PHYSICAL REVIEW A VOLUME 47, NUMBER 4 APRIL 1993

Noise in Mermin's n-particle Bell inequality

Samuel L. Braunstein and A. Mann
Department of Physics, Technion–Israel Institute of Technology, 32 000 Haifa, Israel
(Received 1 December 1992)

We calculate the signal-to-noise ratio in Mermin's n-particle Bell inequality [Phys. Rev. Lett. **65**, 1838 (1990)] based on source noise, data error, and statistical uncertainty. We find that the signal for violation grows exponentially faster than the exponentially growing noise, as long as the noise *per* detector or *per* particle is less than about 14%. Alternately, in the absence of noise a loophole-free test could be made with detector efficiencies as low as 71%.

PACS number(s): 03.65.Bz

Greenberger, Horne, and Zeilinger (GHZ) [1] recently came up with an all-or-nothing argument against local realism (LR) when compared with the ideal outcome of a (theoretically) simple quantum model with three spin-$\frac{1}{2}$ particles. Mermin [2] generalized their argument to n particles and placed it in the form of an inequality so that slight amounts of noise would not invalidate the reasoning. This step was important as it brought the GHZ argument within the realm of experimental testability [3,4]. But how well does the resulting inequality deal with noise? We calculate the signal-to-noise ratio for seeing a violation of this inequality in the presence of three kinds of noise: (i) noise from variations in the quantum state used as source for the correlations; (ii) noise from data-flipping error—the readout randomly flipping the sign of the spin polarization; and (iii) noise from counting statistics due to measuring only a finite number of n-particle coincidences.

Mermin [2] considered n spin-$\frac{1}{2}$ particles entering n spatially separated Stern-Gerlach detectors each oriented along either the x or y axes; detector j ($j = 1,\ldots,n$) measures the value of the spin projection as either $\sigma_x^j = \pm 1$ or $\sigma_y^j = \pm 1$. The world view of LR requires that each of these spin projections have a definite value independent of our having to measure them. Based on reasoning similar to Bell's [5] Mermin showed that if LR were true the *statistical* inequality,

$$-\beta_n \leq \begin{array}{l} \overline{\sigma_y^1\sigma_x^2\cdots\sigma_x^n} + \cdots \\ -\overline{\sigma_y^1\sigma_y^2\sigma_y^3\sigma_x^4\cdots\sigma_x^n} - \cdots \\ +\overline{\sigma_y^1\cdots\sigma_y^5\sigma_x^6\cdots\sigma_x^n} + \cdots \\ -\overline{\sigma_y^1\cdots\sigma_y^7\sigma_x^8\cdots\sigma_x^n} - \cdots \\ +\cdots \end{array} \leq \beta_n \tag{1}$$

would hold, where each line in the center of this inequality contains all distinct permutations of the subscripts that give distinct products, and

$$\beta_n \equiv \begin{cases} 2^{(n-1)/2}\,, & n \text{ odd} \\ 2^{n/2}\,, & n \text{ even} \end{cases}$$

is the LR bound. This inequality involves 2^{n-1} correlation functions, since only configurations with an odd number of detectors oriented along the y axis contribute; an assumption-free test would require that the correlation of each of these configurations be measured.

We start by characterizing the source noise. Mermin's Bell inequality (1) may be written in quantum-mechanical language in the shorthand form

$$-\beta_n \leq \langle \hat{\mathcal{B}} \rangle \leq \beta_n\,,$$

where

$$\hat{\mathcal{B}} \equiv \frac{1}{2i}\left(\prod_{j=1}^{n}\left(\hat\sigma_x^j + i\hat\sigma_y^j\right) - \prod_{j=1}^{n}\left(\hat\sigma_x^j - i\hat\sigma_y^j\right)\right)$$

is the so-called Bell operator [6] associated with the expression in the middle of inequality (1), and $\hat\sigma_x^j$ and $\hat\sigma_y^j$ are two of the Pauli spin matrices associated with the jth particle. We will call

$$\mathcal{S} \equiv \langle \hat{\mathcal{B}} \rangle - \beta_n \tag{2}$$

the *signal* for a violation of the right-hand inequality in (1) (the left-hand inequality may be treated similarly). The size of the violation of this inequality can be completely characterized by the eigenvalues and eigenvectors of the operator $\hat{\mathcal{B}}$ [6]. This particular operator has two nondegenerate eigenvectors

$$|\Psi^{(\pm)}\rangle = \frac{1}{\sqrt{2}}\Big(|\uparrow\cdots\uparrow\rangle \pm i\,|\downarrow\cdots\downarrow\rangle\Big)\,,$$

with eigenvalues $\pm 2^{n-1}$, respectively; all other eigenvectors $|\Psi_j^{(0)}\rangle$ ($j = 1,\ldots,2^n-2$) are degenerate with eigenvalue zero. Thus, either of the states $|\Psi^{(\pm)}\rangle$ will yield a maximal violation that grows exponentially faster than the locally realistic bound; however, as we are concentrating here on the right-hand inequality of (1) we will assume that our pre-noise source of states generates $|\Psi^{(+)}\rangle$.

Source noise will mix with $|\Psi^{(+)}\rangle$ either the negative eigenvalue state or a zero eigenvalue state, thus reducing the size of the violation. Since it is of no importance which zero eigenvalue state is mixed in, it is sufficient to consider a random admixture of them,

$$\sum_{j=1}^{2^n-2} |\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}| = \hat{\mathbb{1}} - |\Psi^{(-)}\rangle\langle\Psi^{(-)}| - |\Psi^{(+)}\rangle\langle\Psi^{(+)}|\,,$$

where $\hat{1}$ is the identity operator in 2^n dimensions. In this way we may without loss of generality study source states of the form

$$\begin{aligned}\hat{\rho} &= [1-\kappa_1-(2^n-2)\,\kappa_2]\,|\Psi^{(+)}\rangle\langle\Psi^{(+)}| \\ &\quad +\kappa_1\,|\Psi^{(-)}\rangle\langle\Psi^{(-)}| + \kappa_2\sum|\Psi_j^{(0)}\rangle\langle\Psi_j^{(0)}| \\ &= (1-\delta_1-\delta_2)\,|\Psi^{(+)}\rangle\langle\Psi^{(+)}| \\ &\quad +\delta_1\,|\Psi^{(-)}\rangle\langle\Psi^{(-)}| + 2^{-n}\delta_2\,\hat{1}\;,\end{aligned}$$

where we have made two choices of parametrization. The second parametrization $\hat{\rho}(\delta_1,\delta_2)$ for our source of states admits a simple interpretation with δ_1 the probability of "phase mixing" (i.e., loss of phase sensitivity in generating $|\Psi^{(+)}\rangle$) and δ_2 the probability for "random state mixing" (i.e., mixing with an arbitrary quantum state).

From the above eigenvalues we see that the expectation of the Bell operator for this noisy source will be

$$2^{n-1}\,(1-2\delta_1-\delta_2)\;. \tag{3}$$

Thus, the violation still grows exponentially with increasing n (however, we have not included any *cost* to keeping the state pure to within a fixed probability as n increases; no doubt at least δ_1 should grow initially exponentially with n).

Now consider the effect of "data-flipping error" [7]. Ideal detectors record ±1 in correspondence with the spin polarization of the incident particle; however, for an imperfect detector there may be an error that randomly flips the sign of the output. Focus attention on the effect of this error in just one detector. If q is the probability for this flipping to occur, changing the sign, then there is also a probability $1-q$ for the product of the detector outputs to remain unflipped. In this way the correlation will be reduced on average by the factor $1-2q$. Further, since the data-flipping error at every detector is statistically independent if all detectors suffer from the same magnitude q of this error, the correlation will be reduced on average by $(1-2q)^n$.

In this way we can see that the expectation of the Bell operator, for a source characterized by $\hat{\rho}(\delta_1,\delta_2)$ and in the presence of a data flipping probability q at each detector, will be given by [cf. Eq. (3)]

$$\langle\hat{\mathcal{B}}\rangle = 2^{n-1}(1-2q)^n\,(1-2\delta_1-\delta_2)\;, \tag{4}$$

since this expectation can be written as a sum of correlation functions, each of which is reduced by the same factor.

In the absence of source noise the signal for violation $\mathcal{S}$ will be positive (denoting violation) when

$$q \le \begin{cases}\left(1-2^{1/2n}/\sqrt{2}\right)/2\,, & n\text{ odd}\\ \left(1-2^{1/n}/\sqrt{2}\right)/2\,, & n\text{ even}\end{cases} \tag{5}$$
$$\xrightarrow[n\to\infty]{} 14.64\%\;.$$

Figure 1 shows the largest value of the data-flipping error allowable from Eq. (5) versus n, the number of detectors used; this behavior is seen to be almost independent of n for $n \gtrsim 5$, so that the violation of inequality (1) is quite

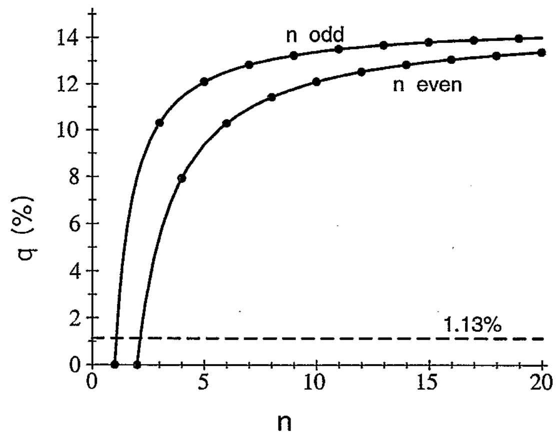

FIG. 1. Plot of maximum data-flipping error q (in %) from Eq. (5), vs n, which still violates inequality (1) in the absence of source noise and counting statistics; the curves for odd and even n are shown separately. The horizontal dashed line is the average data-flipping error $q \simeq 1.13\%$ which Aspect, Grangier, and Roger [8] achieved experimentally.

robust. The horizontal dashed line at a data-flipping error of $q \simeq 1.13\%$ was achieved by Aspect, Grangier, and Roger [8] in their test of Bell inequalities from the early 1980s.

Before calculating the signal-to-noise ratio for a violation we shall consider the expected statistical uncertainty that can be attained for a *single* correlation function given only a finite number M of successful coincidences (which we call "counts"). In determining the correlation between the detectors only the *product* of the detector outputs (each ±1) is relevant—we call this product the "apparatus output;" if m of the M counts yield a positive apparatus output, then we intuitively estimate the correlation between the detectors to be

$$(2m-M)/M\;. \tag{6}$$

We want, however, to place a value on our uncertainty in this estimate, and to do so we must base this estimate on some statistical criterion; here we will use the criterion of maximum likelihood to estimate the most likely probability $p_{\rm est}$ for seeing a positive apparatus output in a single count—from this we can estimate the correlation via $C_{\rm est} = 2p_{\rm est} - 1$.

Bayes's theorem tells us that the probability p for seeing a positive apparatus output given that we have already measured m such positive counts out of M is distributed according to

$$\text{Prob}(p\,|\,m,M) \propto p^m(1-p)^{M-m}\;, \tag{7}$$

if we have no inital prejudice for the value of p. The most likely value for p (the maximum likelihood estimate inferred from our data) is found by elementary calculus to be $p_{\rm est} = m/M$, and so $C_{\rm est} = 2m/M - 1$, which agrees with Eq. (6).

The advantage we have gained over the intuitive approach, by the use of this formalism, is to be able to determine the experimental uncertainty for the correlation of Eq. (6) based solely on the data. In particular, the noise figure we shall quote for this uncertainty will be based on the standard deviation [9] of the distribution of Eq. (7), which was inferred from the measurements. Using the integral identity [10]

$$\int_0^1 dt\, t^k(1-t)^\ell = \left[(k+\ell+1)\binom{k+\ell}{k}\right]^{-1},$$

we calculate the variance of p over Eq. (7) as

$$(\Delta p)^2 = \frac{(M-m+1)(m+1)}{(M+2)^2(M+3)};$$

the experimentally quoted uncertainty for the observed correlation of Eq. (6) would then be $2\Delta p$.

We are interested here in the typical experimental results, so we calculate the average values for the observed correlation and its uncertainty. With an overall probability p_0 for obtaining a positive apparatus output, the distribution of outcomes will be

$$\mathrm{Prob}(m\,|\,p_0, M) = \binom{M}{m} p_0^m (1-p_0)^{M-m}.$$

Thus, the average experimental *estimate* of the correlation is just $\overline{C_{\rm est}} = 2p_0 - 1$, and the average experimental estimate of the variance is

$$(\Delta C_{\rm est})^2 \equiv 4\overline{(\Delta p)^2} = \frac{M(M-1)\left(1-\overline{C_{\rm est}}^2\right)+4(M+1)}{(M+2)^2(M+3)}. \quad (8)$$

We note that for $M \gtrsim 1/p_0(1-p_0)$ this expected uncertainty in the measured correlation simplifies to

$$(\Delta C_{\rm est})^2 \simeq \frac{1}{M}\left(1-\overline{C_{\rm est}}^2\right). \quad (9)$$

This variance agrees with Fisher's theorem [11] for the asymptotic behavior of maximum likelihood estimation. Further, this limit agrees with the Braunstein and Caves [12] calculation of the variance of the estimate for the correlation; by contrast, our result in Eq. (8) is a calculation of the average estimate of the variance of the correlation, and gives a better prediction of achievable experimental sensitivities (it includes the extra statistical noise for *estimating* the variance from the data).

Alternately, for $M \lesssim 1/p_0(1-p_0)$ the variance is independent of p_0 with

$$(\Delta C_{\rm est})^2 \simeq \frac{4(M+1)}{(M+2)^2(M+3)}.$$

The reason for this independence is clear: suppose $1-p_0 \ll 1$, then for fewer than $M \simeq 1/(1-p_0)$ counts almost every apparatus output will be positive; it is only when there is some non-negligible chance to see a negative apparatus output that we can begin to distinguish the size of p_0. A similar argument holds for $p_0 \ll 1$.

Figure 2 shows a log-log plot of the noise $\Delta C_{\rm est}$ [solid line, from Eq. (8)] versus sample size M with $p_0 = 0.99$,

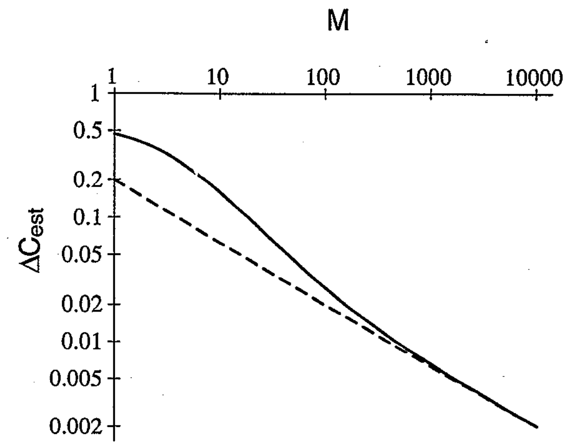

FIG. 2. Log-log plot showing the convergence of the statistical noise $\Delta C_{\rm est}$ [solid line, from Eq. (8)] with increasing number of counts M for $p_0 = 0.99$; the ordinary $1/\sqrt{M}$ behavior of counting statistics [dashed line, from Eq. (9)] is attained only for $M \gtrsim 1/p_0(1-p_0) \simeq 100$.

and the asymptotic result [dashed line, from Eq. (9)] of ordinary counting statistics; we are able to see a "knee" around $M \simeq 1/p_0(1-p_0) \simeq 100$ corresponding to where the asymptotic $1/\sqrt{M}$ behavior usually associated with counting statistics turns on. Indeed, we see that for a small number of counts the asymptotic "counting statistics" result of Eq. (9) leads to too ambitious a figure for the uncertainty achievable.

Finally, combining the above results for source noise, data-flipping error, and statistical uncertainty, we calculate the signal-to-noise ratio for violation of Mermin's inequality (1). The signal $\mathcal{S}$ has already been calculated in Eq. (2) with $\langle\hat{\mathcal{B}}\rangle$ given by Eq. (4). The total variance $\mathcal{N}^2$ for this signal will be simply the sum of the variances for each of the 2^{n-1} correlation functions that need to be measured; if we assume that we divide our resources so as to accumulate the same number M of counts for each of the detector configurations (requiring $N = 2^{n-1}M$ counts in all), then the statistical noise will be $2^{(n-1)/2}\Delta C_{\rm est}$, with $\Delta C_{\rm est}$ given by Eq. (8), and by consistency with the value of $C_{\rm est}$

$$p_0 = \tfrac{1}{2}\left[1+(1-2q)^n(1-2\delta_1-\delta_2)\right].$$

Let us concentrate on the most extreme case where only a single count $M = 1$ is used for each correlation function. In this case $\Delta C_{\rm est} = \sqrt{2}/3$ and the signal-to-noise ratio is given by

$$R \equiv \frac{\mathcal{S}}{\mathcal{N}} \simeq \frac{3}{2^{n/2}} \frac{2^{n-1}(1-2q)^n(1-2\delta_1-\delta_2)-\beta_n}{\sqrt{1-(1-2q)^{2n}(1-2\delta_1-\delta_2)^2}}.$$

Thus, if we neglect source noise, to be able to see a signal-to-noise ratio of at least R we need a data-flipping error of less than

$$q \le \begin{cases} \frac{1}{2}\left[1-\frac{1}{\sqrt{2}}\left(\frac{2}{3}R+\sqrt{2}\right)^{1/n}\right], & n \text{ odd} \\ \frac{1}{2}\left[1-\frac{1}{\sqrt{2}}\left(\frac{2}{3}R+2\right)^{1/n}\right], & n \text{ even} \end{cases} \quad (10)$$

$$\xrightarrow[n\to\infty]{} 14.64\%,$$

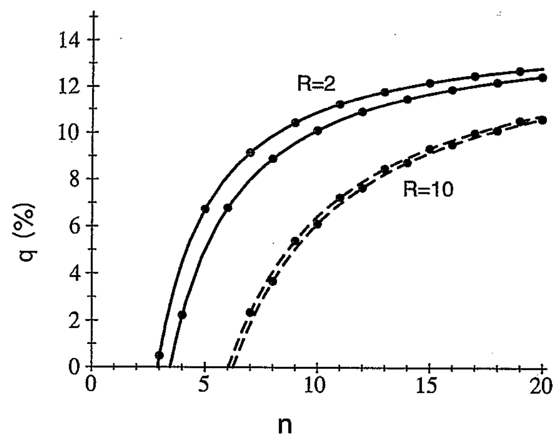

FIG. 3. Plot of the maximum data-flipping error q (in %) from Eq. (10), vs n, which violates inequality (1) in the absence of source noise for a signal-to-noise ratio $R = 2$ (solid lines) and $R = 10$ (dashed lines), based on a *single* count $M = 1$ per correlation function; in each case the curves for odd and even n are drawn separately.

in place of the bound in inequality (5). Figure 3 shows the maximum data-flipping error (in the absence of source noise) that can yield a signal-to-noise error of $R = 2$ or $R = 10$, given that only one count $M = 1$ is used for each of the correlation functions.

Up till now we have ignored issues relating to detector inefficiency, which may allow loopholes in our test. The problem may be transcribed to one involving n spin-1 particles where all particles are detected [13], by assigning 0 to the detector outputs that fail to detect a spin-$\frac{1}{2}$ particle. In this way the same Bell inequality (1) applies and the correlation functions are simply reduced by the product of detector efficiencies η^n (assuming all efficiencies equal), so the expectation of the Bell operator becomes

$$\langle \hat{\mathcal{B}} \rangle = 2^{n-1}\eta^n(1-2q)^n\,(1-2\delta_1-\delta_2) \ .$$

In the absence of noise a positive signal $\mathcal{S}$ for violation of inequality (1) will occur for detector efficiencies satisfying

$$\eta \geq \begin{cases} 2^{1/2n}/\sqrt{2}\,, & n \text{ odd} \\ 2^{1/n}/\sqrt{2}\,, & n \text{ even} \end{cases}$$
$$\underset{n\to\infty}{\longrightarrow} 70.71\% \ .$$

Except for $n = 4$ this is the least restrictive loophole-free condition on detector efficiencies for any test of local realism, the closest being $\eta \geq 82.84\%$ [13], which also requires extra symmetries be checked.

We have calculated the signal-to-noise ratio in Mermin's n-particle Bell inequality, including source noise, data error, and statistical uncertainty. We found that the signal for violation grows exponentially faster than the exponentially growing noise, as long as the noise *per* detector or *per* particle is less than about 14%. Finally, we found that this Bell inequality allows a weakened condition on the detector efficiency needed to perform a loophole-free test of local realism; this condition is $\eta \gtrsim 71\%$.

This work was supported in part by the Fund for the Promotion of Research at the Technion, the Technion VPR Fund, and the Lady Davis Fellowship Trust.

[1] D. M. Greenberger, M. A. Horne, and A. Zeilinger, in *Bell's Theorem, Quantum Theory and Conceptions of the Universe*, edited by M. Kafatos (Kluwer, Dordrecht, 1989), p. 73.
[2] N. D. Mermin, Phys. Rev. Lett. **65**, 1838 (1990).
[3] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, Am. J. Phys. **58**, 1131 (1990).
[4] B. Yurke and D. Stoler, Phys. Rev. Lett. **68**, 1251 (1992).
[5] J. S. Bell, Physics (N.Y.) **1**, 195 (1964).
[6] S. L. Braunstein, A. Mann, and M. Revzen, Phys. Rev. Lett. **68**, 3259 (1992).
[7] N. D. Mermin and G. M. Schwarz, Found. Phys. **12**, 101 (1982).
[8] A. Aspect, P. Grangier, and G. Roger, Phys. Rev. Lett. **49**, 91 (1982).
[9] A standard deviation carries more information about the tail of a distribution than a 68% confidence interval, and so is superior here for characterizing noise.
[10] *Table of Integrals, Series, and Products*, edited by I. S. Gradshteyn and I. M. Ryzhik, (Academic, London, 1980), p. 950.
[11] H. Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, NJ, 1946), pp. 500–504.
[12] S. L. Braunstein and C. M. Caves, Ann. Phys. **202**, 22 (1990).
[13] A. Garg and N. D. Mermin, Phys. Rev. D **35**, 3831 (1987).