Multipartite nonlocal quantum correlations resistant to imperfections

Harry Buhrman
CWI and University of Amsterdam, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands

Peter Høyer and Hein Röhrig
Department of Computer Science, University of Calgary, 2500 University Drive N.W., Calgary AB, Canada T2N 1N4

Serge Massar
Laboratoire d'Information Quantique, Université Libre de Bruxelles, CP 165/59, Avenue F. D. Roosevelt 50, 1050 Bruxelles, Belgium and QUIC, Ecole Polytechnique, CP 165, Université Libre de Bruxelles, 1050 Brussels, Belgium
(Received 27 October 2004; revised manuscript received 17 May 2005; published 18 January 2006)

We use techniques for lower bounds on communication to derive necessary conditions in terms of detector efficiency or amount of superluminal communication for being able to reproduce with classical local hidden-variable theories the quantum correlations occurring in Einstein-Podolsky-Rosen (EPR) experiments in the presence of noise. We apply our method to an example involving n parties sharing a Greenberger-Horne-Zeilinger-type state on which they carry out local measurements. For this example, we show that for local hidden-variable theories to reproduce the quantum correlations, the amount of superluminal classical communication c and the detector efficiency η are constrained by $\eta 2^{-c/n} \leqslant O(n^{-1/6})$. This result holds even if the classical models are allowed to make an error with constant probability.

DOI: 10.1103/PhysRevA.73.012321 PACS number(s): 03.67.Hk, 03.65.Ud

I. INTRODUCTION

Forty years ago, Bell [1] showed that the correlations between the outcomes of measurements carried out on entangled quantum systems cannot be reproduced by a local classical theory (often called a local hidden-variable theory). Since then, extensive work has been carried out on quantum nonlocality, both on the experimental and theoretical aspects. On the theory side, research on quantum nonlocality has branched out into many different and complementary directions.

One important direction of investigation is the search for qualitatively different types of quantum nonlocality. Of particular interest is the discovery of the Greenberger-Horne-Zeilinger (GHZ) paradox [2,3]. In this and related examples, correlations are characterized as nonlocal by the pattern of zero and nonzero joint probabilities. This property has been called "pseudotelepathy," because in every run of the experiment, the parties appear to agree clandestinely on a subset of admissible outputs. It should be contrasted with other examples where nonlocality is implied by the values of these joint probabilities.

An important advance was to show that quantum nonlocality subsists even in the presence of noise as first demonstrated by Clauser *et al.* [4]. This is essential since every experimental test will necessarily be affected by imperfections: the best experiments to date have error rates of the order of a few percent. Much additional work has been devoted to understanding the resistance of quantum nonlocality to imperfections.

In experiments involving entangled photons, there is one particular kind of imperfection that plays a central role, namely the low efficiency of single-photon detectors. A single-photon detector will register the presence of a photon with probability η and will not register the presence of the photon with probability $1-\eta$. For instance, as one goes from visible to infrared wavelengths, η decreases from more than 50% to 10%. Detector inefficiency can be thought of as a specific type of noise. This imperfection was first discussed by Pearle [5] and remains to this day one of the major hurdles to overcome in order to carry out a loophole-free test of quantum nonlocality. Examples show that there are quantum correlations that are highly insensitive to detector inefficiency, but are much more sensitive to other kinds of noise, see Massar [6]; therefore, this kind of imperfection should be studied independently of other kinds of noise.

Note that the complementary error, namely detectors clicking when they should not, can also occur. We consider this error as a general noise, as it cannot be distinguished from other types of noise such as nonmaximally entangled states.

The development of quantum information theory over the past 10 years (see Ref. [7] for a review) has brought a breath of fresh air to the study of quantum nonlocality and important new questions have been raised. Bell had shown that quantum correlations could not be reproduced classically without "superluminal" communication between the parties. But Brassard *et al.* [8] and Steiner [9] initiated the study of how much superluminal communication is required to reproduce the correlations. This question is closely related to quantum communication complexity, in which one enquires whether certain distributed communication tasks can be solved using less quantum communication than is required classically.

Remarkably, the amount of classical communication required to reproduce the quantum correlations and the minimum detector efficiency required to close the detection loophole are closely related quantities as demonstrated by Gisin and Gisin [10] and Massar [6]. In many cases, quantum correlations that require a lot of communication to be reproduced classically cannot be simulated classically without communication, even when the actual detectors are very inefficient.

Another question that has been raised in the context of quantum information theory concerns the asymptotic limit when the size of the entangled system grows. Does the gap between classical and quantum correlations grow, and if so, at what rate? Brassard *et al.* [8] showed that in the bipartite case, the amount of communication required to classically reproduce the quantum correlations can increase exponentially with the number of entangled bits shared by the parties. And it follows from the results by Buhrman *et al.* [11] that there are quantum correlations for n parties each holding a two-dimensional subsystem, such that the amount of communication that must be broadcast in a classical simulation increases logarithmically with the number of parties. Unfortunately, these asymptotic results have only been proved in the total absence of noise.

The only prior asymptotic results in quantum communication complexity that hold in the presence of noise concern multiround quantum communication protocols, such as the appointment-scheduling problem of Buhrman *et al.* [12] or the example due to Raz [13]. It appears that these results cannot be mapped to results concerning quantum nonlocality, whereas communication complexity problems with a single round of communication and nonlocal quantum correlations can generally be mapped one onto the other.

The present work lies at the intersection of these different lines of enquiry. Specifically, we concentrate on the generalization of the GHZ paradox to n parties previously considered by Buhrman *et al.* [11]; their bounds on the GHZ-inspired multiparty communication problem was only proved in the absence of noise. We extend it to the noisy case.

Denote by c the number of bits communicated (via a possibly superluminal channel) in order to reproduce the correlations, and by η^*, the maximum detector efficiency for which a local classical model exists. Our central result can be summarized as

$$\eta^* 2^{-c/n} = O(n^{-1/6}). \qquad (1)$$

Equation (1) holds even if the classical models are allowed to make an error with constant probability.

The present work builds upon the earlier results of Refs. [11,14]. As in these references, we rely heavily on techniques and ideas from the field of communication complexity. The reader unfamiliar with these notions may consult the book by Kushilevitz and Nisan [15] for an introduction to classical communication complexity and Buhrman and Röhrig [16] for a survey of quantum communication complexity.

The remainder of this paper is organized as follows. In Sec. II, we define precisely the concepts of nonlocality used in this paper. In Sec. III, we introduce the combinatorial notions of rectangles and advantage and prove a general relation between η, c, and the error probability ϵ, which depends on the maximum size of high-advantage rectangles. This general result is of interest in its own right and could be of use when studying other instances of quantum nonlocality that exhibit pseudotelepathy. In Sec. IV, we apply the general bound to the GHZ paradox; the proof of Eq. (1) is based on an addition theorem for cyclic groups proved in Sec. V. Finally, we discuss our results and open problems in Sec. VI.

II. NONLOCALITY DEFINITIONS

Consider the following situation. There are n spatially separated parties; party i receives an input $x_i \in \{1,\dots,k\}$ and produces an output $a_i \in \{1,\dots,l\}$. With $x=(x_1,\dots,x_n)$ and $a=(a_1,\dots,a_n)$, let $P(a|x)$ denote the probability of output a given input x. The inputs are distributed according to the probability distribution $\mu(x)$. We formalize this situation as follows.

Definition 1. A (n,k,l) *correlation problem with input distribution* μ is a family of probability distributions $P(\cdot|x)$ on the "outputs" $\{1,\dots,l\}^n$, for each "input" $x \in \{1,\dots,k\}^n$ with $\mu(x)>0$. We denote the support of μ by $D \coloneqq \{x:\mu(x)>0\}$.

Note that in nonlocality experiments, the distribution μ should be a product distribution, otherwise the parties would have trouble selecting x according to μ when the measurements take place in timelike separated regions. However, the mathematical proofs given below, and, in particular, the example of Sec. IV, are based on nonproduct distributions. The way to get around this is the following: During the nonlocality experiment, the inputs are distributed according to a product distribution μ_0, for instance, the uniform distribution. Then when analyzing the data, one first throws away part of the data in such a way that, for the data that is kept, the inputs are distributed according to the desired distribution μ. This can only make the task harder for the local hidden variable theory, since it does not know beforehand which runs will be kept and which will be thrown away. Hence, from now on, we let μ be an arbitrary (possibly nonproduct) distribution.

We are interested in correlation problems obtained from measurements on multipartite entangled quantum states. We define these as follows.

Definition 2. A (n,k,l) *measurement scenario* is a correlation problem in which the parties share an entangled state $|\psi\rangle$; each input $x_i \in \{1,\dots,k\}$ determines a positive operator valued measure (POVM) $\hat{x}_i = \{\hat{x}_i^1,\dots,\hat{x}_i^l\}$ with $\hat{x}_i^j \geq 0$, $\sum_{j=1}^{l} \hat{x}_i^j = 1_i$. If the measurement of party i produces outcome $\hat{x}_i^j$, then it outputs $a_i=j$. The probability $P_{QM}(a|x)$ to obtain outcome a given input x is

$$P_{QM}(a|x) = \langle\psi|\hat{x}_1^{a_1} \otimes \cdots \otimes \hat{x}_n^{a_n}|\psi\rangle.$$

Our aim is to study what classical resources are required to reproduce such measurement scenarios. Let us first consider classical models in which the parties cannot communicate after they have received the inputs. Such models are called *local*. The best the parties can do in this case is to randomly select in advance a deterministic strategy. This motivates the following definition.

Definition 3. A *deterministic local hidden-variable (lhv) model* is a family of functions $\lambda=(\lambda_1,\dots,\lambda_n)$ from the inputs to the outputs: $\lambda_i:\{1,\dots,k\}\to\{1,\dots,l\}$. Each party outputs $a_i=\lambda_i(x_i)$.

A *probabilistic lhv model* (or just *lhv model*) is a probability distribution $\nu(\lambda)$ over all deterministic lhv models for given (n,k,l).

Thus, in probabilistic lhv models, the parties first randomly choose a deterministic lhv model λ using the probability distribution ν. Each party then outputs $a_i = \lambda_i(x_i)$. Hence, the probability distributions for probabilistic lhv models are of the form

$$P(a|x) = \sum_{\lambda} \nu(\lambda) P_{\lambda_1}(a_1|x_1) \ldots P_{\lambda_n}(a_n|x_n),$$

where $P_{\lambda_i}(a_i|x_i) = \delta_{a_i,\lambda_i(x_i)}$ is the probability that party i on input x_i outputs a_i in the deterministic lhv model $\lambda = (\lambda_1, \ldots, \lambda_n)$.

We also consider classical models with communication. In such models, the parties may communicate over a possibly superluminal classical broadcast channel in order to reproduce the quantum correlations P_{QM}. Different communication models exist depending on the resources used. A first possibility is that the parties do not have access to randomness. In this case, we are dealing with a *deterministic classical model with communication*. The second possibility is that the parties have access to a local source of randomness only. In this case, we are dealing with a *classical model with local randomness*. The third possibility is that the parties all have access to the same random bits. This can be implemented by having the parties communicate random bits to each other before the protocol starts. In this case, we say we are dealing with a *classical model with shared randomness*. The resource generally provided to parties in lhv models is shared randomness, see Definition 3; hence in what follows, we shall focus on this case. We formalize these definitions as follows. See Ref. [15] for an explanation why this technical definition coincides with the intuitive picture just given.

Definition 4. Consider n parties who each receive an input $x_i \in \{1, \ldots, k\}$, communicate over a classical broadcast channel, and each produce an output $a_i \in \{1, \ldots, l\}$.

A *deterministic classical model with communication* is a rooted "communication protocol" tree $\mathcal{P}$; each internal node u is labeled with the party $i_u \in \{1, \ldots, n\}$ whose turn it is to broadcast a message; each edge e from u to a descendant is labeled with a set $\mathcal{X}_e \subseteq \{1, \ldots, k\}$ so that the $\mathcal{X}_e$ form a partition of $\{1, \ldots, k\}$; each leaf v is labeled with a lhv model λ_v. An execution of the protocol on input x starts at the root of tree; until a leaf is reached, the execution proceeds from node u to the descendant of u that is reached via the edge e with $x_{i_u} \in \mathcal{X}_e$. It is understood that the choice of the edge is broadcast to all parties so that all parties know at each moment at which node the execution is. When the execution has reached the leaf v, each party i outputs $\lambda_{v,i}(x_i)$ and the execution terminates. If there are m leaves and if the number of children of the nodes on the path from the root to the final leaf is $t_1, \ldots, t_m$, the number of bits broadcast is $c = \lceil \log_2 t_1 \rceil + \cdots + \lceil \log_2 t_m \rceil$.

A *classical model with shared randomness* is an arbitrary probability distribution $\nu(\mathcal{P})$ over deterministic classical models. An execution of such a model first probabilistically selects a deterministic model and then evaluates the deterministic model.

In a *classical model with local randomness*, the distribution $\nu(\mathcal{P})$ is constrained to be a product distribution.

Of course, a classical model that always uses zero bits of communication is just a lhv model.

Definition 5. For a correlation problem P with input distribution μ, we denote by $D(P)$, $R(P)$, and $R^{\text{pub}}(P)$, respectively, the minimum number of bits that must be broadcast in order to perfectly reproduce the correlations P when the parties are deterministic, have local randomness only, or have shared randomness.

Where the choice of the correlation problem P is clear from the context, we drop it and write D, R, and R^{pub}.

Clearly, $D(P) \geq R(P) \geq R^{\text{pub}}(P)$. Since the results of quantum measurements are inherently random, it is, in general, impossible to reproduce the quantum correlations using deterministic lhv models or using deterministic models with communication. Thus, $D(P)$ is meaningless when trying to simulate quantum measurement scenarios. However, deterministic models are a very useful tool for studying the probabilistic models because properties of *all* deterministic models necessarily also hold for *all* probabilistic models, since the probabilistic models are just probabilistic mixtures of deterministic models. Note also that Massar *et al.* [17] showed that $R(P)$ can be infinite when P arises from a quantum measurement scenario.

In general, classical models cannot reproduce the quantum correlations P_{QM} unless communication is possible, the detector efficiency η is sufficiently small, or the error probability is sufficiently large.

Let us consider now the situation where the detectors are inefficient. In this case, we enlarge the space of outputs to $a_i \in \{1, \ldots, l\} \cup \{\perp\}$, where $a_i = \perp$ is the event that the ith detector does not produce an output ("click"). We suppose that each measurement $\hat{x}_i$ has probability η of giving a result and a probability $1-\eta$ of not giving a result. Whether a detector clicks or does not click is independent of the other detectors. This affects the probabilities in a more structured way than simply decreasing the probability that all detectors click simultaneously. This issue has been discussed by and [18]; for simplicity, we will consider here only the two cases that all detectors click (which occurs with probability η^n) or that at least one detector does not click. We define detector efficiency accordingly.

Definition 6. Let $P(\cdot|x)$ be a fixed (n,k,l) correlation problem with input distribution μ. Let

$$C := \{a : \forall \ \ i \ a_i \neq \ \perp\}$$

denote the output vectors where all detectors click. With slight abuse of notation, we also use C as the indicator random variable of the event $a \in C$. We define the *detection efficiency* η of the correlations to be the expectation

$$\eta := \left(\mathrm{E}_{\mu}\left[\sum_a P(a|x) C \right] \right)^{1/n}.$$

Note that here the atomic events are tuples (x,a) of an input and an output vector with a joint distribution of the form Prob[input x and output a] $= \mu(x) P(a|x)$. The expectation above is over the marginal distribution μ of the inputs.

We are also interested in the possibility that the lhv model makes errors. A natural measure of the error rate is the total variation distance:

Definition 7. Suppose that some classical model produces a probability distribution $P(a|x)$, which should approximate the probability distribution produced by a measurement scenario $P_{\mathrm{QM}}(a|x)$. The *total-variation distance* is a measure for how much these two distributions differ

$$\varepsilon_{\mathrm{var}} := \mathrm{E}_\mu\left[\sum_a |P_{\mathrm{QM}}(a|x) - P(a|x)|\frac{C}{\eta^n}\right].$$

The inclusion of the factor C/η^n takes care of the possible finite efficiency of the detectors, assumed to be the same for $P_{\mathrm{QM}}(a|x)$ and for $P(a|x)$.

We will be particularly interested in quantum correlations that exhibit pseudotelepathy, which have $P_{\mathrm{QM}}(a|x)=0$ for some a and x. For such correlations, it is convenient to define the error probability as follows.

Definition 8. Let

$$F := \{(a,x): P_{\mathrm{QM}}(a|x) = 0\}$$

and again we also denote by F the indicator random variable of the event $P_{\mathrm{QM}}(a|x)=0$. The *error probability* is

$$\varepsilon := \mathrm{E}_\mu\left[\sum_a P(a|x)F\frac{C}{\eta^n}\right].$$

Thus, ϵ is the probability to observe in one run an event that cannot occur in the quantum-mechanical model. It is evident that

$$\varepsilon_{\mathrm{var}} \geqslant \varepsilon.$$

For a (n,k,l) correlation problem $P(\cdot|x)$ with input distribution μ, we denote by η^* the maximum detector efficiency of any lhv model that reproduces the quantum correlations and by η^*_ϵ the maximum detector efficiency of any lhv model that reproduces the quantum correlations up to error ϵ. Similarly, we can define D_ϵ, R_ϵ, $R^{\mathrm{pub}}_\epsilon$ as the amounts of communication required to reproduce the correlation problem P in the presence of error. We are interested in η^*_ϵ and by $R^{\mathrm{pub}}_\epsilon$.

We can map every communication model with c bits of communication and with shared randomness into a model with inefficient detectors with efficiency $\eta^n=2^{-c}$: the shared randomness determines the conversation between the parties. Thus, they all agree on the conversation. Each party i checks whether its input x_i is compatible with the conversation and, if yes, produces output a_i according to the communication model and otherwise produces no output, i.e., $\perp$. The total probability that all detectors click is equal to the probability that x belongs to the conversation. Since each input belongs to one and only one conversation, the probability that all detectors click is equal to one over the number of conversations. Note that, in this model, the probability that a specific detector, say detector i, clicks may depend on the input x_i. However, the probability that all detectors click is independent of the input. In summary, we have:

Theorem 1. Consider lhv models where the probability that all detectors click is independent of the input, but where the probability that each detector clicks, say detector i, may depend on its input x_i. Then there exists a lhv model if the probability η^n that all detectors click is at most $2^{-R^{\mathrm{pub}}}$. This implies that in these models,

$$(\eta^*_\varepsilon)^n \geqslant 2^{-R^{\mathrm{pub}}_\varepsilon}. \tag{2}$$

This result was given in Ref. [14] in the absence of error, but it also holds when errors are present.

III. COMBINATORIAL BOUNDS

We now introduce some definitions and notation that allow us to state and then prove our result concerning a general relation between c, η, and ϵ. We are concerned with pseudotelepathy type correlations for which there are some $P(a|x)$ that vanish.

Definition 9. Let $P(\cdot|x)$ be a fixed (n,k,l) correlation problem with input distribution μ. We define the sets of inputs that admit output a as

$$\mathrm{adm}(a) := \{x: P(a|x) > 0\}$$

for all $a \in C$. Moreover, for a set $S \subseteq \{1,\ldots,k\}^n$ of inputs and a specific output $a \in \{1,\ldots,l\}^n$, the *a advantage of S* is

$$\mathrm{adv}_a(S) := \frac{\mu(S \cap \mathrm{adm}(a))}{\mu(S)}$$

for all $a \in C$.

Intuitively, a correlation problem should be easy to simulate classically if $\mathrm{adv}_a(S)$ is large for large sets S, at least if S has simple structure. This is because for the parties to agree to output a, they only need to check if their input is in S, an easy task if S is large and has a simple structure. To make this intuition precise, we introduce the notion of a rectangle, which is an essential tool in communication complexity.

For sets $A_1,\ldots,A_n$, a subset R of the Cartesian product $A_1\times\cdots\times A_n$ is called a *rectangle* if there are $R_1\subseteq A_1,\ldots,R_n\subseteq A_n$ such that $R=R_1\times\cdots\times R_n$, i.e., R is a Cartesian product itself. The importance of rectangles is that for a deterministic lhv model $\lambda=(\lambda_1,\ldots,\lambda_n)$, the set $R_\lambda(a) :=\{x:\lambda(x)=a\}$ of all inputs x leading to output a is a rectangle: $R_\lambda(a)=\lambda_1^{-1}(a_1)\times\cdots\times\lambda_n^{-1}(a_n)$.

We now have the formal framework to state our central technical theorem, which relates the quality parameters of a classical model to certain properties of the correlation problem. Here, the quality parameters of the classical model are the amount of communication c, the detector efficiency η, and the error ϵ. We bound them in terms of the attainable advantage δ (see Definition 9) of rectangles R that have μ weight larger than some threshold r; recall that by Definition 6, C denotes the set of outputs where all detectors click.

Theorem 2. Let P be a fixed (n,k,l) correlation problem with input distribution μ. If for some δ $(0\leqslant\delta\leqslant 1)$, all rectangles R with $\mathrm{adv}_a(R)\geqslant\delta$ have $\mu(R)\leqslant r$ for every $a\in C$, then for every classical model with shared randomness $\nu(\mathcal{P})$ with c bits of communication holds

$$\frac{1}{2^c}\eta^n\left(1-\varepsilon\frac{1}{1-\delta}\right)\leqslant l^n r.$$

The fact that we have all the parameters of the classical model in a single inequality quantifies how the individual loopholes of covert communication, detector inefficiency, and errors can be traded one for the other.

We prove Theorem 2 by deriving bounds for $\eta^n(1-\epsilon)$, the probability that in the classical model, all detectors click and produce a correct output, and for $\eta^n\epsilon$, which is the probability of all detectors clicking but an error occurring. We split the events into originating in large and small rectangles and bound those cases separately.

Proof of Theorem 2. Let $R_{\mathcal{P},v,a}$ denote the set of inputs x for which the deterministic protocol $\mathcal{P}$ terminates in leaf v and outputs a. Every $R_{\mathcal{P},v,a}$ is a rectangle. Let $L := \{(\mathcal{P},v,a) : \mathrm{adv}_a(R_{\mathcal{P},v,a}) \geq \delta\}$ denote the set of protocols, leafs, and outputs where the corresponding rectangle has large advantage. Then, we can bound $\eta^n(1-\epsilon)$ as follows:

$$\begin{aligned}\eta^n(1-\varepsilon) &= \sum_{\mathcal{P},x} \nu(\mathcal{P})\mu(x)C(1-F)\\ &= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu[R_{\mathcal{P},v,a} \cap \mathrm{adm}(a)]\\ &= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})\mathrm{adv}_a(R_{\mathcal{P},v,a}),\end{aligned}$$

where the v range over the leafs of $\mathcal{P}$ and the a over $\{1,\ldots,l\}^n$. The first equality comes from Definition 8, the second equality expresses inputs x as preimages of outputs a from the protocol $\mathcal{P}$ and selects the ones where the protocol does not make an error. By Definition 9, we can rewrite the success probability in terms of the advantage of the rectangle $R_{\mathcal{P},v,a}$, yielding the third equality.

Now we can split this sum into large-advantage and small-advantage triples $(\mathcal{P},v,a)$ and bound their weight with

$$\eta^n(1-\varepsilon) \leq \sum_{(\mathcal{P},v,a)\in L} \nu(\mathcal{P})r + \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})\delta.$$

Finally, we use the fact that there are at most $2^c d^n$ triples $(\mathcal{P},v,a)$ for each fixed $\mathcal{P}$

$$\eta^n(1-\varepsilon) \leq 2^c d^n r + \delta \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a}).$$

In a similar way, we can bound the probability $\eta^n\epsilon$ of observing an error

$$\begin{aligned}\eta^n\varepsilon &= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(x)CF\\ &= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a} \cap [\{1,\ldots,k\}^n \setminus \mathrm{adm}(a)])\\ &= \sum_{\mathcal{P},v,a} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})[1-\mathrm{adv}_a(R_{\mathcal{P},v,a})]\\ &\geq 0 + \sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a})(1-\delta)\\ &= (1-\delta)\sum_{(\mathcal{P},v,a)\notin L} \nu(\mathcal{P})\mu(R_{\mathcal{P},v,a}),\end{aligned}$$

Combining the bounds on $\eta^n(1-\epsilon)$ and $\eta^n\epsilon$, we obtain

$$\eta^n(1-\varepsilon) \leq 2^c l^n r + \frac{\delta}{1-\delta}\eta^n\varepsilon,$$

which implies Theorem 2. ■

IV. APPLICATION TO THE GHZ CORRELATIONS

In this section, we instantiate Theorem 2 by proposing a concrete experiment and showing that for a judicious choice of the threshold parameters δ and r, large rectangles do have small advantage. The experiment is formalized by a measurement scenario as per Definition 2; in this measurement scenario, each of the n parties has a two-dimensional quantum system. The overall state of the n qubits is

$$|\psi\rangle = \frac{|0^n\rangle + |1^n\rangle}{\sqrt{2}}, \tag{3}$$

where $|i^n\rangle = |i\rangle \otimes \cdots \otimes |i\rangle$ with n terms in the product. Each party receives as input $x_i \in \{0,\ldots,k-1\}$. Each party then measures its qubit in the basis

$$|\phi_\pm\rangle = \frac{|0\rangle \pm e^{\pi i x_i/k}|1\rangle}{\sqrt{2}}. \tag{4}$$

If the qubit is projected onto state $|\phi_+\rangle$, then party i outputs $a_i=0$ and if the qubit is projected onto state $|\phi_-\rangle$, party i outputs $a_i=1$.

We call an input $x=(x_1,\ldots,x_n)$ *valid* if it satisfies

$$\left(\sum_{i=1}^n x_i\right) \bmod k = 0, \tag{5}$$

and we let $D \subset \mathbb{Z}_k^n$ denote the set of all valid inputs. Let $G : \mathbb{Z}_k^n \to \{0,1\}$ denote the Boolean function on the valid inputs defined by

$$G(x) = \frac{1}{k}\left[\left(\sum_{i=1}^n x_i\right) \bmod 2k\right].$$

The function G can be viewed as computing the $(1+\log_2 k)$th least significant bit of the sum of the x_i.

It is easy to check that the outputs of the quantum measurement are correlated as follows: If Eq. (5) holds, then

$$\left(\sum_{i=1}^n a_i\right) \bmod 2 = \frac{1}{k}\left[\left(\sum_{i=1}^n x_i\right) \bmod 2k\right] = G(x). \tag{6}$$

Hence, if each party broadcasts its measurement outcome, then each party can locally compute $G(x)$.

Lemma 3. In the model with prior entanglement and classical broadcast communication, the communication complexity of computing $G(x)$ is $O(n)$.

Moreover, the above measurement scenario will exactly reproduce the following $(n,k,2)$ correlation problem (see Definition 1):

Definition 10. Let $\mu(x)$ be the uniform distribution on the valid inputs x, which satisfy Eq. (5),

$$\mu(x) := \begin{cases} \frac{1}{k^{n-1}} & x \text{ valid} \\ 0 & x \text{ invalid} \end{cases}$$

and let

$$P(a|x) := \begin{cases} \frac{1}{2^{n-1}} & \text{if } G(x) = a_1 + \cdots + a_n \bmod 2 \\ 0 & \text{otherwise.} \end{cases}$$

for all $a \in \{0,1\}^n$ and $x \in D$.

A simple classical strategy for reproducing these correlations is for every party to broadcast its input. In particular, this means that for $k=n^{1/6}$ (the value we will use below) the communication problem and the correlation problem can both be solved exactly with $O(n \log_2 n)$ bits of communication.

Note that for $n=3$ and $k=2$, the above correlations constitute the GHZ paradox as formulated by [3]. The case $k=2$ and arbitrary n was studied by Mermin [19] and recently revisited by Brassard *et al.* [20,21]. In Buhrman *et al.* [11] and our earlier research [14], the case where the number of settings k is a power of 2 was considered. In Ref. [11], it was shown that the number of bits c that the parties must broadcast classically in order to reproduce exactly the correlations from Definition 10 is $c=\Omega(n \log_2 n)$ when $k=\Omega(n)$. And in Ref. [14], it was shown that the maximum detector efficiency η^* for which a local classical model can reproduce exactly the correlations from Definition 10 decreases as $1/n$. Furthermore, the classical strategy described above shows that, in the absence of noise, these results are essentially optimal.

We will now show that this optimality continues to hold in the presence of noise and that the classical strategy described above remains close to optimal in the presence of noise. Specifically we will show:

Theorem 4. Let μ be the uniform distribution on valid inputs. Take $k \geqslant n^{1/6}$. Then, the number c of bits broadcast, the efficiency η, and the error ϵ of every lhv model ν are constrained by

$$\frac{1}{2^{c/n}}\eta\left\{1-\varepsilon\left[2+O\left(\frac{1}{n^{1/6}}\right)\right]\right\}^{1/n} = O\left(\frac{1}{n^{1/6}}\right).$$

For fixed ϵ and large n, this implies Eq. (1). As a side note, we also obtain a corresponding bound on the communication complexity of G; previously, only bounds on the on error-free protocols were known.

Corollary 1. Every bounded-error randomized public coin protocol for $G:\mathbb{Z}_k^n \to \{0,1\}$ with $k \geqslant n^{1/6}$ requires $\Omega(n \log_2 n)$ bits of communication.

We now turn to the proof of Theorem 4. We have already defined a correlation problem; to make use of Theorem 2, it remains to show that this correlation problem has for certain δ and r the property

$$\mathrm{adv}_a(R) \geqslant \delta \Rightarrow \mu(R) \leqslant r \tag{7}$$

for all rectangles R and all $a \in C$. It turns out we can show this for $\delta=1/2+O(1/n^{1/6})$ and $r=n^{-1/6(n-n^{5/6}-1)}$.

We do this in two steps with the help of the notion of involved parties: We say a rectangle $R=A_1\times\cdots\times A_n \subseteq \{1,\ldots,k\}^n$ involves m parties if m of the n subsets A_i have size at least 2. Clearly, every rectangle involving at most m parties can have size at most k^m. In particular, we will use:

Lemma 5. Every rectangle R involving at most $n^{5/6}$ parties satisfies $\log_2|R| \leqslant n^{5/6} \ln k$.

Another notion we need is the bias, which measures how much a rectangle of input values is biased toward output $G(x)=0$ or $G(x)=1$: We say a rectangle R has bias of at most δ if

$$|G^{-1}(1) \cap D \cap R| \leqslant (1+\delta)|G^{-1}(0) \cap D \cap R|$$

and

$$|G^{-1}(0) \cap D \cap R| \leqslant (1+\delta)|G^{-1}(1) \cap D \cap R|.$$

The bias is closely related to the advantage: For every a we have $\mathrm{adm}(a)\cap D = G^{-1}(a_1+\cdots+a_n \bmod 2)\cap D$. Therefore, if μ is a distribution that is uniform on D, then R has bias at most δ if and only if it has a advantage of at most $(1+\delta)/(2+\delta)$ for every a.

Below, Lemma 6 states that rectangles involving many parties are almost unbiased; combined with Lemma 5, it gives rise to implication (7). The proof of Lemma 6 is based on an addition theorem for cyclic groups and is given in Sec. V.

Lemma 6. Every rectangle involving at least $n^{5/6}$ parties has bias at most $O(1/n^{1/6})$.

Proof of Theorem 4. Lemma 6 implies that each rectangle involving at least $n^{5/6}$ parties can have a advantage of at most $1/2+O(1/n^{1/6})$ for any a. Conversely, rectangles with a advantage greater than $1/2+O(1/n^{1/6})$ must involve less than $n^{5/6}$ parties. By Lemma 5, such a rectangle R has size less than $k^{n^{5/6}}$ and, thus,

$$\mu(R) = |R|/k^{n-1} \leqslant k^{n^{5/6}-n+1} = n^{-1/6(n-n^{5/6}-1)}.$$

Plugging these values into Theorem 2, we obtain

$$\frac{1}{2^c}\eta^n\left\{1-\varepsilon\left[2+O\left(\frac{1}{n^{1/6}}\right)\right]\right\} \leqslant 2^{-\frac{1}{6}n\log_2 n+O(n)},$$

which is equivalent to the claim in the statement of the theorem. ■

V. AN ADDITION THEOREM

When proving Theorem 4 in Sec. IV, we postponed the proof of Lemma 6; this section is devoted to it. The main ingredient is a theorem for cyclic groups that provides a bound on the bias of multisets obtained by the elementwise addition of subsets from the group. We begin by introducing the notions necessary for the precise statement of this addition theorem.

Let $\mathbb{Z}_T$ denote the additive cyclic group of order T. Let $\mu_A(x)$ denote the multiplicity of an element x in the multiset A. For multisets A and B of $\mathbb{Z}_T$, let $A+B$ denote the multiset $\{a+b \mid a \in A, b \in B\}$.

Definition 11. We say a multiset A of $\mathbb{Z}_T$ has bias at most ϵ with respect to a subgroup $H \leqslant \mathbb{Z}_T$ if $\mu_A(a) \leqslant (1+\epsilon)\mu_A(a+h)$ for all $a \in A$ and all $h \in H$.

The following theorem is the main technical result for proving Lemma 6; at the end of this section, we quickly prove the lemma by an appropriate instantiation of the theorem's parameters.

Theorem 7. Let $A_1, \ldots, A_r$ be subsets of $\mathbb{Z}_T$, each of size at least 2, with $r \geqslant T^3$ and $T=2^t$ a power of 2. Then, the multiset $A_1+A_2+\cdots+A_r$ has bias at most $O(T^{3/2}/r^{1/2})$ with respect to the subgroup $\{0, 2^{t-1}\}$.

This theorem is derived by a sequence of simple reductions to the following observation: We may generate an almost uniformly distributed random number between 0 and $K-1$ by flipping a fair coin K^2 times and counting the number of heads mod K.

We begin with a simple fact about adding two multisets:

Lemma 8. For multisets A and B over $\mathbb{Z}_T$, if A has bias at most ϵ with respect to some subgroup H, then so does $A+B$. In particular, the multiset $A+\{d\}$ has the same bias as A.

The following lemmas build one on top of the previous, culminating in the proof of Theorem 7.

Lemma 9. Let $f:\{0,1\}^s \to \mathbb{Z}_K$ be defined by

$$f(a_1, \ldots, a_s) = \left(\sum_{i=1}^{s} a_i\right) \bmod K.$$

If $s \geqslant K^2$, then $|f^{-1}(x)| \leqslant (1+4K/\sqrt{s})|f^{-1}(y)|$ for all $x, y \in \mathbb{Z}_K$.

Proof. First suppose $x \leqslant y$. Then,

$$\begin{aligned}
|f^{-1}(x)| &= \sum_i \binom{s}{x+iK} \\
&= \sum_{i:y+iK<s/2} \binom{s}{x+iK} + \sum_{i:y+iK\geqslant s/2} \binom{s}{x+iK} \\
&\leqslant \sum_{i:y+iK<s/2} \binom{s}{y+iK} + \sum_{i:y+iK\geqslant s/2} \binom{s}{x+iK+K} \\
&\quad + \binom{s}{s/2} \\
&\leqslant \sum_{i:y+iK<s/2} \binom{s}{y+iK} + \sum_{i:y+iK\geqslant s/2} \binom{s}{y+iK} + \binom{s}{s/2} \\
&= |f^{-1}(y)| + \binom{s}{s/2}.
\end{aligned}$$

Similarly, if $x > y$, then still $|f^{-1}(x)| \leqslant |f^{-1}(y)| + \binom{s}{s/2}$. Thus, for all $y \in \mathbb{Z}_K$, we have that $|f^{-1}(y)|$ is within $\binom{s}{s/2}$ of the average value of $2^s/K$. Hence,

$$\binom{s}{\frac{s}{2}} \leqslant \frac{4}{5}\frac{2^s}{K}\frac{K}{\sqrt{s}} \leqslant \frac{4}{5}\left[|f^{-1}(y)| + \binom{s}{\frac{s}{2}}\right]\frac{K}{\sqrt{s}},$$

from which follows:

$$\binom{s}{\frac{s}{2}} \leqslant \frac{4}{5\frac{\sqrt{s}}{K} - 4}|f^{-1}(y)|.$$

■

Lemma 10. Let $B_1 = \cdots = B_s = \{0, b\}$ be s identical size-2 subsets of $\mathbb{Z}_T$, with $s \geqslant T^2$. Then the multiset $B_1+B_2+\cdots+B_s$ has bias at most $4|H|/s^{1/2}$ with respect to the subgroup $H=\langle b \rangle$.

Proof. Set $K=|H|$ and define function $f:\{0,1\}^s \to \mathbb{Z}_K$ by $f(a_1, \ldots, a_s) = (\Sigma_{i=1}^{s} a_i) \bmod \mathrm{K}$. Then, we may generate the multiset $B_1+B_2+\cdots+B_s$ as $b \cdot f(\{0,1\}^s)$. Applying Lemma 5 gives that f is almost unbiased on $\mathbb{Z}_K$ and, hence, bf is almost unbiased with respect to H. ■

Lemma 11. Let $B_1, \ldots, B_r$ be size-2 subsets of $\mathbb{Z}_T$, with $r \geqslant T^3$. There exists a nontrivial subgroup $H \leqslant \mathbb{Z}_T$ such that $B_1+B_2+\cdots+B_r$ has bias at most $4T^{3/2}/r^{1/2}$ with respect to H.

Proof. First suppose $0 \in B_i$ for all i. There exists some nontrivial element $b \in \mathbb{Z}_T$ such that $B_i=\{0,b\}$ for s of the subsets, with $s \geqslant r/T \geqslant T^2$. Applying Lemma 6 on these s subsets yields a multiset of bias at most $4|\langle b \rangle|/s^{1/2} \leqslant 4T^{3/2}/r^{1/2}$ with respect to $\langle b \rangle$. By Lemma 8, adding the remaining $r-s$ subsets to this multiset does not increase the bias.

In general, we do not have that $0 \in B_i$ for all i. In this case, observe that by Lemma 8, adding any offset to a multiset does not change its bias; thus, we may reduce to the former case by adding an appropriate offset d_i to subset B_i such that $0 \in B_i+\{d_i\}$ for each i. ■

Proof of Theorem 7. Let $B_i \subseteq A_i$ be a random size-2 subset of A_i for each i. By Lemma 11, the subrectangle $R'=B_1 \times \cdots \times B_r$ is almost unbiased with respect to some nontrivial subgroup H'. Since H' is nontrivial, it contains $H=\{0, 2^{t-1}\}$, and hence, R' is also almost unbiased with respect to H. By this selection process, every $(a_1, \ldots, a_r) \in A_1 \times \cdots \times A_r$ has the same probability of being selected, and hence, R itself is almost unbiased with respect to H. ■

Proof of Lemma 6. Set $t=\frac{1}{6}\log_2 n$ and $T=2^t$. Consider any rectangle $R=A_1 \times \cdots \times A_n$ involving at least $r \geqslant n^{5/6}=T^5$ parties. By Theorem 4, the multiset $A_1+\cdots+A_n$ has bias at most $O(T^{3/2}/r^{1/2}) \subseteq O(1/n^{1/6})$ with respect to $\{0, 2^{t-1}\}$. Hence, rectangle R has bias at most $O(1/n^{1/6})$, too. ■

VI. CONCLUSIONS

We studied experiments for validating quantum nonlocality in the presence of noise and with imperfect detectors. In Theorem 2, we derived a general bound relating the resources required by classical models to simulate quantum correlations to combinatorial properties of the quantum mechanical correlations.

The classical resources we considered were detector efficiency, error rate, and (possibly superluminal) communication. Clearly, it is easier for a classical model to simulate a less efficient or noisy quantum mechanical detector. It is also easier for the classical model to simulate the quantum correlations if it is allowed some communication between parties.

We studied the tradeoffs between these different quantities for the specific case where the quantum correlations exhibit the property of pseudotelepathy, which arises when the probabilities $P_{QM}(a|x)$ of output a given input x vanishes for some pairs a,x. This allows us to consider the discrete pattern of admissible and inadmissible outputs in relation to the inputs. We characterize this pattern by the maximum size of rectangles that have most of their support on inputs x for which a given outcome a can occur. Theorem 2 characterizes in a single inequality how the maximum size of these rectangles limits the detector efficiency, error rate, and (possibly superluminal) communication, as well as the tradeoffs between these quantities.

We illustrated Theorem 2 on the generalization of the GHZ paradox to n parties previously considered as a quantum communication complexity problem [11]. We suppose that there is a fixed (i.e., independent of n) nonzero probability ϵ for an error to occur. Denote by c the number of bits communicated (via a possibly superluminal channel) in order to reproduce the correlations. We show that

$$c = \Omega(n \log_2 n).$$

Denote by η^* the maximum detector efficiency for which a local classical model exists. We show that

$$\eta^* = O(n^{-1/6}).$$

In fact, the superluminal communication and detection efficiency can be traded one for the other: We combine the above two results into the following bound:

$$\eta^* 2^{-c/n} = O(n^{-1/6}).$$

This bound sheds new light on the relation between these two quantities, which was previously discussed in [6,14]. To our knowledge, our result constitutes the first noise-resistant example in which the degree to which the quantum correlations are nonlocal increases with the size of the entangled system. As such, it constitutes a significant advance in our understanding of quantum communication complexity and of quantum nonlocality.

There are several directions in which one may wish to improve the result Eq. (1). The first concerns the evaluation of the right-hand side of this relation. A detailed investigation of the proof shows that the right-hand side becomes nontrivial only for values of n that exceed a few hundred. Therefore, our result will not be useful for the moderate values of n, say, $n \leq 10$, which may be attainable by real-world experiments in the next few years. It would be interesting to try to improve Eq. (1) so as to make it relevant for small values of n. Can the gap between the result in the absence of noise [when the right-hand side is $O(n^{-1})$] and the result in the presence of noise be closed?

Another question concerns our notion of error, which is not entirely appropriate to a multiparty setting: one expects that each party may induce an error independently of the other parties. Thus it would be more natural to consider that the probability of an error goes as $\epsilon = 1 - \delta^n$. We do not know whether a constraint of the form Eq. (1) holds in this case.

Notwithstanding the above directions in which improvements are possible, there is a specific sense in which the above result can be shown to be close to optimal. Consider n parties who share an entangled state $|\psi\rangle$ of dimension 2^n. Each party's system is two-dimensional, i.e., each party has a single qubit. Fix a total-variation distance ϵ_{var}. Then for any measurement scenario involving local measurements on the quantum state $|\psi\rangle$, the amount of (superluminal) communication required to reproduce these correlations up to total-variation distance ϵ_{var} is at most of order $n \log_2 n$, and the maximal detector efficiency η^* for which these correlations are local is of order n^{-c} for some constant c. This result will be reported elsewhere [22]. It shows that the example considered above is close to maximally nonlocal, at least if one restricts oneself to a large number of parties each possessing a single qubit.

ACKNOWLEDGMENTS

H.B. and H.R. supported in part by the EU fifth framework projects QAIP, No. IST-1999-11234, and RESQ, No. IST-2001-37559. P.H. supported in part by the Alberta Ingenuity Fund and the Pacific Institute for the Mathematical Sciences. S.M. supported in part by the EU fifth framework projects EQUIP, No. IST-1999-11053, and RESQ, No. IST-2001-37559, by the IUAP program of the Belgian Federal Government through Grant No. V-18, and by the Action de Recherche Concertée de la Communautée Française de Belgique under Grant No. 00/05-251.

[1] J. S. Bell, Physics (Long Island City, N.Y.) **1**, 195 (1965).
[2] D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's Theorem, Quantum Theory, and Conceptions of the Universe*, edited by M. Kafatos, Fundamental Theories of Physics Vol. 37 (Kluwer, Dordrecht, 1989), pp. 69–72.
[3] N. D. Mermin, Phys. Rev. Lett. **65**, 3373 (1990).
[4] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[5] P. Pearle, Phys. Rev. D **2**, 1418 (1970).
[6] S. Massar, Phys. Rev. A **65**, 032121 (2002).
[7] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2000).
[8] G. Brassard, R. Cleve, and A. Tapp, Phys. Rev. Lett. **83**, 1874 (1999).
[9] M. Steiner, Phys. Lett. A **270**, 239 (2000).
[10] N. Gisin and B. Gisin, Phys. Lett. A **260**, 323 (1999).
[11] H. Buhrman, W. van Dam, P. Høyer, and A. Tapp, Phys. Rev. A **60**, 2737 (1999).
[12] H. Buhrman, R. Cleve, and A. Wigderson, in Proceedings of 30th ACM Symposium on Theory of Computing, Dallas, Texas 1998, pp. 63–68, quant-ph/9802040.

[13] R. Raz, in Proceedings of 40th IEEE FOCS, 1999, pp. 358–367.
[14] H. Buhrman, P. Høyer, S. Massar, and H. Röhrig, Phys. Rev. Lett. **91**, 047903 (2003).
[15] E. Kushilevitz and N. Nisan, *Communication Complexity* (Cambridge University Press, Cambridge, 1997).
[16] H. Buhrman and H. Röhrig, in *Mathematical Foundations of Computer Science 2003*, edited by B. Rovan and P. Vojtas, Lecture Notes in Computer Science, Vol. 2747 (Springer, New York, 2003), pp. 1–20, URL http://www.springerlink.com/link.asp?id=1kq4u9qq8c5qpm3h
[17] S. Massar, D. Bacon, N. Cerf, and R. Cleve, Phys. Rev. A **63**, 052305 (2001).
[18] S. Massar and S. Pironio, Phys. Rev. A **68**, 062109 (2003).
[19] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).
[20] G. Brassard, A. Broadbent, and A. Tapp, in *Proceedings of Workshop on Algorithms and Data Structures 2003*, edited by F. Dehne, J. R. Sack, and M. C. Smid, Lecture Notes in Computer Science, Vol. 2748 (Springer, New York, 2003), pp. 1–11.
[21] G. Brassard, A. Broadbent, and A. Tapp, Quantum Inf. Comput. **5**, 538 (2005).
[22] S. Massar and A. Winter (unpublished).