SIAM J. COMPUT.
Vol. 39, No. 4, pp. 1562–1580

SIMULATING QUANTUM CORRELATIONS WITH FINITE COMMUNICATION[*]

ODED REGEV[†] AND BEN TONER[‡]

Abstract. Assume Alice and Bob share some bipartite d-dimensional quantum state. A well-known result in quantum mechanics says that by performing two-outcome measurements, Alice and Bob can produce correlations that cannot be obtained locally, i.e., with shared randomness alone. We show that by using only two bits of communication, Alice and Bob can classically simulate any such correlations. All previous protocols for exact simulation required the communication to grow to infinity with the dimension d. Our protocol and analysis are based on a power series method, resembling Krivine's bound on Grothendieck's constant, and on the computation of volumes of spherical tetrahedra.

Key words. quantum entanglement, communication complexity

AMS subject classification. 68Q25

DOI. 10.1137/080723909

1. Introduction.

Quantum correlations. Consider the following game [12]. Alice receives a bit a, and Bob receives a bit b, both chosen uniformly at random. Their task is to output one bit each in such a way that the XOR of the bits they output is equal to AND(a, b). In other words, they should output the same bit, except when both input bits are 1. Notice that no communication is allowed between them. A moment's reflection shows that their best strategy is to always output, say, 0. This allows them to win on three of the four possible inputs. It is also not difficult to show that equipping them with a shared source of randomness cannot help: the average success probability over the four possible questions will always be at most 75% (simply because one can always fix the shared randomness so as to maximize the average success probability). This bound of 75%, known as the Clauser–Horne–Shimony–Holt (CHSH) inequality, is the simplest example of a *Bell inequality* [6, 12].

A remarkable and well-known fact is that if Alice and Bob are allowed to share *quantum entanglement*, then they can win the game with probability $\approx$ 85%, no matter which questions are asked. Indeed, sharing entanglement allows remote parties to realize correlations that are impossible to obtain classically, without imparting them with the ability to communicate instantaneously. This distinction is one of the most peculiar aspects of quantum theory and required many years to be properly understood [15, 6].

In this paper we address the topic of quantum correlations from a communication complexity perspective. Namely, we are asking how many bits of communication are

[*] Received by the editors May 12, 2008; accepted for publication (in revised form) August 17, 2009; published electronically December 4, 2009.
http://www.siam.org/journals/sicomp/39-4/72390.html

[†] School of Computer Science, Tel Aviv University, Tel Aviv 69978, Israel. This author's research was supported by an Alon Fellowship, by the Binational Science Foundation, by the Israel Science Foundation, by the EU project QAP funded by the IST directorate as contract 015848, and by a European Research Council (ERC) starting grant.

[‡] School of Physics, The University of Melbourne, Victoria 3010, Australia (bentoner@bentoner.com). Part of this work was completed while the author was at CWI (Amsterdam) and Caltech. This author's research was supported by NSF grants PHY-0456720 and CCF-0524828, by ARO grant W911NF-05-1-0294, by the Dutch BSIK/BRICKS project, by EU project QAP, and by NWO VICI project 639-023-302.

needed to explain the phenomenon of quantum correlations. More precisely, we consider the following communication complexity problem, corresponding to the quantum mechanical scenario of a shared bipartite quantum state ρ and local two-outcome measurements $\mathbf{A}$ and $\mathbf{B}$, with the goal being to simulate the correlation (i.e., the parity) of the measurement results.

PROBLEM 1 SIMULATING QUANTUM CORRELATIONS.

Alice's input:	A $d^2 \times d^2$ positive semidefinite matrix ρ with trace 1 representing an operator on the space $\mathbb{C}^d \otimes \mathbb{C}^d$ and a $d \times d$ Hermitian matrix $\mathbf{A}$ with ± 1 eigenvalues.
Bob's input:	The (same) matrix ρ and a $d \times d$ Hermitian matrix $\mathbf{B}$ with ± 1 eigenvalues.
Alice's output:	A bit $\alpha \in \{-1, 1\}$.
Bob's output:	A bit $\beta \in \{-1, 1\}$.
Goal:	The correlation $\mathrm{E}[\alpha\beta]$ should satisfy $\mathrm{E}[\alpha\beta] = \mathsf{Tr}\left(\mathbf{A} \otimes \mathbf{B} \cdot \rho\right)$.

As we discuss in the open problems paragraph, this problem is a special case of the problem of simulating local measurements on quantum states, in which the goal is to simulate the entire output (as opposed to just the correlation) as well as to handle m-outcome measurements for $m > 2$.

In fact, this communication complexity problem can be stated in an entirely classical and much simpler language which we shall adopt from now on. The equivalence between the two formulations was established by Tsirelson [31] and will be described in Appendix A.

PROBLEM 2 SIMULATING QUANTUM CORRELATIONS (CLASSICAL FORMULATION).

Alice:	Receives as input a unit vector $\vec{a} \in \mathbb{R}^n$ and outputs a bit $\alpha \in \{-1, 1\}$.
Bob:	Receives as input a unit vector $\vec{b} \in \mathbb{R}^n$ and outputs a bit $\beta \in \{-1, 1\}$.
Goal:	The correlation $\mathrm{E}[\alpha\beta]$ should satisfy $\mathrm{E}[\alpha\beta] = \langle \vec{a}, \vec{b} \rangle$.

So if $\vec{a} = \vec{b}$, Alice and Bob must always output the same bit, whereas if $\vec{a} = -\vec{b}$, they must always output opposite bits. If, say, $\vec{a}$ is orthogonal to $\vec{b}$, then their outputs should be uncorrelated.

To see how the game described in the beginning of this section fits into this problem, consider the special case in which Alice's input is either the vector $\vec{a}_0 = (1, 0)$ or the vector $\vec{a}_1 = (0, 1)$ and Bob's input is either $\vec{b}_0 = \frac{1}{\sqrt{2}}(1, 1)$ or $\vec{b}_1 = \frac{1}{\sqrt{2}}(1, -1)$. Notice that $\langle \vec{a}_i, \vec{b}_j \rangle$ is $-\frac{1}{\sqrt{2}}$ if $i = j = 1$ and $\frac{1}{\sqrt{2}}$ otherwise. Therefore, if we are able to simulate quantum correlations in this case, then we can win the game with probability $\frac{1}{2} + \frac{1}{2\sqrt{2}} \approx 85\%$. Using our earlier observations, it follows that even when using shared randomness, one cannot solve Problem 2 without any communication, i.e., at least one bit of communication is required.

Previous work. The problem of simulating quantum correlations was introduced independently by several authors, including [23, 29, 8]. It is also closely related to a communication complexity problem introduced by Kremer, Nisan, and Ron [20]. Early work concentrated on the special case of dimension $n = 3$, which turns out to correspond to the case of a shared EPR pair. The protocol in [8] solves this special case with 8 bits of communication; this was subsequently improved to just one bit [30] (see also [10]). Up to now, the best known protocol for the general case of Problem 2 required $\lfloor n/2 \rfloor$ bits of one-way communication [4].

There has also been considerable work on other variants of the question. For instance, one might consider bounds on the *average* communication, as opposed to the worst case communication as we do here. The previous best result in this direction is by Degorre, Laplante, and Roland, who have shown that $(\log n)/2 + O(1)$ bits of communication suffice on average (over the shared randomness of Alice and Bob), but in their protocol the communication in the worst case is unbounded [13, 14]. Another variant of the question allows for an additive error of at most ε in the correlations. In this case there is a straightforward protocol that uses $O(\varepsilon^2 \log(1/\varepsilon))$ bits of communication, independent of n [20].

Our main result improves on all previous work by showing a solution to Problem 2 using a finite amount of communication, independent of the dimension n.

THEOREM 1.1. *There is a public-coin protocol for exactly simulating quantum correlations using two bits of one-way communication.*

We note that the shared randomness is essential: there is no exact private-coin protocol which has bounded communication in the worst case [22]. We also mention that the marginal distributions produced by our protocol (as well as all our intermediate protocols) are uniform. This can be verified from the description of the protocols. Alternatively, note that one can always obtain uniform marginals without changing the joint correlation by simply taking a shared random bit $r \in \{-1, 1\}$ and asking both Alice and Bob to multiply their outputs by r.

The main question left open in the preliminary version of this work was whether the theorem is tight, i.e., whether there is a one bit protocol for the problem. This question has recently been resolved by Vértesi and Bene [32], who showed that no one bit protocol exists. In section 5 we describe some of our own attempts to prove such a result. Although our attempts were unsuccessful, the approaches we describe might be of interest in the future.

Proof outline and techniques. We will start in section 2 by describing some basic protocols. All of these protocols have the property that for any input $\vec{a}$, $\vec{b}$, $\mathrm{E}[\alpha\beta] = h(\langle \vec{a}, \vec{b} \rangle)$ for some function $h : [-1, 1] \to [-1, 1]$, i.e., the correlation between the outputs depends only on the inner product between the input vectors. (In fact, any protocol can be transformed into one that has this property: by using the shared randomness, Alice and Bob can apply a random orthogonal rotation to their inputs; then, it is not difficult to prove that the resulting distribution on inputs depends only on the inner product between the original inputs.)

Our goal, of course, is to come up with a protocol whose "correlation function" h is simply $h(x) = x$. We therefore analyze the correlation functions of our basic protocols. The main part of the analysis is based on the calculation of areas of spherical triangles in four-dimensional space (a topic that was also at the heart of Karloff and Zwick's work on the approximation of MAX3SAT [18]). Unfortunately, as we will see in section 2, none of our basic protocols achieves $h(x) = x$ (see Figures 1 and 2 for plots of some of the correlation functions relative to the desired $h(x) = x$).

Instead, we will show in section 3 that one can take a protocol whose correlation function h is "strong enough" in some precise sense and transform it into another protocol whose correlation function is the desired $h(x) = x$. To complete the proof, we will show in section 4 that the correlation function of our "2-bit orthant protocol" is strong enough.

The transformation shown in section 3 is the heart of our construction. The idea is to carefully choose a mapping C from $\mathbb{R}^n$ to another (infinite-dimensional) Hilbert space with the property that for any vectors $\vec{a}$ and $\vec{b}$, $\langle C(\vec{a}), C(\vec{b})\rangle = f(\langle \vec{a}, \vec{b}\rangle)$ for some function $f : [-1, 1] \to [-1, 1]$. Then, in the transformed protocol, Alice and Bob simply run the original protocol on inputs $C(\vec{a})$ and $C(\vec{b})$. Clearly, this results in a protocol with correlation function $g(x) = h(f(x))$, where h is the correlation function of the original function. In order to achieve the desired correlation function $g(x) = x$, we need to choose f to be h^{-1} (assuming it is well defined of course). Intuitively speaking, the purpose of C is to slightly weaken the correlation function so that it matches the desired $h(x) = x$.

The main effort, therefore, is in constructing a mapping C with the property that $\langle C(\vec{a}), C(\vec{b})\rangle = h^{-1}(\langle \vec{a}, \vec{b}\rangle)$. To demonstrate how such a thing can be achieved, assume, for simplicity, that we have $h^{-1}(x) = x^3$. Then we can choose C to be the mapping $\vec{v} \mapsto \vec{v} \otimes \vec{v} \otimes \vec{v}$, where $\otimes$ denotes the tensor operation. It then follows from the definition that for any vectors $\vec{a}$ and $\vec{b}$,

$$\langle \vec{a} \otimes \vec{a} \otimes \vec{a}, \vec{b} \otimes \vec{b} \otimes \vec{b}\rangle = \langle \vec{a}, \vec{b}\rangle^3,$$

as required. In reality, the function h^{-1} will be much more involved, and we will construct C based on its power series expansion.

The idea of using a power series combined with a mapping C as above originates in Krivine's work on Grothendieck's constant [21]. More recently, Alon and Naor [2] showed that Krivine's method can be interpreted as an algorithmic rounding technique for a certain family of semidefinite programs, and this has since been extended in a series of papers (see, e.g., [11, 3, 1]). As far as we know, our result is the first application of Krivine's method to communication complexity.

Open problems. The problem we consider in this paper is a special case of the more general problem of simulating local measurements on quantum states. Here, as in our problem, Alice and Bob are given (the classical description of) a bipartite quantum state ρ on $\mathbb{C}^d \otimes \mathbb{C}^d$. In addition, Alice is given an m-outcome measurement A, and Bob is given an m-outcome measurement B. The goal is for Alice and Bob to output indices $\alpha, \beta \in [m]$ that are distributed as if they actually performed the measurements A and B on ρ.

The complexity of this general problem is still not well understood, even for very special cases. Note that we do *not* resolve the $m = 2$ special case in this paper: although our protocol gives the correct correlations, it generates uniform marginal distributions and not those predicted by quantum theory. The only case that is essentially resolved is the case $d = m = 2$ [9, 30]. Beyond that, no exact protocol with bounded worst-case communication is known, even for $(d, m) = (2, 3)$ or $(d, m) = (3, 2)$. Let us mention some other known results. First, Brassard, Cleve, and Tapp have established that $\Omega(d)$ bits of communication are necessary for exact simulation of d-outcome measurements on the maximally entangled state in $\mathbb{C}^d \otimes \mathbb{C}^d$ [8]. Massar et al. have shown that there is an exact private-coin protocol for this problem that uses $O(d \log d)$ bits of communication on average [22]. On the other hand, the same authors have shown that any exact protocol with bounded worst-case communication requires an unbounded number of public coins [22]. Finally, Shi and Zhu have shown that it is possible to simulate the required distribution to within variational distance ε using $O(m^6/\varepsilon^2 \log(m/\varepsilon))$ bits of communication [28].

Outline. In section 2 we present our basic protocols and calculate their correlation functions. None of these protocols has the right correlation function. Then, in section 3 we show a general technique to take any protocol with a "strong enough" correlation function, and transform it into one that achieves the right correlation function $h(x) = x$. This part is based on Krivine's power series method. Finally, in section 4 we complete the proof by showing that one of our basic protocols, the "2-bit orthant protocol," indeed has a strong enough correlation function. In subsection 4.1 we show a slightly better protocol that communicates roughly 1.82 bits on average. There is a discussion of lower bounds in section 5.

2. Basic protocols.

2. Basic protocols. In this section we present a number of basic communication protocols and calculate their correlation functions. None of these protocols has the right correlation function, but later we will show how to modify them so that the right correlation function is obtained.

2.1. Protocol with no communication. The following simple protocol uses no communication and is included for completeness. It is based on the "random hyperplane" idea used in [6, 17, 16].

PROTOCOL 1.

Random Variables:	Alice and Bob share a unit vector $\vec{\lambda} \in \mathbb{R}^n$ chosen uniformly at random from the unit sphere.
Alice:	Alice outputs $\alpha = \operatorname{sign}(\langle \vec{a}, \vec{\lambda} \rangle)$.
Bob:	Bob outputs $\beta = \operatorname{sign}(\langle \vec{b}, \vec{\lambda} \rangle)$.

LEMMA 2.1. *The output of Protocol* 1 *satisfies*

$$\mathrm{E}[\alpha\beta] = \frac{2}{\pi} \arcsin(\langle \vec{a}, \vec{b} \rangle).$$

In other words, its correlation function is $h(x) = \frac{2}{\pi}\arcsin(x)$.

Proof. Let $\vec{\mu}$ denote the projection of $\vec{\lambda}$ on the space spanned by $\vec{a}$ and $\vec{b}$, normalized to be of norm 1. By symmetry, $\vec{\mu}$ is distributed uniformly on the unit circle in that two-dimensional space. Therefore,

$$\Pr(\alpha \neq \beta) = \Pr(\operatorname{sign}(\langle \vec{\mu}, \vec{a} \rangle) \neq \operatorname{sign}(\langle \vec{\mu}, \vec{b} \rangle)) = \frac{1}{\pi} \arccos(\langle \vec{a}, \vec{b} \rangle).$$

It follows that

$$\mathrm{E}[\alpha\beta] = 1 - 2\Pr(\alpha \neq \beta) = \frac{2}{\pi} \arcsin(\langle \vec{a}, \vec{b} \rangle),$$

as required. □

2.2. The majority protocol. We now present a natural extension of the protocol in the previous section. This protocol, which we call the "majority" protocol, will not be used in what follows; instead, we will later describe a more efficient protocol. We present the majority protocol because its analysis is somewhat simpler, and so it may be useful for generalizing our technique to simulate stronger correlation functions.

The majority protocol, given as Protocol 2, is parameterized by a fixed even integer $k \geq 0$ and uses k bits of one-way communication. The idea is essentially to repeat the naive random half-space procedure from the last section $k + 1$ times, and then output bits α, β so that their product is equal to the majority of the $k + 1$ products $\alpha_0\beta_0, \ldots, \alpha_k\beta_k$ of the outputs of the individual protocols. The naive way of implementing this would require sending $k + 1$ bits from Alice and Bob. Namely,

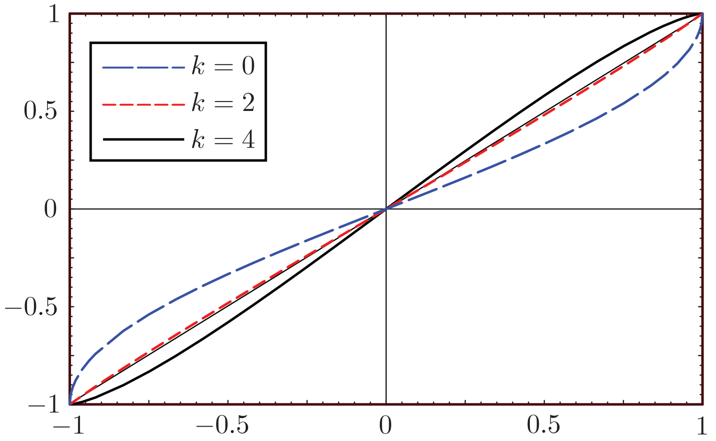

FIG. 1. *Correlation functions obtained by the majority protocol relative to the line $h(x)=x$.*

Alice outputs 1 and sends $\alpha_0,\dots,\alpha_k$ to Bob who outputs $\mathrm{MAJ}(\alpha_0\beta_0,\alpha_1\beta_1,\dots,\alpha_k\beta_k)$. Instead, a simple trick allows Protocol 2 to use only k bits: Alice outputs α_0 and sends $\alpha_1,\dots,\alpha_k$ to Bob, who outputs $\mathrm{MAJ}(\beta_0,\alpha_0\alpha_1\beta_1,\dots,\alpha_0\alpha_k\beta_k)$.

As k grows, the correlation function produced by the protocol becomes stronger, as shown in Figure 1. It turns out that $k=4$ bits are sufficient to be able to simulate quantum correlations. The proof of this fact is omitted, since we will instead use the more efficient "orthant" protocol described in subsection 2.3 below.

PROTOCOL 2.

Random Variables: Alice and Bob share $k+1$ unit vectors $\vec{\lambda}_i\in\mathbb{R}^n$ for $i=0,1,\dots,k$, chosen independently and uniformly at random from the unit sphere.

Alice: Let $\alpha_i=\mathrm{sign}(\langle\vec{a},\vec{\lambda}_i\rangle)$ for $i=0,1,\dots,k$. Alice outputs $\alpha=\alpha_0$ and sends to Bob the k bits $\alpha_0\alpha_1,\dots,\alpha_0\alpha_k$.

Bob: Let $\beta_i=\mathrm{sign}(\langle\vec{b},\vec{\lambda}_i\rangle)$ for $i=0,1,\dots,k$. Bob outputs

$$\beta=\mathrm{MAJ}(\beta_0,\alpha_0\alpha_1\beta_1,\dots,\alpha_0\alpha_k\beta_k),$$

where MAJ is the majority function.

Let $g_k^{\mathrm{MAJ}}:[0,1]\to[-1,+1]$ be defined by

$$g_k^{\mathrm{MAJ}}(p)=1-2\sum_{i=0}^{k/2}\binom{k+1}{i}(1-p)^i p^{k+1-i}.$$

Let

$$h_k^{\mathrm{MAJ}}(x)=g_k^{\mathrm{MAJ}}\left(\frac{1}{\pi}\arccos(x)\right).$$

LEMMA 2.2. *The correlation function of Protocol 2 is h_k^{MAJ}.*

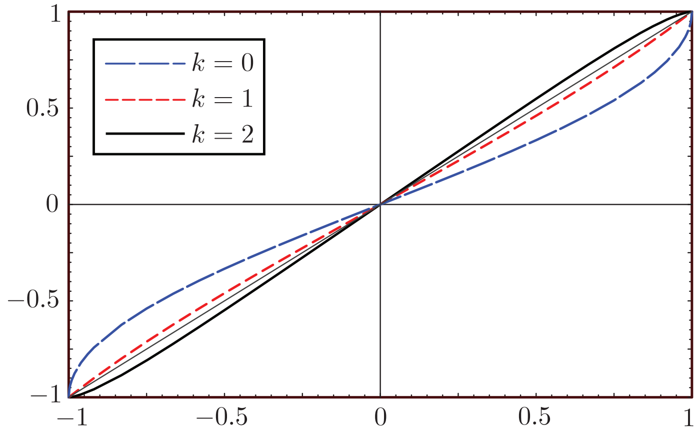

FIG. 2. *Correlation functions obtained by the orthant protocol relative to the line $h(x)=x$.*

Proof. Because the unit vectors $\vec{\lambda}_i$ are chosen independently, the events $\alpha_i=\beta_i$ for $i=1,\ldots,k+1$ are independent. Let $p=\Pr(\alpha_i\neq\beta_i)=(1-\mathrm{E}[\alpha_i\beta_i])/2$, which is independent of i. By Lemma 2.1,

$$p=\frac{1}{\pi}\arccos(\langle\vec{a},\vec{b}\rangle).$$

Thus

$$\Pr(\alpha\beta=-1)=\sum_{i=0}^{k/2}\binom{k+1}{i}(1-p)^i p^{k+1-i}.$$

Note that $\mathrm{E}[\alpha\beta]=1-2\Pr(\alpha\beta=-1)$, which completes the proof. □

2.3. The orthant protocol. In this section we present a more efficient protocol that, in some sense, seems to give the strongest possible correlations. The protocol is parameterized by an integer $k\geq 0$ and uses k bits of one-way communication. We call it the "orthant protocol," since it is based on the partitioning of $(k+1)$-dimensional space into its 2^{k+1} orthants (where an orthant is the higher-dimensional analogue of the two-dimensional quadrant). As we shall see below, the correlation function achieved by this protocol is determined by certain areas on the surface of the sphere in $k+2$ dimensions. Such questions seem difficult in general (see [18]). Luckily, for our purposes it suffices to consider the low-dimensional cases $k=0,1,2$, since the $k=2$ protocol already yields correlations that are strong enough. The correlation functions produced by the protocol are shown in Figure 2.

The protocol is given as Protocol 3. Roughly speaking, Alice and Bob start by projecting their vectors onto a random $(k+1)$-dimensional subspace. Alice then sends to Bob the orthant inside the $(k+1)$-dimensional space in which her vector lies, and Bob uses the half-space determined by this orthant to determine his output. To be more precise, instead of a random orthogonal projection, we use here a random Gaussian matrix G. This leads to a much cleaner analysis; moreover in the limit of large n, the two distributions are essentially the same. We also use the same trick used in the majority protocol to reduce the communication from the naive $k+1$ bits to k bits.

Protocol 3.

Random Variables: Alice and Bob share a random $(k+1)\times n$ matrix G, each of whose entries is an independent standard normal variable, i.e., a normal variable with mean 0 and variance 1.

Alice: Let $\alpha_i = \mathrm{sign}((G\vec{a})_i)$ for $i = 0, 1, \ldots, k$, and let $c_i = \alpha_0\alpha_i$ for $i = 1, 2, \ldots, k$. Alice outputs α_0 and sends to Bob the k bits $c_1, \ldots, c_k$.

Bob: Bob outputs

$$\beta = \mathrm{sign}\left[\langle G\vec{b}, (1, c_1, \ldots, c_k)\rangle\right].$$

We now analyze the correlation function given by this protocol. For any unit vectors $\vec{a}, \vec{b} \in \mathbb{R}^n$, the output of the protocol satisfies

$$\begin{aligned}\mathrm{E}[\alpha_0 \cdot \beta] &= \mathrm{E}[\mathrm{sign}[\alpha_0 \cdot \langle G\vec{b}, (1, c_1, \ldots, c_k)\rangle]]\\ &= \mathrm{E}[\mathrm{sign}[\langle G\vec{b}, (\alpha_0, \alpha_1, \ldots, \alpha_k)\rangle]],\end{aligned}$$

where expectations are taken over the choice of G. The expression inside the last expectation is $+1$ or -1, depending on whether $G\vec{b}$ is in the half-space defined by the center of the orthant containing $G\vec{a}$. By symmetry, it is enough to consider the positive orthant, and hence the above is equal to

$$2^{k+2}\Pr\left[\sum_{i=0}^{k}(G\vec{b})_i \geq 0 \text{ and for all } 0 \leq i \leq k, (G\vec{a})_i \geq 0\right] - 1. \tag{1}$$

We now claim that the joint distribution of $G\vec{a}$ and $G\vec{b}$ is a $(2k+2)$-dimensional Gaussian variable with mean 0 and covariance matrix

$$M = \begin{pmatrix} I & \rho I \\ \rho I & I \end{pmatrix},$$

where each I is a $(k+1)\times(k+1)$ identity matrix and ρ denotes the inner product $\langle \vec{a}, \vec{b}\rangle$. To see this, notice that by the rotational invariance of the Gaussian distribution, we can assume that $\vec{a} = (1, 0, \ldots, 0)$ and $\vec{b} = (\rho, \sqrt{1-\rho^2}, 0, \ldots, 0)$. The claim now follows by using the fact that the first two columns of G are two independent $(k+1)$-dimensional standard Gaussians, i.e., a Gaussian with mean 0 and covariance I.

Our next observation is that the probability in (1) depends only on the sum of coordinates of $G\vec{b}$. We therefore define the real random variable Z to be $\sum_{i=0}^{k}(G\vec{b})_i$. The joint distribution of $G\vec{a}$ and Z is given by a $(k+2)$-dimensional Gaussian with mean 0 and covariance matrix

$$M' = AMA^t = \begin{pmatrix} 1 & 0 & \cdots & 0 & \rho \\ 0 & 1 & \cdots & 0 & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \rho \\ \rho & \rho & \cdots & \rho & k+1 \end{pmatrix},$$

where A is the linear transformation taking $(G\vec{a}, G\vec{b})$ to $(G\vec{a}, Z)$. We therefore see that the probability in (1) is exactly the probability that a vector sampled from a Gaussian distribution with mean 0 and covariance matrix M' is in the positive orthant.

By the Cholesky decomposition, we can write $M' = C^t C$ for the $(k+2) \times (k+2)$ matrix

$$C = \begin{pmatrix} 1 & 0 & \cdots & 0 & \rho \\ 0 & 1 & \cdots & 0 & \rho \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & \rho \\ 0 & 0 & \cdots & 0 & \sqrt{(k+1)(1-\rho^2)} \end{pmatrix}.$$

It is easy to see that

$$C^{-1} = \begin{pmatrix} 1 & 0 & \cdots & 0 & -\dfrac{\rho}{\sqrt{(k+1)(1-\rho^2)}} \\ 0 & 1 & \cdots & 0 & -\dfrac{\rho}{\sqrt{(k+1)(1-\rho^2)}} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -\dfrac{\rho}{\sqrt{(k+1)(1-\rho^2)}} \\ 0 & 0 & \cdots & 0 & \dfrac{1}{\sqrt{(k+1)(1-\rho^2)}} \end{pmatrix}.$$

Since $(C^{-1})^t M' C^{-1} = I$, applying the linear transformation $(C^{-1})^t$ to a Gaussian random variable with mean 0 and covariance matrix M' transforms it into a standard Gaussian variable. Under this transformation, the positive orthant, which is the cone spanned by the standard basis vectors, becomes the cone spanned by the rows of C^{-1}. We conclude that the probability in (1) is exactly the probability that a vector sampled from a standard Gaussian distribution is in the cone spanned by the rows of C^{-1}. By the spherical symmetry of the standard Gaussian distribution, we can equivalently ask for the relative area of the sphere $S^{k+1} \subset \mathbb{R}^{k+2}$ that is contained inside the cone spanned by the rows of C^{-1}.

The case $k = 0$. We can now compute the probability in (1) for each of $k = 0, 1, 2$. We start with the simplest case of $k = 0$. Here, we are interested in the relative length of the circle S^1 contained in the cone spanned by the rows of C^{-1}. Obviously, this is given by the angle between the two vectors divided by 2π, which is $\arccos(-\rho)/(2\pi)$. Hence by (1) the correlation function in this case is simply

$$h_0^{\mathrm{ORT}}(\rho) := \frac{2}{\pi} \arccos(-\rho) - 1 = \frac{2}{\pi} \arcsin(\rho).$$

We could also obtain this result by noting that the $k = 0$ protocol is essentially identical to the one from section 2.1.

The case $k = 1$. We now analyze the more interesting case $k = 1$. Here, we are interested in the relative area of the sphere S^2 contained in the cone spanned by the three rows of C^{-1}. The intersection of S^2 with a cone spanned by three vectors is known as a *spherical triangle*; see Figure 3. Its area, as given by Girard's formula (see, e.g., [7, page 278]), is $\alpha_1 + \alpha_2 + \alpha_3 - \pi$, where $\alpha_1, \alpha_2, \alpha_3$ are the three angles of the triangle (as measured on the surface). In more detail, if v_1, v_2, v_3 are the vectors spanning the cone, then α_1 is the angle between the two vectors obtained by projecting v_2 and v_3 on the plane orthogonal to v_1 (and similarly for α_2 and α_3). In our case, the cone is spanned by $v_1 = (\sqrt{2(1-\rho^2)}, 0, -\rho)$, $v_2 = (0, \sqrt{2(1-\rho^2)}, -\rho)$, and $v_3 = (0, 0, 1)$. Clearly $\alpha_3 = \pi/2$, and a short calculation shows that $\alpha_1 = \alpha_2 =$

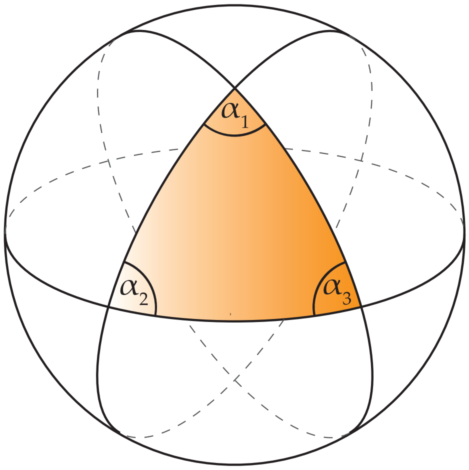

FIG. 3. *A spherical triangle.*

$\arccos(-\rho/\sqrt{2})$. Plugging this into Girard's formula and using the fact that the area of the sphere is 4π, we obtain that the relative area of S^2 contained in the cone spanned by the rows of C^{-1} is $(2\arccos(-\rho/\sqrt{2}) - \pi/2)/(4\pi)$. Hence by (1) the correlation function in this case is

$$h_1^{\mathrm{ORT}}(\rho) := \frac{4}{\pi}\arccos\left(-\frac{\rho}{\sqrt{2}}\right) - 2 = \frac{4}{\pi}\arcsin\left(\frac{\rho}{\sqrt{2}}\right).$$

The case $k = 2$. We finally arrive at the most important case $k = 2$. Here we are considering *spherical tetrahedra*, defined as the intersection of S^3 with a cone spanned by four vectors. Unlike the case of spherical triangles, no closed formula is known for the volume of a spherical tetrahedron (see [18] for further discussion and references). Fortunately, there is a simple formula for the *derivative* of the volume, as we describe in what follows.

We start with some preliminaries on spherical tetrahedra, closely following Appendix A in [18]. A spherical tetrahedron is defined by four unit vectors $v_0, v_1, v_2, v_3 \in S^3$ forming its vertices. For $0 \le i < j \le 3$, let $\theta_{ij} = \arccos(\langle v_i, v_j\rangle)$ be the angle between v_i and v_j. Equivalently, θ_{ij} is the spherical length of the edge ij. Another set of six parameters associated with a spherical tetrahedron are its *dihedral angles* λ_{ij}, $0 \le i < j \le 3$, describing the angle between the two faces meeting at the edge ij. They are defined as

$$\lambda_{01} = \arccos\frac{\langle v_0 \wedge v_1 \wedge v_2, v_0 \wedge v_1 \wedge v_3\rangle}{|v_0 \wedge v_1 \wedge v_2||v_0 \wedge v_1 \wedge v_3|}$$

and similarly for the other five dihedral angles, where the *high-dimensional inner product* is defined as

$$\langle a_1 \wedge a_2 \wedge a_3, b_1 \wedge b_2 \wedge b_3\rangle = \det\begin{pmatrix} \langle a_1, b_1\rangle & \langle a_1, b_2\rangle & \langle a_1, b_3\rangle \\ \langle a_2, b_1\rangle & \langle a_2, b_2\rangle & \langle a_2, b_3\rangle \\ \langle a_3, b_1\rangle & \langle a_3, b_2\rangle & \langle a_3, b_3\rangle \end{pmatrix}$$

and the *high-dimensional norm* is given by

$$|a_1 \wedge a_2 \wedge a_3| = \langle a_1 \wedge a_2 \wedge a_3, a_1 \wedge a_2 \wedge a_3\rangle^{1/2}.$$

Finally, in order to compute the volume of a spherical tetrahedron, we use a formula due to Schläfli [26], which says that for every $0 \le i < j \le 3$,

$$\frac{\partial \mathrm{Vol}}{\partial \lambda_{ij}} = \frac{\theta_{ij}}{2},$$

where $\mathrm{Vol} = \mathrm{Vol}(\lambda_{01}, \lambda_{02}, \lambda_{03}, \lambda_{12}, \lambda_{13}, \lambda_{23})$ is the volume of a spherical tetrahedron with the given edge lengths.

Our goal is to compute the volume of the spherical tetrahedron whose vertices are the rows of C^{-1} normalized to be of norm 1:

$$\begin{aligned} v_0 &= \left(\sqrt{3-3\rho^2}, 0, 0, -\rho\right)/\sqrt{3-2\rho^2},\\ v_1 &= \left(0, \sqrt{3-3\rho^2}, 0, -\rho\right)/\sqrt{3-2\rho^2},\\ v_2 &= \left(0, 0, \sqrt{3-3\rho^2}, -\rho\right)/\sqrt{3-2\rho^2},\\ v_3 &= (0,0,0,1)\,. \end{aligned}$$

From this it easily follows that

$$\theta_{03} = \theta_{13} = \theta_{23} = \arccos(-\rho/\sqrt{3-2\rho^2}) \quad \text{and} \quad \theta_{01} = \theta_{02} = \theta_{12} = \arccos(\rho^2/(3-2\rho^2)).$$

Moreover, a straightforward calculation reveals that

$$\lambda_{03} = \lambda_{13} = \lambda_{23} = \pi/2 \quad \text{and} \quad \lambda_{01} = \lambda_{02} = \lambda_{12} = \arccos(-\rho/\sqrt{3})$$

and that the derivative of the latter term as a function of ρ is $(3-\rho^2)^{-1/2}$. By using Schläfli's formula and integrating along ρ, we obtain that the volume of our spherical tetrahedron is

$$\int_{-1}^{\rho} 3 \cdot (1/2) \arccos(\sigma^2/(3-2\sigma^2)) \cdot (3-\sigma^2)^{-1/2} d\sigma,$$

where we used that for $\rho = -1$ this volume is 0. Since the total area of S^3 is $2\pi^2$, we obtain using (1) that the correlation function in this case is

$$h_2^{\mathrm{ORT}}(\rho) := \frac{12}{\pi^2} \int_{-1}^{\rho} \frac{\arccos(\sigma^2/(3-2\sigma^2))}{\sqrt{3-\sigma^2}} d\sigma - 1. \tag{2}$$

3. Simulation of the joint correlation.

In this section we describe how to take any protocol whose correlation function is "strong enough" and use it to solve Problem 2. This section and the next one rely on some basic facts from the theory of real analytic functions which can be found in, e.g., [19]. As we said earlier, the idea is to carefully choose a mapping C such that when Alice and Bob apply the protocol on $C(\vec{a})$ and $C(\vec{b})$, the resulting correlation function will be correct.

Protocol 4 Transformed protocol.

Alice and Bob map their vectors to $C(\vec{a})$ and $C(\vec{b})$ and run the original protocol on these vectors.

Fix some arbitrary correlation function $h : [-1,1] \to [-1,1]$. We now give sufficient conditions on h under which the required transformation C exists. First, we require that $h(1) = 1$ and that h is odd, continuous, and monotonically increasing. Moreover, we require that its series expansion about 0,

$$h(x) = \sum_{k=0}^{\infty} c_{2k+1} x^{2k+1}, \tag{3}$$

converges to $h(x)$ on the interval $(-1, 1)$, which implies that h is (real) analytic on $(-1, 1)$. Finally, we require that $c_1 > 0$ and $c_{2k+1} \leq 0$ for all $k > 0$.

In section 4 we will show that the orthant protocol with $k = 2$ satisfies these properties. A crucial fact for our protocol is that under the above requirements on h, the power series of h^{-1} converges on $[-1, 1]$, and all its coefficients are nonnegative. This is shown in the following lemma.

LEMMA 3.1. *If h satisfies the above conditions, then h^{-1} has a power series expansion*

$$h^{-1}(x) = \sum_{k=0}^{\infty} d_{2k+1} x^{2k+1} \tag{4}$$

that converges on $[-1, 1]$ and satisfies $d_{2k+1} \geq 0$ for all $k \geq 0$.

Proof. We first notice that under the above conditions, the power series in (3) converges to h also at the endpoints $-1, 1$. This is easy to prove and follows from Abel's lemma, which says that if all but finitely many of the coefficients of a power series are nonnegative (or nonpositive), then the value of the series at the endpoints is given by the limit of its values as we approach the endpoint. We hence see that $\sum_{k=0}^{\infty} c_{2k+1} = 1$.

It also follows easily that the inverse function $h^{-1} : [-1, 1] \to [-1, 1]$ is well defined and is odd. Moreover, by the real analytic inverse function theorem (see [19, Theorem 1.5.3]), h^{-1} is analytic on $(-1, 1)$ and hence has a series expansion about $x = 0$, as in (4). In order to analyze this series, we use a known formula for the coefficients of an inverse function (see, e.g., [24, equation (4.5.12)]):

$$d_k = \frac{1}{k c_1^k} \sum_{\ell_1, \ell_2, \ldots} \frac{(k)(k+1)\cdots(k-1+\ell_1+\ell_2+\cdots)}{\ell_1! \ell_2! \ell_3! \cdots} \left(-\frac{c_2}{c_1}\right)^{\ell_1} \left(-\frac{c_3}{c_1}\right)^{\ell_2} \cdots,$$

where the sum runs over nonnegative integers satisfying $\ell_1 + 2\ell_2 + 3\ell_3 + \cdots = k - 1$. Since in our case every term in the sum is nonnegative, it follows that $d_k \geq 0$ for all k, as required.

It remains to show that the series converges on $[-1, 1]$. In fact, it is sufficient to show that the series converges on $(-1, 1)$: convergence at the endpoints $-1, 1$ would follow by Abel's lemma, as before. We do this by showing that $d_k \leq 1/k$, as this immediately implies that the power series converges on the interval $(-1, 1)$. Using the above formula, we get that for all $k > 0$, d_{2k+1} is equal to

$$\frac{1}{(2k+1) c_1^{2k+1}} \sum_{\ell_2, \ell_4, \ldots} \frac{(2k+1)(2k+2)\cdots(2k+\ell_2+\ell_4+\cdots)}{\ell_2! \ell_4! \cdots} \left(-\frac{c_3}{c_1}\right)^{\ell_2} \left(-\frac{c_5}{c_1}\right)^{\ell_4} \cdots,$$

where the sum runs over nonnegative integers satisfying $\ell_2 + 2\ell_4 + 3\ell_6 + \cdots = k$. We extend the sum to all nonnegative integers $\ell_2, \ell_4, \ldots$, obtaining

$$d_{2k+1} \le \frac{1}{(2k+1)c_1^{2k+1}} \sum_{m=0}^{\infty} \binom{2k+m}{m} \sum_{\ell_2,\ell_4,\ldots} \frac{m!}{\ell_2!\ell_4!\cdots} \left(-\frac{c_3}{c_1}\right)^{\ell_2} \left(-\frac{c_5}{c_1}\right)^{\ell_4} \cdots \tag{5}$$

$$= \frac{1}{(2k+1)c_1^{2k+1}} \sum_{m=0}^{\infty} \binom{2k+m}{m} \left(\frac{-c_3 - c_5 - \cdots}{c_1}\right)^{m}, \tag{6}$$

where the inner sum in (5) is over all indices $\ell_2 + \ell_4 + \cdots = m$ and we used the multinomial theorem to obtain (6). By our observation above, $-\sum_{k=1}^{\infty} c_{2k+1} = c_1 - 1$. Set $z = 1 - 1/c_1$. Then $0 \le z < 1$ and

$$\begin{aligned} d_{2k+1} &\le \frac{1}{2k+1}(1-z)^{2k+1} \sum_{m=0}^{\infty} \binom{2k+m}{m} z^m \\ &= \frac{1}{2k+1}(1-z)^{2k+1}(1-z)^{-(2k+1)} = \frac{1}{2k+1}, \end{aligned}$$

since the sum is just the negative binomial series. We conclude that the power series converges to h^{-1} on the interval $(-1, 1)$, which also implies convergence at the endpoints by Abel's lemma. □

The transformation C is obtained by applying the following lemma to h^{-1}.

LEMMA 3.2. *Let $f : [-1, 1] \to [-1, 1]$ be a function with a power series expansion $f(x) = \sum_{k=0}^{\infty} d_k x^k$ that converges on $[-1, 1]$ and satisfies $d_k \ge 0$ for all $k \ge 0$ and $f(1) = 1$. Then for any $n \ge 1$, there exists a transformation $C : S^{n-1} \to S^{\infty}$ such that for all $\vec{a}, \vec{b} \in S^{n-1}$, $\langle C(\vec{a}), C(\vec{b})\rangle = f\big(\langle \vec{a}, \vec{b}\rangle\big)$.*

Proof. Define

$$C(\vec{v}) = \bigoplus_{k=0}^{\infty} \sqrt{d_k}\, \vec{v}^{\,\otimes k},$$

where $\vec{v}^{\,\otimes k}$ denotes the vector $\vec{v} \otimes \vec{v} \otimes \cdots \otimes \vec{v}$ with k tensor factors. Note that this is well defined, since $d_k \ge 0$ for all k. By definition, for any $\vec{a}, \vec{b} \in S^{n-1}$ we have

$$\langle C(\vec{a}), C(\vec{b})\rangle = \sum_{k=0}^{\infty} d_k \langle \vec{a}^{\otimes k}, \vec{b}^{\otimes k}\rangle = \sum_{k=0}^{\infty} d_k \langle \vec{a}, \vec{b}\rangle^k = f\big(\langle \vec{a}, \vec{b}\rangle\big),$$

which in particular implies that $C(\vec{a})$ is a unit vector for any $\vec{a} \in S^{n-1}$. □

Remark 3.3. By Schoenberg's theorem [27, Theorem 2], the conditions on f in Lemma 3.2 are in fact necessary for the transformation C to exist. It is also known that if we are only interested in a transformation C for a particular value of n (rather than for all $n \ge 1$), it is sufficient (and necessary) to require that f, in addition to satisfying $f(1) = 1$, has a nonnegative convergent series expansion in Gegenbauer polynomials [21, 27].

THEOREM 3.4. *Protocol* 4 *is well defined and solves Problem* 2.

Proof. By Lemma 3.1, we can apply Lemma 3.2 to h^{-1} to obtain the transformation C. Since C maps unit vectors to unit vectors, Protocol 4 is well defined. Moreover, its output satisfies

$$\mathrm{E}[\alpha\beta] = h\big(\langle C(\vec{a}), C(\vec{b})\rangle\big) = h \circ h^{-1}\big(\langle \vec{a}, \vec{b}\rangle\big) = \langle \vec{a}, \vec{b}\rangle,$$

as required. □

Remark 3.5. As mentioned in Remark 3.3, the properties of h^{-1} in the conclusion of Lemma 3.1 are not just sufficient, but are also necessary for Protocol 4 to be well defined. It is therefore natural to ask if the conditions on h in Lemma 3.1 are also necessary. It turns out that they are not: take, for example, h such that $h^{-1}(x) = 0.9x+0.1x^3$, calculate that $h(x) = 1.11x-0.15x^3+0.06x^5+O(x^7)$, and notice that the coefficient of x^5 is positive. So is there a necessary and sufficient condition? We do not know. Looking at Figure 3, one might be tempted to replace the condition that $c_1 > 0$ and $c_{2k+1} \leq 0$ for all $k > 0$ with the weaker condition that $h(x) \geq x$ for $x \geq 0$. But this condition is not sufficient: the function $h(x) = x+0.1x^3-0.1x^5$ satisfies $h(x) \geq x$ for $x \geq 0$ (as well as our other requirements), but $h^{-1}(x) = x - 0.1x^3 + O(x^5)$.

4. Analysis of the power series. In this section, we show that the orthant protocol with $k = 2$ satisfies the requirements listed in section 3 and hence can be used to simulate quantum correlations with only two bits of communication. A similar but much more involved analysis holds also for the majority protocol with $k = 4$ and implies a protocol with four bits of communication. We omit this analysis, since the orthant protocol is superior in all respects.

LEMMA 4.1. *Let $h(x) = h_2^{ORT}(x)$ be as given in* (2). *Then $h(1) = 1$, and h is odd, continuous, and monotonically increasing. Moreover, it is (real) analytic on* $(-1, 1)$*, and its power series about $x = 0$,*

$$h(x) = \sum_{k=0}^{\infty} c_{2k+1}x^{2k+1},$$

satisfies $c_1 > 0$ and $c_{2k+1} < 0$ for all $k > 0$.

Proof. The first four conditions are obvious. Moreover, being composed of analytic functions, h is easily seen to be analytic on $(-1, 1)$. We now show that the coefficients c_{2k+1} have the right sign. From the derivative

$$h'(x) = \frac{12\arccos(x^2/(3-2x^2))}{\pi^2\sqrt{3-x^2}},$$

it follows that $c_1 = h'(0) = 2\sqrt{3}/\pi > 0$. To show that the rest of the coefficients are negative, consider the second derivative

$$h''(x) = -\frac{24x}{\pi^2}H_1(x^2)H_2(x^2),$$

where

$$H_1(t) = \frac{1}{(3-t)^{3/2}(3-2t)}, \tag{7}$$

$$H_2(t) = \sqrt{\frac{3(3-t)}{1-t}} - \frac{3-2t}{2}\arccos\left(\frac{t}{3-2t}\right).$$

It is clear from (7) that all the coefficients in the power series of H_1 about 0 are positive. Therefore, it is sufficient to show that all the coefficients in the power series of H_2 are positive. We calculate $H_2(0) = 3 - 3\pi/4 > 0$ and $H_2'(0) = (3 + \pi)/2 > 0$. Next, we calculate

$$H_2''(t) = \frac{\sqrt{3}}{2}\left(7 + \frac{3}{1-t}\right)\frac{1}{(3-2t)(1-t)^{3/2}\sqrt{3-t}}.$$

Hence all coefficients in the power series of H_2 are positive, as required. □

4.1. The 1.82 bit protocol. In this section we observe that the amount of communication can be lowered to 1.82 bits on average by performing the $k = 1$ orthant protocol with probability

$$p := \frac{8 - 2\pi}{8 + \left(\sqrt{6} - 2\right)\pi} \approx 0.18$$

and the $k = 2$ orthant protocol the remainder of the time. This is the largest value of p for which our protocol works. One could possibly obtain a slightly better protocol by directly modifying the $k = 2$ orthant protocol, but the analysis seems to get too complicated.

LEMMA 4.2. *Let $h(x) = ph_1^{ORT}(x) + (1-p)h_2^{ORT}(x)$. Then $h(1) = 1$, and h is odd, continuous, and monotonically increasing. Moreover, it is (real) analytic on $(-1, 1)$, and its power series about $x = 0$,*

$$h(x) = \sum_{k=0}^{\infty} c_{2k+1} x^{2k+1},$$

satisfies that $c_1 > 0$ and $c_{2k+1} \le 0$ for all $k > 0$.

Proof. The first five properties are easy to verify as before. A short calculation shows that $c_1 > 0$. Then calculate

$$h''(x) = -\frac{24x}{\pi^2} H_1(x^2) H_2(x^2),$$

where

$$\begin{aligned}
H_1(t) &= \frac{1}{(2-t)^{3/2}},\\
H_2(t) &= -p\frac{\pi}{6} + (1-p)\frac{1}{(3-t)^{3/2}} H_3(t),\\
H_3(t) &= (2-t)^{3/2}\left[\frac{1}{3-2t}\sqrt{\frac{3(3-t)}{1-t}} - \frac{1}{2}\arccos\left(\frac{t}{3-2t}\right)\right].
\end{aligned}$$

It is clear that the coefficients c_{2k+1} will be nonpositive for all $k > 0$ if $H_1(t)$ and $H_2(t)$ have series expansions with nonnegative coefficients. This is clear for $H_1(t)$. For $H_2(t)$, first notice that $H_2(0) = 0$ (this explains our choice of p) and then note that it has a series expansion with nonnegative coefficients if $H_3(t)$ does. We calculate $H_3(0) = (4 - \pi)/\sqrt{2} > 0$ and differentiate to obtain

$$H_3'(t) = \frac{3\sqrt{2-t}}{4}\left[\frac{2\sqrt{3-t}\,(5-4t)}{\sqrt{3}(3-2t)^2(1-t)^{3/2}} + \arccos\left(\frac{t}{3-2t}\right)\right].$$

From this we calculate $H_3'(0) = (20 + 9\pi)/12\sqrt{2} > 0$ and then differentiate again, finding

$$H_3''(t) = \frac{3\sqrt{3}}{4\sqrt{2-t}} H_4(t),$$

where

$$H_4(t) = \frac{79 - 157t + 85t^2 + 11t^3 - 20t^4 + 4t^5}{(3-2t)^3(1-t)^{5/2}\sqrt{3-t}} - \frac{1}{2\sqrt{3}}\arccos\left(\frac{t}{3-2t}\right).$$

We calculate $H_4(0) = (316 - 27\pi)/108\sqrt{3} > 0$, differentiate once more, and find that

$$H_4'(t) = \frac{1}{54\sqrt{1-t}(3-t)^{3/2}} \left[1297 + \frac{9399t}{(3-2t)} + \frac{30696t^2}{(3-2t)^2} + \frac{66116t^3}{(3-2t)^3} + \frac{115080t^4}{(3-2t)^4} + \frac{59616t^5}{(3-2t)^4(1-t)} + \frac{8748t^6}{(3-2t)^4(1-t)^2} + \frac{540t^7}{(3-2t)^4(1-t)^3} \right],$$

which we have written in a form that makes it clear that $H_4(t)$ has a series expansion with nonnegative coefficients. Tracing backwards through the proof, we conclude that the coefficients in the series expansion of $h(x)$ about $x = 0$ have the desired property. □

5. Lower bounds. As mentioned in the introduction, it has recently been shown by Vértesi and Bene [32] that our main theorem is tight, i.e., no one bit protocol exists for Problem 2. Here we describe some of our own attempts to prove this, which, although unsuccessful, might shed further light on the complexity of the problem.

Consider the special case when $\vec{a}$ and $\vec{b}$ are both uniformly distributed in S^{n-1} but constrained so that $\langle \vec{a}, \vec{b} \rangle = \pm(1-\varepsilon)$, with $0 < \varepsilon \ll 1$. Given a one bit protocol $\mathcal{P}$, define the function

$$B_n(\varepsilon) = 2 - \mathrm{E}_{\langle \vec{a}, \vec{b} \rangle = 1-\varepsilon} \left[\mathrm{E}_{\mathcal{P}}[\alpha\beta] \right] + \mathrm{E}_{\langle \vec{a}, \vec{b} \rangle = -1+\varepsilon} \left[\mathrm{E}_{\mathcal{P}}[\alpha\beta] \right],$$

where the outer expectations are taken over all vectors $\vec{a}, \vec{b} \in S^{n-1}$ with inner product $\pm(1-\varepsilon)$ and the inner expectations are taken over any shared randomness used by the protocol $\mathcal{P}$. If $\mathcal{P}$ solves Problem 2, then we necessarily have $\mathrm{E}_{\mathcal{P}}[\alpha\beta] = \langle \vec{a}, \vec{b} \rangle$, and hence $B_n(\varepsilon) = 2\varepsilon$ for all n and ε. Our earlier analysis shows that for the orthant protocol (Protocol 3) with $k = 1$, as ε goes to 0, $B_n(\varepsilon)$ approaches $8\varepsilon/\pi > 2\varepsilon$. We conjecture that for sufficiently large n and sufficiently small ε, all one bit protocols satisfy $B_n(\varepsilon) > 2\varepsilon$ (and therefore do not solve Problem 2). In fact, we conjecture that the orthant protocol with $k = 1$ is optimal for $B_n(\varepsilon)$, i.e., we conjecture that for all one bit protocols, $\lim_{n\to\infty,\varepsilon\to 0} B_n(\varepsilon)/\varepsilon \geq 8/\pi$.

One approach to prove these conjectures is the following. First, since we are only interested in minimizing the value of $B_n(\varepsilon)$ and not in obtaining the correct correlations, we can restrict attention to deterministic protocols. Any deterministic protocol partitions Alice's sphere S^{n-1} into four sets $R(\alpha, c)$, depending on which bit α she outputs and which bit c she sends. Once we have specified Alice's strategy, we can assume Bob acts optimally to minimize $B_n(\varepsilon)$. The contribution to $B_n(\varepsilon)$ comes from regions near which $R(-1, c)$ meets $R(+1, c)$, since it is in these areas that Bob cannot tell whether Alice outputs +1 or −1 and hence cannot correlate his answer perfectly with Alice's. Therefore, in order to prove the conjectures, one should argue that any protocol $\mathcal{P}$ must have local regions where $R(-1, c)$ meets $R(+1, c)$ and that the way these regions meet in the $k = 1$ orthant protocol is optimal. Formalizing this notion would seem to require topological arguments, perhaps an extension of the Borsuk–Ulam theorem.

In another attempt to shed light on the problem, we show now how to extend a lower bound of Barrett, Kent, and Pironio [5], who improved on an earlier result of Pironio [25]. Barrett, Kent, and Pironio showed that if we examine the transcript of communication between Alice and Bob of any protocol for Problem 2, then with probability 1 (over the shared randomness used by the protocol) the transcript must show some communication. In other words, it cannot be the case that Alice and Bob sometimes output results using shared randomness alone. But this leaves open the possibility that, say, Alice almost always sends the same message to Bob.

Here, we show a lower bound on the (min)entropy of the communication transcript. More specifically, we show an upper bound on the maximum probability with which a transcript can appear in a protocol for Problem 2.

PROPOSITION 5.1. *There exists a distribution on inputs such that in any protocol that solves Problem 2, no transcript can appear with probability greater than* $(3-\sqrt{2})/2 \approx 0.79$ *when applied to this input distribution.*

Proof. Let P be a protocol that solves Problem 2. As mentioned in the introduction, P in particular allows us to solve the following problem with probability $\frac{1}{2}+\frac{1}{2\sqrt{2}} \approx 0.85$: Alice and Bob receive bits a and b, respectively, and their task is to output one bit each in such a way that the XOR of the bits they output is equal to $\mathrm{AND}(a,b)$. Assume the bits a and b are chosen uniformly at random, and consider the resulting distribution on transcripts created by P. Consider the most likely transcript T, and let p denote the probability with which it occurs. We now construct a protocol P' with no communication as follows. Alice checks whether the transcript T is consistent with her input. If so, she outputs a bit as in P; if not, she outputs a random bit. Bob does the same. Note that with probability p, P' behaves identically to P. With probability $1-p$, however, at least one of the parties detects that the transcript is not consistent with his or her input and outputs a random bit. By the definition of the problem, in this case the success probability is $\frac{1}{2}$. Therefore, the overall success probability of P' is at least

$$\frac{1}{2}+\frac{1}{2\sqrt{2}}-\frac{1-p}{2},$$

which must be at most $\frac{3}{4}$ by the CHSH inequality. □

Appendix A. A classical reformulation. The equivalence between Problem 1 and Problem 2 is due to Tsirelson [31]. Here we sketch only the easy direction of this equivalence: a solution to Problem 2 implies a solution to Problem 1.

Let ρ be a state on $\mathbb{C}^d \otimes \mathbb{C}^d$, and let A and B be $d \times d$ Hermitian matrices whose eigenvalues are in $\{-1,1\}$. The goal is for Alice and Bob to output bits α and β such that

$$\mathrm{E}[\alpha\beta] = \mathsf{Tr}\left(A \otimes B \cdot \rho\right).$$

Let $a_1,\ldots,a_{d^2} \in \mathbb{C}$ be the d^2 entries of the matrix $A \otimes \mathbb{1}_B \sqrt{\rho}$, and similarly let $b_1,\ldots,b_{d^2} \in \mathbb{C}$ be those of $\mathbb{1}_A \otimes B\sqrt{\rho}$. Let $n=2d^2$, and define the n-dimensional real vectors

$$\begin{aligned}\vec{a} &= (\mathrm{Re}\, a_1,\ldots,\mathrm{Re}\, a_{d^2},\mathrm{Im}\, a_1,\ldots,\mathrm{Im}\, a_{d^2}),\\ \vec{b} &= (\mathrm{Re}\, b_1,\ldots,\mathrm{Re}\, b_{d^2},\mathrm{Im}\, b_1,\ldots,\mathrm{Im}\, b_{d^2}).\end{aligned}$$

Then

$$\langle \vec{a},\vec{a}\rangle = \sum_{j=1}^{d^2} |a_j|^2 = \mathsf{Tr}(A \otimes \mathbb{1}_B \rho A \otimes \mathbb{1}_B) = \mathsf{Tr}(A^2 \otimes \mathbb{1}_B \cdot \rho) = \mathsf{Tr}(\rho) = 1,$$

and similarly $\langle \vec{b},\vec{b}\rangle = 1$. Moreover,

$$\langle \vec{a},\vec{b}\rangle = \sum_{j=1}^{d^2} a_j \cdot b_j^* = \mathsf{Tr}(A \otimes \mathbb{1}_B \cdot \rho \cdot \mathbb{1}_A \otimes B) = \mathsf{Tr}\left(A \otimes B \cdot \rho\right).$$

Hence, Alice and Bob can use $\vec{a}$ and $\vec{b}$ as input to Problem 2 in order to solve Problem 1.

Acknowledgments. Part of this work was done while we were visiting the Institut Henri Poincaré as part of the program "Quantum information, computation and complexity," and we would like to thank the organizers for their efforts. We thank Aram Harrow for discussions about lower bounds, Falk Unger for assistance with the proof of Lemma 3.1, and Peter Harremoës for discussions about Schoenberg's theorem.

REFERENCES

[1] N. Alon, K. Makarychev, Y. Makarychev, and A. Naor, *Quadratic forms on graphs*, Invent. Math., 163 (2006), pp. 499–522.

[2] N. Alon and A. Naor, *Approximating the cut-norm via Grothendieck's inequality*, SIAM J. Comput., 35 (2006), pp. 787–803. Preliminary version in STOC'04.

[3] S. Arora, E. Berger, E. Hazan, G. Kindler, and S. Safra, *On non-approximability for quadratic programs*, in Proceedings of the 46th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2005, pp. 206–215.

[4] D. Bacon and B. F. Toner, *How to simulate quantum correlations*, unpublished.

[5] J. Barrett, A. Kent, and S. Pironio, *Maximally nonlocal and monogamous quantum correlations*, Phys. Rev. Lett., 97 (2006), 170409.

[6] J. S. Bell, *On the Einstein-Podolsky-Rosen paradox*, Physics, 1 (1964), pp. 195–200.

[7] M. Berger, *Geometry* II, Universitext, Springer-Verlag, Berlin, 1987.

[8] G. Brassard, R. Cleve, and A. Tapp, *Cost of exactly simulating quantum entanglement with classical communication*, Phys. Rev. Lett., 83 (1999), pp. 1874–1877.

[9] N. J. Cerf, N. Gisin, and S. Massar, *Classical teleportation of a quantum bit*, Phys. Rev. Lett., 84 (2000), 2521.

[10] N. J. Cerf, N. Gisin, S. Massar, and S. Popescu, *Simulating maximal quantum entanglement without communication*, Phys. Rev. Lett., 94 (2005), 220403.

[11] M. Charikar and A. Wirth, *Maximizing quadratic programs: Extending Grothendieck's inequality*, in Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 2004, pp. 54–60.

[12] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett., 23 (1969), pp. 880–884.

[13] J. Degorre, S. Laplante, and J. Roland, *Simulating quantum correlations as a distributed sampling problem*, Phys. Rev. A (3), 72 (2005), 062314.

[14] J. Degorre, S. Laplante, and J. Roland, *Classical simulation of traceless binary observables on any bipartite quantum state*, Phys. Rev. A (3), 75 (2007), 012309.

[15] A. Einstein, P. Podolsky, and N. Rosen, *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev., 47 (1935), pp. 777–780.

[16] M. X. Goemans and D. P. Williamson, *Improved approximation algorithms for maximum cut and satisfiability problems using semidefinite programming*, J. Assoc. Comput. Mach., 42 (1995), pp. 1115–1145. Preliminary version in STOC'94.

[17] A. Grothendieck, *Résumé de la théorie métrique des produits tensoriels topologiques*, Bol. Soc. Mat. São Paulo, 8 (1953), pp. 1–79.

[18] H. Karloff and U. Zwick, *A 7/8-approximation algorithm for MAX 3SAT?*, in Proceedings of the 38th Annual IEEE Symposium on Foundations of Computer Science (FOCS), 1997, pp. 406–415.

[19] S. G. Krantz and H. R. Parks, *A Primer of Real Analytic Functions*, 2nd ed., Birkhäuser Advanced Texts: Basler Lehrbücher., Birkhäuser, Boston, 2002.

[20] I. Kremer, N. Nisan, and D. Ron, *On randomized one-round communication complexity*, Comput. Complexity, 8 (1999), pp. 21–49. Preliminary version in STOC'95.

[21] J. L. Krivine, *Constantes de Grothendieck et fonctions de type positif sur les sphères*, Adv. Math., 31 (1979), pp. 16–30.

[22] S. Massar, D. Bacon, N. J. Cerf, and R. Cleve, *Classical simulation of quantum entanglement without local hidden variables*, Phys. Rev. A (33), 63 (2001), 052305.

[23] T. Maudlin, *Bell's inequality, information transmission, and prism models*, in Philosophy of Science Association 1992, Vol. 1, D. Hull, M. Forbes, and K. Okruhlik, eds., Philosophy of Science Association, Bloomington, IN, 1992, pp. 404–417.

[24] P. M. Morse and H. Feshbach, *Methods of Theoretical Physics, Part* I, McGraw-Hill, New York, 1953.

[25] S. Pironio, *Violations of Bell inequalities as lower bounds on the communication cost of nonlocal correlations*, Phys. Rev. A (3), 68 (2003), 062102.

[26] L. Schläfli, *On the multiple integral* $\int^n dx\, dy \dots dz$*, whose limits are* $p_1 = a_1x + b_1y + \cdots + h_1z > 0$*,* $p_2 > 0, \dots, p_n > 0$*, and* $x^2 + y^2 + \cdots + z^2 < 1$, Quart. J. Math, 2 (1858), pp. 269–300. Continued in Vol. 3 (1860), pp. 54–68 and pp. 97–108.

[27] I. J. Schoenberg, *Positive definite functions on spheres*, Duke Math. J., 9 (1942), pp. 96–108.

[28] Y. Shi and Y. Zhu, *Tensor norms and the classical communication complexity of nonlocal quantum measurement*, SIAM J. Comput., 38 (2008), pp. 753–766. Preliminary version in FOCS'05.

[29] M. Steiner, *Towards quantifying non-local information transfer: Finite-bit non-locality*, Phys. Lett. A, 270 (2000), pp. 239–244.

[30] B. F. Toner and D. Bacon, *Communication cost of simulating Bell correlations*, Phys. Rev. Lett., 91 (2003), 187904.

[31] B. S. Tsirelson, *Quantum analogues of the Bell inequalities. The case of two spatially separated domains*, J. Soviet Math., 36 (1987), pp. 557–570.

[32] T. Vértesi and E. Bene, *Lower bound on the communication cost of simulating bipartite quantum correlations*, arXiv:0904.1390.