocol solving the hidden matching problem, even with small error probability, and involving only one-way communication from Alice to Bob needs messages of length at least about $\sqrt{n}$. Thus we have an exponential separation between classical one-way protocols and quantum one-way protocols.

Variants of the hidden matching problem have been used recently to obtain other quantum-classical separations. For example, Gavinsky, Kempe, *et al.* (2008) showed a $\log_2 n$ qubits vs $\sqrt{n}$-classical-bit separation for one-way protocols for a Boolean function derived from the hidden matching problem (while HM itself is a rela-

[6]This is due to an effect called the "birthday paradox" or "birthday problem." It states that if we throw roughly $\sqrt{n}$ balls into n bins at random, then probably there will be a bin containing at least two balls.

tional problem). Gavinsky (2008a) used another variant of HM to exhibit a relational problem where quantum one-way protocols are exponentially more efficient than classical two-way protocols.

H. Inner product

In the previous sections we gave examples of quantum-classical separations. The parameters were different, but in each case we showed that there was a quantum protocol for the problem at hand that required far less communication than the best classical protocols. Could this always be the case? Could quantum communication complexity be much more efficient for every communication complexity problem? The answer to this is negative—in fact, for most communication complexity problems, quantum communication does not help much.

An important example is the inner product function $[\mathrm{IP}(x,y)=x\cdot y=\Sigma_{i=1}^{n}x_iy_i(\mathrm{mod}\ 2)]$. All protocols, both classical and quantum, need to send about n bits or qubits to solve this. We sketch the proof of Cleve *et al.* (1998) here for the case of errorless quantum protocols with qubit communication and without entanglement; the proof for the more general case of entanglement is slightly more complicated. The proof uses the IP protocol to communicate Alice's n-bit input to Bob and then invokes Holevo's theorem to conclude that many qubits must have been communicated in order to achieve this. Suppose Alice and Bob have some protocol P for IP. They can use this to compute the following mapping:[7]

$$|x\rangle|y\rangle \mapsto |x\rangle(-1)^{x\cdot y}|y\rangle. \tag{16}$$

Now suppose Alice starts with an arbitrary n-bit state $|x\rangle$ and Bob starts with the uniform superposition $(1/\sqrt{2^n})\Sigma_{y\in\{0,1\}^n}|y\rangle$. If they apply the above mapping, the final state becomes

$$|x\rangle\frac{1}{\sqrt{2^n}}\sum_{y\in\{0,1\}^n}(-1)^{x\cdot y}|y\rangle.$$

If Bob applies a Hadamard transform to each of his n qubits, then he obtains the basis state $|x\rangle$, so Alice's n classical bits have been communicated to Bob. Holevo's theorem now implies that the IP protocol must communicate n qubits (which can trivially be achieved). The same argument can, with a minor modification, be made to work even if Alice and Bob share unlimited prior entanglement, yielding a lower bound of $n/2$ qubits (which can trivially be achieved using dense coding). With some more technical complication, the same idea gives a $\frac{1}{2}(1-2\epsilon)^2n$ lower bound for ϵ-error protocols (Cleve *et al.*, 1998). The constant factor in this bound was subsequently improved to the optimal $\frac{1}{2}$ by Nayak and Salzman (2002).

[7]This is an oversimplification of matters: in order to get the map of Eq. (16) one first needs to construct a new protocol P^{-1}, which is the reverse of the original communication protocol P. This can be done without error because the original protocol is without error. Combining protocols P and P^{-1} one can obtain map (16). If protocol P uses c qubits of communication, protocol P^{-1} also uses c qubits, and the protocol for obtaining state (16) uses $2c$ qubits. But the crucial point is that still at most c qubits are sent from Alice to Bob since P^{-1} is the reverse of P. Holevo's theorem imposes a lower bound on the communication from Alice to Bob, and hence we get a lower bound of n qubits on c.

IV. NONLOCALITY AND COMMUNICATION COMPLEXITY

A. Converting communication complexity to nonlocality

In Sec. II we introduced several simple nonlocality scenarios. Then in Sec. III we introduced communication complexity and gave several problems for which there are large, sometimes exponential, separations between the classical and quantum communication complexities. In this section we put together these two approaches and derive from the communication complexity problems new nonlocality problems which are very hard, sometimes exponentially hard, to solve in a classical model. In particular, we present nonlocality problems based on the distributed Deutsch-Jozsa problem and on the hidden matching problem. In Sec. VII we return to these nonlocality problems and discuss these newly developed tests in the context of experimental errors.

1. Mapping one-way quantum communication complexity to nonlocality

In this section we use the following mapping which, when applicable, is very powerful.

Consider a communication complexity problem where the number q of qubits exchanged in the quantum communication model with one-way communication from Alice to Bob is less than the number c of bits required to solve the problem classically when the parties have shared randomness. Further suppose that—due to some symmetry of the problem—it can be solved if Alice starts with an arbitrary basis state $|k\rangle$ (the value of k being known beforehand to both Alice and Bob) as follows: she carries out a transformation $U_A(x)$ on this state (that depends on her input x but does not depend on k), sends it to Bob who carries out a transformation $U_B(y)$ (that depends on his input y but does not depend on k), and then measures in the computational basis. The probability of finding result ℓ is thus $|\langle\ell|U_B(y)U_A(x)|k\rangle|^2$. From the knowledge of ℓ, k, and y, Bob can find the value of the function $f(x,y)$.

Now consider the following process: Alice and Bob share a maximally entangled state $|\psi\rangle=2^{-q/2}\Sigma_{i=0}^{2^q-1}|i\rangle|i\rangle$; Alice carries out a local transformation $U_A(x)^T$ (where T means transposition in the $|i\rangle$ basis); she measures in the computational basis. Bob carries out the transformation $U_B(y)$; he measures in the computational basis. Suppose that Alice obtains outcome k and Bob obtains outcome ℓ. The probability of finding these joint outcomes is

$$P(k,\ell|x,y) = |\langle \ell|\langle k|U_B(y)U_A(x)^T|\psi\rangle|^2 = 2^{-q}|\langle \ell|U_B(y)U_A(x)|k\rangle|^2$$

(the last equality is easy to check). If Alice now sends to Bob the outcome k of her measurement (which requires q bits), then Bob can compute $f(x,y)$. Thus this constitutes a solution of the communication complexity problem in the entanglement model with half the communication that would be required if they had used the trivial mapping based on teleportation. More importantly, the correlations $P(k,\ell|x,y)$ are nonlocal since they could not be obtained in a classical model with shared randomness without at least $c-q>0$ bits of classical communication.

B. Nonlocal version of the distributed Deutsch-Jozsa problem

The above mapping can be applied to the distributed Deutsch-Jozsa problem from Sec. III.D. We describe here the result of the mapping.

Nonlocal DJ problem. Let n be a power of 2: $n=2^m$. Alice and Bob receive n-bit inputs x and y that satisfy the DJ promise: either $x=y$ or x and y differ in exactly $n/2$ positions. The task is for Alice and Bob to provide outputs $a,b\in\{0,1\}^{\log_2 n}$ such that when $x=y$ then $a=b$ and when x and y differ in exactly $n/2$ positions then $a\neq b$.

They achieve this as follows:

(1) Alice and Bob share the maximally entangled state $(1/\sqrt{n})\Sigma_{i=0}^{n-1}|i\rangle|i\rangle$.

(2) Alice and Bob both apply locally a conditional phase to obtain $(1/\sqrt{n})\Sigma_{i=0}^{n-1}(-1)^{x_i}(-1)^{y_i}|i\rangle|i\rangle$.

(3) Alice and Bob both apply a Hadamard transform: $(1/n\sqrt{n})\Sigma_{a=0}^{n-1}\Sigma_{b=0}^{n-1}[\Sigma_{i=1}^{n}(-1)^{x_i+y_i+i\cdot(a\oplus b)}]|a\rangle|b\rangle$.

(4) Alice and Bob measure in the computational basis.

For every a, the probability that both Alice and Bob obtain the same result a is

$$\left|\frac{1}{n\sqrt{n}}\sum_{i=0}^{n-1}(-1)^{x_i+y_i}\right|^2,$$

which is $1/n$ if $x=y$ and 0 otherwise. Hence this solves the problem.

Note that if Alice then communicates the result of her measurement to Bob (using $\log_2 n$ bits), he could solve the distributed Deutsch-Jozsa problem since he could then check whether $k=\ell$ or $k\neq\ell$. But we know that solving the distributed Deutsch-Jozsa problem requires at least $0.007n$ bits. Thus we have a nonlocality problem that can be solved if Alice and Bob share $\log_2 n$ ebits but which requires about $0.007n$ bits to be solved in a classical model with shared randomness and classical communication. Note that this very large lower bound on the amount of classical communication would disappear in the bounded-error setting where we allow the correlations $P(a,b|x,y)$ to differ slightly from the ideal correlations.

C. Nonlocal version of the hidden matching problem

The same mapping can be applied to the hidden matching problem to yield a nonlocality problem.

Nonlocal HM problem. Assume that $n=2^m$, so we can index the numbers between 1 and n with m-bit strings.

Alice receives a string $x\in\{0,1\}^n$. Bob receives a perfect matching M on $\{1,\dots,n\}$ (i.e., a partition into $n/2$ disjoint pairs).

Alice must give as output some $k\in\{0,1\}^m$. Bob must give as output a matching $(i,j)\in M$ and $\ell\in\{0,1\}^m$.

Alice and Bob's output must satisfy $i\cdot(k\oplus\ell)+j\cdot(k\oplus\ell)=x_i+x_j \bmod 2$ (recall that $a\cdot b=\Sigma_i a_i b_i$ is the inner product between bit strings a and b and $a\oplus b$ is the bitwise XOR of a and b: the ith bit of $a\oplus b$ is $a_i\oplus b_i$).

Note that if at the end of the protocol Alice sends k to Bob at a cost of $m=\log_2 n$ classical bits, then Bob has enough information to compute the triple $(i,j,x_i\oplus x_j)$, i.e., to solve the hidden matching problem as defined in Sec. III.G. But we know that classical one-way communication from Alice to Bob needs about $\sqrt{n}$ bits to solve the hidden matching problem. Therefore the correlations in the nonlocal HM problem themselves can only be reproduced if Alice sends Bob at least about $\sqrt{n}$ bits of communication (if we are restricted to one way).

We now show that Alice and Bob can obtain the correlations of the nonlocal HM problem using local measurements on $m=\log_2 n$ ebits. The initial state is

$$\frac{1}{\sqrt{n}}\sum_{i\in\{0,1\}^m}|i\rangle|i\rangle.$$

Alice adds the phases $(-1)^{x_i}$. Bob views M as an orthogonal decomposition of the space $\mathbb{C}^n$ into $n/2$ two-dimensional subspaces. For instance, the projector for the subspace corresponding to $(i,j)\in M$ would be $P_{ij}=|i\rangle\langle i|+|j\rangle\langle j|$. Bob applies this measurement on the state he received and obtains the label of some random $(i,j)\in M$. This projects the joint state to

$$\frac{1}{\sqrt{2}}[(-1)^{x_i}|i\rangle|i\rangle+(-1)^{x_j}|j\rangle|j\rangle].$$

Now they both apply Hadamard transforms to each of their m qubits. This gives the state

$$\frac{1}{\sqrt{2}}\left(\frac{(-1)^{x_i}}{n}\sum_{k,\ell\in\{0,1\}^m}(-1)^{i\cdot k+i\cdot\ell}|k\rangle|\ell\rangle + \frac{(-1)^{x_j}}{n}\sum_{k,\ell\in\{0,1\}^m}(-1)^{j\cdot k+j\cdot\ell}|k\rangle|\ell\rangle\right) = \frac{1}{n\sqrt{2}}\sum_{k,\ell\in\{0,1\}^m}[(-1)^{x_i+i\cdot(k\oplus\ell)}+(-1)^{x_j+j\cdot(k\oplus\ell)}]|k\rangle|\ell\rangle.$$

Both parties measure their half of the state in the computational basis. They obtain m-bit strings k and ℓ, respectively, satisfying $x_i+i\cdot(k\oplus\ell)=x_j+j\cdot(k\oplus\ell) \pmod 2$, since the other k,ℓ pairs have amplitude 0. This gives $i\cdot(k\oplus\ell)+j\cdot(k\oplus\ell)=x_i+x_j \pmod 2$.

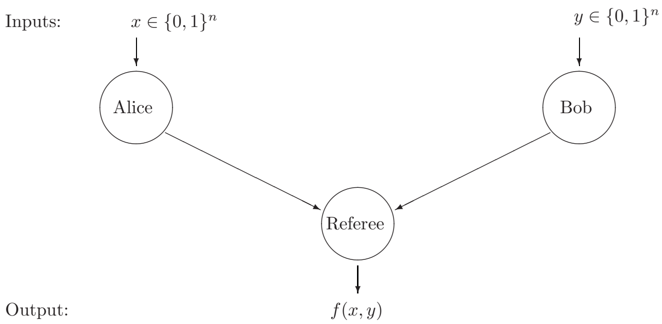

FIG. 5. The simultaneous message passing variant of the communication complexity scenario: Alice and Bob receive n-bit strings x and y, respectively, as input and their communication is restricted to each sending one message to a third party, called the referee. From these messages, the referee computes some function $f(x,y)$ as the output of the protocol. There are tasks of this form where communication in terms of quantum messages is exponentially more efficient than communication in terms of classical messages.

V. QUANTUM FINGERPRINTING AND THE SIMULTANEOUS MESSAGE PASSING MODEL

We now describe a model, called the simultaneous message passing (SMP) model, which is neither a nonlocality test nor the full-fledged communication complexity scenario, yet that is relevant to both. The basic structure is shown in Fig. 5.

Alice and Bob each receive an n-bit input (x and y, respectively). In this scenario, they do not have any shared resources such as shared randomness or an entangled state, but they do have local randomness. They each are required to send a single message to a third party, called the referee. The referee, upon receiving message m_A from Alice and m_B from Bob, should output the value of some (Boolean) function $f(x,y)$. The goal is to compute $f(x,y)$ with a minimum amount of communication from Alice and Bob to the referee. This scenario was introduced by Yao (1979) for the setting where m_A and m_B are classical messages consisting of bits. We compare this classical model to the corresponding quantum version, where m_A and m_B consist of qubits. We shall see that for the very natural problem of equality, where $f(x,y)=1$ if and only if $x=y$, there is an exponential savings in communication when qubits are used instead of classical bits. Classically, the problem of the bounded-error communication complexity of equality in the SMP model was open for almost 20 years until Newman and Szegedy (1996) exhibited a lower bound of about $\sqrt{n}$ bits. This is tight since Ambainis (1996) constructed a bounded-error protocol for this problem where the messages are $O(\sqrt{n})$ bits long (we describe a slightly less efficient classical protocol in Sec. V.B). In contrast, Buhrman *et al.* (2001) showed that in the quantum setting this problem can be solved with little communication: only $O(\log_2 n)$ qubits suffice.

A. Quantum fingerprints

In order to construct the efficient quantum SMP protocol for equality, we need to borrow ideas from the efficient classical randomized communication complexity protocol for equality from Sec. III.A. Recall that in that protocol Alice interprets her input x as a polynomial $p_x(t)=\sum_{i=1}^n x_i t^{i-1}$ over some finite field $\mathbb{F}$ of size m (about $3n$), and then she picks a random point $a\in\mathbb{F}$ and sends a and $p_x(a)$ to Bob. The pair $a,p_x(a)$ is called a "fingerprint" of x since it describes characteristics of x that can aid in identifying it. Carrying out this fingerprinting procedure in superposition results in a *quantum fingerprint* of x:

$$|F_x\rangle = \frac{1}{\sqrt{m}}\sum_{a\in\mathbb{F}} |a\rangle|p_x(a)\rangle.$$

Note that $|F_x\rangle$ consists of only $2\log_2 m = 2\log_2 n + O(1)$ qubits.

B. Classical protocol for equality

A nearly optimal[8] $O(\sqrt{n}\log_2 n)$ classical protocol for equality in the SMP model goes as follows. Alice produces a list of $k=O(\sqrt{n})$ random points $a_1,\ldots,a_k$ in $\mathbb{F}$ and sends the list $\{(a_i,p_x(a_i))\}_{i=1}^k$ to the referee. Bob does the same with respect to y, sending $\{(b_i,p_y(b_i))\}_{i=1}^k$ to the referee. By the birthday paradox (see the footnote in Sec. III.G), with constant probability there exist i and j such that both a_i and b_j equal the same field element d. In this case the referee can compare $p_x(d)$ with $p_y(d)$. If $x=y$ then $p_x=p_y$ and hence $p_x(d)=p_y(d)$. On the other hand, if $x\neq y$, then since p_x and p_y are *different* polynomials of degree at most $n-1$, with probability $\geqslant 2/3$, we have $p_x(d)\neq p_y(d)$. The protocol for the referee is now clear: if the lists of Alice and Bob have a point d in common, then the referee outputs 1 if and only if $p_x(d)=p_y(d)$. If there is no point in common (which happens only with small probability) or if $p_x(d)\neq p_y(d)$, then the referee outputs 0.

C. Quantum protocol for equality

We now have everything in place to describe the quantum protocol for equality. Alice sends state $|F_x\rangle$ to the referee and Bob sends $|F_y\rangle$. Note that if the referee now measures $|F_x\rangle$ in the computational basis, then he will find a random point a and the value $p_x(a)$, just like the classical protocol described above. The referee thus needs to do something smarter. The key observation is the following about the inner products between fingerprints:

$$\langle F_x|F_y\rangle = \begin{cases} 1 & \text{if } x=y \\ \leqslant \frac{1}{3} & \text{if } x\neq y. \end{cases} \tag{17}$$

If $x=y$ then clearly $\langle F_x|F_y\rangle=1$. If $x\neq y$ then

[8] Ambainis's protocol [Ambainis (1996)] gets rid of the $\log_2 n$ factor.

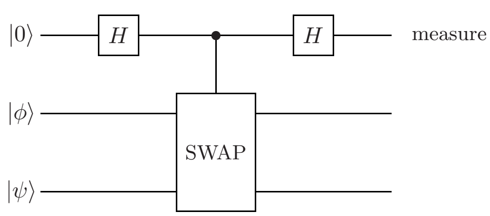

FIG. 6. Quantum circuit to test if $|\phi\rangle = |\psi\rangle$ or $|\langle\phi|\psi\rangle| \leq \frac{1}{3}$.

$$\langle F_x | F_y \rangle = \frac{1}{m} \sum_{i,j \in \mathbb{F}} \langle p_x(i) | \langle i | j \rangle | p_y(j) \rangle = \frac{1}{m} \sum_{i \in \mathbb{F}} \langle p_x(i) | p_y(i) \rangle .$$

Since p_x and p_y are different polynomials of degree at most $n-1$, they have the same value $p_x(i) = p_y(i)$ for at most $n-1$ values of i. Hence the inner product is at most $(n-1)/m \leq \frac{1}{3}$.

When Alice and Bob send their quantum fingerprints to the referee, he has to determine the inner product between the two states he receives. The following test (Fig. 6), sometimes called the SWAP test, accomplishes this task with a small error probability.

This circuit first applies a Hadamard transform to a qubit that is initially $|0\rangle$, then swaps the other two registers conditioned on the value of the first qubit being $|1\rangle$, then applies another Hadamard transform to the first qubit and measures it. Here SWAP is the operation that swaps the states $|\phi\rangle$ and $|\psi\rangle$: $|\phi\rangle|\psi\rangle \mapsto |\psi\rangle|\phi\rangle$. The referee receives $|\phi\rangle$ from Alice and $|\psi\rangle$ from Bob and applies the test to these two states. An easy calculation reveals that the outcome of the measurement is 1 with probability $(1 - |\langle\phi|\psi\rangle|^2)/2$. Hence if $|\phi\rangle = |\psi\rangle$ then we observe 1 with probability 0, but if $|\langle\phi|\psi\rangle| \leq \frac{1}{3}$ then this probability is $\geq \frac{4}{9}$. Repeating this procedure with several individual fingerprints can make the error probability arbitrarily close to 0.

D. Subsequent work in the SMP model

After the quantum fingerprinting scheme showed the power of quantum communication in the SMP model, a number of further results appeared. Yao (2003) exhibited an efficient protocol for testing if the inputs x and y are at some constant Hamming distance d, while Gavinsky, Kempe, and de Wolf (2006) related quantum fingerprinting to a technique from machine learning which brings out its weaknesses. One can also study the variant of the SMP model where Alice and Bob start with a shared entangled state but can only send classical messages to the referee. Gavinsky, Kempe, Regev, and de Wolf (2006) exhibited a problem based on the hidden matching problem and a quantum protocol that solves it with $O(\log_2 n)$ ebits and $O(\log_2 n)$ classical bits of communication, while any quantum SMP protocol without prior entanglement needs to send at least about $(n/\log_2 n)^{1/3}$ qubits. This shows that entanglement can reduce communication (even quantum communication) exponentially, at least for relational problems in the SMP model.[9] Finally, Gavinsky *et al.* (2008) showed that if Alice's message to the referee is allowed to be quantum, while Bob's message can only be classical, then the quantum advantages over purely classical protocols mostly disappear. In particular, the equality problem requires communication at least $\sqrt{n/\log_2 n}$ in this hybrid case.

VI. OTHER ASPECTS OF QUANTUM NONLOCALITY

A. Nonlocal boxes

In previous sections we studied a hierarchy of resources. In particular, we discussed and compared the correlations $P(a,b|x,y)$ that can be obtained using only shared randomness, by local measurements on entangled states, and finally those that can be obtained if communication between the parties is allowed. In this section we discuss an interesting set of correlations that lies between the last two classes.

To understand these new correlations, note that any correlations $P(a,b|x,y)$ obtained in a local hidden variable model or by local measurements on an entangled state must obey the following properties:

$$\text{Positivity:} \quad P(a,b|x,y) \geq 0, \tag{18}$$

$$\text{Normalization:} \quad \sum_{a,b} P(a,b|x,y) = 1, \tag{19}$$

No signaling:

$$\begin{aligned} &\sum_b P(a,b|x,y) = P(a|x) \text{ is independent of } y, \\ &\sum_a P(a,b|x,y) = P(b|y) \text{ is independent of } x. \end{aligned} \tag{20}$$

The last condition expresses the fact that Bob cannot transmit any information about his input y to Alice, and similarly Alice cannot communicate to Bob any information about her input x. We are interested here in correlations that obey the above three conditions, but that cannot be obtained from local measurements on entangled states.

To illustrate this idea, suppose that Alice and Bob each have some kind of device [introduced independently by Khalfi and Tsirelson (1985) and by Popescu and Rohrlich (1994)] such that Alice can provide an input $x \in \{0,1\}$ to her device and obtain an output $a \in \{0,1,\}$ and Bob can provide an input $y \in \{0,1\}$ to his device and obtain an output $b \in \{0,1\}$, such that the probabilities of the outputs given the inputs obey

[9] Recently, Gavinsky (2008b) extended this to a similar separation in the more standard two-way model.

$$P(a,b|x,y) = \begin{cases} \frac{1}{2} & \text{if } a \oplus b = x \wedge y \\ 0 & \text{otherwise.} \end{cases} \quad (21)$$

Note that, much like the correlations that can be established by use of quantum entanglement, this device is atemporal: Alice gets her output as soon as she feeds in her input, regardless of if and when Bob feeds in his input, and vice versa. Also inspired by entanglement, this is a one-shot device: the correlation appears only as a result of the first pair of inputs fed in by Alice and Bob. This device obeys conditions (1)–(3) above, so it cannot be used to signal. We call it a nonlocal (NL) box [other terminology in use is PR box, in reference to Popescu and Rohrlich (1994)].

With this device Alice and Bob always obtain $a \oplus b = x \wedge y$, whereas we know that for local measurements on entangled quantum states this relation can only be satisfied with probability at most $\cos^2(\pi/8)$ under the uniform distribution on the inputs x and y (see Sec. III.C for a proof). Thus this is an "imaginary" device in the sense that it cannot be realized physically without Alice and Bob's devices being connected by some kind of communication channel. It is, however, an interesting resource to consider since it is "stronger" than correlations that can be obtained from local measurements on entangled states but "weaker" than actual communication.

A systematic study of the properties of correlations obeying the above three conditions was initiated by Barrett, Linden, *et al.* (2005), and it was shown that they obey properties that one thinks of as genuinely quantum, such as monogamy and no cloning (Masanes *et al.*, 2006). They also allow for secure key distribution (Barrett, Hardy, and Kent, 2005).

Because of the apparent "reasonableness" of the nonlocal box, Popescu and Rohrlich (1994) raised the question why such correlations cannot be realized in nature without communication between the parties. The most straightforward answer is the technical proof in Sec. II.C; however, one might seek a more intuitive or philosophical explanation. One possible approach is provided by communication complexity. It was shown by van Dam (2000, 2005) and also noted by one of the authors of the present review (Cleve) that if Alice and Bob have an unlimited amount of nonlocal boxes then all communication complexity problems become trivial.

Suppose Alice and Bob have an unlimited supply of nonlocal boxes, as described in Eq. (21). Suppose Alice receives input $x \in \{0,1\}^n$ and Bob receives input $y \in \{0,1\}^n$. Then communication complexity becomes trivial in the sense that the value of any Boolean function $f(x,y) \in \{0,1\}$ can be computed with certainty with a *single* bit of communication from Alice to Bob.

To prove this, consider an arbitrary function $f:\{0,1\}^n \times \{0,1\}^n \to \{0,1\}$. It can be expressed as a Boolean circuit consisting of NOT and $\wedge$ (AND) gates, with inputs $x_1, \dots, x_n$ and $y_1, \dots, y_n$. The idea is to represent the value of each gate of this circuit in terms of two *shares*, one possessed by Alice and the other by Bob. For a bit a, its representation as shares is any (a', a''), where $a = a' \oplus a''$. Until the end of the protocol, Alice's information about each gate will be just the first bit of its share and Bob's information will be the second bit. They start by constructing shares of the input bits: $(x_i, 0)$ for each of Alice's input bits x_i (Bob does not need to know x_i to construct his share 0) and similarly $(0, y_i)$ for each of Bob's input bits y_i. For each gate in the circuit, if Alice and Bob collectively know the input bits as shares then they can produce the shares for the output bit without any communication. For each NOT gate, Alice merely negates her share (and Bob does nothing to his share). For each $\wedge$ gate, assume that the shares of inputs are (a', a'') and (b', b''). The shares of the output should be (c', c'') such that

$$c' \oplus c'' = (a' \oplus a'') \wedge (b' \oplus b'') = (a' \wedge b') \oplus (a' \wedge b'') \oplus (a'' \wedge b') \oplus (a'' \wedge b''). \quad (22)$$

Consider the four terms arising above. Since Alice possesses a' and b', she can easily compute $a' \wedge b'$, and similarly Bob can compute $a'' \wedge b''$. The difficult terms are $a' \wedge b''$ and $a'' \wedge b'$ because they contain bits that are spread between Alice and Bob and this is where the nonlocal boxes are used. Alice and Bob use one nonlocal box to obtain bits d' and d'' so that $d' \oplus d'' = a' \wedge b''$. They use a second nonlocal box to obtain e' and e'' so that $e' \oplus e'' = a'' \wedge b'$. Then Alice sets her share to $c' = (a' \wedge b') \oplus d' \oplus e'$ and Bob sets his share to $c'' = (a'' \wedge b'') \oplus d'' \oplus e''$. Clearly,

$$\begin{aligned} c' \oplus c'' &= (a' \wedge b') \oplus (d' \oplus e'') \oplus (d' \oplus e'') \oplus (a'' \wedge b'') \\ &= (a' \wedge b') \oplus (a' \wedge b'') \oplus (a'' \wedge b') \oplus (a'' \wedge b'') \end{aligned} \quad (23)$$

as required. At the end, Alice and Bob possess shares for the value of f and Alice sends her 1 bit share to Bob, enabling him to compute the value of f.

Is this result specific to the nonlocal boxes of the form Eq. (21) (in which case it could be viewed as some kind of anomaly in the space of all possible no-signaling correlations) or does it hold for other no-signaling correlations? In particular, does it hold for *noisy* correlations? It was shown by Brassard *et al.* (2006) that the latter is the case if one slightly adapts the definition of what it means for communication complexity to be trivial.

Suppose Alice and Bob have an unlimited supply of noisy nonlocal boxes whose outputs satisfy Eq. (21) with probability $p \geqslant (3+\sqrt{6})/6 \approx 90.8\%$. Then communication complexity becomes trivial in the sense that there exists $q > 1/2$ (possibly depending on p but on no other parameter) such that, for any $n \geqslant 0$, if Alice receives input $x \in \{0,1\}^n$ and Bob receives input $y \in \{0,1\}^n$, then they can find with probability at least q the value of any Boolean function $f(x,y) \in \{0,1\}$ with a single bit of communication from Alice to Bob.

Note that this result does not hold if Alice and Bob share entangled states instead of (noisy) nonlocal boxes. Indeed this follows from the result of Cleve *et al.* (1998), discussed in Sec. III.H, that computing the inner product of two n-bit strings with success probability $q>1/2$ requires $O(n)$ bits of communication even if Alice and Bob have an unlimited supply of entangled particles.

Thus the fact that communication complexity is not trivial (i.e., that some communication complexity problems are hard whereas others are easy) can be viewed as a partial characterization of the nonlocal correlations that can be obtained by local measurements on entangled particles. Is this a complete characterization? In particular, what is the exact noise threshold p where nonlocal boxes with noise p render communication complexity trivial? The current bounds on p are $85.4\% \approx (2+\sqrt{2})/4 \leq p \leq (3+\sqrt{6})/6 \approx 90.8\%$. If the lower bound is the correct one, we would have an interesting answer to the question raised by Popescu and Rohrlich. We leave this as an open problem.

Another related open question arising by analogy with the process of entanglement purification (Bennett, Brassard, *et al.*, 1996) is whether it is possible to "purify" nonlocal boxes. That is, given a supply of nonlocal boxes that work correctly with probability p, is it possible to produce, using only local operations, a nonlocal box with a success probability greater than p? For a first step in this direction, see Forster *et al.* (2009).

B. Bell inequalities and Tsirelson bounds

As discussed in the previous section, there are correlations, such as the nonlocal box, that cannot be reproduced by local measurements on entangled particles, but that nevertheless obey the conditions of positivity, normalization, and no-signaling, see Eqs. (18)–(20). More generally, we would like to understand within the space of all possible correlations $\mathbf{P}=\{P(a,b|x,y)\}$ which ones can be obtained by using only shared randomness (i.e., by local hidden variable models), which ones can be realized by carrying out local measurements on entangled particles, and what are the ultimate limits set by Eqs. (18)–(20).

Answering this question would address the question raised by Popescu and Rohrlich mentioned above and would give us basic insights into communication complexity. Indeed it would allow us to understand quantitatively the differences between shared randomness, shared entanglement, and nonlocal correlations, each of which can be viewed as a different resource for communication complexity. For instance, answering this question can have immediate implications for communication complexity in the entanglement model, at least in the case where Alice and Bob use only one round of communication.

Before addressing this question it is useful to better understand the geometry of nonlocal correlations. To this end we introduce Bell expressions; that is, linear combinations of the correlations

$$C(\mathbf{P}) = \sum_{abxy} c_{abxy} P(a,b|x,y), \tag{24}$$

where c_{abxy} are real numbers. It is easy to show that the space of correlations that can be reproduced using local hidden variables (i.e., using only shared randomness) is a polytope; that is, it can be characterized by a finite number of inequalities, called Bell inequalities, of the form

$$C(\mathbf{P}) \leq C_{\mathrm{LHV}}. \tag{25}$$

To compute the maximum value allowed by local hidden variable (LHV) models, we can restrict ourselves to deterministic models, where $a=a(x)$ is a function of input x and $b=b(y)$ is a function of y. We then have

$$C_{\mathrm{LHV}} = \max_{a(x),b(y)} \sum_{xy} c_{a(x)b(y)xy}.$$

If we consider local measurements on entangled quantum states, then we have bounds of the form

$$C(\mathbf{P}) \leq C_{\mathrm{QM}}, \tag{26}$$

where

$$C_{\mathrm{QM}} = \max \sum_{abxy} c_{abxy} \langle \psi | \Pi_a(x) \otimes \Pi_b(y) | \psi \rangle,$$

where the maximum is taken over all states $|\psi\rangle$ and over all projective measurements $\{\Pi_a(x)\}$ (depending only on x) and projective measurements $\{\Pi_b(y)\}$ (depending only on y). (By projective measurements, we mean a set of projectors $\Pi_a=\Pi_a^2$ that sum to the identity $\Sigma_a \Pi_a = I$.) Recently it has been shown how the quantum value C_{LHV} could be bounded by a hierarchy of semidefinite programs (Navascues *et al.*, 2007), although the issue of whether this hierarchy converges remains open (Scholz and Werner, 2008).

If we impose only the no-signaling conditions, then we have

$$C(\mathbf{P}) \leq C_{\text{no signaling}}, \tag{27}$$

where the right-hand side is the maximum of Eq. (24) subject to Eqs. (18)–(20). Note that Eqs. (18)–(20) define another polytope, the no-signaling polytope, and the maximum value of $C(\mathbf{P})$ will be attained at a vertex of the polytope.

We illustrate the above concepts by a specific kind of Bell expression, called XOR nonlocal games (Cleve *et al.*, 2004). In this particular case, the outputs $a,b \in \{0,1\}$ are bits and we wish them to come as close as possible to satisfying a condition of the form

$$a \oplus b = f(x,y) \tag{28}$$

for all x,y. The most celebrated example is the CHSH case, where x,y are also bits and the condition is $a \oplus b = x \wedge y$ [see Eq. (7)].

In the case of XOR games, we take the constants c_{abxy} in Eq. (24) to have the form

$$c_{abxy} = w_{xy}(-1)^{a\oplus b\oplus f(x,y)} = m_{xy}(-1)^{a\oplus b}, \tag{29}$$

where $w_{xy} \geqslant 0$ can be thought of as the weight we give to the pair of inputs x,y, and $m_{xy} = w_{xy}(-1)^{f(x,y)}$. In the particular case of the CHSH expression, we take $m_{xy} = (-1)^{x\wedge y}$, resulting in the famous CHSH inequality.

When considering LHV theories, it is convenient to define new variables $A_x = (-1)^{a(x)}$ and $B_y = (-1)^{b(y)}$, whereupon the maximum value of the Bell expression reachable by LHV theories is

$$C_{\text{LHV}} = \max_{A_x, B_y \in \{+1,-1\}} \sum_{xy} m_{xy} A_x B_y.$$

In the case of local measurements carried out on entangled quantum states, we can write

$$\sum_{a,b} P(a,b|x,y)(-1)^a(-1)^b = \langle \psi | A_x \otimes B_y | \psi \rangle,$$

where $|\psi\rangle$ is the quantum state shared by Alice and Bob and A_x, B_y are Hermitian operators with eigenvalues in $\{+1,-1\}$. We now use the following result of Tsirelson (1987).

Suppose Alice and Bob measure observables A_x and B_y, both with eigenvalues in $\{+1,-1\}$, on a pure quantum state $|\psi\rangle \in \mathbb{C}^d \otimes \mathbb{C}^d$; then there are real unit vectors $\alpha(x), \beta(y) \in \mathbb{R}^{2d^2}$ such that for all x and y, $\langle \psi | A_x \otimes B_y | \psi \rangle = \alpha(x)\beta(y)$.

Thus we can reexpress the maximal value of C attainable by quantum mechanics as

$$C_{\text{QM}} = \max_{\alpha(x), \beta(y) \in \mathbb{R}^n} \sum_{xy} m_{xy} \alpha(x) \cdot \beta(y).$$

If we impose only the no-signaling conditions, then it is possible to satisfy Eq. (28) for all x,y by choosing $P(ab|xy) = 1/2$ if $a \oplus b = f(x,y)$ and $P(ab|xy) = 0$ if $a \oplus b \neq f(x,y)$. Hence the maximum value of the game is

$$C_{\text{no signaling}} = \sum_{xy} |m_{xy}|.$$

As an illustration, in the case of the CHSH inequality, the results of Sec. II.B can be reexpressed as stating that $C_{\text{LHV}} = 2$, $C_{\text{QM}} = 2\sqrt{2}$, and $C_{\text{no signaling}} = 4$.

Interestingly, the ratio between the LHV values and the quantum value can be bounded independent of the number of inputs x,y and the choice of matrix m_{xy} by Grothendiek's constant K_G, as first noted by Tsirelson (1987):

$$C_{\text{QM}} \leqslant K_G C_{\text{LHV}}.$$

A recent development of this line of work is the realization that, for certain Bell inequalities, a violation larger than a critical value $C(\mathbf{P}) > C_d$ guarantees that if the correlations are obtained by local measurements on an entangled quantum state, then the state belongs to a Hilbert space of dimension at least d^2 (i.e., Alice and Bob's spaces each has a dimension at least d) (Brunner, Pironio, *et al.*, 2008; Wehner *et al.*, 2008; Briët *et al.*, 2009; Vértesi and Pál, 2009). These Bell inequalities can thus be thought of as "dimension witnesses."

C. Classical simulation of quantum correlations and quantum communication

Consider a nonlocality experiment in which Alice and Bob share an entangled quantum state and carry out local measurements on this state or consider a quantum communication protocol in which Alice and Bob carry out several rounds of quantum communication and then carry out measurements on the quantum states. How much classical resources are required to reproduce these quantum experiments? The results from Secs. III and IV show that the classical resources must sometimes be larger, even exponentially larger, than the quantum resources. Is this the worst one can expect? What are good protocols to simulate the quantum experiments with classical resources? In this section we review progress on these questions, presenting a number of scenarios, and what resources can be used to simulate them. Note that we are not of course claiming that nature works as in these simulations but rather we are studying how one could mimic nature with these alternative resources.

1. When no communication is needed

When states are very noisy, it may be possible to simulate local measurements on them using only shared randomness even though the states are entangled. Werner's discovery of a family of states, now known as Werner states, for which such a simulation is possible (Werner, 1989), was one of the founding results of quantum information. Werner's model was restricted to local projective measurements. Later improvements include Acín *et al.* (2006) and Barrett (2002) where it was shown that simulations using only shared randomness can also exist when considering the more general case of local positive operator valued measures (POVMs),[10] which are the most general kind of measurement allowed by quantum mechanics.

2. One-way quantum communication

We first consider the very simple scenario where Alice wants to communicate a single qubit to Bob and Bob wants to carry out a projective measurement on the qubit. We formalize this simple scenario as follows.

a. Simulation of one-way communication of a single qubit and subsequent projective measurement

Alice receives as input a normalized vector $\vec{x} \in \mathbb{R}^3$, with length $\|\vec{x}\| = 1$, which describes the quantum state $\rho = I/2 + \vec{x} \cdot \vec{\sigma}/2$ where $\vec{\sigma} = (X,Y,Z)$ is the vector of nontrivial Pauli matrices from Eq. (14); Bob receives as input a normalized vector $\vec{y} \in \mathbb{R}^3$, which describes his pro-

[10] A POVM is a set $\{A_k\}$ of positive-semidefinite matrices that sum to identity $\Sigma_k A_k = I$. When applied to a quantum system in state ρ, the probability of obtaining measurement outcome k is $\text{Tr}(A_k \rho)$.

jective measurement $\vec{y}\cdot\vec{\sigma}$. Bob must output a bit b, with probabilities satisfying $P(b=0|\vec{x}\vec{y})-P(b=1|\vec{x}\vec{y}) = \mathrm{Tr}(\rho\vec{y}\cdot\vec{\sigma})$.

We can generalize this to the case where Alice sends n qubits to Bob and Bob carries out a POVM on the n qubits.

b. Simulation of one-way communication of n qubits

Alice receives as input the classical description of a quantum state $|\psi\rangle$, for instance, by giving her the values of the coefficients c_i of the state in a standard basis $|\psi\rangle = \Sigma_i c_i|i\rangle$. Bob is given the classical description of a measurement, for instance, by giving him the matrix elements of the POVM elements A_k in the standard basis. The task is for Bob to provide an outcome k, such that the probability of outcome k occurring is $P(k|\psi) = \langle\psi|A_k|\psi\rangle$.

These are communication complexity scenarios where Alice and Bob's inputs are infinite dimensional. If one allows for slight imperfections in the simulation, then one can truncate the description of the matrix elements of $|\psi\rangle$ and A_k and make the number of input bits finite. For instance, on Alice's side, if $|\psi\rangle$ corresponds to the quantum state of n qubits, then one can truncate the number of inputs to $O(n2^n)$ bits [by describing each coefficient c_i with $O(n)$ bits of precision]. If Alice sends her truncated input to Bob, then we have, up to a small error, a classical simulation [using $O(n2^n)$ bits] of any one-way quantum communication protocol in which n qubits are sent from Alice to Bob. One cannot hope to do much better than this since the HM problem of Sec. III.G exhibits an n vs $2^{\Omega(\sqrt{n})}$ gap between the quantum and classical one-way communication complexities (and this was further strengthened to two-way classical communication complexity by Gavinsky (2008a).

3. Entanglement simulation

We can also consider the case where Alice and Bob want to simulate local measurements on entangled quantum particles. The simplest nonlocality scenario occurs when Alice and Bob carry out projective measurements on a single ebit.

a. Simulation of projective measurements on a single ebit

Alice and Bob each receives as input a normalized vector in $\mathbb{R}^3, \vec{x}, \vec{y}$ with $\|\vec{x}\|=\|\vec{y}\|=1$, which describe their projective measurements $\vec{x}\cdot\vec{\sigma}$, $\vec{y}\cdot\vec{\sigma}$. Alice and Bob must each output a bit (a,b, respectively) such that the correlations obey

$$P(a=b|\vec{x},\vec{y}) - P(a\neq b|\vec{x},\vec{y}) = -\vec{x}\cdot\vec{y} = \langle\psi_-|\vec{x}\cdot\vec{\sigma}\otimes\vec{y}\cdot\vec{\sigma}|\psi_-\rangle,$$

where $|\psi_-\rangle = (|0\rangle|1\rangle - |1\rangle|0\rangle)/\sqrt{2}$ and such that the marginals $P(a|\vec{x},\vec{y})$ and $P(b|\vec{x},\vec{y})$ are uniform [i.e., $P(a=0|\vec{x},\vec{y}) = P(a=1|\vec{x},\vec{y}) = 1/2$, etc.].

This can be generalized to the case where Alice and Bob carry out POVMs on arbitrary entangled states of n qubits.

b. Simulation of entangled states of dimension 2^n

Alice and Bob share a classical description of a pure entangled quantum state $|\psi\rangle_{AB}$, where Alice and Bob's systems are each of dimension 2^n. Alice and Bob receive as inputs x,y the classical (infinite-dimensional) descriptions of the measurements they should do (for instance, the inputs could consist in the matrix elements of the POVM elements in a standard basis). Alice and Bob must provide outputs a,b such that the joint probability $P(a,b|x,y)$ equals the probability of getting measurement outcomes a and b when measurements x and y are carried out on state $|\psi_{AB}\rangle$.

If we have a simulation of one-way quantum communication, then we can transform it into a simulation of entanglement. To see this, note that one can rewrite the joint probabilities as $P(a,b|x,y) = P(a|x)P(b|x,y,a)$. The simulation is then as follows: Alice chooses a according to the probability distribution $P(a|x)$; she then sends Bob sufficient information so that he can choose an output b distributed according to $P(b|x,y,a)$. It is easy to show that for this second task [producing b distributed according to $P(b|x,y,a)$] it suffices for Alice to send Bob the measurement outcome and to describe to him the state onto which his system is projected after Alice's measurement.[11] Using this correspondence, we thus have a protocol which provides, up to a small error, a classical simulation [using $O(n2^n)$ bits of one-way communication] of any measurement on entangled states of n qubits.

4. Exact classical simulations

Remarkably it is also possible, at least in some cases, to perfectly simulate the quantum communication or quantum entanglement scenarios with *finite* classical communication. In such perfect simulations we do not tolerate any error. Of course such exact simulations are in principle not necessary if one wants to interpret the results of real experiments, as any real experiment will always have small imperfections. But these exact simulations are interesting for at least two reasons. On the one hand, they show that perfectly simulating quantum systems is not much more costly than approximately simulating them. On the other hand, these exact simulations have quite interesting structure.

Exact classical simulations of quantum correlations were first independently reported by Maudlin (1992), Brassard *et al.* (1999), and Steiner (2000). Here we review the subsequent works on this topic.

We first consider a weak model, where the *average* amount of classical communication is bounded (but in

[11]We can assume without loss of generality that Alice's POVM elements all have rank 1, which implies that conditional on the measurement outcome, Bob's state is pure.

the worst case the amount of classical communication may be infinite). This model was first used by Maudlin (1992) and Steiner (2000) in the context of classical simulation of a single ebit. In Massar *et al.* (2001) this approach was generalized to the simulation of communicating n qubits or the simulation of POVM measurements on n ebits using $O(n2^n)$ bits of two-way classical communication on average.

A stronger and more interesting model is when the amount of classical communication is bounded (even in the worst case). This model was introduced by Brassard *et al.* (1999). The simulations were improved and as shown by Toner and Bacon (2003), the classical simulation of projective measurements on a single ebit could be realized with a single bit of classical communication from Alice to Bob, and the communication of a single qubit could be simulated with 2 bits of communication. Note that these simulations use an infinite amount of shared randomness, a requirement that was shown by Massar *et al.* (2001) to be necessary when the amount of communication is bounded (in the worst case).

An even stronger model for the simulation of entanglement is for Alice and Bob to use nonlocal boxes as a resource rather than classical communication. Indeed, as discussed in Sec. VI.A, 1 bit of classical communication can be used to realize a nonlocal box, but a nonlocal box cannot be used to communicate. It was shown by Cerf *et al.* (2005) that simulating projective measurements on a single ebit could be carried out with the use of a single nonlocal box. A unified approach to protocols simulating a single ebit with 1 bit of communication or with one nonlocal box was presented by Degorre *et al.* (2005).

VII. IMPLEMENTATIONS

A. Inefficient detectors

1. The detection loophole

In this section we consider nonlocality scenarios, but put a constraint on the quantum model. We suppose that any measurement on a quantum system gives the results predicted by quantum mechanics with probability η and does not give any result with probability $1-\eta$.

The motivation for considering this model is that most quantum communication experiments use photons. Photons are very practical because they can be quite easily produced, manipulated, transmitted over long distances, and measured. Unfortunately photons get absorbed during transmission (in commercial optical fibers, photons have approximately 50% probability of being absorbed after traveling 15 km) and single-photon detectors have limited efficiency: they will sometimes not detect a photon even though it is present. These effects can be described by the above model.

In most nonlocality experiments to date, the detector efficiency η was so low that the correlations could be explained by a classical model using shared randomness and no communication (a local hidden variable model). This is called the detection loophole (Pearle, 1970). Thus, for instance, in the CHSH experiment the correlations can be explained by a local hidden variable model if $\eta \leq 2/(\sqrt{2}+1) \simeq 0.8284$. A detector efficiency better than this bound has not (yet) been achieved in experiments involving only photons.

One solution to the above problem is technological: one should use a quantum system on which measurements can be carried out with high efficiency. In this respect atoms or ions are particularly interesting because measurements on these systems can be carried out with essentially 100% efficiency. Thus experiments involving two entangled ions have been carried out in which the detection loophole was closed (Rowe *et al.*, 2001). However, these experiments have not yet allowed both the detection loophole and the locality loophole (i.e., carrying out both measurements at spatially separated locations) to be closed simultaneously.

Instead of improved technology, another solution to this problem is to develop new nonlocality tests that demonstrate nonlocality with low detector efficiency. As we discuss in the following, the communication problems and protocols developed in the previous sections can, in principle, be used to build such tests.

2. Communication complexity and the detection loophole

Communication complexity suggests that by increasing the dimension d of the entangled system under study, one can decrease exponentially (in d) the required efficiency of the detectors. Indeed, it appears that in many cases the minimum number c of bits of classical communication required to reproduce the quantum correlations is related to the minimum efficiency of the detectors required for the correlations to be nonlocal by $\eta \geq 2^{-O(c)}$. That there should be a relation between c and η was first noted by Gisin and Gisin (1999) and further studied by Massar (2002) and Buhrman *et al.* (2003).

To understand this relation we compare two classical schemes:

- In the first scheme, which was discussed in Secs. II and IV, the detectors have 100% efficiency and the parties have shared randomness and may exchange up to c bits of classical communication.
- In the second scheme, the parties have shared randomness, and each party has a detector of efficiency η. This means that each party will with probability η give an output and with probability $1-\eta$ produce no output. The detectors are assumed to be independent, so that the probability that both detectors give an output is η^2. In physics terminology this would be called a local hidden variable model with detector efficiency η. (We also consider below the case where one of the detectors has efficiency η and the other always gives a result, i.e., is 100% efficient.)

These two schemes can be related in a number of ways. The simplest relation is that any classical protocol with c bits of communication can be mapped into a classical protocol with no communication but with detector efficiency $\eta^2=2^{-c}$.

This mapping is very simple: Alice and Bob use shared randomness r, which is uniformly distributed over all possible conversations. Each party checks whether r is a conversation that is consistent with their input. If it is then they give the corresponding output, if it is not then they do not give any output. The probability that both Alice and Bob give an output is at least 2^{-c}.

This protocol is not perfect since the probability that the parties give an output may differ from one party to the other or from one input to the other. What is interesting is that in a number of cases the converse holds: if the quantum correlations cannot be reproduced with less than c bits of communication, then they can be reproduced without communication only if the detector efficiency η is less than $2^{-\Omega(c)}$.

A first example where this converse occurs is when bounds on c and on the minimum detection efficiency η can be obtained from the size of monochromatic rectangles (see Appendix B). This approach was implicit in Massar (2002) where it was shown that the correlations of the distributed Deutsch-Jozsa problem could not be reproduced by a local hidden variable model if $\eta \geqslant O(n^{3/4})2^{-0.0035n}$ when the inputs consist of n-bit strings, and hence the parties use a maximally entangled system of dimension n. Using the size of monochromatic rectangles was exploited more fully by Buhrman *et al.* (2003) in the context of a multipartite communication complexity problem and then extended by Buhrman *et al.* (2006) to take into account the possibility of errors. In particular, Buhrman *et al.* (2006) showed how one could obtain a lower bound $c \geqslant B_R$ on the minimum amount of communication required to reproduce the correlations, where B_R is a function of the size and discrepancy of rectangles. It then followed that the correlations could be obtained by a local hidden variable model with detectors of efficiency η only if $\eta \leqslant 2^{-B_R/n}$ (where n is the number of parties). If the rectangle lower bound on c is close to tight, then this implies the relation we mentioned above between c and η.

3. Asymmetric detection loophole

Another interesting example arises if we suppose that Alice's detector is inefficient, but that Bob's detector is perfect. This situation is motivated by the experimental situation reported by Moehring *et al.* (2004), where an ion is entangled with a photon. As discussed above, the measurements on the ion can be done with 100% efficiency, whereas those on the photon will be inefficient. The problem in which Alice's detector is inefficient and Bob's detector is perfect was previously investigated from the point of view of the detection loophole by Brunner *et al.* (2007) and Cabello and Larsson (2007) for entangled systems of dimension 2.

We prove in Appendix D that the hidden matching problem is particularly well adapted to this asymmetric scenario. Namely, we show the following.

Suppose Alice and Bob try to implement the hidden matching problem using $\log_2 n$ ebits, as discussed in Sec. III.G. Suppose that Alice's detector has efficiency η whereas Bob's detector has 100% efficiency. Then the correlations obtained by measuring the ebits cannot be reproduced by a classical model without communication if $\eta \geqslant 2^{-\Omega(\sqrt{n})}$, even allowing for a small error probability.

To our knowledge, this is the first time it is shown that an exponentially small detection efficiency can be tolerated when allowing for a small error probability.

B. Present and future experiments

1. Experimental quantum nonlocality

During the past there have been many experiments that studied the correlations exhibited by measurements on entangled quantum particles. Their main aim was to test quantum mechanics by comparing its predictions with those of hidden variable models. The short result is that the predictions of quantum mechanics have always been verified to very high precision. However, up to now some "loopholes" have always been left open, which allow the possibility of explaining the data with—admittedly contrived—local hidden variable models.

We review here how experiments on quantum nonlocality have been improved during the past decades. We then discuss how the insights from communication complexity suggest new experimental challenges. We also discuss experimental realizations of quantum communication complexity.

After the initial experiment by Freedman and Clauser (1972) on the correlations exhibited by entangled photons, the first qualitative advance was the experiments of Aspect that used time-varying analyzers in order to close the locality loophole. Indeed in previous experiments the measurements were kept fixed for long periods of time while experimental results were accumulated; then the measurements were changed and a new set of data was acquired for the new measurement setting. In Aspect's experiment (Aspect, Dalibard, and Roger, 1982) the measurement settings changed periodically in time. In the later experiment of Weihs *et al.* (1998), the measurement settings were chosen at random using a quantum random number generator.

Another important advance of the experiments of Aspect *et al.* (1981, 1982) was a very precise check that the measured correlations coincide with the correlations $P_{\mathrm{QM}}(ab|xy)$ predicted by quantum mechanics for local measurements on a maximally entangled state of two particles (earlier experiments were much more imprecise).

Some other noteworthy advances are nonlocality experiments in which the two particles were separated by a large distance of 10 km (Tittel *et al.*, 1998) and 50 km (Marcikic *et al.*, 2004); nonlocality experiments on bipartite entangled systems of dimension 3 (Vaziri *et al.*, 2002; Thew *et al.*, 2004); and nonlocality experiments on entangled states of three (Pan *et al.*, 2000; Rauschenbeutel

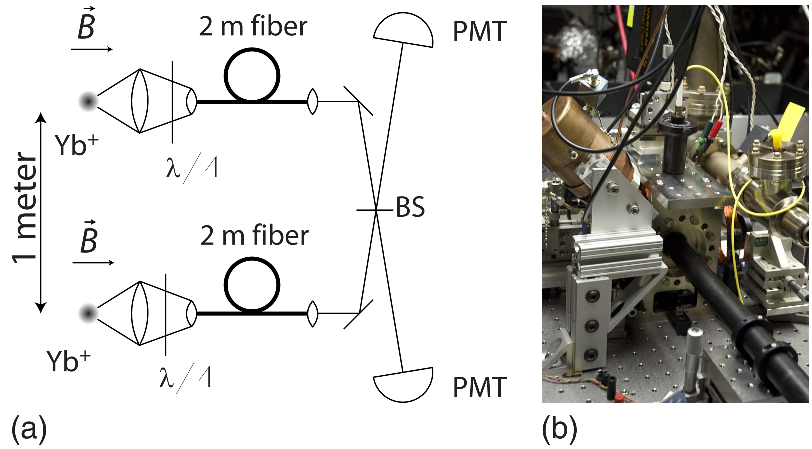

FIG. 7. (Color online) Bell inequality with two remote atomic qubits. The left-hand side is a schematic description of the experiment reported by Matsukevich *et al.* (2008) in which the internal states of two ions separated by about 1 m were entangled. Measurements on the two ions then allowed the violation of the CHSH inequality with the detection loophole closed. A series of laser pulses simultaneously excites both Yb^+ ions in such a way that when they deexcite they emit a photon whose polarization is entangled with the ion. A lens is used to couple the photons into optical fibers. The wave plate ($\lambda/4$) is used for convenience to convert circular polarization into linear polarization. The two photons interfere on a beam splitter (BS) and are detected by photomultiplier tubes (PMTs). Simultaneous detection of a photon by the two PMTs signals that the photons were in a Bell state, thereby realizing entanglement swapping: the two ions are now entangled. The internal states of the ions are then measured, enabling a violation of the CHSH inequality. Note that there are many inefficiencies in this experiment: only a fraction of emitted photons are coupled into the optical fibers and only a fraction of the photons reaching the PMTs are detected. But when two photons are detected, one knows with certainty that the two ions are entangled. The right-hand side is a photograph of one of the ion traps. The other trap is similar and located about 1 m away on the same optical table. Both courtesy of Monroe and Matsukevich.

et al., 2000) and four particles (Sackett *et al.*, 2000; Zhao *et al.*, 2003).

In the above experiments the detection loophole was not closed. This means that the raw data acquired during the experiment could be explained by a local hidden variable model. It was only by making the physically very reasonable assumption that the events in which the detector gives a click are independent of the measurement settings and measurement results (known in the physics literature as the "fair sampling assumption") that these experiments could be assumed to be in contradiction with local hidden variable models.

There have now been two experiments involving ions in which the detection loophole has been closed. In the first, the two entangled ions were separated by about 3 μm (Rowe *et al.*, 2001); in the second, presented in more detail in Fig. 7, the two entangled ions were separated by about 1 m (Matsukevich *et al.*, 2008). In view of these advances, closing both the locality and detection loopholes simultaneously does not seem out of reach.

From the point of view of communication complexity, closing the detection loophole is more important than closing the locality loophole. Indeed, if the detection loophole is not closed, it means that the raw data can be explained by a model without communication. On the other hand, if the detection loophole is closed, then, by sharing the entanglement, the parties have a resource that could only be reproduced classically by communication between the parties. The same is true in other applications of quantum nonlocality: closing the detection loophole (but not necessarily the locality loophole) allows one to increase the security of quantum key distribution (Acin *et al.*, 2007).

C. Future nonlocality experiments

The progress in quantum communication complexity points the way toward new tests of quantum nonlocality which use not one ebit, as in the CHSH test, but many ebits. Ideas for these new tests come from the entanglement-based Deutsch-Jozsa problem discussed in Sec. IV, the entanglement-based hidden matching problem discussed in Sec. III.G, recent work of Gavinsky (2008b), and also the (nonconstructive) results on three-party correlations reported by (Perez-Garcia *et al.*, 2008). There are at least two motivations for such experiments. First, they could be more robust against experimental imperfections (such as the detection loophole or errors) than nonlocality tests used at present. Second, they could illustrate the efficiency of quantum mechanics over classical mechanics as experiments on a small number e of ebits could only be reproduced classically using an exponentially large (in e) amount c of classical communication.

These nonlocality experiments on many ebits can be characterized by several parameters. In particular, these would include the number e of ebits involved or equivalently the dimension $d=2^e$ of the entangled quantum system; the minimum detector efficiency η required for the correlations to be nonlocal; the amount ϵ of errors that can be tolerated; and the amount c of classical communication that would be required to reproduce the quantum correlations. In general, for any given nonlocality test, we can expect trade-offs between η, ϵ, and c.

Note that the proposals inspired by communication complexity typically are asymptotic results that deal with the limit where the number of ebits tends toward infinity: $e \to \infty$. However, real experiments will deal with small values of e. For instance, if we think of the detection loophole, one should recall that this is only a problem for experiments dealing with entangled photons. On the other hand, the Hilbert space of a single photon can be larger than 2. One can thus effectively manipulate more than one qubit while manipulating only a single photon. This is potentially an interesting opportunity. Indeed it would be interesting to devise nonlocality experiments that tolerate inefficient detectors (say, $\eta < 10\%$) in Hilbert spaces of moderate dimension (say, $d=10$). If one could devise such a nonlocality experiment, there would be a strong incentive to realize it experimentally. Indeed whereas experiments involving entangled atoms or ions may be the short-term solution to solving the detection loophole, such experiments are much slower and much more expensive than experiments involving photons only. Numerical searches for such a nonlocality experiment have been undertaken but unsuccessfully so far (Massar *et al.*, 2002).

In summary, quantum communication complexity suggests the possibility of new nonlocality experiments on a moderate number of ebits that either are resistant to imperfections or require very large amounts of classical resources to reproduce classically. Realizing such experiments will require further progress on the theoretical and experimental sides.

1. Experimental communication complexity

The experimental situation concerning communication complexity proper is less advanced. Indeed, in order to carry out any nontrivial experimental demonstration of communication complexity, one needs to take into account the limited efficiency of detectors which has been such a plague for nonlocality experiments. In this respect, the first convincing communication complexity experiment to date is that reported by Trojek *et al.* (2005) in which six parties, materialized by wave plates along a beam on an optical table, carried out the communication complexity problem proposed by Cleve and Buhrman (1997) and Buhrman *et al.* (1999, 2001) but in the version proposed by Galvão (2001), which does not use entanglement. In this experiment the limited efficiency of detectors was explicitly taken into account. Experiments that studied the entanglement-based version of this problem while explicitly taking into account the limited efficiency of the detectors have also been reported (Zhang *et al.*, 2007) based on the proposal of Cabello and López-Tarrida (2005).

Another protocol which has been studied experimentally is quantum fingerprinting which in the SMP model performs exponentially better than classical protocols (see Sec. V.C). The possibility of realizing such an experiment on a small scale involving one or a few photons has been discussed by de Beaudrap (2004) and Massar (2005), and later performed using photons (Horn *et al.*, 2005) and in NMR (Du *et al.*, 2006).

In the future we may expect further proof-of-principle experiments of quantum communication complexity involving the exchange of more qubits and larger distance between the parties. Good candidates for such experiments are Raz's communication complexity problem, the hidden matching problem and its extensions, and quantum fingerprinting.

VIII. CONCLUSION

A. Open questions

Quantum communication complexity and quantum nonlocality are by now mature fields. But many questions remain open. Here we collect a few.

(1) *Additional natural problems in quantum communication complexity*. Find additional problems—if possible, natural problems that could have potential applications—for which quantum communication is much more efficient than classical commmunication.

(2) *How much entanglement is needed to get a reduction of communication: Equivalence of quantum communication and entanglement models of communication complexity*. In the entanglement model of communication complexity, the parties have an unlimited supply of entanglement and use it to reduce the amount of classical communication. How much entanglement is really needed? In classical communication complexity with shared randomness Newman's theorem (Newman, 1991) states that, if we allow a small increase in the error probability, the parties need only have $O(\log_2 n)$ shared random bits (where n is the size of the inputs). Does something similar hold when we replace shared randomness by entanglement? Answering this question would essentially establish whether the quantum communication and the entanglement models of communication complexity are equivalent.

(3) *Are most quantum states useful for communication complexity*? It was recently shown by Gross *et al.* (2009) that most n-qubit states (with respect to the uniform measure) are not useful—they are typically too entangled—in the measurement-based version of quantum computation. Are most states useful for communication complexity? For two parties the answer is yes, as we can work in the Schmidt basis. But consider three parties sharing a random state of $3n$ qubits (each party having n qubits). How useful are most states for communication complexity (asymptotically as n tends to infinity)?

(4) *Find new nonlocal games qualitatively different from existing ones*. In particular, consider the following more specific subquestions:

- For two-party XOR games, the ratio between the classical and quantum values of the game is bounded by a constant. However, in Perez-Garcia *et al.* (2008) showed—using a nonconstructive proof—that this is not the case for three-party games. Can one exhibit an explicit example of this type?

- Find Bell inequalities involving rather small systems, say where the dimension of each party's Hilbert space is less than $d=10$, which allow for very small detector efficiencies.

(5) *Nonlocal boxes and communication complexity*. As discussed in Sec. VI.A, nonlocal boxes are an interesting resource to consider from the point of view of communication complexity. In this regard, two interesting questions are the following:

- First, what is the noise threshold below which nonlocal boxes make communication complexity trivial [see Brassard *et al.* (2006) for a formulation of this problem]? Is this threshold the maximum value $p=(2+\sqrt{2})/4$ attainable by local measurements on entangled quantum systems?
- Second, is it possible to amplify nonlocal correlations in the sense that given a large number of devices that will produce correlations $P(ab|xy)$ corresponding to PR boxes with noise p, is it possible to use the devices in such a way as to produce correlations with a lower value of p? First results in this direction can be found in Brunner and Skrzypczyk (2009) and Forster *et al.* (2009).

(6) *Simulation of quantum correlations and quantum communication*. In this context, some questions that come to mind are the following:

- Exact simulation of more than one qubit or ebit using bounded classical communication (in the worst case) or nonlocal boxes. Some preliminary results on this topic have been obtained in the particular case where Alice and Bob carry out measurements with binary outcomes (Degorre *et al.*, 2007; Regev and Toner, 2007).
- The simulation of nonmaximally entangled states using nonlocal boxes. This appears to be much harder than the simulation of maximally entangled states [see Brunner *et al.* (2005) and Brunner, Gisin, *et al.* (2008) for some first results].
- The simulation of multipartite nonlocal correlations.

B. What have we learned from quantum communication complexity?

Communication complexity is a task for which quantum information can beat classical information. Such tasks are rare, and finding more potential applications of quantum information is important.

Unfortunately most quantum communication complexity problems are extremely sensitive to noise, highly contrived, or do not offer exponential gains over the best classical protocols (in which case the advantages of quantum communication will probably be more than offset by the lower cost and higher speed of classical communication). The most interesting proposal so far is maybe the SMP model without shared randomness (a somewhat contrived model) where equality (a very natural problem) can be solved exponentially more efficiently using quantum communication. Thus there is the tantalizing possibility that some time in the future, quantum communication complexity could be used in practical applications.

Independently of whether quantum communication complexity ever finds some real-world applications, the results obtained so far have important conceptual implications. First, they offer new insights into the power of quantum information and, in particular, of quantum computing. Indeed the basic aim of computer science, taken in a wide sense, is to accomplish a task using the minimum amount of resources. In the usual formulation, the resource that we want to minimize is the running time of the computer. This is the most important application of quantum computing as Shor's algorithm suggests that a quantum computer would allow exponential speedups. But in this context it is very difficult—if not impossible—to prove that quantum computers are more powerful than classical computers. The advantage of quantum computation can, however, be proven in simpler contexts such as the black-box model of quantum computing, where the resource that is quantified is the number of calls to an oracle, or communication complexity where the resource that is quantified is the amount of communication. The existence of these models where it can be rigorously shown that quantum information offers important advantages over classical information reinforces our confidence that quantum computers are much more powerful than classical computers for certain tasks.

Second, the study of quantum communication complexity has led to the proposal of new tests of quantum mechanics. Indeed from Bell onwards it was known that if one wants to replace quantum mechanics by a classical model, this classical model would have to use faster than light signaling. The discovery of fast quantum algorithms suggested that such a classical model would use an exponentially larger amount of resources. Quantum communication complexity has now advanced to the point where it may be possible to propose experiments in which one can prove that a classical simulation would require exponentially more resources than are used quantum mechanically.

In summary, quantum communication complexity is now a mature field that has led to some fundamental insights into the nature of computation and the foundations of physics.

ACKNOWLEDGMENTS

H.B. and R.d.W. were partially supported by a Vici and a Vidi grant, respectively, from The Netherlands Organization for Scientific Research (NWO) and by the

European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate under Contract No. 015848. R.C. was partially supported by Canada's NSERC, CIFAR, QuantumWorks, MITACS, and the U.S. ARO. S.M. was partially supported by the Interuniversity Attraction Poles Programme, Belgian State, Belgian Science Policy under Grant No. IAP6-10 and by the EU project QAP under Contract No. 015848.

APPENDIX A: NAYAK'S PROOF OF A CONSEQUENCE OF HOLEVO'S BOUND

Here we prove that if we are encoding n bits in d-dimensional quantum states, then the average recovery probability is at most $d/2^n$. Therefore, an exact procedure requires $d \geqslant 2^n$ and thus at least n qubits.

Let $\rho_0, \ldots, \rho_{2^n-1}$ be the d-dimensional states that encode the elements of $\{0,1\}^n$ (which we identify with $\{0,1,\ldots,2^n-1\}$ in the obvious way). Let $E_0, \ldots, E_{2^n-1}$ be the measurement operators applied for decoding (they sum to the d-dimensional identity). The probability of successfully recovering $x \in \{0,1\}^n$ from its encoding is $\mathrm{Tr}(E_x \rho_x)$. Therefore, we can bound the success probability for a uniformly random $x \in \{0,1\}^n$ by

$$\frac{1}{2^n}\sum_{x=0}^{2^n-1}\mathrm{Tr}(E_x\rho_x) \leqslant \frac{1}{2^n}\sum_{x=0}^{2^n-1}\mathrm{Tr}(E_x) = \frac{1}{2^n}\mathrm{Tr}\left(\sum_{x=0}^{2^n-1}E_x\right) = \frac{1}{2^n}\mathrm{Tr}(I) = \frac{d}{2^n}. \tag{A1}$$

The first inequality follows because the density operator ρ_x is positive semidefinite and has trace 1; therefore it can be unitarily diagonalized: $U^*\rho_x U = D$, where D is diagonal with diagonal entries that are non-negative and sum to 1. Because the trace is invariant under cyclic permutations of the matrices, we now have

$$\begin{aligned}\mathrm{Tr}(E_x\rho_x) &= \mathrm{Tr}(U^*E_xUU^*\rho_xU) \\ &= \mathrm{Tr}(U^*E_xUD) \leqslant \mathrm{Tr}(U^*E_xUI) = \mathrm{Tr}(E_x).\end{aligned}$$

APPENDIX B: RECTANGLES AND THE LOWER BOUND FOR DISTRIBUTED DEUTSCH-JOZSA

Separations between quantum and classical communication complexities always require two things: an efficient quantum protocol for some problem and a *lower bound* on the communication of all classical protocols solving that same problem. In this appendix we give some tools for lower bounding classical communication complexity, leading eventually to the lower bound on classical protocols for the distributed Deutsch-Jozsa problem mentioned in Sec. III.D.

1. Rectangles

Consider some communication complexity problem $f: X \times Y \to \{0,1\}$, where Alice starts with an input $x \in X$ and Bob starts with an input $y \in Y$. We start by introducing the crucial combinatorial notion for classical lower bounds. A *rectangle* is a set $R \subseteq X \times Y$ that is of the form $R = A \times B$ with $A \subseteq X$ and $B \subseteq Y$. For example, if $n=2$ and $A=\{00,01\}$, $B=\{01,10\}$ then $R=A\times B$ $=\{(00,01),(00,10),(01,01),(01,10)\}$ is a rectangle. The following result is a fundamental property of classical deterministic protocols.

Lemma 1. If a deterministic protocol has communication c, then there exist 2^c rectangles $R_1, \ldots, R_{2^c}$ that partition, $X \times Y$ such that the protocol gives the same output a_i for each $(x,y) \in R_i$.

We omit the easy proof of this lemma, which is by induction on c. For example, suppose there is only one k-bit message m going from Alice to Bob and then Bob returns the 1-bit output. Then the 2^{k+1} rectangles would be of the form $R_{m,a} = A_m \times Y_{m,a}$, with $m \in \{0,1\}^k$ and $a \in \{0,1\}$, where A_m is the set of x's for which Alice sends k-bit message m and $Y_{m,a}$ is the set of y's for which Bob returns output a when receiving message m. Note that if our protocol computes f correctly, then the rectangles are "monochromatic": the protocol returns the same answer $f(x,y)$ for all $(x,y) \in R_i$.

As a simple application of this we prove the so-called "rank lower bound." Consider some communication complexity problem $f: X \times Y \to \{0,1\}$. Let M_f be the $|X| \times |Y|$ matrix whose entries are defined by $M_f(x,y) = f(x,y)$. This is called the communication matrix of f. It can be viewed as a two-dimensional truth table. We use rank(f) to denote the rank of this matrix over the field of real numbers. For example, the communication matrix for the equality function is the $2^n \times 2^n$ identity matrix, which has 1's on its diagonal and 0's elsewhere. Hence $\mathrm{rank}(\mathrm{EQ}) = 2^n$.

Suppose we have some c-bit deterministic protocol that computes f. We know that this partitions the input space $X \times Y$ into rectangles $R_1, \ldots, R_{2^c}$. Since each 1-input (x,y) occurs in exactly one 1-rectangle, we have

$$M_f = \sum_{i:a_i=1} R_i,$$

where we view R_i as a $2^n \times 2^n$ matrix with 1's on its elements and 0's elsewhere. Note that R_i is a matrix of rank 1. Hence, using $\mathrm{rank}(A+B) \leqslant \mathrm{rank}(A) + \mathrm{rank}(B)$, we get

$$\begin{aligned}\mathrm{rank}(M_f) = \mathrm{rank}\left(\sum_{i:a_i=1} R_i\right) &\leqslant \sum_{i:a_i=1} \mathrm{rank}(R_i) \\ &= \sum_{i:a_i=1} 1 \leqslant 2^c.\end{aligned}$$

But that means that a lower bound on the rank of M_f implies a lower bound on the communication! In particular, it follows that for the equality problem, the communication c needs to be at least n.

2. Randomized protocols

In a randomized protocol, Alice and Bob may flip coins and the protocol has to output the right value $f(x,y)$ with probability $\geq 2/3$ for all $(x,y) \in \mathcal{D}$. We can fix these coins to obtain a deterministic protocol. Suppose randomized protocol A uses c bits of communication and has success probability $2/3$ on all inputs. Let $A(x,y,r_A,r_B)=1$ if the protocol gives the correct output $f(x,y)$ on input x,y using coin flips r_A for Alice and r_B for Bob, and $A(x,y,r_A,r_B)=0$ otherwise. For each x,y, a good randomized protocol satisfies

$$\mathbb{E}_{r_A,r_B}[A(x,y,r_A,r_B)] \geq 2/3,$$

where the expectation is taken over uniformly chosen strings r_A and r_B. Now let $\mu: X \times Y \to [0,1]$ be an input distribution. Then also

$$\mathbb{E}_{\mu,r_A,r_B}[A(x,y,r_A,r_B)] \geq 2/3,$$

where the expectation is taken over r_A, r_B, and x,y picked according to μ. By the averaging principle, there exists a way to fix r_A and r_B such that the success probability (under μ) of the resulting deterministic protocol is at least $2/3$. Accordingly, if we want to impose a lower bound on the randomized communication complexity of a function, it suffices to find some "hard" input distribution μ and to show that all deterministic protocols that have error at most $1/3$ under that distribution need a lot of communication.

The reason why the step to deterministic protocols is helpful is that deterministic protocols partition the input space into rectangles as seen before. Suppose we can show that all "large" rectangles in the communication matrix have roughly as many 0's as 1's in them (weighed according to μ). Then the protocol will make a large error on all large rectangles. Conversely, if we know the protocol does not make a large error, most of its rectangles must have been "small." But that can only be if there are many rectangles. Since the number of rectangles is 2^c, the communication c must have been large. This idea leads to the following lower bound method. The *discrepancy* of rectangle $R=A\times B$ under μ is the difference between the weight of the 0's and the 1's in that rectangle:

$$\delta_\mu(R) = |\mu[R \cap f^{-1}(1)] - \mu[R \cap f^{-1}(0)]|.$$

The discrepancy of f under μ is the maximum over all rectangles:

$$\delta_\mu(f) = \max_R \delta_\mu(R).$$

If f has small discrepancy, that means that all large rectangles are roughly balanced. Suppose a deterministic protocol partitions the input space into rectangles $R_1, \ldots, R_{2^c}$. Suppose it has success probability $1/2+\epsilon$. The best bias (difference between success and failure probabilities) that the protocol can achieve on rectangle R_i is $\delta_\mu(R_i)$ by giving the output with highest weight in that rectangle. The success probability is $\Sigma_i \mu[R_i \cap f^{-1}(a_i)]$ and the error probability is $\Sigma_i \mu[R_i \cap f^{-1}(1-a_i)]$, where a_i is the majority value of f on the pairs $(x,y) \in R_i$, weighted according to μ. Hence we have

$$2\epsilon \leq \sum_{i=1}^{2^c} \mu[R_i \cap f^{-1}(a_i)] - \sum_{i=1}^{2^c} \mu[R_i \cap f^{-1}(1-a_i)]$$
$$\leq \sum_{i=1}^{2^c} \delta_\mu(R_i) \leq 2^c \delta_\mu(f).$$

This is a lower bound on the communication: $c \geq \log_2[2\epsilon/\delta_\mu(f)]$. Accordingly, a distribution μ where $\delta_\mu(f)$ is small gives a lower bound on the communication of deterministic protocols for f under μ, and then the same lower bound applies to randomized protocols.

3. Discrepancy of the inner product function

To illustrate the discrepancy lower bound technique, we now consider the inner product function, defined by $\mathrm{IP}(x,y)=x\cdot y \pmod 2$. We show that its discrepancy under the uniform distribution is very small. We analyze the $2^n \times 2^n$ matrix M whose (x,y) entry is $(-1)^{x\cdot y}$. This is just the communication matrix for IP, with 0's replaced by 1's and 1's replaced by -1's. Lindsey's lemma shows that large rectangles in M are quite balanced.

Lemma 2 (Lindsey). For every rectangle $R=A\times B$, the absolute value of the sum of the M entries in that rectangle is at most $\sqrt{|A|\cdot|B|\cdot 2^n}$.

Proof: It is easy to see that M is symmetric and $M^2 = 2^n I$. This implies, for any vector v,

$$\|Mv\|^2 = v^T M^T M v = 2^n v^T v = 2^n \|v\|^2,$$

where the norm is the usual Euclidean vector length. Let $v_A \in \{0,1\}^{2^n}$ and $v_B \in \{0,1\}^{2^n}$ be the characteristic (column) vectors of sets A and B. The sum of the M entries in R is $\Sigma_{a\in A, b\in B} M_{ab} = v_A^T M v_B$. We can bound this using Cauchy-Schwarz inequality:

$$|v_A^T M v_B| \leq \|v_A\| \cdot \|M v_B\| = \|v_A\| \cdot \sqrt{2^n} \|v_B\|$$
$$= \sqrt{|A| \cdot |B| \cdot 2^n}.$$

■

Let $\mu(x,y)=1/2^{2n}$ be the uniform input distribution. Note that the discrepancy of the rectangle R under μ is exactly the difference of $+1$'s and -1's in R divided by 2^{2n}. By Lindsey's lemma, this is $\delta_\mu(R) \leq \sqrt{|A|\cdot|B|}/2^{3n/2}$. Because $|A|,|B| \leq 2^n$, it follows that the discrepancy of the inner product function under the uniform distribution is $\delta_\mu(\mathrm{IP}) \leq 2^{-n/2}$. Hence we get an $n/2$ lower bound on the randomized communication complexity of IP.

4. The lower bound for the distributed Deutsch-Jozsa problem

Recall the distributed Deutsch-Jozsa problem from Sec. III.D. Buhrman *et al.* (1998) used a combinatorial result of Frankl and Rödl (1987) to prove the following classical lower bound.

Theorem 3. Every deterministic classical protocol that solves the distributed Deutsch-Jozsa problem needs to communicate at least $0.007n$ bits.

Proof: Suppose there is a c-bit deterministic classical protocol for the problem. Each c-bit conversation corresponds to a rectangle $R = A \times B$, with $A, B \subseteq \{0,1\}^n$, such that the protocol has the same conversation and output if and only if $(x,y) \in R$. Since there are at most 2^c possible conversations, the protocol partitions $\{0,1\}^n \times \{0,1\}^n$ in at most 2^c different such rectangles. Now consider all n-bit strings x with Hamming weight $n/2$ (i.e., $n/2$ ones and $n/2$ zeros). There are $\binom{n}{n/2} \approx 2^n/\sqrt{n}$ of those. Since every (x,x) pair must occur in some rectangle and there are only 2^c rectangles, there is a rectangle $R = A \times B$ that contains at least $2^n/(\sqrt{n}2^c)$ different such (x,x) pairs. Let $S = \{x : |x| = n/2, (x,x) \in R\}$ be the set of such x. Since R contains some (x,x) pairs (on which the protocol outputs 1) and the protocol has the same output for all inputs in R, R cannot contain any 0-inputs. This implies that the Hamming distance of every pair $x, y \in S$ is different from $n/2$, for otherwise (x,y) would be a 0-input in R. Viewing the strings x in S as characteristic vectors of sets, it is easy to see that the size of the intersection of $x, y \in S$ is never $n/4$. Thus we have a set system S of at least $2^n/\sqrt{n}2^c$ sets over an n-element universe, such that the size of the intersection of any two sets in S is not $n/4$. However, by Corollary 1.2 of Frankl and Rödl (1987), such a set system can have at most 1.99^n elements, so we have

$$\frac{2^n}{\sqrt{n}2^c} \leqslant |S| \leqslant 1.99^n.$$

This implies $c \geqslant \log_2(2^n/\sqrt{n}1.99^n) \geqslant 0.007n$. ■

APPENDIX C: RAZBOROV'S LOWER BOUND FOR THE QUANTUM COMMUNICATION COMPLEXITY OF INTERSECTION

While the previous section discussed some basic methods for lower bounding classical communication complexity, here we focus on methods to lower bound quantum communication complexity (sometimes with prior entanglement).

1. The Kremer-Razborov-Yao lemma and its consequences

The following lemma is due to Razborov (2003) (Proposition 3.3) and is similar to earlier statements by Yao (1993) and Kremer (1995). It can intuitively be viewed as a quantum analog of the rectangle decomposition of classical protocols explained in Appendix B.1. We skip the easy proof, which is by induction on q.

Lemma 4 (Kremer-Razborov-Yao). Let $|\Psi\rangle$ denote the (possibly entangled) starting state of a quantum protocol that communicates q qubits of communication and has binary output. For all inputs x of Alice and y of Bob, there exist linear operators $A_h(x), B_h(y)$, for all $h \in \{0,1\}^{q-1}$, each with operator norm (i.e., largest singular value) at most 1, such that the acceptance probability (i.e., probability of output 1) of the protocol is

$$P(x,y) = \left\| \sum_{h \in \{0,1\}^{q-1}} [A_h(x) \otimes B_h(y)] |\Psi\rangle \right\|^2,$$

where the norm is the usual Euclidean vector length.

Consider the special case where the protocol starts without entanglement, so we can write $|\Psi\rangle = |\Psi_A\rangle|\Psi_B\rangle$. In this case we can rewrite the acceptance probabilities as

$$\begin{aligned}
P(x,y) &= \left\| \sum_{h \in \{0,1\}^{q-1}} [A_h(x) \otimes B_h(y)] |\Psi_A\rangle|\Psi_B\rangle \right\|^2 \\
&= \langle\Psi_A|\langle\Psi_B| \left(\sum_{h \in \{0,1\}^{q-1}} [A_h(x) \otimes B_h(y)] \right)^* \\
&\quad \times \left(\sum_{h' \in \{0,1\}^{q-1}} [A_{h'}(x) \otimes B_{h'}(y)] \right) |\Psi_A\rangle|\Psi_B\rangle \\
&= \sum_{h,h' \in \{0,1\}^{q-1}} \langle\Psi_A|A_h(x)^* A_{h'}(x)|\Psi_A\rangle \\
&\quad \times \langle \Psi_B|B_h(y)^* B_{h'}(y)|\Psi_B\rangle.
\end{aligned}$$

Let $a(x)$ be the 2^{2q-2}-dimensional row vector with (h,h') entry equal to $\langle\Psi_A|A_h(x)^* A_{h'}(x)|\Psi_A\rangle$, and similarly define column vector $b(y)$ with entries $\langle\Psi_B|B_h(y)^* B_{h'}(y)|\Psi_B\rangle$, then the last expression is just the scalar product $a(x)b(y)$. If we now define A to be the $|X| \times 2^{2q-2}$ matrix with rows $a(x)$, and B the $2^{2q-2} \times |Y|$ matrix with columns $b(y)$, then we have proved the following lemma.

Lemma 5. Consider a quantum communication protocol (without prior entanglement) on input set $X \times Y$, which communicates q qubits, with acceptance probabilities denoted by $P(x,y)$, with P as the corresponding $|X| \times |Y|$ matrix. There exist $|X| \times 2^{2q-2}$ matrix A and $2^{2q-2} \times |Y|$ matrix B, both with entries of absolute value at most 1, such that $P = AB$.

Note that the rank of matrix P is at most 2^{2q-2} since $\mathrm{rank}(AB) \leqslant \min(\mathrm{rank}(A), \mathrm{rank}(B))$. This allows us to generalize the classical rank lower bound from Appendix B.1 to the quantum domain. If we have a q-qubit protocol that computes some function $f: X \times Y \to \{0,1\}$ with success probability 1, then $P(x,y)$ equals $f(x,y)$, and the $|X| \times |Y|$ matrix P is actually the communication matrix M_f, whose (x,y) entry is $f(x,y)$. Hence we obtain a lower bound $q \geqslant \mathrm{rank}(P)/2 + 1 = \mathrm{rank}(M_f)/2 + 1$ on the quantum communication of protocols with success probability 1. Similarly, one can obtain lower bounds on the bounded-error quantum communication complexity by applying a lower bound on the rank needed for a matrix P that is close to the matrix of function values at each entry [since an ϵ-error protocol satisfies $|P(x,y) - f(x,y)| \leqslant \epsilon$ for all inputs].

Finally, we note without proof that one can also use the discrepancy method (Appendix B.2) to obtain a lower bound on quantum communication complexity

(Kremer, 1995) even for protocols with prior entanglement (Linial and Shraibman, 2007). Since the inner product function has very small discrepancy (Appendix B.3), we thus have another way of showing a linear lower bound for it, different from the one explained in Section III.H.

2. Translation from protocols to polynomials

The following key lemma is implicit in Razborov's paper (Razborov, 2003); the presentation given here is taken from Klauck *et al.* (2007). It allows us to translate the average acceptance probability of a q-qubit protocol (as a function of the intersection size i of the inputs x and y, viewed as subsets of $\{1,\ldots,n\}$) to a polynomial in i of degree roughly q. Accordingly, efficient protocols give low-degree polynomials.

Razborov's proof relies on the following linear algebraic notions. The operator norm $\|A\|$ of a matrix A is its largest singular value σ_1 (not to be confused with the Euclidean vector norm of Lemma 4). The trace inner product—also known as the Hilbert-Schmidt inner product—between A and B is $\langle A,B\rangle=\mathrm{Tr}(A^*B)$. The trace norm is $\|A\|_{\mathrm{tr}}=\max\{|\langle A,B\rangle|:\|B\|=1\}=\Sigma_i\sigma_i$, the sum of all singular values of A. The Frobenius norm is $\|A\|_F=\sqrt{\Sigma_{ij}|A_{ij}|^2}=\sqrt{\Sigma_i\sigma_i^2}$.

Lemma 6. Consider a quantum communication protocol (without prior entanglement) on n-bit inputs x and y, which communicates q qubits, with acceptance probabilities denoted by $P(x,y)$. Define

$$P(i)=\mathbb{E}_{|x|=|y|=n/4,|x\wedge y|=i}[P(x,y)],$$

where the expectation is taken uniformly over all x,y that each has weight $n/4$ and intersection i. For every $d\leqslant n/4$ there exists a degree-d polynomial q such that $|P(i)-q(i)|\leqslant 2^{2q-(d/4)}$ for all $i\in\{0,\ldots,n/8\}$.

Proof. We only consider the $\mathcal{N}=\binom{n}{n/4}$ strings of weight $n/4$. Let P denote the $\mathcal{N}\times\mathcal{N}$ matrix of the acceptance probabilities on these inputs. By Lemma 5, we can write $P=AB$, where A is an $\mathcal{N}\times 2^{2q-2}$ matrix with each entry at most 1 in absolute value and similarly for B. Note that $\|A\|_F,\|B\|_F\leqslant\sqrt{\mathcal{N}2^{2q-2}}$. By the Cauchy-Schwarz inequality for unitarily invariant norms [Bhatia (1997), p. 95], we have

$$\|P\|_{\mathrm{tr}}\leqslant\|A\|_F\cdot\|B\|_F\leqslant\mathcal{N}2^{2q-2}.$$

Let μ_i denote the $\mathcal{N}\times\mathcal{N}$ matrix corresponding to the uniform probability distribution on $\{(x,y):|x\wedge y|=i\}$. These "combinatorial matrices" have been well studied (Knuth, 2003). Note that $\langle P,\mu_i\rangle$ is the expected acceptance probability $P(i)$ of the protocol under that distribution. One can show that the different μ_i commute; thus they have the same eigenspaces $E_0,\ldots,E_{n/4}$ and can be simultaneously diagonalized by some orthogonal matrix U. For $t\in\{0,\ldots,n/4\}$, let $(UPU^T)_t$ denote the block of UPU^T corresponding to E_t and let $a_t=\mathrm{Tr}[(UPU^T)_t]$ be its trace. Then we have

$$\sum_{t=0}^{n/4}|a_t|\leqslant\sum_{j=1}^{\mathcal{N}}|(UPU^T)_{jj}|\leqslant\|UPU^T\|_{\mathrm{tr}}=\|P\|_{\mathrm{tr}}\leqslant\mathcal{N}2^{2q-2},$$

where the second inequality is a property of the trace norm.

Let λ_{it} be the eigenvalue of μ_i in eigenspace E_t. Knuth (2003) gave an exact combinatorial expression for λ_{it}. We will not state this explicitly here but just note that λ_{it} is a degree-t polynomial in i and that $|\lambda_{it}|\leqslant 2^{-t/4}/\mathcal{N}$ for $i\leqslant n/8$. Now consider the high-degree polynomial p defined by

$$p(i)=\sum_{t=0}^{n/4}a_t\lambda_{it}.$$

This satisfies

$$\begin{aligned}p(i)&=\sum_{t=0}^{n/4}\mathrm{Tr}[(UPU^T)_t]\lambda_{it}=\langle UPU^T,U\mu_iU^T\rangle=\langle P,\mu_i\rangle\\&=P(i).\end{aligned}$$

Let q be the degree-d polynomial obtained by removing the high-degree parts of p:

$$q(i)=\sum_{t=0}^{d}a_t\lambda_{it}.$$

Then P and q are close on all integers i between 0 and $n/8$:

$$\begin{aligned}|P(i)-q(i)|=|p(i)-q(i)|&=\left|\sum_{t=d+1}^{n/4}a_t\lambda_{it}\right|\leqslant\frac{2^{-d/4}}{\mathcal{N}}\sum_{t=0}^{n/4}|a_t|\\&\leqslant 2^{-d/4+2q}.\end{aligned}$$

■

3. The quantum lower bound for intersection

Now suppose we have a q-qubit protocol for the intersection problem, say with error probability at most 1/3 on every input x,y. Our goal is to show that q is at least about $\sqrt{n}$. Since the protocol outputs 1 with high probability if and only if x and y intersect in at least one point, we know the following about the quantity $P(i)=\mathbb{E}_{|x|=|y|=n/4,|x\wedge y|=i}[P(x,y)]$: $P(0)\in[0,1/3]$ and $P(i)\in[2/3,1]$ if $i\in\{1,\ldots,n\}$.

This $P(i)$ is only defined on integers, but by Lemma 6 we can approximate it up to some small additive error ϵ using a polynomial q of degree $d=8q+\lceil 4\log_2(1/\epsilon)\rceil$. Then we know $q(0)\in[-\epsilon,1/3+\epsilon]$ and $q(i)\in[2/3-\epsilon,1+\epsilon]$ if $i\in\{1,\ldots,n/8\}$. However, the following result of Ehlich and Zeller (1964) and Rivlin and Cheney (1966) says that such a polynomial q must have degree about $\sqrt{n}$.

Theorem 7 (Ehlich and Zeller; Rivlin and Cheney). Let $p:\mathbb{R}\to\mathbb{R}$ be a polynomial such that $b_1\leqslant p(i)\leqslant b_2$ for every integer $0\leqslant i\leqslant N$, and the derivative p' satisfies

$|p'(x)| \geq c$ for some real $0 \leq x \leq N$. Then the degree of p is at least $\sqrt{cN/(c+b_2-b_1)}$.

It thus follows that the original protocol must have communicated at least about $\sqrt{n}$ qubits. Razborov gave essentially tight lower bounds not just for the intersection problem but for any communication problem that depends only on the size of the intersection of the inputs x and y. This combines Lemma 6 with a polynomial degree lower bound due to Paturi (1992). The lower bound proof we gave here only applies to quantum protocols that do not start with an entangled state, but Razborov showed the same lower bound for protocols with prior entanglement at the expense of some more technical complication. Recently, an alternative proof was obtained by Sherstov (2008).

APPENDIX D: ASYMMETRIC DETECTION EFFICIENCY

Here we prove the results stated in Sec. VII.A.3 concerning the connection between asymmetric experiments where a single detector is inefficient and classical protocols with perfect detectors that use one-way communication, i.e., where all communication takes place from Alice to Bob.

Suppose that in order to reproduce the quantum correlations using one-way communication from Alice to Bob and shared randomness, $c_{\epsilon'}$ bits of communication are required to reproduce the correlations with error ϵ'. More precisely, the error is measured as the total variational distance between the predictions of quantum theory $P_{\rm QM}(ab|xy)$ and the output $P_{\rm class}(ab|xy)$ of the classical protocol:

$$\text{error} = \max_{xy} \sum_{ab} |P_{\rm class}(ab|xy) - P_{\rm QM}(ab|xy)|.$$

Also suppose that there exists a protocol that uses only shared randomness (a local hidden variable model) in which Alice's detector has efficiency η_ϵ and Bob's detector is perfect, and that reproduces the quantum correlations with error ϵ. More precisely the fact that Alice's detector has efficiency η means that $P(\perp b|xy) = \eta$ independently of b,x,y, where $\perp$ corresponds to Alice's detector not giving a result. The error is measured as the total variational distance between the predictions of quantum theory $P_{\rm QM}(ab|xy)$ (when the detectors are 100% efficient) and the predictions $P_{\rm LHV}(ab|xy)$ of the local hidden variable (LHV) model. We divide the latter by η to take into account that Alice's detector gives a result with probability η:

$$\text{error} = \max_{xy} \sum_{ab} \left| \frac{P_{\rm LHV}(ab|xy)}{\eta} - P_{\rm QM}(ab|xy) \right|.$$

Then we have the following.

Theorem 8. With the above hypothesis, we have $\eta_\epsilon \leq O((-\ln \epsilon) 2^{-c_{2\epsilon}})$.

To prove this, we use the LHV model with detection efficiency η_ϵ to construct a classical protocol with communication. The LHV uses shared randomness r. Alice and Bob share k independently chosen instances of the shared randomness $r_1, r_2, \ldots, r_k$. Alice checks whether she should give an output for at least one value of the shared randomness. This occurs with probability $1-(1-\eta)^k$. If so, she sends Bob the index j of the shared randomness r_j for which she gives an output (using $\log_2 k$ bits of communication), and they give the corresponding output. If there is no instance of the shared randomness for which Alice should give an output in the LHV model, Alice gives a random output and sends Bob a random index j. This occurs with probability $(1-\eta)^k$, and in this case Alice and Bob's results will most likely be completely different from those predicted by quantum mechanics. The error probability in the model with communication is thus $P(\text{error}) \leq [1-(1-\eta)^k]\epsilon + (1-\eta)^k \leq \epsilon + (1-\eta)^k$. We take $k = \ln \epsilon / \ln(1-\eta)$, then the error is bounded by

$$P(\text{error}) \leq \epsilon + (1-\eta)^{\ln \epsilon / \ln(1-\eta)} = 2\epsilon.$$

But we know that to produce the correlations with error 2ϵ we need at least $c_{2\epsilon}$ bits of one-way communication, hence $k \geq 2^{c_{2\epsilon}}$. Therefore $-\ln(1-\eta) \leq (-\ln \epsilon) 2^{-c_{2\epsilon}}$, which implies the result.

Note that the above mapping does not hold when both Alice and Bob's detectors are inefficient since if they try the above procedure, they will need to find a value of the shared randomness r_j for which both their detectors produce an output, i.e., solve an instance of the intersection problem.

We apply this result to the hidden matching problem. As mentioned in Sec. III.G, this problem can be solved using $\log_2 n$ ebits and $\log_2 n$ bits of classical communication from Alice to Bob; but if classical communication from Alice to Bob is considered, then at least $\Omega(\sqrt{n})$ bits of communication are required, even allowing for a small error probability. This implies that the correlations obtained by measuring the ebits can only be reproduced using at least $\Omega(\sqrt{n})$ bits of classical communication from Alice to Bob, even allowing for a small error probability. The above result then shows that these correlations remain nonlocal (i.e., cannot be reproduced by a classical model without communication) if Bob's detector has 100% efficiency and Alice's detector has efficiency $\eta \geq 2^{-\Omega(\sqrt{n})}$, even allowing for a small error probability.

REFERENCES

Aaronson, S., and A. Ambainis, 2003, *Proceedings of the 44th IEEE FOCS* (IEEE, Los Alamitos, CA), pp. 200–209.

Acin, A., N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, 2007, Phys. Rev. Lett. **98**, 230501.

Acín, A., N. Gisin, and B. Toner, 2006, Phys. Rev. A **73**, 062105.

Ambainis, A., 1996, Algorithmica **16**, 298.

Aravind, P. K., 2004, Am. J. Phys. **72**, 1303.

Aspect, A., J. Dalibard, and G. Roger, 1982, Phys. Rev. Lett. **49**, 1804.

Aspect, A., P. Grangier, and G. Roger, 1981, Phys. Rev. Lett. **47**, 460.

Aspect, A., P. Grangier, and G. Roger, 1982, Phys. Rev. Lett.

49, 91.
Barrett, J., 2002, Phys. Rev. A **65**, 042302.
Barrett, J., L. Hardy, and A. Kent, 2005, Phys. Rev. Lett. **95**, 010503.
Barrett, J., N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, 2005, Phys. Rev. A **71**, 022101.
Bar-Yossef, Z., T. S. Jayram, and I. Kerenidis, 2004, *Proceedings of the 36th ACM STOC* (ACM, New York), pp. 128–137.
Bell, J. S., 1965, Physics (N.Y.) **1**, 195.
Bennett, C., G. Brassard, C. Crépeau, R. Jozsa, A. Peres, and W. Wootters, 1993, Phys. Rev. Lett. **70**, 1895.
Bennett, C., and S. Wiesner, 1992, Phys. Rev. Lett. **69**, 2881.
Bennett, C. H., H. J. Bernstein, S. Popescu, and B. Schumacher, 1996, Phys. Rev. A **53**, 2046.
Bennett, C. H., and G. Brassard, 1984, *Proceedings of the IEEE International Conference on Computers, Systems and Signal Processing* (IEEE, Los Alamitos, CA), pp. 175–179.
Bennett, C. H., G. Brassard, S. Popescu, B. Schumacher, J. A. Smolin, and W. K. Wootters, 1996, Phys. Rev. Lett. **76**, 722.
Bhatia, R., 1997, *Matrix Analysis*, Graduate Texts in Mathematics Vol. 169 (Springer-Verlag, New York).
Brassard, G., 2003, Found. Phys. **33**, 1593.
Brassard, G., A. Broadbent, and A. Tapp, 2005, Found. Phys. **35**, 1877.
Brassard, G., H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, 2006, Phys. Rev. Lett. **96**, 250401.
Brassard, G., R. Cleve, and A. Tapp, 1999, Phys. Rev. Lett. **83**, 1874.
Brassard, G., P. Høyer, M. Mosca, and A. Tapp, 2002, *Quantum Computation and Quantum Information: A Millennium Volume*, AMS Contemporary Mathematics Series Vol. 305 (AMS, Providence, RI), pp. 53–74.
Briët, J., H. Buhrman, and B. Toner, 2009, e-print arXiv:0901.2009.
Brukner, Č., M. Żukowski, J.-W. Pan, and A. Zeilinger, 2004, Phys. Rev. Lett. **92**, 127901.
Brunner, N., N. Gisin, S. Popescu, and V. Scarani, 2008, Phys. Rev. A **78**, 052111.
Brunner, N., N. Gisin, and V. Scarani, 2005, New J. Phys. **7**, 88.
Brunner, N., N. Gisin, V. Scarani, and C. Simon, 2007, Phys. Rev. Lett. **98**, 220403.
Brunner, N., S. Pironio, A. Acin, N. Gisin, A. Méthot, and V. Scarani, 2008, Phys. Rev. Lett. **100**, 210503.
Brunner, N., and P. Skrzypczyk, 2009, Phys. Rev. Lett. **102**, 160403.
Buhrman, H., R. Cleve, and W. van Dam, 2001, SIAM J. Comput. **30**, 1829.
Buhrman, H., R. Cleve, J. Watrous, and R. de Wolf, 2001, Phys. Rev. Lett. **87**, 167902.
Buhrman, H., R. Cleve, and A. Wigderson, 1998, *Proceedings of the 30th ACM STOC* (ACM, New York), pp. 63–68.
Buhrman, H., P. Høyer, S. Massar, and H. Röhrig, 2006, Phys. Rev. A **73**, 012321.
Buhrman, H., P. Høyer, S. Massar, and H. Röhrig, 2003, Phys. Rev. Lett. **91**, 047903.
Buhrman, H., W. van Dam, P. Høyer, and A. Tapp, 1999, Phys. Rev. A **60**, 2737.
Cabello, A., 2001a, Phys. Rev. Lett. **87**, 010403.
Cabello, A., 2001b, Phys. Rev. Lett. **86**, 1911.
Cabello, A., 2005, Phys. Rev. Lett. **95**, 210401.
Cabello, A., and J.-A. Larsson, 2007, Phys. Rev. Lett. **98**, 220402.
Cabello, A., and A. J. López-Tarrida, 2005, Phys. Rev. A **71**, 020301(R).
Cerf, N. J., N. Gisin, S. Massar, and S. Popescu, 2005, Phys. Rev. Lett. **94**, 220403.
Cirel'son, B. S. T., 1980, Lett. Math. Phys. **4**, 93.
Clauser, J. F., M. A. Horne, A. Shimony, and R. A. Holt, 1969, Phys. Rev. Lett. **23**, 880.
Cleve, R., and H. Buhrman, 1997, Phys. Rev. A **56**, 1201.
Cleve, R., P. Høyer, B. Toner, and J. Watrous, 2004, *Proceedings of the 19th IEEE Complexity* (IEEE, Los Alamitos, CA), pp. 236–249.
Cleve, R., W. van Dam, M. Nielsen, and A. Tapp, 1998, *Proceedings of the First NASA QCQC Conference*, Lecture Notes in Computer Science Vol. 1509 (Springer, Berlin), pp. 61–74.
Collins, D., and N. Gisin, 2004, J. Phys. A **37**, 1775.
Collins, D., N. Gisin, N. Linden, S. Massar, and S. Popescu, 2002, Phys. Rev. Lett. **88**, 040404.
de Beaudrap, J. N., 2004, Phys. Rev. A **69**, 022307.
Degorre, J., S. Laplante, and J. Roland, 2005, Phys. Rev. A **72**, 062314.
Degorre, J., S. Laplante, and J. Roland, 2007, Phys. Rev. A **75**, 012309.
Deutsch, D., and R. Jozsa, 1992, Proc. R. Soc. London, Ser. A **439**, 553.
Du, J., P. Zou, X. Peng, D. K. Oi, L. C. Kwek, and A. Ekert, 2006, Phys. Rev. A **74**, 042319.
Ehlich, H., and K. Zeller, 1964, Math. Z. **86**, 41.
Einstein, A., B. Podolsky, and N. Rosen, 1935, Phys. Rev. **47**, 777.
Forster, M., S. Winkler, and S. Wolf, 2009, Phys. Rev. Lett. **102**, 120401.
Frankl, P., and V. Rödl, 1987, Trans. Am. Math. Soc. **300**, 259.
Freedman, S. J., and J. F. Clauser, 1972, Phys. Rev. Lett. **28**, 938.
Galvão, E. F., 2001, Phys. Rev. A **65**, 012318.
Gavinsky, D., 2008a, *Proceedings of the 40th ACM STOC* (ACM, New York), pp. 95–102.
Gavinsky, D., 2008b, e-print arXiv:0901.0956.
Gavinsky, D., J. Kempe, and R. de Wolf, 2006, *Proceedings of the 21st IEEE Complexity* (IEEE, Los Alamitos, CA), pp. 288–298.
Gavinsky, D., J. Kempe, I. Kerenidis, R. Raz, and R. de Wolf, 2008, SIAM J. Comput. **38**, 1695.
Gavinsky, D., J. Kempe, O. Regev, and R. de Wolf, 2006, *Proceedings of the 38th ACM STOC* (ACM, New York), pp. 594–603.
Gavinsky, D., O. Regev, and R. de Wolf, 2008, Chicago J. Theor. Comput. Sci. **2008**, 7.
Gisin, N., and B. Gisin, 1999, Phys. Lett. A **260**, 323.
Greenberger, D. M., M. Horne, and A. Zeilinger, 1989, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos (Kluwer, Dordrecht), pp. 69–72.
Gross, D., S. T. Flammia, and J. Eisert, 2009, Phys. Rev. Lett. **102**, 190501.
Grover, L. K., 1996, *Proceedings of the 28th ACM STOC* (ACM, New York), pp. 212–219.
Holevo, A. S., 1973, Probl. Peredachi Inf. **9**, 3 [Probl. Inf. Transm. **9**, 177 (1973)].
Horn, R. T., S. A. Babichev, K.-P. Marzlin, A. I. Lvovsky, and B. C. Sanders, 2005, Phys. Rev. Lett. **95**, 150502.
Hromkovič, J., 1997, *Communication Complexity and Parallel Computing* (Springer-Verlag, New York).

Kalyanasundaram, B., and G. Schnitger, 1992, SIAM J. Discrete Math. **5**, 545, earlier version in Structures'87.

Khalfi, L. A., and B. S. Tsirelson, 1985, in *Symposium on the Foundations of Modern Physics*, edited by P. Lahti and P. Mittelstaedt (World Scientific, Singapore), pp. 441–460.

Klauck, H., R. Špalek, and R. de Wolf, 2007, SIAM J. Comput. **36**, 1472, earlier version in FOCS'03.

Knuth, D. E., 2003, *Selected Papers on Discrete Mathematics*, CSLI Lecture Notes Vol. 106 (Stanford University, Stanford).

Kremer, I., 1995, MS thesis (Hebrew University).

Kushilevitz, E., and N. Nisan, 1997, *Communication Complexity* (Cambridge University Press, Cambridge).

Linial, N., and A. Shraibman, 2007, *Proceedings of the 39th ACM STOC* (ACM, New York), pp. 699–708.

Marcikic, I., H. de Riedmatten, W. Tittel, H. Zbinden, M. Legré, and N. Gisin, 2004, Phys. Rev. Lett. **93**, 180502.

Masanes, L., A. Acin, and N. Gisin, 2006, Phys. Rev. A **73**, 012112.

Massar, S., 2002, Phys. Rev. A **65**, 032121.

Massar, S., 2005, Phys. Rev. A **71**, 012310.

Massar, S., D. Bacon, N. J. Cerf, and R. Cleve, 2001, Phys. Rev. A **63**, 052305.

Massar, S., S. Pironio, J. Roland, and B. Gisin, 2002, Phys. Rev. A **66**, 052112.

Matsukevich, D. N., P. Maunz, D. L. Moehring, S. Olmschenk, and C. Monroe, 2008, Phys. Rev. Lett. **100**, 150404.

Maudlin, T., 1992, in *PSA Proceedings of the 1992 Biennial Meeting of the Philosophy of Science Association*, edited by D. Hull, M. Forbes, and K. Okruhlik (Philosophy of Science Association, East Lansing, MI), Vol. 1, pp. 404–417.

Mermin, N. D., 1990a, Phys. Rev. Lett. **65**, 3373.

Mermin, N. D., 1990b, Phys. Today **43** (6), 9.

Mermin, N. D., 1993, Rev. Mod. Phys. **65**, 803.

Moehring, D. L., M. J. Madsen, B. B. Blinov, and C. Monroe, 2004, Phys. Rev. Lett. **93**, 090410.

Navascues, M., S. Pironio, and A. Acin, 2007, Phys. Rev. Lett. **98**, 010401.

Nayak, A., 1999, *Proceedings of the 40th IEEE FOCS* (IEEE, Los Alamitos, CA), pp. 369–376.

Nayak, A., and J. Salzman, 2002, *Proceedings of the 34th ACM STOC* (ACM, New York), pp. 698–704.

Newman, I., 1991, Inf. Process. Lett. **39**, 67.

Newman, I., and M. Szegedy, 1996, *Proceedings of the 28th ACM STOC* (ACM, New York), pp. 561–570.

Nielsen, M. A., and I. L. Chuang, 2000, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge).

Pan, J.-W., D. Bouwmeester, M. Daniell, H. Weinfurter, and A. Zeilinger, 2000, Nature (London) **403**, 515.

Paturi, R., 1992, *Proceedings of the 24th ACM STOC* (ACM, New York), pp. 468–474.

Pearle, P. M., 1970, Phys. Rev. D **2**, 1418.

Perez-Garcia, D., M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge, 2008, Commun. Math. Phys. **279**, 455.

Popescu, S., and D. Rohrlich, 1994, Found. Phys. **24**, 379.

Rauschenbeutel, A., G. Nogues, S. Osnaghi, P. Bertet, M. Brune, J.-M. Raimond, and S. Haroche, 2000, Science **288**, 2024.

Raz, R., 1999, *Proceedings of the 31st ACM STOC* (ACM, New York), pp. 358–367.

Razborov, A., 1992, Theor. Comput. Sci. **106**, 385.

Razborov, A., 2003, Izv. Russ. Acad. Sci., Math. **67**, 145.

Regev, O., and B. Toner, 2007, *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS'07)* (IEEE, Los Alamitos, CA), pp. 384–394.

Rivlin, T. J., and E. W. Cheney, 1966, SIAM J. Numer. Anal. **3**, 311.

Rowe, M. A., D. Kielpinski, V. Meyer, C. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, 2001, Nature (London) **409**, 791.

Sackett, C. A., D. Kielpinski, B. E. King, C. Langer, V. Meyer, C. J. Myatt, M. Rowe, Q. A. Turchette, W. M. Itano, D. J. Wineland, and C. Monroe, 2000, Nature (London) **404**, 256.

Scholz, V. B., and R. F. Werner, 2008, e-print arXiv:0812.4305.

Schrödinger, E., 1935, Proc. Cambridge Philos. Soc. **31**, 555.

Schrödinger, E., 1936, Proc. Cambridge Philos. Soc. **32**, 446.

Schumacher, B., 1995, Phys. Rev. A **51**, 2738.

Shannon, C. E., 1948, Bell Syst. Tech. J. **27**, 379.

Sherstov, A., 2008, *Proceedings of the 40th ACM STOC* (ACM, New York), pp. 85–94.

Shor, P. W., 1997, SIAM J. Comput. **26**, 1484, earlier version in FOCS'94.

Steiner, M., 2000, Phys. Lett. A **270**, 239.

Thew, R. T., A. Acín, H. Zbinden, and N. Gisin, 2004, Phys. Rev. Lett. **93**, 010503.

Tittel, W., J. Brendel, H. Zbinden, and N. Gisin, 1998, Phys. Rev. Lett. **81**, 3563.

Toner, B. F., and D. Bacon, 2003, Phys. Rev. Lett. **91**, 187904.

Trojek, P., C. Schmid, M. Bourennane, C. Brukner, M. Żukowski, and H. Weinfurter, 2005, Phys. Rev. A **75**, 050305(R).

Tsirelson, B. S., 1987, J. Sov. Math. **36**, 557.

van Dam, W., 2000, Ph.D. thesis (University of Oxford).

van Dam, W., 2005, e-print arXiv:quant-ph/0501159.

van Dam, W., R. D. Gill, and P. D. Grünwald, 2005, IEEE Trans. Inf. Theory **51**, 2812.

Vaziri, A., G. Weihs, and A. Zeilinger, 2002, Phys. Rev. Lett. **89**, 240401.

Vértesi, T., and K. F. Pál, 2009, Phys. Rev. A **79**, 042106.

Wehner, S., M. Christandl, and A. Doherty, 2008, Phys. Rev. A **78**, 062112.

Weihs, G., T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, 1998, Phys. Rev. Lett. **81**, 5039.

Werner, R. F., 1989, Phys. Rev. A **40**, 4277.

Werner, R. F., and M. M. Wolf, 2001, Quantum Inf. Comput. **1** (3), 1.

Yao, A. C.-C., 1979, *Proceedings of the 11th ACM STOC* (ACM, New York), pp. 209–213.

Yao, A. C.-C., 1993, *Proceedings of the 34th IEEE FOCS* (IEEE, Los Alamitos, CA), pp. 352–360.

Yao, A. C.-C., 2003, *Proceedings of the 35th ACM STOC* (ACM, New York), pp. 77–81.

Zhang, J., X.-H. Bao, T.-Y. Chen, T. Yang, A. Cabello, and J.-W. Pan, 2007, Phys. Rev. A **75**, 022302.

Zhao, Z., T. Yang, Y.-A. Chen, A.-N. Zhang, M. Żukowski, and J.-W. Pan, 2003, Phys. Rev. Lett. **91**, 180401.