Information and quantum nonseparability

B. W. Schumacher
Department of Physics, Kenyon College, Gambier, Ohio 43022
(Received 16 December 1988)

An information-theoretic inequality analogous to the well-known result of Bell [Physics **1**, 195 (1964)] is formulated using the concept of information distance. This inequality, like Bell's, is true for all local-hidden-variable theories, but not for quantum mechanics. The metric space structure of this new inequality suggests a reformulation of familiar Bell inequalities in terms of a "covariance distance." Quantum nonseparability can be demonstrated through violations of these inequalities even in cases where the correlation between two systems is extremely weak. The connection between nonseparability and complementarity is also briefly discussed in this paper.

PACS number(s): 03.65.Bz

I. INTRODUCTION

In a famous paper, Bell[1] demonstrated that no local-hidden-variable theory can reproduce all of the statistical predictions of quantum theory. In particular, he showed that highly correlated states of quantum systems have properties that are fundamentally nonclassical. Bell's argument is notable for its simplicity. Suppose that a quantum system consists of a pair of spin-$\frac{1}{2}$ particles. The pair of spins constitutes the simplest possible compound quantum system. The spins are prepared in a singlet state

$$|\psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\downarrow\rangle - |\downarrow\uparrow\rangle) .$$

Observables are taken to be compounds of the individual spins measured in units of $\hbar/2$, so that each observable can have the values $+1$ or -1. In the singlet state, the expectation values $\langle A\rangle = \langle B\rangle = 0$, so that the covariance

$$\text{cov}(A,B) = \langle (A - \langle A\rangle)(B - \langle B\rangle)\rangle = \langle AB\rangle .$$

For example, if A and B represent parallel components of the spin $\langle AB\rangle = -1$.

The predictions of quantum theory are probabilistic, but the nature of the indeterminacy in the theory is not immediately obvious. It might be supposed that the results of the spin measurements are determined by the values of unknown "hidden" variables. Further, it might seem reasonable to expect that these variables should function "locally"—that is, that the result of the spin measurement on one system is not influenced by the choice of spin observable measured on the other system. Under these innocent sounding assumptions, Bell demonstrated the following inequality:

$$|\langle AB\rangle - \langle AC\rangle| \le 1 + \langle BC\rangle .$$

This inequality can, in fact, be violated in the case of a singlet state, given a judicious choice of spin observables. In other words, quantum mechanics is incompatible with the idea that measurement results are due to the values of local-hidden variables.

The profoundly nonclassical character of the correlations in such quantum states has been termed "nonseparability".[2] Experiments to date give strong evidence that the Bell inequality is violated in nature exactly as predicted by quantum mechanics.[3] Thus, every local-hidden-variable theory can be excluded as the possible "machinery" behind quantum mechanics.

The remarkable scope of Bell's proof has led to intense discussion about the nature of quantum correlations. In this paper, I hope to further this discussion by shedding light on the structure of the Bell inequality. I will use the notion of "information distance" to derive an information-theoretic inequality that, like the Bell inequality, holds for local-hidden-variable theories but is violated in certain quantum-mechanical situations. This information-theoretic result is closely related to an inequality derived by Braunstein and Caves,[4] but has a geometrical meaning.

The geometrical interpretation of this new inequality will suggest a new understanding of Bell's result in terms of a "covariance distance." Finally, I will show that the nonclassical properties of quantum correlations are evident not only in a highly correlated state like the singlet $|\psi\rangle$, but in every pure state having any nonzero degree of correlation (that is, every state that is not a product of subsystem states). Quantum nonseparability is the rule, not the exception.

II. COVARIANCE BELL INEQUALITIES

In this section I will discuss a variant of the Bell inequality first proposed by Clauser, Horne, Shimony, and Holt.[5] Consider, as before, a system composed of two spins denoted S_1 and S_2. The observables are components of one spin or the other in units of $\hbar/2$. The system is prepared in a singlet state. On S_1 we perform a measurement either of observable A or B, and on S_2 we measure either C or D. The spectrum of each observable is the set $\{+1,-1\}$. The measurement results are assumed to be governed by some hidden variables that function locally. The values of these variables are statistically distributed in some fashion, and this distribution of values gives rise to the randomness in the measurement results.

Since the results of all possible measurements are supposed to be determined by hidden variables, it is meaningful to consider quantities constructed from the results of several different measurements. For instance, one can construct

$$M = A(C-D)+B(C+D) .$$

M is constructed so that $M=\pm 2$. It follows that

$$-2 \leq \langle M \rangle = \langle AC \rangle - \langle AD \rangle + \langle BC \rangle + \langle BD \rangle \leq +2 ,$$

where the expectations are averages over the distribution of the values of the hidden variables. If the assumption of locality holds, then each term in $\langle M \rangle$ can be experimentally determined. Each pair of observables is measured on an arbitrarily selected subensemble of an ensemble of similarly prepared systems. Since the choice of observable on S_1 does not affect the results of S_2 measurements (and vice versa) the observed subensemble statistics will probably represent accurately the statistics of the hidden variables. The terms in $\langle M \rangle$ individually involve observables on different subsystems; the simultaneous measurement of such pairs of observables affords no obstacle.

Now examine the predictions of quantum mechanics for the singlet state. Let θ be the angle between the directions along which the components A and C of the spins are measured. Then the probabilities of the four possible outcomes of the joint measurement of A and C are

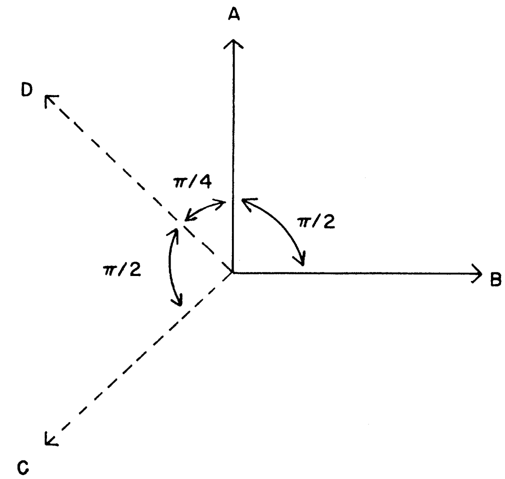

FIG. 2. Spin measurement axes yielding $\langle AC \rangle - \langle AD \rangle + \langle BC \rangle + \langle BD \rangle = 2\sqrt{2}$ for a singlet state.

$$p(++) = p(--) = \tfrac{1}{2}\sin^2\left[\frac{\theta}{2}\right] ,$$

$$p(+-) = p(-+) = \tfrac{1}{2}\cos^2\left[\frac{\theta}{2}\right] ,$$

where, for example, $p(+-)$ is the probability that the result of the A measurement is $+1$ and the C measurement is -1. The covariance $\langle AC \rangle = -\cos\theta$.

Consider the directions for A, B, C, and D shown in Fig. 1. When the spin observables for the two systems are chosen in this fashion, $\langle AC \rangle = \langle BC \rangle = \langle BD \rangle = -1/\sqrt{2}$ and $\langle AD \rangle = 1/\sqrt{2}$. Thus, $\langle M \rangle = -2\sqrt{2}$, which violates the covariance inequality derived above.

The meaning of this fact is fairly easy to see. Evidently the quantum-mechanical anticorrelation of "nearby" spin observables (i.e., those pairs for which θ is small) is stronger than the positive correlation of "distant" spin observables can permit in a hypothetical local-hidden-variable theory. The reverse is also true. Consider the spin observables indicated in Fig. 2. In this case, $\langle M \rangle = +2\sqrt{2} \not\leq 2$. Hence, the positive correlation of distant observables is greater than can be accounted for in a local-hidden-variable theory, given the strong anticorrelation of nearby observables.

Finally, it is useful to point out the role of the locality assumption. The covariance inequality itself follows from the hidden variable assumption alone. Locality is an auxiliary condition which ensures that the covariances in $\langle M \rangle$ (which are defined relative to the distribution of hidden variables) are the same as the statistical covariances measured by an experimenter. Some such auxiliary condition must be imposed if the inequality is to be testable.

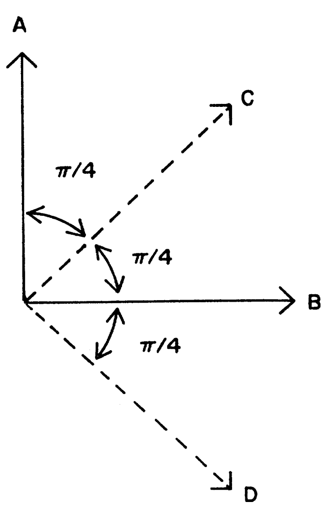

FIG. 1. Spin measurement axes yielding $\langle AC \rangle - \langle AD \rangle + \langle BC \rangle + \langle BD \rangle = -2\sqrt{2}$ for a singlet state.

III. INFORMATION DISTANCE

Information theory was developed by Shannon in 1948 as a mathematical description of communication.[6] Since that time, its concepts have found wider application, recently in various aspects of quantum theory.[7] The fundamental quantity of information theory is the *information*, or *entropy*, associated with a probability distribution,

$$H(X) = -\sum_i p(x_i)\ln p(x_i) \ .$$

(In this definition $0\ln 0 = 0$.) The information $H(X)$ satisfies the following properties.

(1) $H(X) \geq 0$, with equality if one outcome has probability unity.

(2) $H(X) \leq \ln N$, where N is the number of possible outcomes for X, with equality if each outcome has probabilities $1/N$.

(3) $H(X,Y) \geq H(X)$, where $H(X,Y)$ is the *joint* information for X and Y.

(4) $H(X,Y) \leq H(X)+H(Y)$, with equality if X and Y are independent.

$H(X)$ is a measure of the uncertainty associated with the probability distribution over X. It corresponds to the quantity of additional information required to specify the value of X, given the *a priori* probability distribution. Unlike other measures of uncertainty such as the covariance, $H(X)$ depends only on the probabilities of distinct events, not on their numerical values. Because it requires less mathematical "structure" for its definition and interpretation, the information is in some sense a more fundamental description of uncertainty.

Shannon's information can be used to define several information-theoretic quantities of interest. For example, the *conditional* information is $H(X|Y) = H(X,Y) - H(Y)$. This is a measure of the average uncertainty in X that remains after Y is known. If X and Y are not independent, then knowledge of Y will tend to reduce the uncertainty in X. This suggests a natural measure of the mutual dependence of X and Y, called the *correlation* information (or *mutual* information),

$$\begin{aligned} H(X{:}Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X,Y) \ . \end{aligned}$$

Two variables are correlated in an information-theoretic sense if a knowledge of one provides information about the other. Both correlation and anticorrelation (in a covariance sense) corresponds to a positive correlation information.

Zurek[8] has used information-theoretic quantities to define an *information distance* $\delta(X,Y)$ between two random variables X and Y,

$$\begin{aligned} \delta(X,Y) &= H(X|Y) + H(Y|X) \\ &= H(X,Y) - H(X{:}Y) \\ &= 2H(X,Y) - H(X) - H(Y) \ . \end{aligned}$$

The information distance is a symmetric measure of the lack of correlation between X and Y. Zurek's information distance satisfies the axioms for the abstract concept of a metric, provided that we note that it measures the distance between *equivalence classes* of variables. If X and Y are variables that provide exactly the same information—for example, if Y is a one-to-one function of X—then the information distance $\delta(X,Y)=0$.

With this proviso in mind, it is instructive to verify that δ is a metric. The properties of positive definiteness and symmetry follow directly from the definition. Only the triangle inequality remains to be proven. Suppose that X, Y, and Z are random variables with a joint probability distribution, and note that

$$\begin{aligned} &H(X,Y,Z) \geq H(X,Z) \ , \\ &2H(X,Y,Z) - H(X) - H(Z) \geq \delta(X,Z) \ . \end{aligned}$$

Now add the term $H(X,Y) - H(X,Y) + H(Y,Z) - H(Y,Z) = 0$ to the left-hand side of the inequality, obtaining

$$H(X|Y,Z) + H(Y|X) + H(Z|X,Y) + H(Y|Z) \geq \delta(X,Z) \ .$$

Since $H(X|Y,Z) \leq H(X|Y)$ and $H(Z|X,Y) \leq H(Z|Y)$,

$$\delta(X,Y) + \delta(Y,Z) \geq \delta(X,Z),$$

which is the desired inequality.

From the triangle inequality one can prove a whole host of additional metric relations, such as the following "quadrilateral inequality":

$$\delta(W,X) + \delta(X,Y) + \delta(Y,Z) \geq \delta(W,Z) \ .$$

In fact, for any set $\{A,B,C,\ldots,Y,Z\}$ of random variables, it is easy to prove the polygon inequality

$$\delta(A,B) + \delta(B,C) + \cdots + \delta(Y,Z) \geq \delta(A,Z) \ .$$

IV. INFORMATION-THEORETIC BELL INEQUALITIES

Consider once again the pair of spins in a singlet state, with spin observables A and B defined on S_1 and C and D defined on S_2. If the observed probabilities of the various measurement outcomes were in fact due to some statistical distribution of deterministic hidden variables, then a joint distribution would exist over all measurement results for all four observables. According to the quadrilateral inequality mentioned above,

$$\delta(A,C) + \delta(B,C) + \delta(B,D) \geq \delta(A,D) \ .$$

If the hidden variables are local in character, then the various probabilities involved in the above inequality can be determined by observing the statistics of actual pairwise measurements of the observables. Thus, in a local-hidden-variable theory this inequality is subject to experimental test.

This inequality, like the Bell inequality, is violated in quantum mechanics. Consider the observables A and C, which represent components of spin along axes separated by an angle θ. Then the information distance $\delta(A,C)$ is just

$$\delta(A,C) = 2f\left[\frac{\theta}{2}\right] ,$$

where $f(\phi)=-\cos^2\phi \ln\cos^2\phi-\sin^2\phi \ln\sin^2\phi$. Choose the four observables as indicated by the four axes in Fig. 3. Then

$$\delta(A,C)=\delta(B,C)=\delta(B,D)=2f\left[\frac{\pi}{16}\right]=0.323\ ,$$

$$\delta(A,D)=2f\left[\frac{3\pi}{16}\right]=1.236\ .$$

Since $0.323+0.323+0.323<1.236$, the information-theoretic Bell inequality is violated for this combination of measurements on the singlet state.

In other words, the information distance defined by Zurek is not a metric at all in the quantum-mechanical case. A diagram of the situation is given in Fig. 4. A and D are "further apart" (in information distance) than the information quadrilateral inequality permits, given the "closeness" of A to C, C to B, and B to D. Since information distance measures the lack of correlation between observables, we can say that A and D are less correlated than would be possible in a local-hidden-variable theory, or, conversely, that the nearby pairs of observables are more correlated than would be possible.

In a similar way, Braunstein and Caves[4] derived an inequality based on the conditional information

$$H(A|B)+H(B|C)+H(C|D)\geq H(A|D)\ .$$

Since the information distance is a "symmetrized" form of the conditional information, it is clear that this inequality is closely related to the ones described in this paper. In fact, this conditional information inequality implies the information-distance quadrilateral inequality; thus, the quantum violation of the quadrilateral inequality discussed here implies the quantum violation of the conditional information inequality discussed by Braunstein and Caves.

Why does quantum mechanics violate information-theoretic Bell inequalities? The key assumption in the derivation of the triangle inequality for information distance was the existence of joint probability distributions over *triples* of random events. Quantum theory does not provide such distributions in the situation described here,

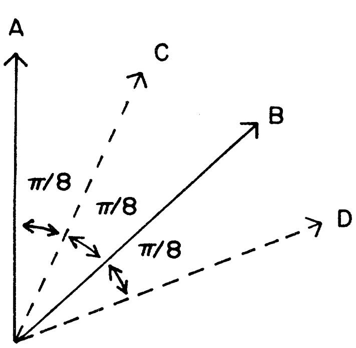

FIG. 3. Spin measurement axes yielding a violation of the information distance quadrilateral inequality for a singlet state.

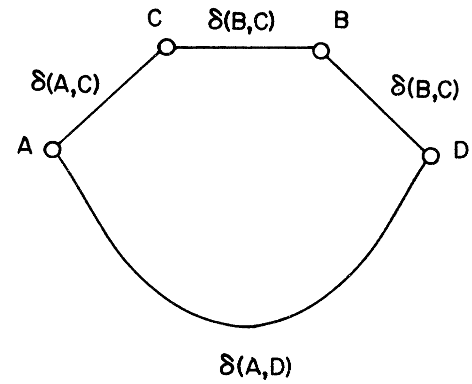

FIG. 4. Schematic representation of quantum nonseparability. The information distance $\delta(A,D)$ is greater than is allowed by the (classical) metric properties of information distance.

since this would involve joint probability distributions for incompatible observables (A and B on S_1, for example). The violation of this inequality is an expression of complementarity.

Further information-theoretic Bell inequalities can be constructed as metric inequalities for information distance. For instance, if parallel components of spin are measured on S_1 and S_2, the information distance between the observables will be zero. Thus, a degenerate quadrilateral inequality can be written $\delta(A,B_2)+\delta(B_1,C)\geq\delta(A,C)$, where B_1 and B_2 are parallel spin observables on the two subsystems. We can also construct experimentally testable polygon inequalities for larger sets of observables. The failure of the triangle inequality due to complementarity will lead to the possibility of violating these further inequalities in quantum mechanics.

V. COVARIANCE DISTANCE

These considerations give a simple geometric interpretation to quantum nonseparability as the failure of the information distance triangle inequality due to complementarity. I will now show that it is possible to develop a similiar geometric understanding of the familiar covariance Bell inequalities. Let X and Y be two random variables that take on the values $+1$ and -1. Define the *covariance distance* $\Delta(X,Y)=1-\langle XY\rangle$. (Strictly speaking, $\langle XY\rangle$ is the covariance of X and Y only if one of the variables has an expectation value of zero. However, the terminology seems natural.) The covariance distance is clearly a symmetric function of X and Y. It is also positive-definite,

$$\begin{aligned}\Delta(X,Y)&=1-\langle XY\rangle\\&=1-p(++)-p(--)+p(+-)+p(-+)\\&=2p(+-)+2p(-+)\\&\geq 0\ .\end{aligned}$$

[Here $p(+-)$ is the probability that $X=+1$ and $Y=-1$, etc.]

As before, proof of the triangle inequality follows from considering three random variables X, Y, and Z, each variable taking on values of ± 1, with a joint probability distribution $p(xyz)$. Then

$$\Delta(X,Y)=2[p(+-+)+p(+--)+p(-++)+p(-+-)] ,$$

$$\Delta(Y,Z)=2[p(++-)+p(-+-)+p(+-+)+p(--+)] ,$$

$$\Delta(X,Z)=2[p(++-)+p(+--)+p(-++)+p(--+)] .$$

The triangle inequality can now be checked directly,

$$\begin{aligned}\Delta(X,Y)+\Delta(Y,Z)&=2[p(++-)+p(+--)+p(-++)+p(--+)]\\&\quad+4[p(+-+)+p(-+-)]\\&\geq \Delta(X,Z) .\end{aligned}$$

From the triangle inequality one can derive all of the polygon inequalities. Among these is the quadilateral inequality

$$\Delta(A,C)+\Delta(B,C)+\Delta(B,D)\geq \Delta(A,D) ,$$

which reduces to the form

$$\langle AC\rangle-\langle AD\rangle+\langle BC\rangle+\langle BD\rangle\leq +2 .$$

This can be recognized as one of the covariance Bell inequalities discussed above, which can be violated in quantum mechanics. All of the usual Bell inequalities can be cast in the form of metric space inequalities using covariance distance.

VI. BEYOND THE SINGLET STATE

In elementary discussions like this one, quantum violations of Bell inequalities are usually presented using the singlet state of a pair of spins. Nonseparability, however, is not a property peculiar to the singlet state. The singlet state represents a "maximally correlated" state, so that the nonclassical character of quantum correlations is particularly obvious; but such states comprise a set of measure zero among the set of all possible states of the spins. If Bell inequalities were only violated in maximally correlated states, the violation would be of little real significance, since the slightest variation from maximal correlation (which any actually achievable state must have) would hide it. Nonseparability, to be observable, must be "robust."

Of course, this is, in fact, the case. Since the quantum violations of the Bell inequality are fairly large, small departures from maximal correlation do not bring the correlations into the classical domain. Clauser, Horne, Shimony, and Holt,[5] for example, considered realizable experiments involving the imperfect correlation that are observable between photons produced in an atomic cascade. Because of the complications of this realistic situation (e.g., the recoil of the atom), the photon pair might not even be in a pure state.

In this paper, a different situation will be considered. We suppose that the pair of spins is prepared in a pure state other than the singlet state. Small departures from the exact singlet state would produce correspondingly small differences in the covariance and information distances between the various observables. Covariance and information-theoretic Bell inequalities should therefore be violated for pure quantum states which lie close to, but are not exactly equal to, maximally correlated states such as the singlet.

How close, then, is close enough? It is clear that not every quantum state of a pair of spins will show nonseparability. For example, consider a product state $|\psi\rangle=|\psi_1\rangle\otimes|\psi_2\rangle$. Since no correlation at all will be present between the results of S_1 measurements and S_2 measurements, it will not be possible to construct a violation of any Bell inequality. How strongly must the two spins be correlated before it is possible?

Consider a state of the form

$$|\psi(\alpha)\rangle=\cos\alpha|a_1a_2\rangle-\sin\alpha|b_1b_2\rangle ,$$

where $|a_1\rangle$ and $|b_1\rangle$ form a set of basis states for S_1, as do $|a_2\rangle$ and $|b_2\rangle$ for S_2. The parameter α determines the degree of correlation of the state $|\psi(\alpha)\rangle$: For $\alpha=0$ the state is a product state, and for $\alpha=\pi/4$ the state is maximally correlated. It turns out that any state of the pair of spins can be written in this way with a suitable choice of basis states with $0\leq\alpha\leq\pi/4$ (Ref. 9); the singlet state, for example, corresponds to $|a_1\rangle=|b_2\rangle=|\uparrow\rangle$, $|a_2\rangle=|b_1\rangle=|\downarrow\rangle$, and $\alpha=\pi/4$.

Observables for each spin are taken to have eigenvalues ± 1, as usual, and are specified by writing the eigenstates in the $|a\rangle,|b\rangle$ basis. It will suffice for our purposes to consider only real linear combinations of $|a\rangle$ and $|b\rangle$,

$$|+\rangle=\cos\theta|a\rangle+\sin\theta|b\rangle ,$$

$$|-\rangle=-\sin\theta|a\rangle+\cos\theta|b\rangle .$$

θ is the angle in Hilbert space between the correlation basis $\{|a\rangle,|b\rangle\}$ and the observable basis $\{|+\rangle,|-\rangle\}$. If we fix θ_1 and θ_2 for the observables on S_1 and S_2, we find that

$$\langle +_1+_2|\psi(\alpha)\rangle=\cos\alpha\,\cos\theta_1\cos\theta_2+\sin\alpha\,\sin\theta_1\sin\theta_2 ,$$

$$\langle +_1-_2|\psi(\alpha)\rangle=-\cos\alpha\,\cos\theta_1\sin\theta_2+\sin\alpha\,\sin\theta_1\cos\theta_2 ,$$

$$\langle -_1+_2|\psi(\alpha)\rangle=-\cos\alpha\,\sin\theta_1\cos\theta_2+\sin\alpha\,\cos\theta_1\sin\theta_2 ,$$

$$\langle -_1-_2|\psi(\alpha)\rangle=\cos\alpha\,\sin\theta_1\sin\theta_2+\sin\alpha\,\cos\theta_1\cos\theta_2 .$$

The probabilities are $p(+_1+_2)=|\langle +_1+_2|\psi(\alpha)\rangle|^2$, etc.

Fix some angle θ and define observables A and B on S_1 and C and D on S_2 by the angles

$$\theta_A=-\theta_D=\theta ,$$

$$\theta_B=\theta_C=0 .$$

The structure of $|\psi(\alpha)\rangle$ defines a "joint correlation axis" (really a pair of axes, one for each spin). Observables B

and C lie along this joint axis, while A and D are tilted by a spatial angle $\theta/2$ (and thus a Hilbert-space angle θ) on opposite sides of the joint axis. The covariance distances can now be calculated,

$$\Delta(B,C)=0 ,$$

$$\Delta(A,C)=\Delta(B,D)=2\sin^2\theta ,$$

$$\Delta(A,D)=4(\cos\alpha+\sin\alpha)^2\cos^2\theta\sin^2\theta .$$

The quadrilateral inequality for covariance distance, which is a covariance Bell inequality, states that

$$\Delta(A,C)+\Delta(B,C)+\Delta(B,D)\geq\Delta(A,D) ,$$

$$4\sin^2\theta\geq 4(1+2\cos\alpha\sin\alpha)\cos^2\theta\sin^2\theta ,$$

$$1\geq(1+\sin2\alpha)\cos^2\theta .$$

(θ is assumed to be nonzero.) For a given correlation parameter α in the range $(0,\pi/4]$, it is always possible to find some small nonzero value for θ which will violate this inequality. Thus, it is possible to find sets of observables that violate a covariance Bell inequality no matter how small α is, provided that it is not zero.

When α is near to zero, the range of values of θ which give Bell inequality violations is quite small. Nevertheless, the range is finite for any $\alpha>0$. Numerical calculations also show that it is also possible to find sets of observables which violate information-theoretic Bell inequalities even if the correlation parameter α is exceedingly close to zero.

VII. REMARKS

In this paper, I have aimed to clarify the Bell inequalities in several ways, with the hope of better understanding the significance of their failure in quantum mechanics. The essential form of these inequalities is geometrical, as polygon inequalities in a metric space of random variables. The distance function can be either covariance distance or information distance, leading to inequalities of various kinds. In each case, quantum-mechanical situations can be constructed that contradict the inequality.

The proofs of the polygon inequalities in a metric space rest on the triangle inequality. For covariance distance and information distance, the proof of the triangle inequality rests in turn upon the existence of joint probability distributions. In quantum mechanics, complementarity precludes the existence of joint distributions over incompatible observables. Complementarity is thus the source of the violation of the Bell inequalities and of the general quantum property of nonseparability.

This property is not a strange, special property of only a few states like the singlet state. Instead, nonseparability can be seen, through the violation of Bell inequalities, in almost all quantum states. Only product states, which have no correlations at all, are "non-nonseparable." Just as complementarity is a very general situation, describing almost any pair of observables, so also nonseparability is a very general property, true for almost any state of a compound system.

ACKNOWLEDGMENTS

I would like to acknowledge a number of helpful discussions with F. M Kronz and the continuing encouragement of Professor John A. Wheeler. I also wish to thank W. K. Wootters and W. H. Zurek for drawing my attention to important earlier work in this subject.

[1]J. S. Bell, Physics (N.Y.) **1**, 195 (1964).
[2]B. d'Espagnat, *Conceptual Foundations of Quantum Mechanics* (Addison-Wesley, Reading, MA, 1971).
[3]A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **47**, 460 (1981).
[4]S. L. Braunstein and C. M. Caves, Phys. Rev. Lett. **61**, 662 (1988).
[5]J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. **23**, 880 (1969).
[6]C. E. Shannon, Bell Syst. Tech. J. **27**, 379 (1948).
[7]For example, see W. K. Wootters, Ph.D. thesis, University of Texas, Austin, 1980.
[8]W. H. Zurek, Nature **341**, 119 (1989).
[9]E. Schmidt, Math. Ann. **63**, 433 (1906).