A Loophole-Free Bell’s Inequality Experiment

Paul G. Kwiat, Aephraim M. Steinberg, and Raymond Y. Chiao

Department of Physics, U.C. Berkeley
Berkeley, CA 94720

Philippe H. Eberhard
Lawrence Berkeley Laboratory, U.C. Berkeley
Berkeley, CA 94720

Abstract

The proof of Nature’s nonlocality through Bell-type experiments is a topic of longstanding interest. Nevertheless, no experiments performed thus far have avoided the so-called “detection loophole,” arising from low detector efficiencies and angular-correlation difficulties. In fact, most, if not all, of the systems employed to date can never close this loophole, even with perfect detectors. In addition, another loophole involving the non-rapid, non-random switching of various parameter settings exists in all past experiments. We discuss a proposal for a potentially loophole-free Bell’s inequality experiment. The source of the EPR-correlated pairs consists of two simultaneously-pumped type-II phase-matched nonlinear crystals and a polarizing beam splitter. The feasibility of such a scheme with current detector technology seems high, and will be discussed. We also present a single-crystal version, motivated by other work presented at this conference.

In a separate experiment, we have measured the absolute detection efficiency and time response of four single-photon detectors. The highest observed efficiencies were 70.7 ± 1.9% (at 633 nm, with a device from Rockwell International) and 76.4 ± 2.3% (at 702 nm, with an EG&G counting module). Possible efficiencies as high as 90% were implied. The EG&G devices displayed sub-nanosecond time resolution.

1 Introduction

It is now well known that quantum mechanics (QM) yields predictions which are inconsistent with the seemingly innocuous concepts of locality and reality. This was first shown by Bell in 1964 [1, 2] for the case of two quantum-mechanically entangled particles, e.g., particles in a singlet-like state, which do not possess definite polarizations even though they are always orthogonally polarized. As implied by Einstein, Podolsky, and Rosen (EPR) [3], it is straightforward to construct local realistic models that explain certain features predicted by QM (e.g., the total anti-correlation between detectors measuring the same polarization component of the two particles). The quantum mechanical contradiction with local realism becomes apparent only by considering situations of non-perfect correlations (i.e., measuring the polarization components at intermediate, non-orthogonal angles). More recently, Greenberger, Horne, and Zeilinger (GHZ) [4] and Mermin [5] have shown that QM and local realism are incompatible even at the level of perfect correlations,
for certain states of three or more particles. Hardy has also presented a clever gedanken experiment using electron-positron annihilation to achieve a contradiction with local realism without the need for inequalities [6], and has recently proposed an optical analog which may allow a feasible experimental implementation [7]. Unfortunately, none of these ingenious extensions and generalizations of the work of Bell reduces the experimental requirements for a completely unambiguous test. In fact, all of them seem to mandate even stronger constraints on any real experiment than do the original two-particle inequalities. One exception is the recent discovery that the detection efficiency requirement can be reduced by employing a state of two particles that are not maximally-entangled, i.e., with an unequal superposition of the two terms [8].

Experimental tests of Bell’s inequalities have been extended to new systems, some relying on energy-time or phase-momentum entanglement [9-12]; nevertheless, regardless of the type of entanglement employed, no test of Bell’s inequalities to date has been incontrovertible, due to several loopholes which severely reduce the true impact such an experiment might yield. One of these, the angular correlation problem, has essentially been solved by turning to down-conversion light sources over cascade sources, leaving the fast-switching loophole and the detection loophole. The former concerns the space-like separation of the different parts of the experiment. Clearly, no claims about nonlocality can be made if the pre-detector analyzers are varied so slowly that a signal traveling at the speed of light could carry the analyzer-setting information back to the source or to the other analyzer before a pair was produced or detected. To close this loophole, the analyzers’ settings should be rapidly and randomly changed. Only one Bell-type experiment, that of Aspect et al. [13], has made any attempt at all to address this locality condition, but even in that experiment the loophole remains. Although the experiment used rapidly-varying analyzers, the variation was not random, and it has been argued that the rapidity of the polarization switching was not sufficient to disprove a causal connection between the analyzer and the source [14, 15].

The detection loophole arises from the non-unity detection efficiency in any real experiment (efficiencies in past tests were at best 10%), so that only a fraction of the emitted correlated pairs is detected. If the efficiency is sufficiently low, then it is possible for the subensemble of detected pairs to give results in agreement with quantum mechanics, even though the entire ensemble satisfies Bell’s inequalities. Due to the non-existence of adequate detectors, experiments have so far employed an additional assumption, equivalent to the fair-sampling assumption that the fraction of detected pairs is representative of the entire ensemble [16, 17]. In order to experimentally close this loophole, one must have detectors with sufficiently high single-photon detection efficiencies. Formerly, it was believed that $\approx 83\% = 2\sqrt{2} - 2$ was the lower efficiency limit. However, one of us (P. H. E.) has shown that by using a non-maximally entangled state (i.e., one where the magnitudes of the probability amplitudes of the contributing terms are not equal), one may reduce the detector requirement to $\approx 67\%$, in the limit of no background [8].

2 High-efficiency single-photon detectors

The highest single-photon detection efficiencies to date have been observed using avalanche photodiodes in the Geiger mode; until recently these have been limited to about 40%. We have measured efficiencies as high as 76%, and there are indications that these may be improved to 80% or even 90% [18, 19]. The technique used to measure the absolute efficiency of a single-photon detector is now fairly well known. It was proposed by Klyshko [20], and first used by Rarity et al. [21]
FIG. 1. A simplified schematic of the setup used to measure absolute quantum efficiencies. The 10-cm-long KDP crystal is pumped by the 351-nm line from an argon-ion laser. The smaller iris and interference filter on the path of the trigger photon serve (through phase-matching and energy conservation constraints) to define the path of the conjugate photons, which are all collected by the bottom detector (modulo losses en route). The outputs of the detectors are amplified and fed into the START and STOP channels of a Time to Amplitude Converter/Single Channel Analyzer. The coincidence rate output, as well as the two singles rates, are measured with a counter and stored on a computer. By comparing the coincidence and trigger singles rates, the efficiency of the bottom detector may be determined.

to characterize a silicon avalanche photodiode. The pairs of photons produced in spontaneous parametric down-conversion are highly correlated in time, and reasonably well collimated (i.e., constraining the direction of one photon of a pair determines within a few milliradians the direction of the other). One photon of each pair is directed to a “trigger” detector, and the collection optics are arranged to catch all of the “conjugate” photons with the detector whose efficiency is to be measured (see Fig. 1). The singles count rates at each detector ($R_t$ and $R_c$) are measured, as well as the rate of coincidence counts ($R_{t,c}$) between the detectors. In the ideal limit of no accidental counts (arising from photons from different pairs “accidentally” arriving within the coincidence timing window) and no background events (from unwanted external light, dark counts within the detector, or electronic noise), the efficiency of the “conjugate” detector is simply the ratio of coincidence rate to the “trigger” detector singles rate: $\eta_c = R_{t,c}/R_t$. In the presence of accidentals, $A$, and trigger detector background, $BG$, the formula is modified slightly:

$$\eta_c = \frac{R_{t,c} - A}{R_t - BG}.$$  (1)

In practice our correlated photon pairs resulted from pumping a potassium di-hydrogen phosphate (KDP) crystal, cut for type-I phase-matching. The down-converted photons typically exited
the crystal at a few degrees with respect to the axis of the pump beam. Using irises and filters to select the trigger photons, we were able to measure the efficiency at the wavelength pairs 702-702nm and 632nm-788nm (the energies of the down-converted photons must sum to the energy of the parent ultraviolet photon at 351nm). We examined four single-photon detectors: two Single Photon Counting Modules (SPCM-200-PQ, EG&G), and two Solid State PhotoMultipliers (SSPM, Rockwell International Corp.). The former devices use Geiger-mode silicon avalanche photodiodes specially manufactured to have a very low "k", the ratio of hole- to electron-ionization coefficients [22]; our devices were also custom modified to employ a high overbias voltage of 30V. The SSPMs are also silicon devices, but operate using impurity-band-to-conduction-band impact-ionization avalanches, yielding a very sensitive response in the infrared. The avalanches are localized within areas several microns in size, and do not in general lead to device breakdown, so that these devices are capable of distinguishing between single-, double-, etc. photon detections [23].

The highest observed efficiencies were 70.9±1.9% (with an SSPM, at 632 nm), and 76.4±2.3% (with an SPCM, at 702 nm). We believe these to be the highest reported single-photon detection efficiencies in the visible spectrum; they are important for quantum cryptography and loophole-free tests of Bell’s inequalities, as well as more prosaic applications such as photon correlation spectroscopy and velocimetry. It is important to note that associated with each of the detectors there are other sources of loss, which may yet be improved. Notably, the SPCM detector is housed in a can with uncoated glass windows, and the detector surface itself was broadband anti-reflection coated. Using multi-layer, wavelength-specific coatings, it should be possible to essentially eliminate losses at these interfaces, implying a detection efficiency of ≥ 82%. Moreover, a higher overbias is expected to increase the efficiency even further. The plastic optical fibers used to couple the light into the SSPMs were measured (after dismantling the apparatus) to possess unexpected losses—correcting for all of these losses would suggest SSPM efficiencies as high as 93±7%. (Since then, relative measurements of SSPM efficiencies of 90±5% have been observed.) Work is currently underway to improve the fiber coupling scheme. In addition to high efficiency, a useful single-photon detector must have a low level of background, or noise. The SPCMs are internally cooled to about -30°C, and have rather small active areas (only (0.1mm)²); their dark count rates are correspondingly low, typically 65s⁻¹. The SSPMs are cryogenically cooled to 6 K, but have much larger active areas ((1nm)²); typical dark count rates are 7,000s⁻¹.

Previous measurements of the time correlation of the photon pairs have shown that they are emitted within 40 fs of each other [24]. Therefore, they can be used to accurately measure the intrinsic time resolution of single-photon detectors, by mapping out the coincidence rate as a function of electronic delay time. We found that the time profile for coincidences between the SSPM and the SPCM consisted of a main 3.5ns-FWHM peak preceded by a smaller peak by 11ns. A similar time profile between two SPCM’s displayed only one peak, with 300ps FWHM. Afterpulses were detected in the SPCM’s at a level of less than 10⁻⁴ of the counting rate, with an exponential falloff time-constant of 4.5µs.

3 Proposed EPR-source

Even with unit-efficiency detectors, the down-conversion schemes used until now are inadequate for a completely unambiguous test of Bell’s inequalities, because they must perforce discard counts. In the simplest of the down-conversion Bell-inequality experiments [25, 26], non-collinear corre-
FIG. 2. Schematic of a novel arrangement in which a loophole-free test of Bell’s inequalities is feasible. a) An ultraviolet pump photon may be spontaneously down-converted in either of two nonlinear crystals, producing a pair of orthogonally-polarized photons at half the frequency. One photon from each pair is directed to each output port of a polarizing beam splitter. When the outputs of both crystals are combined with an appropriate relative phase $\delta$, a true singlet- or triplet-like state may be produced. By using a half waveplate to effectively exchange the polarizations of photons originating in crystal 2, one overcomes several problems arising from non-ideal phase-matching. An additional mirror is used to direct the photons oppositely to separated analyzers. b) A typical analyzer, including a half waveplate (HWP) to rotate the polarization component selected by the analyzing beam splitter, and precision spatial filters to select only conjugate pairs of photons. In an advanced version of the experiment, the HWP could be replaced by an ultrafast polarization rotator (such as a Pockels or Kerr cell) to close the space-like-separation loophole.

lated photons were directed through equal path lengths to opposite sides of a 50-50 beam splitter, aligned so that the transmitted mode of one photon coincided with the reflected mode of the conjugate photon, and vice versa. A half waveplate prior to the beam splitter was used to rotate the polarization of one of the photons (which were initially horizontally polarized) by $90^\circ$. Coincidence rates between detectors looking at the two output ports were recorded, as a function of the orientation of polarizers at the detectors. In only measuring coincidence rates, the experimenters were able to effectively create a singlet-like state by discarding cases where both photon exited the same port of the beam splitter. However, it should be stressed that because of these discarded terms, the detection efficiency is inherently limited to 50% (unless the detector can reliably distinguish one photon from two), and no indisputable test of Bell’s inequalities is possible. A similar problem arises in the single-crystal Bell-inequality experiment presented at this conference [27], as well as in experiments based on energy-time entanglement [11] and phase-momentum entanglement [12]. We propose here a setup which should permit for the first time closure of the loopholes. A very different experiment by Edward Fry is just underway [28], using for the first time atoms from dissociated mercury dimers as the correlated particles. The advantage is that detection efficiencies of 95% are possible by photoionizing the atoms and detecting the photoelectrons.

A schematic of our proposed source is shown in Fig. 2a [29]. Two nonlinear crystals are simultaneously pumped by a coherent pump beam to induce spontaneous parametric down-conversion; the pumping intensity can be independently varied at each crystal. The crystals are cut for type-II collinear degenerate phase matching (i.e., the down-converted photons are collinear and orthogonally polarized, with spectra at roughly twice the pump wavelength). For example, we envisage using 1-cm long crystals of beta-barium borate (BBO), pumped by the 325-nm line from a HeCd laser incident at $54^\circ$ to the optic axis. For clarity, we first assume a monochromatic pump beam (at frequency $2\omega_o$), and a single-mode treatment of the down-converted photons. Then the state after the crystals is

$$|\Psi> = \sqrt{1-|A|^2}|\text{vac}> + \frac{A}{\sqrt{1+|f|^2}}(|H,V>_{\text{crystal 1}} + f|H,V>_{\text{crystal 2}}) ,$$

(2)
where we have omitted higher order terms (for the very unlikely case in which more than one pump photon down-converts; by reducing the pump intensity, the contribution of these terms can be made as small as desired). \( A \) includes the down-conversion efficiency into the modes we are considering, and also the pump field strength; \( f \) represents a possible attenuation of the pump beam incident on crystal 2. The state (2) describes a photon pair [one photon polarized horizontally (H), the other vertically (V)] originating with probability amplitude \( A/\sqrt{1 + |f|^2} \) in crystal 1 and with probability \( Af/\sqrt{1 + |f|^2} \) in crystal 2. We now combine the modes from the two crystals at a polarizing beam splitter. For an ideal polarizing beam splitter, incident \( p \)-polarized light (horizontal in Fig. 2) is completely transmitted, while incident \( s \)-polarized light [vertical (out of the plane of the paper) in Fig. 2] is completely reflected; therefore, one photon of each pair will be directed to output port 3, while the conjugate photon is directed to output port 4. Including a phase shift \( \delta = 2\omega_0 \Delta x/c \) (where \( \Delta x \), the difference in path lengths, may be varied by moving one of the mirrors slightly) between the two non-vacuum terms of (2), we then have

\[
|\Psi > \approx (|V >_3 |H >_4 + f e^{i\delta} |H >_3 |V >_4 ) ,
\]

where we have omitted the (predominant, but uninteresting) vacuum term and the prefactor \( A/\sqrt{1 + |f|^2} \). For the balanced case \( (f = 1) \), and for \( \delta = 180^\circ \), (3) reduces to the familiar singlet-like state. (In practice, one should probably use a triplet-like state \( [\delta = 0] \), as this is far less sensitive to cross-talk effects in the polarizing beam splitter—see below.) Note that this contains no non-coincidence terms that must be intentionally discarded to prepare a singlet-like state. The source may thus find application in quantum cryptography [30, 31], as it doubles the signal-to-noise ratio of most previous down-conversion EPR schemes.

With the above source of correlated particles, one can now perform a polarization test of Bell's inequalities. Polarization analysis is performed using an additional polarizing beam splitter after each output port of the interferometer, and examining one or both channels of each analyzer with high efficiency detectors (see Fig. 2b). "Rotation" of these analyzers can be effectively accomplished by using a half waveplate before each one to rotate the polarization of the light. If the detectors are far separated from each other and from the source, and one uses some rapid, random means to rotate the light before the analyzers, (such as a Pockels or Kerr cell, whose voltage is controlled by a random signal), then one can close the space-like separation loophole. The signal could be derived, for instance, from the decay of a radioactive substance, or even from the arrival of starlight. Note that since the down-converted photons are emitted within tens of femtoseconds of one another [24] (unlike the photons in an atomic cascade), the limiting time factors will be the detector resolution (expected to be less than 10 ns) and the switching time (which can also be on the order of nanoseconds).

4 Other considerations

It has been shown that by using a non-maximally entangled state (i.e., one where the magnitudes of the probability amplitudes of the contributing terms are not equal), one may reduce the detector efficiency requirement [8]. The basic idea is that by making one term of (3) have a greater amplitude than the other, one effectively polarizes the source. For example, if \( f \) is a
real number $\leq 1$, then a photon travelling to port 4 possesses a net horizontal polarization, while a photon travelling to port 3 appears somewhat vertically-polarized. By appropriately choosing the polarization-analyzer angles, one may reduce the contributions of the singles rates, while still violating a Bell's inequality, even for $\eta$ as low as 67%. It should be noted, however, that the magnitude of the violation is reduced as $f$ is reduced, increasing the relative importance of undesirable background counts. A background level (e.g., from any stray light or dark counts) of less than 1% is desired.

One problem that arises in the above scheme is the effect of walkoff of the down-converted photons in the birefringent parent crystal. While the birefringence of the nonlinear crystal is essential for achieving phase-matching, it also results in a relative displacement of the two down-converted photons: they propagate in the same direction after exiting the crystal, but are separated by a distance $d = L \tan \rho$, where $L$ is the propagation distance inside the crystal, and $\rho$ is the intra-crystal angle between the ordinary and extraordinary beams. Consequently, after the polarizing beam splitter, the position of a detected photon partially labels its origin, degrading coherence. Remarkably, insertion of an extra half waveplate after one of the crystals to rotate the polarization by 90° avoids this problem. Photons from either crystal exiting port 3 of the beam splitter would be initially extraordinary-polarized; photons exiting port 4 of the beam splitter would be initially ordinary-polarized. A similar situation occurs due to longitudinal walkoff: After propagation through some length of the birefringent down-conversion crystal, one of the down-converted photons will “pull ahead” of its conjugate. In the absence of the half waveplate, one could in principle determine from which crystal a given pair of photons originated by examining the timing of the coincident detection. This distinguishability of contributing paths removes quantum interference, just as in the transverse walkoff situation considered above. (This same effect was discussed by Sergienko et al. at this conference [32].) Fortunately, the extra half waveplate also removes this longitudinal walkoff effect. With the waveplate, an infinitely-fast detector looking at the originally extraordinary-polarized photons would always trigger before the detector looking at the originally ordinary-polarized photons; hence, interference remains. The waveplate also removes difficulties arising from finite bandwidth and vector phase-matching considerations [29].

For a plane-wave pump, the phase-matching constraints for down-conversion imply that with careful spatial filtering, one can in principle collect only conjugate pairs of photons from the crystals (i.e., essentially no unpaired photons, aside from stray light). Once we allow a more realistic, gaussian-mode pump, then this is no longer possible. For identical, finite-sized collection irises, there will always exist situations where one photon is detected while the other is not (even aside from the problem of inefficient detectors). This effect is mitigated by collecting over a larger solid angle. In particular, in order to keep the losses less than 2%, we must employ irises which accept light out to 30 times the pump divergence angle. For example, if we employed a 325-nm pump with a beam waist radius ($1/e^2$) of 3.5 mm, then we would need to accept all half-angles up to 1 milliradian. (In practice, this could be accomplished by use of a precision spatial filter system in each output port; see Fig. 2b.)

We have also investigated the effects of using various non-ideal optical elements. One of these investigations has led us to a novel interference effect. In considering an imperfect recombining polarizing beam splitter, we discovered that the singlet-like state is much more sensitive than the triplet-like state to beam-splitter “cross talk”, when $f = 1$. By “cross talk” we mean the situation where photons exit the wrong port of the beam splitter (i.e., some fraction of the incident $p$-
FIG. 3. The two processes that can lead to the final state of both photons going toward the same detector are indistinguishable; thus, we add their probability amplitudes. We assume that $f = 1$, and for simplicity that the transmission and reflection amplitudes are $1/\sqrt{2}$ and $i/\sqrt{2}$, respectively. When $\delta = 0$ (triplet-state), there is destructive interference.

polarized light is reflected, while some fraction of the incident s-polarized light is transmitted. Whenever this happens, we once again have a situation in which both photons can leave the same output port of the beam splitter – just as in the experiment discussed at the beginning of Sect. 3, this reduces the effective efficiency, making it more difficult to violate Bell’s inequalities. However, in the special case for which the percentage of cross-talk for the two polarizations is equal (one example is the case of a non-polarizing 50-50 beam splitter), there is quantum interference which prevents the two photons from exiting the same port of the beam splitter; the interference only exists for the triplet arrangement ($\delta = 0^\circ$), not the singlet case ($\delta = 180^\circ$) [see Fig. 3]. This is, in some sense, the complementary effect to one in some one-crystal experiments (see, for instance, [24, 33]), in which the two-photons from a down-conversion crystal always take the same port of a 50-50 beam splitter. It would be interesting to demonstrate this phenomenon experimentally.

Using a scheme that was motivated by two of those presented at this conference [34, 35], it is possible to produce this EPR-source using only one down-conversion crystal. The key is to reflect the pump beam back through the crystal (see Fig. 4). Although this scheme has the advantage that it reduces the number of optical elements (most importantly the number of crystals) it does have a few difficulties not present in the 2-crystal approach. First, one must use an isolator to prevent the pump from coupling back into the laser. Next, the return path must be kept short compared to the pump coherence length (typically 10 cm for a HeCd laser), to maintain coherence between the processes in which the down-converted pair originated from a left-going or a right-going pump photon. Finally, one must find very good dichroic mirrors to separate the pump beam.
FIG. 4. Proposed one-crystal scheme to produce EPR-state.

from the down-converted photons. One solution is to use high-quality prisms (not shown in Fig. 4).

In conclusion, with the setup described herein it should be possible to produce an indisputable violation of a Bell’s inequality. We have examined the effects of background, imperfect polarizing beam splitters, and phase-distorting optics. For λ/20-flatness optics, a background level of 1%, and custom-selected polarizing beam splitters (with an extinction ratio of at least 500:1), numerical calculation predicts that a violation should be possible as long as the net detection efficiency is greater than 82.6%. Naturally, all optics would be anti-reflection coated to minimize reflection losses; including a 0.25% loss for each interface, and the 2% loss from the gaussian nature of the beam, this means that the bare detector efficiency needs to be at least 89.6%, which may be achievable in light of our recent measurements [18, 19]. Of course, for a safety margin, one would like it to be even higher.

5 Acknowledgments

We would like to acknowledge the invaluable assistance of Joshua Holden. One of us (P.H.E.) was supported by the U.S. Department of Energy, Contract No. DE-AC03-76SF00098. The rest were supported by the Office of Naval Research under Grant No. N00014-90-J-1259.
References
