Closing loopholes in experimental Bell tests

Moses Misplon

Carleton College, Department of Physics, Northfield, MN

(Dated: 16 April 2017)

Ever since John Bell proposed an experimental test of locality and realism in 1964, physicists have sought a conclusive experimental disproof of these seemingly fundamental principles. However, for fifty years the unique technical challenges of experimental Bell tests resulted only in incomplete Bell tests implemented across a wide variety of experimental platforms, each with remaining "loopholes" limiting its applicability. Only in 2015 were the three key loopholes, efficiency, communication, and freedom-of-choice, all closed in the first loophole-free Bell tests, which drew on technological advancements and the lessons of earlier Bell test implementations.

I. BELL TESTS AND ORIGINS

Bell tests, based on the work of J. S. Bell, provide an experimental test of two natural philosophical assumptions that implicitly underlay all physics prior to quantum mechanics: realism and locality. A *realistic* theory of a physical system assigns a definite probability distribution to the outcome of any measurement that could be conducted on that system without disturbing it, whether the measurement is conducted or not. Realism thus captures the common intuition that physical reality should exist even when it is not actively being observed. A *local* theory restricts the outcome of any measurement to depend only on events in the measurement's relativistic past. Locality amounts to physical influence never traveling at superluminal speeds, which seems to be a natural consequence of relativity, under which information cannot travel at superluminal speeds. Although the principles of realism and locality each seem natural and irrefutable, Bell's Theorem implies that a single hypothetical experiment, the Bell test, is capable of simultaneously disproving all local realistic physical theories. While the Bell test was proposed in 1964 and the first implementations followed immediately thereafter, the combination of technological limitations and the unique demands placed by the test slowed progress for over fifty years. Only in 2015 were the first conclusive experimental Bell tests reported.

Bell's disproof of local realism ironically originated in the attempts of early quantum physicists to emphasize realism's importance. In a famous 1935 paper,[1] Einstein, Podolsky, and Rosen (EPR) argued that quantum mechanics was an incomplete theory because it lacked properties equivalent to realism, which the authors deemed essential for any physical theory. The EPR authors supported this surprising conclusion by demonstrating that the quantum-mechanical theory governing a pair of entangled particles was not realistic. Bohm later simplified this EPR gedankenexperiment from continuous observables to discrete two-state observables.[2] The EPR-Bohm gedankenexperiment considers two spin-1/2 particles prepared in the singlet spin state (the maximally entangled spin state)

$$\frac{1}{\sqrt{2}}\left(|\uparrow_{k,1}\rangle|\downarrow_{k,2}\rangle + |\downarrow_{k,1}\rangle|\uparrow_{k,2}\rangle\right), \tag{1}$$

which has the same representation (up to phase) when expressed in terms of the spin-up and spin-down states associated with any axis k.

The gendakenexperiment then supposes that the particles are separated as shown in Figure 1 and sent to spin measurement apparatus at locations A and B, and considers

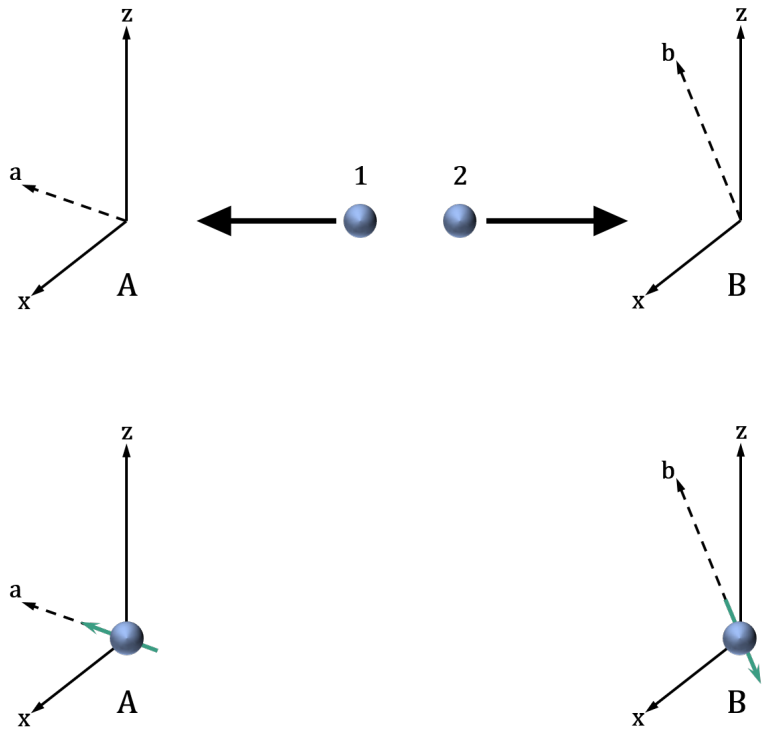


FIG. 1: Schematic of the EPR-Bohm gendankenexperiment. Two spin-1/2 particles 1 and 2 are prepared in the singlet spin state $\frac{1}{\sqrt{2}}\left(|\uparrow_1\rangle|\downarrow_2\rangle + |\downarrow_1\rangle|\uparrow_2\rangle\right)$ and sent along the y-axis in opposite directions to distant measurement apparatus A and B. The apparatus at A measures the spin of particle 1 along the axis a and the identical apparatus at B measures 2 along the axis b, where a and b each can be chosen to be any axis in the xz-plane. A particle will always be measured to have spin $\pm 1/2$ (spin-up or -down) along any axis selected. When the same measurement settings are selected ($a = b$), measurements on the singlet state are perfectly anticorrelated: exactly one particle will be measured in each spin state.

what can be inferred about the spin of particle 1 before it is measured at A. By the EPR authors' logic, a measurement of the arbitrarily-distant particle 2 cannot physically affect the actual state of particle 1. Thus, in any realistic theory, particle 1's spin state must simultaneously have all properties implied by any possible measurement of particle 2. Because measurements along the same axis of two particles in a singlet spin state are perfectly anticorrelated, measuring particle 2 along the axis a completely determines the spin of particle 1 along axis a (if 2 is spin-up, then 1 is spin-down, and vice versa). Therefore, in any realistic theory, particle 1 must have a definite value for its spin with respect to

every possible measurement axis a in the xz-plane. However, in quantum mechanics, the spin operators for the x and z spin components are noncommuting, and thus particle 1 cannot have definite values for its spin along both the x and z axes. Thus the EPR authors concluded that quantum mechanics is not realistic.

The EPR authors, persuaded of quantum mechanics' correctness by the overwhelming experimental evidence, believed the resolution of this "paradox" lay in a more complex realistic and deterministic theory for which quantum mechanics' statistical predictions were simplifying averages, much as statistical mechanics is a consequence of classical mechanics. This overarching theory would include additional variables that could not be directly observed, and a lack of knowledge of the values of these "hidden variables" in any particular experiment would result in the appearance of quantum mechanics' probabilistic behavior. This appealing resolution was gradually undermined by theorists, who throughout the 1940s and 1950s attempted to demonstrate that no realistic theory could replicate the quantum-mechanical predictions for entangled two-particle systems.[3] Finally, Bell[4] brought clarity to these efforts by recognizing that for the EPR authors to argue that the choice of measurement on particle 2 had no influence on particle 1 it was necessary to assume that the propagation of physical influence is not instantaneous, and thus making explicit the assumption of locality implicit in the EPR paper.

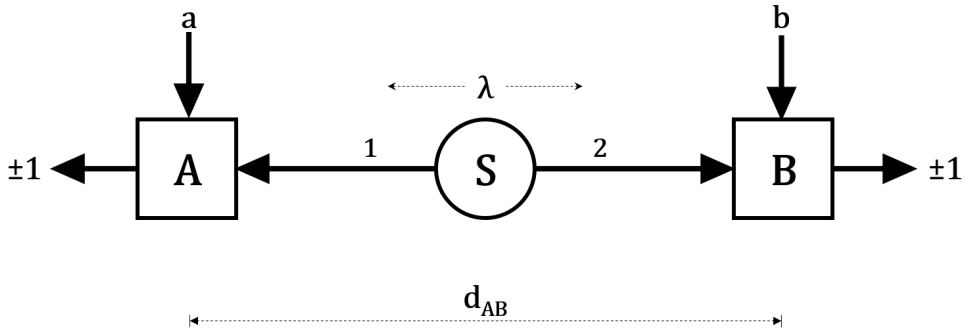


FIG. 2: The Bell test conceptual model derived from the EPR-Bohm gedankenexperiment. A source S emits particles 1 and 2 with shared information λ which interact with apparatus A and B respectively. The apparatus, separated by distance d_{AB}, each accept a parameter (a and b respectively) and produce an output ± 1 dependent on this parameter and λ.

Bell then considered an abstracted model of the EPR-Bohm gendankenexperiment (Figure 2), and analyzed it in the context of an arbitrary local realistic theory (LRT). Bell's

generalized experiment consists of a sequence of trials; in each, a common source S emits two particles, known as a Bell pair. In a LRT, the state of the particles could be described by an arbitrarily complex set of dynamic "hidden" variables; however, by the assumption of realism, the outcomes of all future measurements on the pair must be encoded in the initial values of these variables. Bell let λ represent this shared initial state of the Bell pair, and $\rho(\lambda)$ be the distribution of λ values for Bell pairs emitted by S. The particles carry the information about λ as they travel in opposite directions from S to measurement apparatus A and B, each of which outputs one of two measurement outcomes, ± 1. Each apparatus has one adjustable parameter (a and b respectively) that controls what measurement is performed; in the EPR-Bohm gendankenexperiment, this parameter is the axis along which the spin component is measured. Thus the measurement outcome at apparatus A is a function $A(a, b, \lambda) = \pm 1$ of the measurement settings and the hidden variables. Under the assumption of locality, information about b or B cannot influence the outcome A if the setting selection and measurement at B are spacelike-separated from both the selection and measurement at A; by this and the symmetric argument for B,

$$A(a, b, \lambda) = A(a, \lambda) = \pm 1 \tag{2}$$

$$B(a, b, \lambda) = B(b, \lambda) = \pm 1. \tag{3}$$

The infinite-time correlation $E(a, b)$ in measurements over many trials with fixed measurements settings a at A and b at B is defined by

$$E(a, b) \equiv \int A(a, \lambda) B(b, \lambda) \rho(\lambda) d\lambda. \tag{4}$$

Recalling that in the EPR-Bohm spin system, pairs of particles can be prepared such that if their spins are measured along the same axis, the results are always opposing, Bell concluded more generally that measurements on his Bell pair using the same measurement setting ($a = b = q$) were perfectly anticorrelated, so that for any q,

$$E(q, q) = -1. \tag{5}$$

Via mathematical manipulation of just (2), (3), (4), and (5), Bell derived the first of the Bell inequalities,

$$|E(q, r) - E(q, s)| \leq 1 + E(r, s), \tag{6}$$

for any settings q, r, and s (see Appendix A for this derivation). Because the only assumptions of the derivation were locality and realism, this inequality holds for the predictions of any LRT. Bell demonstrated that for particular measurement settings, quantum-mechanical predictions for the EPR-Bohm spin system violate this bound. Thus quantum mechanics is inconsistent with any local realistic theory, a result known as Bell's Theorem.

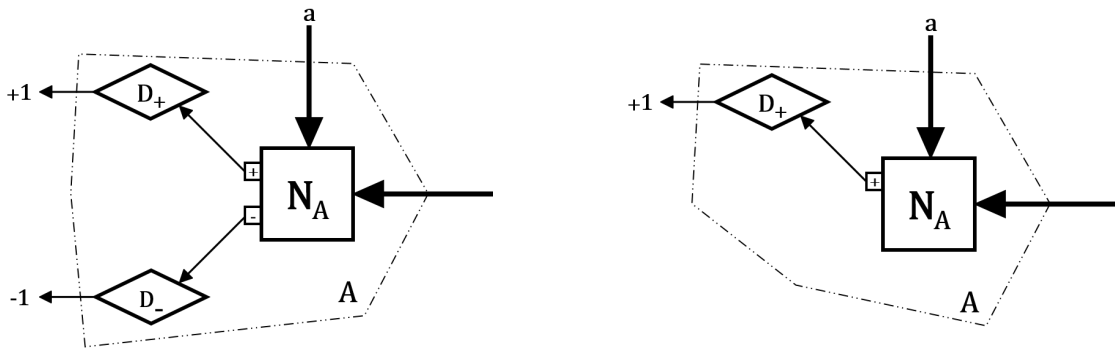
The reliability of quantum mechanics was firmly established by 1964, but counter-intuitively Bell's Theorem did not immediately disprove local realism. The accuracy of quantum mechanics had been verified in many other contexts but never in a Bell-type experiment, and so it remained possible that quantum mechanics failed to accurately predict separated measurements on entangled particles. Thus Bell and others proposed that Bell tests, experiments designed to measure correlations in two-particle systems and detect violations of the Bell inequalities, be conducted to disprove local realism once and for all. A Bell test assumes local realism as its null hypothesis; concludes that under this hypothesis a particular physical system, the experimental platform, must obey a Bell inequality; measures that physical system; and from statistically significant violations of the Bell inequalities can reject the null hypothesis, disproving local realism. While Bell's proposed test was easy to describe, it took over fifty years before a conclusive Bell test was finally conducted.

II. OBSTACLES TO EXPERIMENTAL BELL TESTS: ASSUMPTIONS AND LOOPHOLES

The delay was due to the challenge of translating Bell's idealized test into a feasible experiment without making supplemental assumptions that were empirically unverified. The power of the Bell test is the minimality of its assumptions: by setting bounds on experimental outcomes relying only on locality and realism, Bell's test was capable of disproving *every* physical theory that included locality and realism as principles. However, in practice this universality is easily diminished when experimentalists must make additional assumptions in order to interpret their results as a violation of the Bell inequality or demonstrate that the Bell inequality actually applies to their experiment. If these supplemental assumptions cannot be independently empirically verified, then they are added to the null hypothesis of the Bell test, restricting the test to only be capable of disproving LRTs that also obey the additional assumptions. Moreover, physicists have been able to explicitly construct

exotic LRTs that violate even the most commonsense assumptions.[5–7] Thus any unverified assumption in an experimental Bell test introduces a loophole that limits the conclusiveness of the results by allowing some LRTs to remain effectively untested.

Some assumptions were unnecessarily introduced by Bell himself. First, Bell mistakenly assumed perfect anticorrelation (5) although perfect correlations are impossible to verify empirically in finite time. Bell's setup also implicitly assumed a particular type of measurement apparatus. Most measurement apparatus can be conceptually divided into an analyzer, which separates incoming particles according to the measurement outcome, and detector(s) which count the particles received from an output of the analyzer. In many experimental contexts, it is easier to implement a filtering or one-channel analyzer, in which particles measured $+1$ pass through the analyzer but particles measured -1 are blocked, than a two-channel analyzer, in which all particles are transmitted by the analyzer, with the particles measuring $+1$ and -1 separated into two different output channels. There is no functional difference between the resulting two- or one-channel Bell tests, shown in Figure 3, in the case of perfect efficiency, where "no detection" can be taken to mean -1, but given the experimental reality of inefficiency it is not possible to infer -1 outcomes directly from one-channel data. Therefore, a one-channel experimental test cannot measure the correlations of Bell's original inequality (6).



(a) Two-channel Apparatus

(b) One-Channel Apparatus

FIG. 3: Schematic comparison of apparatus for two-channel and one-channel Bell tests. In both cases, source particles enter from the right into analyzer N_A. In the two-channel case, the two outputs of N_A are routed to detectors D_+ and D_-, which report detections as $+1$ and -1 outcomes respectively. In the one-channel case, only a $+$ output is available, and all detections are reported as $+1$ outcomes.

To correct for these limitations, Clauser et. al. improved Bell's derivation to produce two new Bell inequalities suitable for experimental contexts. The CHSH inequality[5]

$$S = |E(q,r) + E(q',r) + E(q',r') - E(q,r')| \leq 2, \tag{7}$$

which applies only to two-channel experiments, upper-bounds any LRT's predictions for the Bell quantity S; for certain settings (q, q', r, r') quantum mechanics predicts S up to $2\sqrt{2}$, violating this bound. The CH inequality[6]

$$S' = \frac{|p_{AB}(q,r) + p_{AB}(q',r) + p_{AB}(q',r') - p_{AB}(q,r')|}{p_A(q') + p_B(r)} \leq 1, \tag{8}$$

where $p_{AB}(a,b)$ is the probability of photons being detected in the $+1$ channels of both A and B given analyzer settings a and b and $p_A(a)$ is the probability of a photon being detected at the $+1$ channel of A given analyzer setting a, has similar properties but can be used in either one- or two-channel Bell tests.

While these loopholes could be eliminated with more careful derivation, less tractable loopholes remained. These stemmed from two primary sources: assumptions necessary to convert experimental data into the probabilities used in the Bell inequalities and assumptions of independence that are not justified by locality. The three loopholes recognized as most critical are the efficiency loophole (also known as the detection or fair sampling loophole), the communication loophole (also known as the locality loophole), and the freedom-of-choice loophole.

II.1. From Counts to Correlations: Statistics and Inefficiency

The raw output of an experimental Bell test, as shown in Figure 4, is a finite list of tuples (a, b, A, B) representing individual trials. However, the CHSH and CH inequalities bound the infinite-trial probabilities of particular measurement outcomes given measurement settings. The translation from finite counts to probabilities is common in experimental contexts; ordinarily the probability $p(V)$ of an event V is estimated by the rate of occurrence

$$R(V) \equiv \frac{\#\text{ of trials in which } V \text{ occurred}}{\#\text{ of trials}}, \tag{9}$$

where in the limit of infinite trials $R(V)$ approaches $p(V)$. In an experimental context, not every attempted trial will produce usable data. In a Bell test in particular, some particles

will fail to transmit through the analyzer, and the detectors will fail to detect all particles that reach them. Inefficiency in each component of the Bell test combines to generate an overall efficiency defined by

$$\eta \equiv \frac{\#\text{ of successful trials}}{\#\text{ of trials in which at least one particle enters its analyzer}}. \tag{10}$$

In the absence of any known physical effects that could cause systematic basis, typical experimental analysis addresses inefficiency by assuming that the successful trials represent a fair sample of all of the trials. The rate $R(V)$ can then be calculated from just the successful trials. In the context of a Bell test, this fair sampling assumption is equivalent to assuming that the transmission and detection probabilities of particles are independent of λ, a, and b.

#	a	b	A	B
1	0	$\frac{\pi}{4}$	$+1$	-1
2	$\frac{\pi}{2}$	$\frac{\pi}{4}$	$+1$	$+1$
3	$\frac{\pi}{2}$	$\frac{\pi}{4}$	-1	$+1$
4	0	$\frac{3\pi}{4}$	-1	-1
5	$\frac{\pi}{2}$	$\frac{3\pi}{4}$	$+1$	-1
⋮	⋮	⋮	⋮	⋮

(a) Ideal (perfectly efficient) Bell test. Every trial can be used to calculate probability estimates.

#	a	b	A	B
1	0	$\frac{\pi}{4}$	$+1$	-1
2	$\frac{\pi}{2}$	$\frac{\pi}{4}$		$+1$
3	$\frac{\pi}{2}$	$\frac{\pi}{4}$	-1	$+1$
4	0	$\frac{3\pi}{4}$	-1	
5	$\frac{\pi}{2}$	$\frac{3\pi}{4}$		-1
⋮	⋮	⋮	⋮	⋮

(b) Realistic Bell test. Trials 2, 4, and 5 cannot be used in calculations of probability estimates.

FIG. 4: Hypothetical example of raw data from a two-channel experimental Bell test. A complete trial is recorded via a tuple (a, b, A, B) containing measurement settings and outcomes.

While fair sampling greatly simplifies analysis, it is not theoretically justified by local realism and cannot be experimentally verified; thus assuming fair sampling (or any other properties of the unsuccessful trials) introduces the efficiency loophole. However, without this assumption, decreases in η will decrease all estimated probabilities because calculations of $R(V)$ will include only a subset of successful trials in the numerator but all trials in the denominator. Thus increasing inefficiency reduces S and S'. For sufficiently low efficiencies,

the expected Bell inequality violation will disappear completely. False positives (or "dark counts"), in which the detector triggers in the absence of the target particle, compound the inefficiency problem. In a typical experiment, the dark counts would be measured by removing the source, the dark rate would be modeled as a Poisson process, and finally the expected dark counts would be subtracted from the measured counts. However, LRTs can be constructed for which the dark rate itself depends on the source, and so to avoid additional assumptions, Bell tests should not correct for dark counts. If no such corrections are made, then S and S' also decrease as the background, defined by

$$\zeta \equiv \frac{\#\text{ of dark counts}}{\#\text{ of total counts}}, \tag{11}$$

increases.

Early attempts to estimate what efficiencies and backgrounds would be required to evade the efficiency loophole considered only the singlet state and found a critical efficiency of $\eta_c = 2(\sqrt{2}-1) \approx 82.8\%$ regardless of the background.[6] More sophisticated analysis by Eberhard[8] determined that, counter-intuitively, non-maximally entangled states improve the critical efficiency, albeit only if the background is low. For any efficiency, Eberhard proved that the optimal state was of the form

$$|\psi_r\rangle = \frac{1}{\sqrt{1+r^2}} \left(|\uparrow_1\rangle |\downarrow_2\rangle + r |\downarrow_1\rangle |\uparrow_2\rangle \right) \tag{12}$$

for some $0 \leq r \leq 1$ (where $r = 0$ is the unentangled state and $r = 1$ is the maximally entangled state). The improvements in critical efficiency offered by these Eberhard states are shown in Figure 5. When the background $\zeta = 0$, Eberhard found that the true critical efficiency was $\eta_c = 2/3$, achieved in the limit as $r \to 0$. Therefore, a Bell test must maintain $\eta > 2/3$, or any analysis which achieves a Bell inequality violation will open the efficiency loophole.

II.2. When Locality Does Not Imply Independence: Communication and Freedom-of-Choice Loopholes

Although physicists were quick to note the efficiency loophole,[5,6] appreciation of the communication and freedom-of-choice loopholes developed more slowly.[6,9–12] In the derivation of the Bell inequalities, locality justifies the independence of several pairs of events: measurement outcomes must be independent of each other, as mathematically expressed in equations

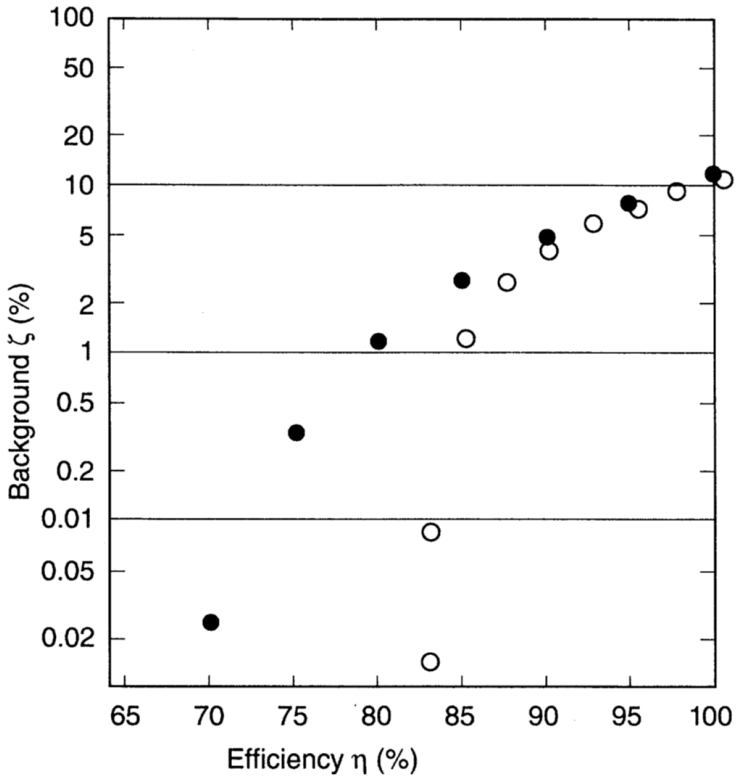


FIG. 5: Plot of the maximum possible background for a given efficiency such that a Bell test with that efficiency and background is theoretically capable of producing a violation of the Bell inequalities without opening the efficiency loophole. The solid circles represent the maximum background possible if using the optimal Eberhard state for the given efficiency, while the hollow circles represent the inferior maximum background when the initial state is fixed to be the singlet state. As efficiency increases, the difference between the two methods decreases because the optimal Eberhard state approaches the singlet state. The region below the solid-circle curve represents the experimental conditions that are necessary for closing the efficiency loophole. Plot from Eberhard.[8]

(2) and (3) and the implicit additional conditions that a, b, and λ are all pairwise independent. If the emission, setting selection, and measurement events are ideally positioned in spacetime, as shown in Figure 6, locality implies all the necessary independence conditions; however, the argument fails when the spacelike separation of events is not enforced in a particular experiment. The consequent supplemental assumptions that these events are still independent are grouped into the communication loophole (assumption of no physical influence of the measurement processes on each other and on the emission event) and the

freedom-of-choice loophole (assumption that the settings selected are not dependent on λ).

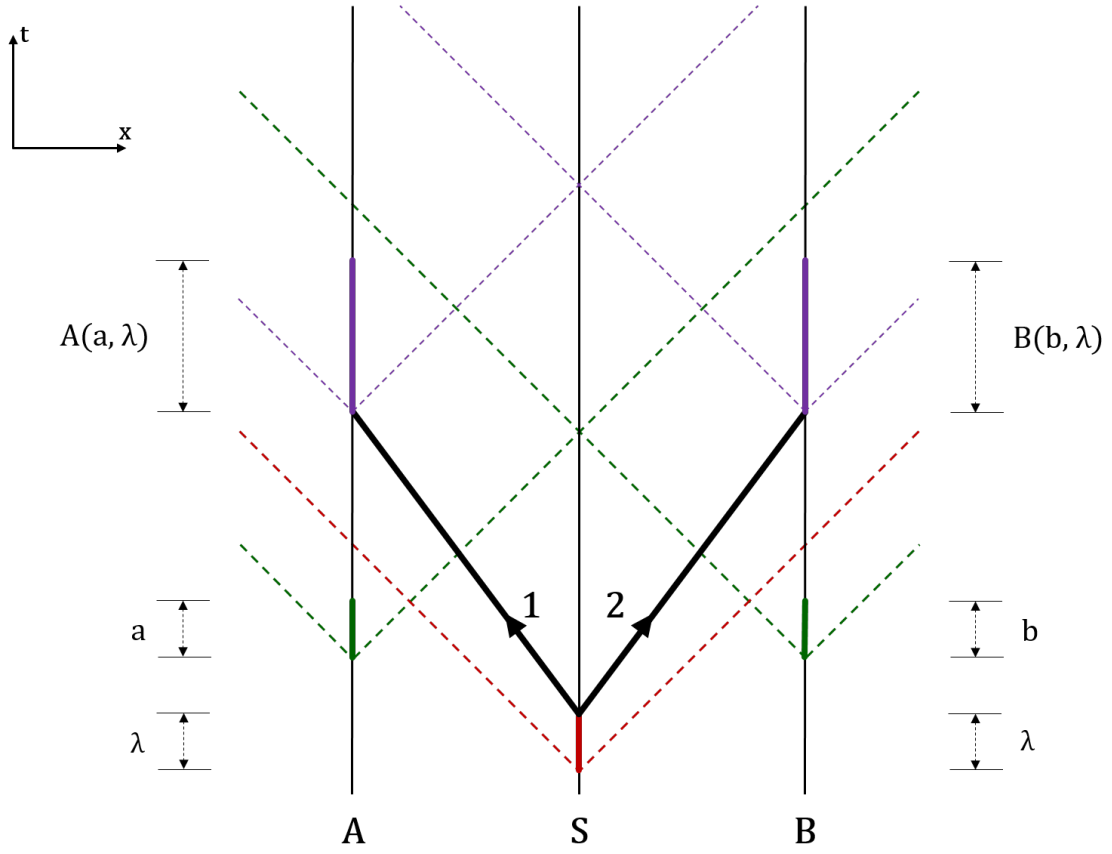
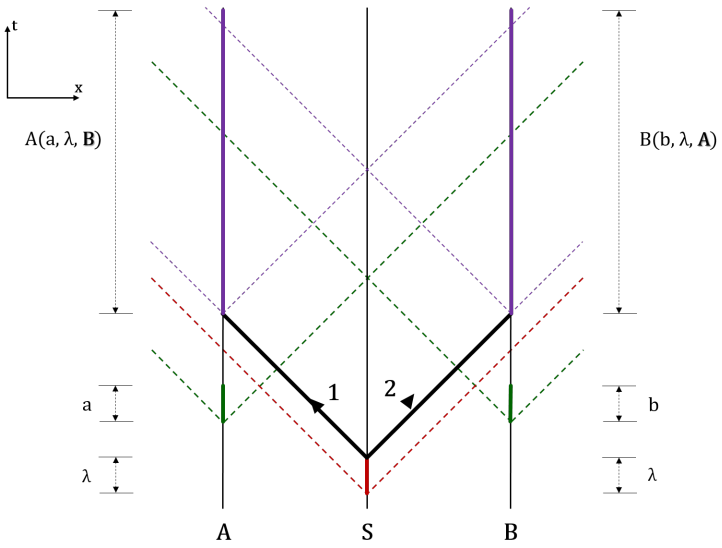
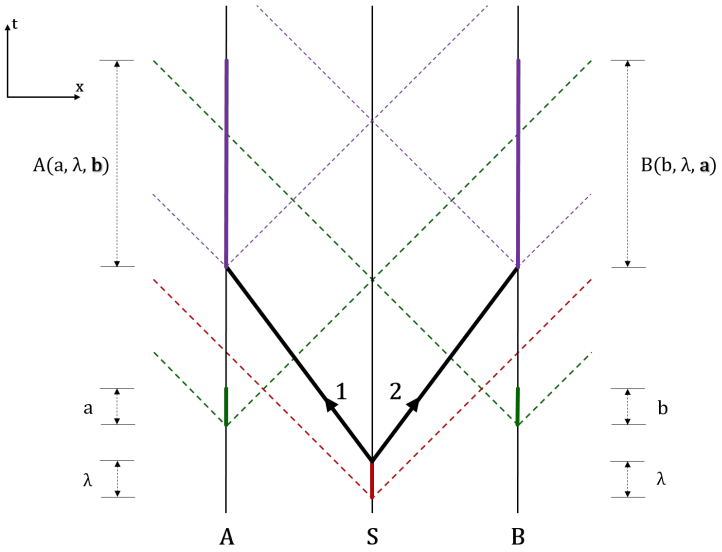


FIG. 6: Minkowski diagram of an ideal symmetric Bell test with no locality-related loopholes. This diagram is drawn in the rest frame of A, S, and B. The dark red interval indicates the emission time of the Bell pair from S (during which λ is determined), while the thick black lines mark the subluminal trajectories of the two emitted particles from S to the apparatus A and B. The green intervals represent the randomized measurement setting selections, while the purple intervals indicate the measurements themselves. The dashed lines represent boundaries of the forward light cones of each event; under the assumption of locality, an event can only physically influence events in its forward light cone. Note that in this experiment free of locality-related loopholes, under locality the setting selections and measurements cannot influence either the emission event or the events at the other measurement apparatus (closing the communication loophole) and the emission event cannot influence the measurement settings (closing the freedom-of-choice loophole).



(a) Communication loophole due to measurement. The measurement intervals are in each other's forward light cones, and thus the measurement outcome at one apparatus can physically influence the measurement at the other.



(b) Communication loophole due to setting selection. The measurement at each apparatus is in the forward light cone of the setting selection at the other apparatus; thus the settings at one apparatus can influence the measurement outcome at the other.

FIG. 7: Minkowski diagram of symmetric Bell tests which close the freedom-of-choice loophole but retain the communication loophole. Diagram elements are as defined as in Figure 6.

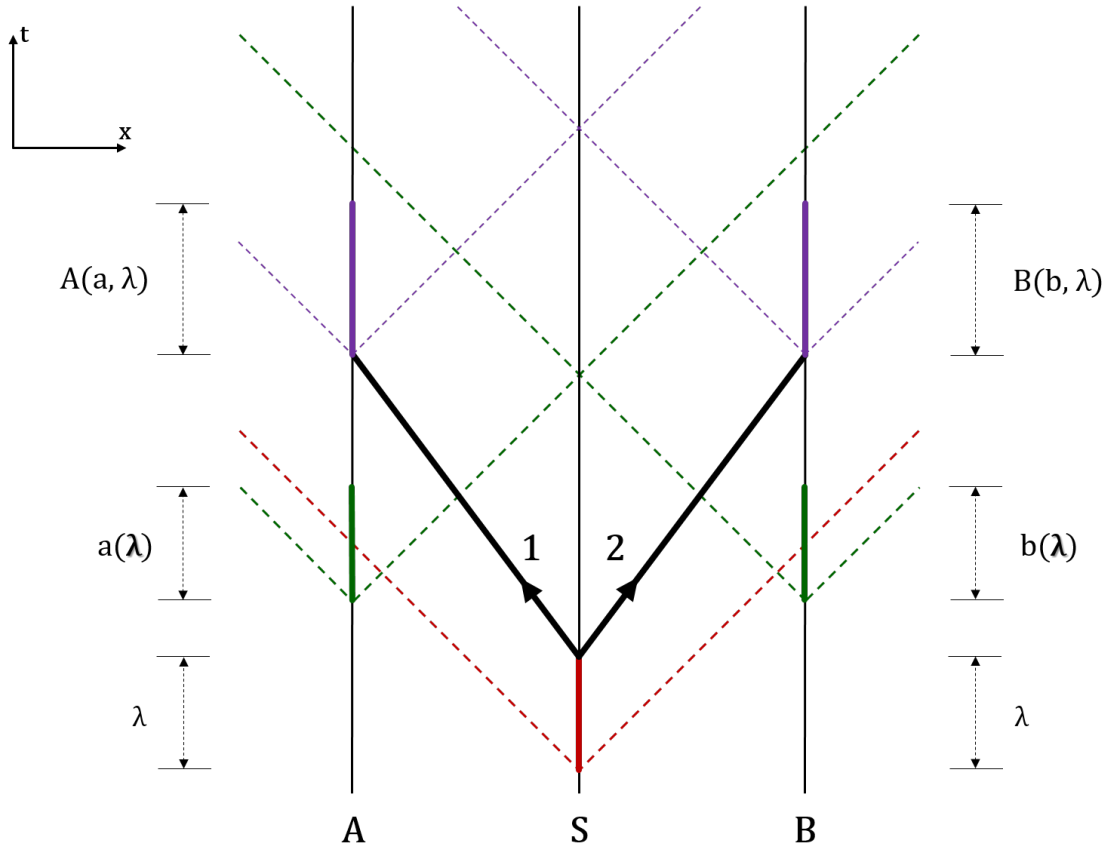


FIG. 8: Minkowski diagram of a symmetric Bell test which closes the communication loophole but retains the freedom-of-choice loophole. Diagram elements are as defined as in Figure 6. Note that the measurement setting selections are in the relativistic future of the emission event, so a and b can be dependent on λ under the assumption of locality; thus the freedom-of-choice loophole is open.

The communication loophole is present if locality permits the setting selection or measurement at one apparatus to physically influence the emission event or the outcome of the measurement at the other apparatus (as shown in Figure 7). To close this loophole, the first moment of the setting selection at B must be spacelike-separated from the last moment of the measurement at A, and vice versa. As an additional subtlety, the setting selection at each apparatus must be, as Bell noted, "free or random." If the process selecting settings is predictable in advance (e.g. a psuedorandom number generator), locality does not preclude LRTs in which this predictive information is immediately transmitted by an unknown mechanism to the source or to the other apparatus and biases the emissions or measurements. Thus for the purposes of the communication loophole, the setting selection interval begins

at the moment the selection can be predicted with accuracy greater than 1/2. However, if the settings are determined too late, it is conversely possible under locality for λ to influence the setting selections, opening the freedom-of-choice loophole as shown in Figure 8. Keeping both loopholes closed simultaneously requires generating truly random numbers in a narrow temporal window. This is especially foreboding because there is no way to empirically verify the randomness or unpredictability of a physical process, and so assumptions must be made about the physics of the random number generator (RNG).

III. THE FIRST EXPERIMENTAL BELL TESTS: ATOMIC CASCADES

While the original gedankenexperiment underpinning the Bell test considered the spins of entangled spin-1/2 particles, the Bell inequalities apply to spacelike-separated and randomized measurements on any two-state, two-particle system. Moreover, many systems are predicted by quantum mechanics to violate the Bell inequalities; essentially any two-particle system with two-state non-commuting observables will be expected to produce Bell inequality violations. In practice, however, Bell tests have proved easier to conduct on some experimental platforms than others. Early experimenters faced the challenge of finding a platform that could be practically implemented using existing technology with sufficient efficiency, spacelike separation, and count rates. While Bell tests were conducted with other platforms, such as proton-proton scattering and positronium annihilation gamma rays (see Appendix B), the most successful and popular platform for the first three decades of experimental Bell tests was polarization of visible light emitted by atomic cascades.[7] This platform was proposed by CHSH,[5] first implemented by Freedman and Clauser,[13] and later reanalyzed by CH.[6]

III.1. Introduction to the Atomic Cascade Platform

An atomic cascade is a sequence of spontaneous transitions in an atom's electronic energy levels. An atom at the top of a cascade will spontaneously traverse down the cascade, emitting photons with energies corresponding to each transition in quick succession. Two-step atomic cascades of the form

$$X_{E_1,J} \rightarrow X_{E_0,J} + \gamma_1 + \gamma_2, \tag{13}$$

where E_1 is the atomic energy at the top of the cascade, E_0 is the ground state energy, γ_1 and γ_2 are the two photons emitted, and the atomic spin quantum number J is the same before and after the cascade, are particularly useful platforms for Bell tests. Conservation of angular momentum then implies that γ_1 and γ_2 have zero net angular momentum, and thus the photons will have orthogonal linear polarizations with respect to any pair of orthogonal axes. The photons emitted by such a two-step cascade form a Bell pair in the linear polarization singlet state $(\uparrow_1\rightarrow_2 + \rightarrow_1\uparrow_2)/\sqrt{2}$. If the lifetime of the intermediate state is short relative to d_{AB} and the photons are emitted from the atom in approximately antiparallel directions, then spacelike-separated measurements of linear polarization can be made on each photon. In some cascades, both photons in the Bell pair have visible wavelength, which allows experimentalists to bring the tools of visible-light optics to bear. Several different atomic transitions have appropriate energy level spacing and lifetimes for such a cascade; the common choice in early Bell tests was the $4p^2\,{}^1S_0 \rightarrow 4p4s^1P_1 \rightarrow 4s^2\,{}^1S_0$ transition in atomic calcium (shown in Figure 9). Other cascades in atomic mercury with similar properties also saw use.[14,15]

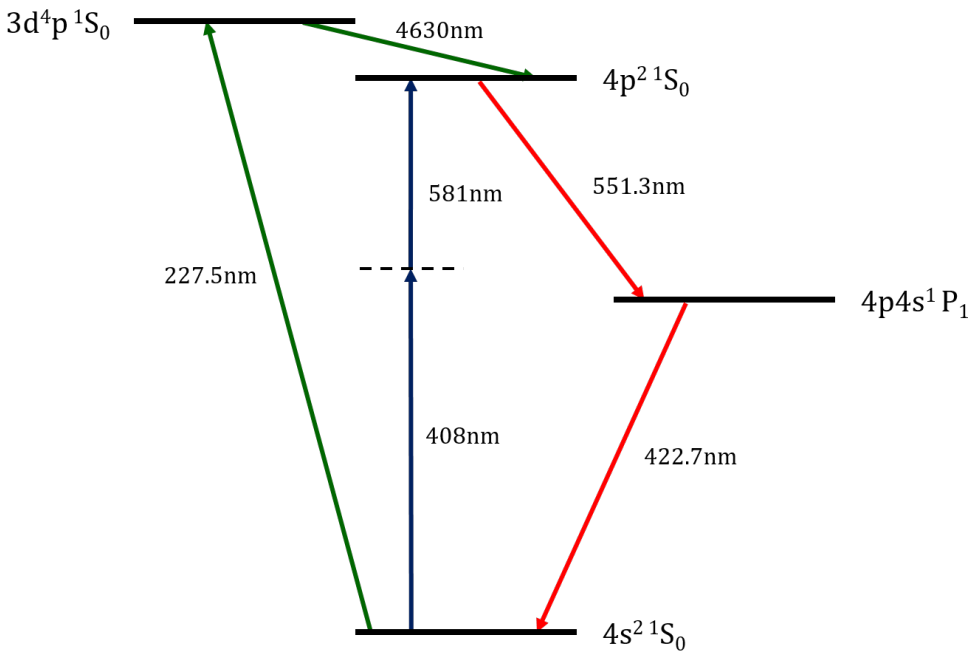


FIG. 9: Partial energy level diagram of atomic calcium. The two cascade transitions (red) are shown, along with the pumping transitions used by Freedman and Clauser[13] (green) and Aspect et. al.[16] (blue).

The atomic cascade platform proved relatively easy to implement, and the use of photons

traveling at nearly the speed of light made it easier to close the communication loophole. Another key advantage of atomic cascades were high-quality analyzers; efficient linear polarizers were readily available for visible light. Unfortunately, the single-photon detectors of the era were photomultipliers with detection efficiencies of $10 - 30\%$, resulting in $\eta \ll 2/3$. While there was hope of solving the detection problem with technological advancement, atomic cascade Bell tests possessed a more fundamental flaw. Under the definition of efficiency given in (10), to achieve high efficiency it is necessary (but not sufficient) for the Bell pair partner of any particle which enters one analyzer to have a high probability of entering the other analyzer. After all, a Bell pair is a trial if at least one particle enters an analyzer, but can only be a successful trial if both particles enter and pass through opposite analyzers. However, the apparatus both only admit particles from some limited solid angle (either because of a physical aperture or an effective aperture imposed by the rest of the apparatus). Therefore, given the direction of emission of one particle in the pair, it must be possible to predict the direction of the other particle with high probability: that is, the paired emissions must have high angular correlation to be useful for an efficient Bell test. However, the atom in an atomic cascade is itself a massive third body that can absorb a great deal of momentum from the Bell pair, and thus the two photons need not be emitted with any particular relative orientation. In fact, the photons' angular correlation is too low to achieve the desired η even if the apparatus itself is perfectly efficient.[6] The efficiency loophole can never be closed in any atomic cascade experiment - only a two-body source will suffice.

III.2. The First Atomic Cascade Bell Test

The first atomic cascade Bell test was performed by Freedman and Clauser a mere eight years after Bell's original paper. As shown in Figure 10, the Freedman implementation was a one-channel experiment with a simple source. A beam of calcium atoms was struck with light filtered to wavelengths close to the energy gap between the ground state and an auxiliary pumping state, which quickly decayed to the top of the cascade (see Figure 9). A filtering "pile of plates" polarizer consisting of a series of thin glass plates analyzed the photons. Each plate, oriented at Brewster's angle with respect to the incoming photons, optimally transmitted photons with linear polarization in the plane of the tilt and reflected

those with orthogonal polarization; additional plates further improved the filter. Freedman and Clauser achieved analyzer efficiencies in excess of 95%; later experiments boosted this efficiency with additional plates.[14] Photons that successfully passed through the analyzer were detected by a photodetector. Freedman measured the efficiencies of their commercially available detectors to be between 13% and 28%, yielding $\eta < 0.15\%$. Moreover, the detectors' high dark rates and the relatively weak source led to substantial background ($\zeta \approx 3\%$). The photodetector counts were sent to a shared coincidence circuit which recorded the counts at A, the counts at B, and the coincident counts (counts at A occurring within some small temporal window of a count at B).

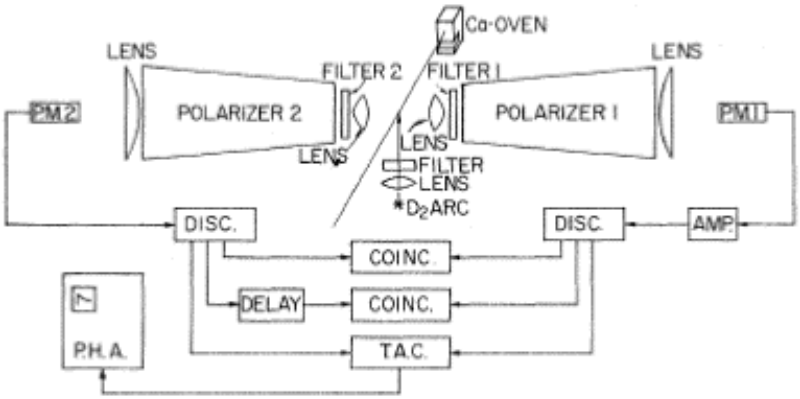


FIG. 10: Design of Freedman and Clauser one-channel calcium cascade Bell test. A beam of calcium atoms from an oven intersected a filtered light beam from a deuterium arc lamp. Photons emitted from the resulting cascade were sent through pile-of-plates polarizers to photomultipliers (PM) for detection; detections from A and B were both recorded separately (DISC.) and fed into a coincidence circuit (COINC.). Figure from Freedman and Clauser.[13]

The Freedman experiment, as well as subsequent similar experiments,[7,14] made no attempt to close the freedom-of-choice loophole or the setting-selection portion of the communication loophole. Setting selection was by manual rotation of the polarizer, and so could occur only infrequently. Also, spacelike separation of the measurement events themselves was not clearly established, leaving the communication loophole fully open. While the photons' arrivals at the apparatus were spacelike-separated, the light's coherence length was on the order of d_{AB}, and so the authors could not claim that the entirety of A's measurement interaction was spacelike-separated from the entirety of B's. The situation with regard to

the efficiency loophole was even more bleak: the three-body problem blocked closure of the efficiency loophole even if massive improvements in detector efficiency were made.

III.3. Improvements to Atomic Cascade Experiments

Although the polarization of atomic cascade photons was already known to be inadequate for a loophole-free test, a series of important advances in source, separation, analyzer, and setting selection design throughout the late 1970s and early 1980s, primarily by Aspect et. al., paved the way for future loophole-free tests.

The earliest modification, a more powerful and controllable laser-pumped source, was developed by Fry and Thompson[15] and later modified by Aspect et. al.[10,16,17] In the Aspect scheme (see Figure 9), two antiparallel laser beams were focused on the calcium beam. The combined energy of the lasers was actively tuned via a feedback loop to the transition from the ground state to the top of the cascade; because the transition was driven by two-photon absorption, a nonlinear process, the pumping was highly sensitive to variations in energy. This sensitivity allowed Aspect et. al. to selectively pump one hyperfine transition, eliminating hyperfine splitting, which tended to reduce polarization correlation and prevent production of the desired polarization singlet state. Moreover, because of the laser's high intensity, small line width, and narrow beam width, Aspect's smaller active region had much greater Bell pair emission rates than Freedman's, reducing the time required to acquire data by over two orders of magnitude (from 80 hours[13] to a few minutes), and reducing the background by a similar factor. This more productive source allowed Aspect et. al. to tolerate increased photon losses, and thus enabled them to increase d_{AB} to 13m of free space, about eight times the coherence length of the photons, finally closing the communication loophole for measurements. Aspect et. al.[17] also demonstrated an efficient two-channel analyzer for atomic cascade Bell tests, replacing filtering "pile-of-plates" polarizers with cubic beam-spliting polarizers of only slightly lower efficiency ($\approx 92\%$).

The most notable contribution of Aspect et. al. to loophole-free Bell tests was the group's time-varying analyzers, which were the first serious attempt to address the settings selection portion of the communication loophole. Although this was the first loophole recognized by Bell in 1964, it took until 1976 for Aspect et. al. to propose,[9] and until 1982 for them to implement,[10] a plausible rapid setting selection mechanism. Finding an optical

polarizer whose axis of polarization could be changed in nanoseconds was extremely difficult: clearly neither pile-of-plates nor cubic polarizers, whose settings were changed by manually rotating the entire polarizer assembly, could be adapted. Linear electro-optical modulators (EOMs), such as Pockels cells, which rotate a photon's linear polarization through an angle proportional to an applied voltage, were a plausible option. These cells are capable of state changes on the order of a nanosecond; in combination with a fixed linear polarizer they could provide a rapidly switchable linear polarizer. However, contemporary EOMs would quickly overheat and fail with rapid repeated switching.

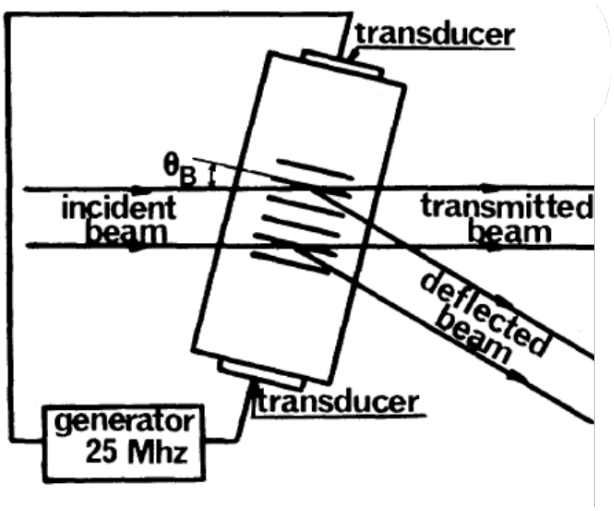


FIG. 11: The optical switch used in Aspect et. al.'s time-varying analyzers. Photons in the incident beam interact with standing waves in water driven by a pair of transducers, and are alternately transmitted or deflected by $2\theta_B$ depending on whether the standing wave is at its minimum or maximum respectively. Here θ_B is the Bragg angle for the medium. Figure from Aspect et. al.[10]

In the face of these limitations, Aspect et. al. insightfully separated switching and polarization entirely by constructing "time-varying analyzers" featuring two fixed polarizers preceded by a two-channel optical switch (Figure 11). The Aspect acoustic-optical switch was a tank of water with standing acoustic waves driven at a fixed frequency ($\approx$ 25MHz) by a pair of electro-acoustic transducers. The tank was placed in the path of a beam of incoming photons at the Bragg angle $\theta_B = 0.005$rad. Photons arriving at a standing wave minimum were directly transmitted, while those arriving at a standing wave maximum were deflected at $2\theta_B$. The narrowly separated outgoing beams were each routed into a fixed filtering linear polarizer, each followed by a detector. This combination of the optical switch, two

fixed linear polarizers, and two detectors is equivalent to a single-channel variable analyzer followed by a single detector. Two of these time-varying apparatus, with slightly different switching frequencies, were required for Aspect et. al.'s Bell test.

The Aspect time-varying analyzers were a major step forward, but their quasi-periodic setting selection left the communication loophole unclosed. The settings did change approximately every 10ns, well below $d_{AB} = 20$ns, so that the setting change at A was spacelike-separated from the measurement at B. However, the settings were not freely selected at the time of each change but instead effectively selected at the moment the predictable transducers were turned on. Therefore, locality alone could not guarantee the independence of the setting selections. Although Aspect et. al. were unable to perform the truly random measurement setting selection required to close the communication loophole, their implementation of rapid switching was the first substantial advance toward achieving this goal.

IV. OPTICAL FIBER AND DOWN CONVERSION: CLOSING LOOPHOLES INDIVIDUALLY

By the mid 1980s, the inherent limitations of atomic cascades as a Bell test platform were becoming ever more apparent: while closing the communication loophole seemed within reach, the efficiency loophole remained firmly open, and was destined to remain that way due to the three-body problem. Throughout the 1990s and 2000s, experimental Bell tests remained primarily photonic, but ventured far beyond the limitations of atomic transitions and free space transmission. New technologies allowed the closure of of all three loopholes, but not in the same Bell test.

IV.1. The Introduction of SPDC

A solution to the three-body problem in photonic Bell tests arrived in 1988 in the form of spontaneous parametric down-conversion (SPDC). In a SPDC interaction, shown in Figure 12, a single photon is split into two photons in a way that conserves energy and phase: thus the wavenumbers k_i and the associated phases ω_i obey the relationships

$$k_0 = k_1 + k_2 \tag{14}$$

$$\omega_0 = \omega_1 + \omega_2. \tag{15}$$

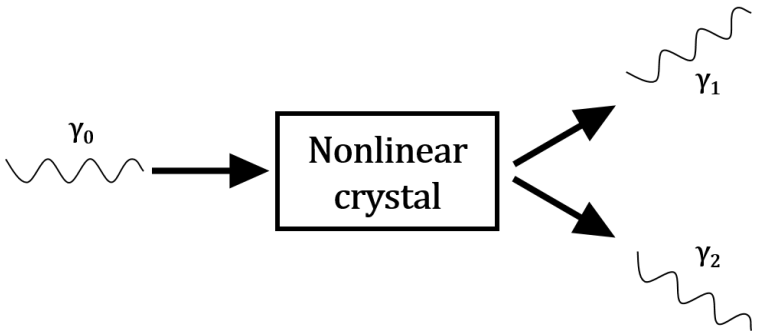


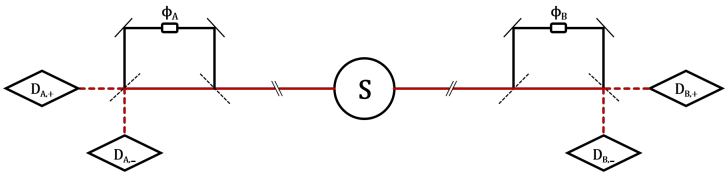
FIG. 12: Schematic of a spontaneous parametric down-conversion (SPDC) interaction. A photon γ_0 incident on a nonlinear crystal has a small probability of splitting into two photons γ_1 and γ_2 in a process that conserves energy and phase. The outgoing photons are emitted at equal angles with respect to the axis of the incident photon.

The ratio of ω_1 to ω_2 and k_1 to k_2 can be controlled by altering the crystal geometry and temperature. The case $k_1 = k_2$ is of particular use in Bell experiments because then SDPC generates two indistinguishable photons, each with half the energy of the incident photon. While sending pulses from a laser into a nonlinear crystal will automatically produce indistinguishable pairs of photons in a two-body system, entangling the pairs for use in a Bell test requires additional manipulation with interferometers. Shih and Alley[18] were the first to demonstrate the value of SPDC sources for Bell tests; their Bell tests incorporating SPDC sources could not address the communication loophole, but did illustrate the feasibility of directing and controlling SPDC sources.

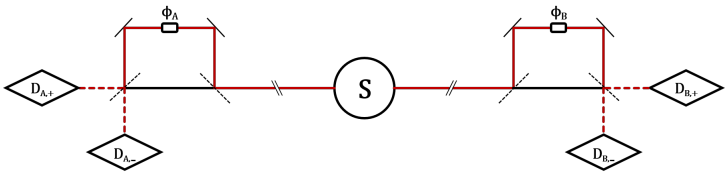
IV.2. Adding Fiber to Close Communication

Throughout the 1990s, single-mode optical fibers, which can transmit individual photons of visible wavelengths over significant distances with both low losses and high state fidelity, dramatically increased feasible d_{AB} in SPDC-sourced experimental Bell tests from meters to kilometers, culminating in the closure of the communication loophole.

Tapster et. al. were among the first to couple Bell test photons into optical fibers for transmission.[19] Their 1994 Bell test maintained $d_{AB} = 4.3$km; a similar test by Tittel et. al. four years later[20] spanned over 10km. The platform for both of these tests, theoretically developed by Franson[21] and shown in Figure 13, is based on time-of-arrival measurements



(a) Paths taken by photons in a short-short pair (each photon takes one of the two dotted paths to a detector with equal probability).



(b) Paths taken by photons in a long-long pair.

FIG. 13: Schematics of an idealized Franson-type Bell test, with indistinguishable pairs of photon paths (short-short (a) and long-long (b)) highlighted. Two photons are emitted from a source S, with their time of emission known only to precision τ_1, but the time between their emissions limited to $\tau_2 \ll \tau_1$. One is sent to apparatus A (left) and the other to apparatus B (right). The apparatus each consist of an unbalanced interferometer with the arms separated by some phase difference ϕ and time difference τ_3 such that $\tau_2 \ll \tau_3 \ll \tau_1$. 50-50 beam-splitters (dashed black lines) ensure each arm is equally likely for a given photon and the detector (D_+ or D_-) reached is independent of the arm traversed. In the case that detections at A and B are separated by time $\ll \tau_3$ it is clear that both photons traversed the same arms of their respective photodetectors, but because the original emission time is known only to τ_1, either arm is possible; the photon pair is in a superposition of long-long and short-short path combinations and is thus entangled. The introduction of a phase difference $\phi_A - \phi_B$ between the long arms offsets the arrival times at A and B. If the difference is greater than τ_2, the long-long combination no longer results in effectively coincident arrival times. The coincidence rate is thus predicted by quantum mechanics to exhibit nonlocal dependence on $\phi_A - \phi_B$ (but not on ϕ_A or ϕ_B); this prediction violates the CH inequality for the system, which turns out to have nearly identical mathematical structure to the traditional photon polarization system.

of emission-time entangled photons passed through unbalanced interferometers that have different relative phases but are otherwise identical. While the Tapster et. al. experiment, whose settings were adjusted by manual insertion of phase plates into the interferometers,

had limited value for closing the communication loophole, Tittel et. al. implemented a time-varying analyzer consisting of two interferometers with different fixed settings, with selection between them made passively by a 50-50 beam splitter. This implicitly relied on the unverifiable assumption that such a beam splitter randomly assigns a direction to each photon; it is possible to construct LRTs in which the splitter behavior instead depends on λ. Still, Tittel et. al.'s record d_{AB}, achieved while simultaneously implementing varying measurement settings, was a substantial advance.

The communication loophole was ultimately closed by a 1998 experiment by Weihs et. al.[22] that used optical fibers but returned to the platform of photon polarization states. SPDC pair production followed by an interferometer at S produced pairs of 702nm photons in the polarization singlet state. The photons were sent via optical fibers to A and B, which were separated by 400m, or 1.3μs, on the Innsbruck University campus. To maintain non-communication, Weihs et. al. selected their settings, the axes of the linear polarization measurements, using a high-speed RNG based on passive beam-splitting of photons emitted by an LED, reducing the interval from the beginning of selection to the end of measurement below this 1.3μs limit. The analyzer itself was EOM-based; due to great technological advances from the 1970s, this analyzer operated for extended periods at a switching rate of up to 30MHz. Weihs et. al. estimated that the measurement setting selection interval was 25ns long, while the measurement itself from the photon entering the apparatus until digital tagging of the outcome took approximately 75ns, so the entire measurement process took only $\approx$ 100ns, much less than the available window of 1.3μs.

To truly close the communication loophole, Weihs et. al. also took new steps to separate the recording of coincidence events at A and B. Previous Bell tests routed signals from each detector back to a central coincidence circuit that tallied coincidence detections for each detector combination in real time. However, this left open the incredible possibility that the coincidence circuits themselves were influenced by subluminally transmitted information about a, b, and λ. Thus Weihs et. al. completely separated data recording at A and B: at each apparatus. Measurement outcomes were digitally tagged with the time and current setting and recorded, and only later compared by time stamp to tally coincident counts. This approach requires precise and pre-synchronized timing at A and B. Weihs et. al. deployed rubidium atomic clocks to keep the timing discrepancies below 20ns.

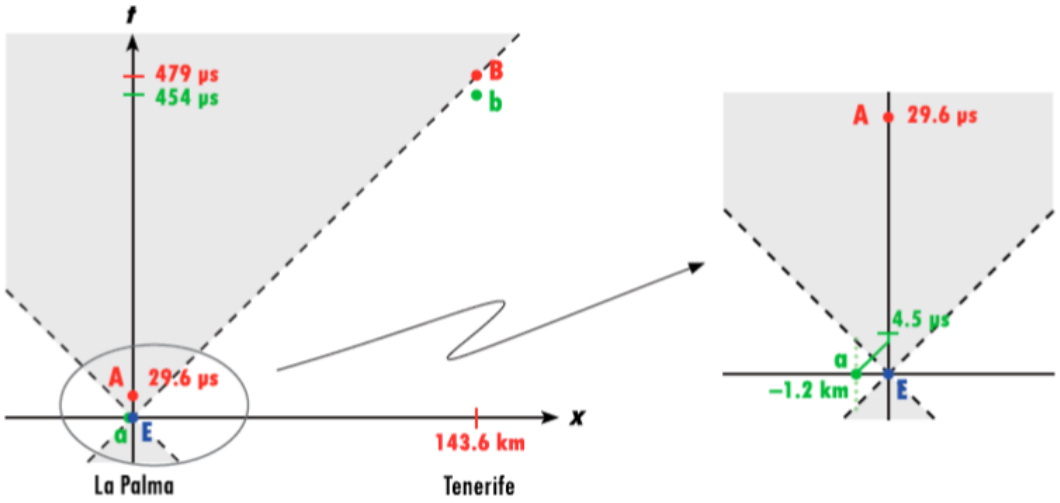
The Innsbruck experiment firmly closed the communication loophole,[23] but failed to

address freedom-of-choice: the measurement selection process began only 100ns before the measurement, long after subluminal information about λ could have arrived from S. The Innsbruck experiment was also a step backward on efficiency, with $\eta \approx 5\%$, far from closing the loophole, although the efficiency loophole was closed around the same time using a completely different experimental platform.[24]

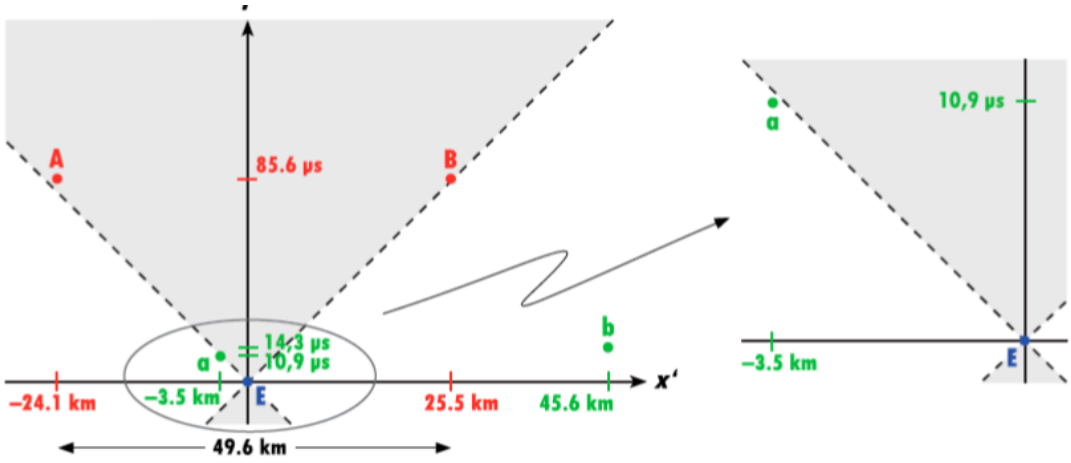
IV.3. Addressing Freedom-of-Choice

The first experiment to seriously address the freedom-of-choice loophole was Scheidl et. al.'s experiment on SPDC-generated polarization-entangled EOM-analyzed photon pairs at the Canary Islands astronomical facilities.[11] The relevant facilities are a transmitter for optical photons on one island and an optical ground station telescope on another which together form a 144km free-space channel for visible-wavelength photons. This free-space channel provided immense spacial separation, but because there was no second similar free-space channel, the resulting Bell test was highly asymmetrical, as diagrammed in Figure 14.

To produce the necessary spacelike separations between events, Scheidl et. al. resorted to several clever tactics. The experimenters inserted a 6km coil of optical fiber between S and A to delay A's measurement and spacelike-separate it from the measurement at B. For more convenient analysis, the delay length was chosen so that the test was symmetric in another reference frame (see Figure 14). A's RNG was placed over a kilometer away from A to spacelike separate the setting selection and emission events, and the settings were transmitted to A by radio. At both A and B, electronic delays in the setting processing ensured only random bits of sufficient "age" were used, so that the physical processes generating the bits were spacelike-separated from both S and the other measurement. Scheidl et. al. found $S = 2.37 \pm 0.02$, a strong violation of the CHSH inequality, under only the supplemental assumption of fair sampling, required because $\eta = 0.00032\%$ due to the free-space channel.



(a) Minkowski diagram in the rest frame of the apparatus and source.



(b) Minkowski diagram in the inertial reference frame in which the measurements A and B are simultaneous; the experiment was adjusted so that the setting selection events a and b are approximately simultaneous in this frame as well.

FIG. 14: Minkowski diagrams of the highly asymmetric Scheidl et. al. Bell test; inserts to the right show magnified views. A and S are collocated (separation of approximately 10m) at the transmitter on La Palma, while B is at the telescope on Tenerife (four orders of magnitude greater separation). The emission event is marked with E; the gray region is the light cone of this emission event. Two different reference frames are presented; because spacelike separation is a property independent of inertial reference frame, Scheidl could use either for their analysis. Note that the measurement setting selections a and b lie outside of E's light cone in both diagrams and cannot be influenced by the emission event under the assumption of locality. Figures from Scheidl et. al.[11]

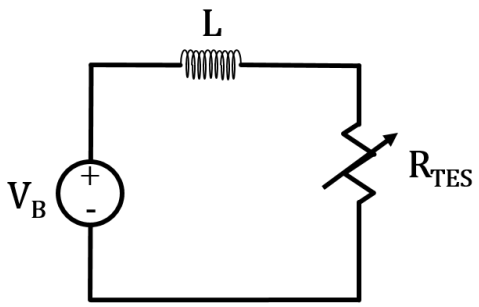
V. A LOOPHOLE-FREE YEAR

After 50 years, the quest for a loophole-free Bell test culminated in 2015, in which independent research teams at Delft University of Technology,[25] the University of Vienna,[26] and the National Institute of Standards and Technology (NIST)[27] conducted three successful loophole-free tests published in the span of a few weeks. All three teams used pairs of photons transmitted via single-mode optical fibers, as well as the same rapid physical RNGs, based on the phase of spontaneous photon emissions from a laser diode, designed, produced, and analyzed by Abellán et. al.[28] However, the teams' tactics for overcoming the efficiency loophole blocking prior photon polarization tests differed. The Vienna and NIST teams stuck with the conventional experimental platform and simply dramatically improved the detection efficiency using more advanced detection methods. In contrast, the Delft team implemented an event-ready scheme in which the photons' entanglement was transfered to the spins of diamond NV centers, and only if this transfer to the more efficiently detectable spins was successful was a trial of the Bell test conducted.

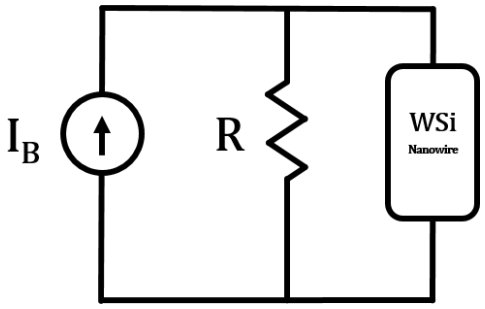
V.1. Vienna and NIST: Superconducting Advances in Detection

Both the Vienna and NIST experiments followed Weihs et. al. in methodology, with S emitting polarization-entangled SPDC-generated visible-wavelength Bell pairs, but reduced the scale somewhat (NIST to $d_{AB} = 185$m and Vienna to $d_{AB} = 58$m), reducing photon losses in transmission. This decreased spacial separation shortened the interval available for space-like separated measurements and setting selections. To compensate, the Vienna and NIST teams relied on accurate and independent time-tagging of measurements as well as the increased speed of the Abellàn RNG, which produces a fresh random bit every 5ns. Both Vienna and NIST experimenters applied single-channel analyzers consisting of a rapidly switchable EOM followed by a filtering polarizer and detected violations of the CH inequality.

The two teams differed in their choices of high-efficiency single-photon detectors. The Vienna team's detectors were transition-edge sensors (TESs) coupled to a series of amplifiers, the first of which was a superconducting quantum interference device (SQUID). A TES, diagrammed in Figure 15a, is a temperature sensor that exploits the sharp increase in resistance in a superconducting film as it transitions from superconducting to regular



(a) Circuit diagram of a model transition edge sensor (TES). A thin tungsten film forms the variable resistor R_{TES}, while the inductor L is used for readout by inductive coupling.



(b) Circuit diagram of a model superconducting nanowire single-photon detector. The detector is current biased with a fixed current I_B; the nanowire is tungsten silicide.

FIG. 15: Model circuit diagrams for the single-photon detectors used in the Vienna and NIST experiments, respectively.

behavior. The absorption of a single photon by the film, which is cooled to just below its critical temperature (for Vienna's tungsten films, approximately 150mK), can provide sufficient energy to trigger the phase transition.[26,29] The resulting decrease in current through the voltage-biased TES results in a spike in the inductance of an inductor placed in series with the film; this inductor is inductively coupled to the SQUID, which amplifies the signal to a level at which classical electronics can further amplify and process it. Encasing the conducting film in a cavity resonant at the target wavelength boosts the absorption probability for a photon entering the TES, resulting in detection efficiencies in excess of 95%. Even better for the Vienna team, it is possible to determine the wavelength of the photon absorbed from the TES output using the relationship

$$E_{photon} = V_B \int |I_S - I(t)|dt, \tag{16}$$

where V_B is the bias voltage, I_S is the current when the TES is in its superconducting state, and $I(t)$ is the actual current. The Vienna team applied this information to screen out detections of dark count photons with other wavelengths, reducing the background. Earlier generations of TES/SQUIDs were insufficient for the Vienna test due to recovery times (time required for the film to cool to the critical temperature and be available for another detection) on the order of 5μs. However, recent advances lowered the recovery time well

below the 1μs per measurement cycle in the Vienna experiment. The Vienna experimenters thus were able to reach $\eta \approx 75\%$, sufficient for a violation of the CH inequality with the appropriate Eberhard state.

The NIST team deployed a different detector that also exploits the superconducting transition, the superconducting nanowire single-photon detector (SNSPD). A typical visible-light or near-infrared SNSPD, diagrammed in Figure 15b, consists of a nanowire cooled well below its critical temperature and biased with a current just below its critical current. An incident photon absorbed by the nanowire will briefly reduce the effective critical current in a small region of the wire, triggering a phase transition from superconducting to regular behavior. The resulting spike in resistance in the nanowire increases current through a parallel resistor; this current is amplified to produce a signal. By 2015, SNSPDs had achieved detection efficiencies of over 91%, equivalent to those of TES/SQUIDs, but with the advantage that SNSPDs can operate at higher temperatures than tungsten TESs (up to nearly 2K) and have shorter recovery times, on the order of 40ns.[30] Unfortunately, SNSPD signals do not provide information about the energy of the incident photon, and so the NIST team was unable to screen out dark counts. However, NIST achieved $\eta \approx 75\%$ as well.

V.2. Delft: NV Centers and Event-Ready Detection

The Delft team took a very different approach to closing the efficiency loophole, shifting measurements to observables that are easier to detect using an event-ready design. In an event-ready setup, shown in Figure 16, the particles are first prepared, then checked to see if they are in the appropriate state. If so, the experiment is deemed to be "event-ready", and only then are measurements conducted. Such a setup can avoid the efficiency loophole even if most pairs fail to trigger the event-ready signal; only the pairs triggering the signal are considered Bell trials. This argument only breaks down if the measurement settings a and b influence the event-readiness of pairs in an λ-dependent way; event-ready setups typically avoid this by spacelike-separating the setting selection from the generation of the event-ready signal.

The Delft experimenters implemented their event-ready scheme using nitrogen vacancy (NV) centers. A NV center (shown in Figure 17) occurs in diamond with nitrogen impurities; the substitution of a nitrogen and a vacancy for a pair of carbon atoms results in a

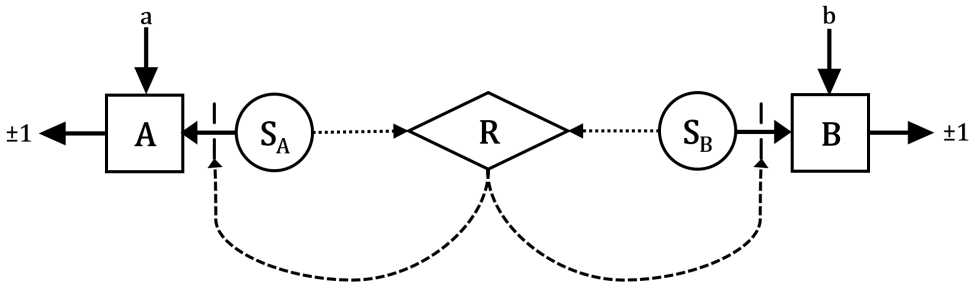


FIG. 16: Schematic of an event-ready Bell test. A particle is prepared at each of S_A and S_B, and information about the preparation is sent to the event-ready detector R. If the pair is in an appropriate state for measurement, R sends a "ready" signal back to both apparatus, triggering measurements at both A and B. If no ready signal is received, a new pair of particles is prepared. Effectively, λ is actually determined at R, not at the sources. Measurement settings a and b are accepted by A and B shortly before they would receive the ready signal, and simply ignored if no ready signal is received.

pair of electrons which have spin-dependent energy levels distinct from the other electronic energy levels of diamond that can be independently manipulated. The spins of NV centers have high fidelity and detection efficiency; however, they are fixed in place and have a low interaction probability, which makes preparing a pair of entangled NV centers a challenge. In a conventional Bell test, these restrictions would be debilitating; however, in an event-ready scheme, even a low probability of successfully preparing a Bell pair is sufficient as long as it is possible to detect when a pair is ready for measurement. In the Delft test, the spins of two fixed NV centers at A and B were entangled via an entanglement-swapping scheme. The spin state of each NV center was entangled with a single photon, which was sent to R via optical fiber. Measurements on this pair of photons after they are interfered at R can determine if the NV center spins are in the singlet state, which triggers the event-ready signal. In the Delft experiment, the event-ready signal was issued only on $6.4 \times 10^{-7}\%$ of attempts, which did not affect the efficiency of the Bell test but did limit the total number of trials.

When A and B received the event-ready signal from R, they performed measurements of the spin components of the NV centers along a randomly selected axis. The measurement apparatus measured the spin along a fixed axis of the NV center via spin-dependent fluorescence. When subjected to microwave frequencies tuned to a transition accessible only from the $|\uparrow\rangle$ state, the $|\uparrow\rangle$ state fluorescences, while $|\downarrow\rangle$ state does not. In the Delft experiment,

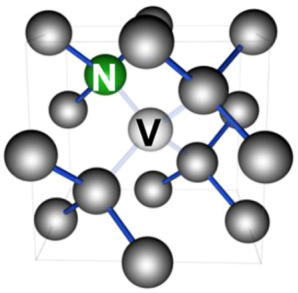


FIG. 17: The structure of a nitrogen vacancy (NV) center in diamond. A NV center consists a nitrogen (N) and vacancy (V) being substituted for two adjacent carbon atoms (dark gray). The Delft experiment contained two negatively charged nitrogen vacancy centers (NV^-); the extra electron creates a pair of free electrons in the vacancy. Figure from Batalov et. al.[31]

the resonant microwave signal was applied for 3.7μs, and thus many photons were emitted by the $|\uparrow\rangle$ state while none were emitted by the $|\downarrow\rangle$ state, making it easy to distinguish between the two even with inefficient photodetectors. To perform a measurement along a different axis, the experimenters applied a pulse to rotate the spin before performing the measurement. Because of the large separation $d_{AB} = 1.28$km made possible by the event-ready scheme, the Delft team had relatively little difficulty ensuring spacelike-separation of the key events. Because of their two-channel analysis and detection method, the Delft experimenters were able to analyze their results using the CHSH inequality.

V.3. Final Results

All three groups found statistically significant (p-value $p < 0.05$) violations of their respective Bell inequalities. Hensen et. al. reported that the Delft experiment was only able to conduct 245 trials using their event-ready entanglement swapping technique, but reached $p = 0.039$. Shalm et. al. found $p = 2.3 \times 10^{-7}$ after an extremely conservative statistical analysis of the NIST experimental results of approximately 175 million trials. Finally, Giustina et. al. reported $p = 3.74 \times 10^{-31}$ from analysis of 3.5 billion trials collected in only an hour at Vienna. Together, these results, derived with the absolute minimum of assumptions, close the communication, efficiency, and freedom-of-choice loopholes and offer nearly the strongest possible experimental disproof of local realism.

VI. CONCLUSION: THE UNCLOSABLE LOOPHOLE

Nevertheless, not even these "loophole-free" tests are truly free of loopholes, and neither will any future experimental Bell test be, no matter how technologically advanced. While the communication and efficiency loopholes can be completely closed, the assumptions underlying the freedom-of-choice loophole can be reduced but never completely eliminated. While locality implies that the pair emission at S does not have a direct physical influence on the measurement selections at A and B if these events are spacelike-separated (as they are in all of the "loophole-free" Bell tests), this does not actually imply the independence of a and b from λ. The past light cones of the emission and the selections certainly overlap, and any event E in the common past of these three events can have a physical influence all three without violating locality, as shown in Figure 18. To avoid the shadow of this common past, all experimental Bell tests must assume that the selections, the emissions, or both are independent of events in sufficiently far back in their own pasts: the three "loophole-free" tests assume in particular that the RNG outputs determining the settings selections are highly unpredictable. However, while there are advanced statistical tests to distinguish randomness from nonrandomness, there is no way to prove that an RNG produces random outputs from these outputs themselves,[28] and thus assumptions about the underlying physical processes must be made. While the designers of the RNG used in all three "loophole-free" tests used as their source a physical process widely believed to be highly unpredictable and conducted an incredibly conservative analysis of the RNG's predictability, their conclusions are not justified by local realism alone.

A commonly proposed solution to this quandary is to separate the selection and emission events by immense distances. For example, the selections could be based on the fluctuations of pulsars at opposite ends of the universe.[12] This could push earliest possible E billions of years into the past; recent experiments along these lines have already reached intervals of over 600 years.[32] However, no matter the scale of the Bell test, the freedom-of-choice loophole remains: a particularly pernicious local realistic theory could account for the outcomes of every future Bell test as a consequence of a single universal E, the big bang itself. Such superdeterministic theories are deeply undesirable because they are arbitrarily complex, with a large subset of physical events simply "hard-coded" into the theory to fool any and every Bell test. Thus almost all physicists reject sufficiently superdeterministic physical theories,

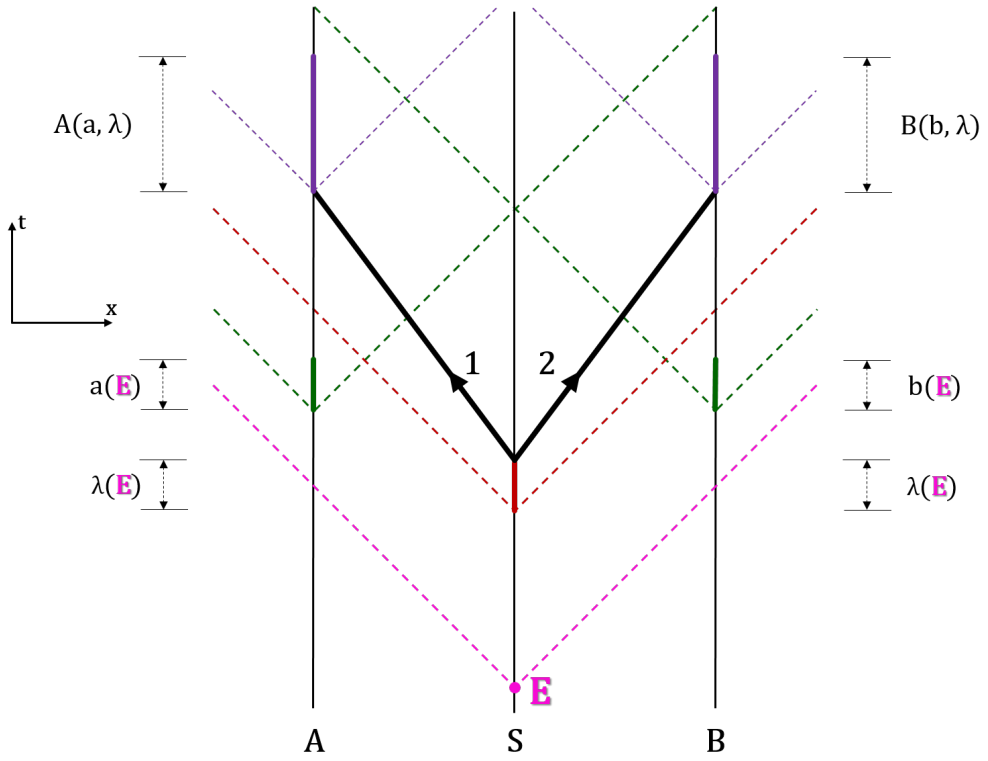


FIG. 18: Minkowski diagram of an ideal symmetric Bell test illustrating the remaining freedom-of-choice loophole. Diagram elements are as defined as in Figure 6. Here E is an arbitrary event in the common past of the setting selections and the Bell pair emission: all three events are in E's forward light cone. Under the assumption of locality, E can physically influence a, b, and λ, generating correlations among the three even though they are all pairwise spacelike-separated.

even those consistent with all experimental evidence, as lacking any useful conception of physical causality and predictability.

Viewed through this lens, the historical trajectory of experimental Bell tests has been to pare away at the vast space of conceivable local realistic theories by disproving subclasses corresponding to different supplemental assumptions one at a time, leaving a smaller and smaller set of increasingly superdeterministic theories behind. 2015's "loophole-free" tests marked the moment at which the only remaining local realistic theories were so deeply rooted in superdeterminism that the physics community could reject them as nonviable. While progress may continue on experimental Bell tests that are even more "loophole-free," Einstein's dream of a human-comprehensible local realistic theory of physics is finally over.

Appendix A: Derivation of the Bell Inequalities

Although the Bell inequalities are unobvious, their derivations from the principles of locality and realism are not particularly mathematically sophisticated. The following simplified form of Bell's derivation of the original inequality relies only on a series of manipulations of inequalities and absolute values across integrals, with the only use of physical principles found in the assumptions (2), (3), and (5). The CHSH and CH inequalities can be derived from (2) and (3) alone using similar techniques.

By the perfect anticorrelation assumption (5), for any measurement setting q

$$-1 = E(q,q) = \int_{\rho(\lambda)} A(q,\lambda)B(q,\lambda)d\lambda. \tag{A1}$$

However for any q and λ, $A(q,\lambda)$ and $B(q,\lambda)$ are ± 1, so the product $A(q,\lambda)B(q,\lambda) = \pm 1 \geq -1$ as well. If a function takes on only values ≥ -1, its integral can have value -1 only if the function has value -1 essentially everywhere (everywhere but a set of measure zero). Therefore, for any q, $A(q,\lambda)B(q,\lambda) = -1$ for all λ except a set with probability 0, and so $A(q,\lambda) = -B(q,\lambda)$ with probability 1.

Equipped with this equality and the fact that $A(a,\lambda)A(a,\lambda) = (A(a,\lambda))^2 = 1$ for all a and λ, consider the difference $E(q,r) - E(q,s)$ for arbitrary q, r, and s.

$$\begin{aligned} E(q,r) - E(q,s) = & \int_{\rho(\lambda)} \left(A(q,\lambda)B(r,\lambda) - A(q,\lambda)B(s,\lambda)\right) d\lambda & \text{(A2)} \\ = & -\int_{\rho(\lambda)} \left(A(q,\lambda)A(r,\lambda) - A(q,\lambda)A(s,\lambda)\right) d\lambda & \text{(A3)} \\ = & -\int_{\rho(\lambda)} A(q,\lambda)A(r,\lambda)[1 - A(r,\lambda)A(s,\lambda)]d\lambda & \text{(A4)} \end{aligned}$$

Taking absolute values of this equality and noting that $|A(q,\lambda)A(r,\lambda)| = |\pm -1| = 1$ for every λ, it follows that

$$\begin{aligned} |E(q,r) - E(q,s)| = & \left| \int_{\rho(\lambda)} A(q,\lambda)A(r,\lambda)[1 - A(q,\lambda)A(s,\lambda)]d\lambda \right| & \text{(A5)} \\ \leq & \int_{\rho(\lambda)} \Big|A(q,\lambda)A(r,\lambda)\Big|\Big|1 - A(r,\lambda)A(s,\lambda)\Big|d\lambda & \text{(A6)} \\ = & \int_{\rho(\lambda)} \left(1 - A(r,\lambda)A(s,\lambda)\right) d\lambda & \text{(A7)} \\ = & 1 + \int_{\rho(\lambda)} A(r,\lambda)B(s,\lambda)d\lambda & \text{(A8)} \end{aligned}$$

$$= 1 + E(r, s), \tag{A9}$$

which is precisely the original Bell inequality (6).

Appendix B: Other Experimental Platforms

While polarization of visible-light photons has been the most common platform for experimental Bell tests over the past fifty years, several other platforms have been proposed or tested. Two early candidates, polarization of gamma rays from positronium annihilation and spins of scattered protons, demonstrated some advantages over the contemporary atomic cascade experiments. However, both used indirect measurement techniques that required extensive supplement assumptions. More recently, the efficiency loophole was first closed by a Bell test conducted on trapped ions; however, this platform was inappropriate for closing other loopholes.

1. Positronium Annihilation

The earliest Bell test, the Kasday et. al. positronium annihilation experiment,[33] avoided the three-body source and inefficient detection of atomic cascades, but suffered from equally serious analyzer limitations.[7] The experimental design itself, diagrammed in Figure 19, was a modification of a twenty-year-old experiment by Wu and Shaknov on the polarization of gamma rays emitted in positronium annihilation. Positrons from a ^{64}Cu source struck and annihilated electrons in a Cu plate. The resulting pair of gamma rays is in the singlet polarization state by spin conservation. Kasday et. al. selected only gamma rays emitted along the axis of the incoming positrons; by conservation of momentum in this post-annihilation two-body system, Bell pairs with one photon emitted along this axis must have the other traveling precisely antiparallel, a theoretically perfect angular anticorrelation. There are not experimentally viable linear polarizers for high energy photons, so instead Kasday et. al. conducted an indirect polarization measurement with Compton polarimeters. These one-channel analyzers consisted of plastic scintillators that Compton-scattered gamma rays into the plane orthogonal to the primary axis of the experiment and lead collimators and detectors that selected and counted gamma rays at a given scattering angle.

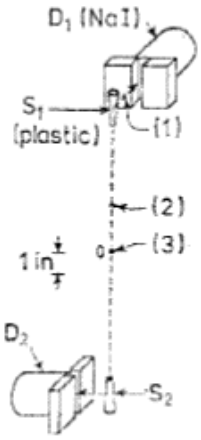


FIG. 19: Design of the Kasday et. al. positronium annihilation Bell test. Positrons from a ^{64}Cu source at (3) annihilated electrons in a separate copper plate (also at (3)). Resulting pairs of gamma rays traveling along the axis (2) scattered off of plastic scintillators S_i and, if scattered at the appropriate angle, passed through lead collimators and were detected by NaI detectors. Figure from Clauser and Shimony,[7] and in turn from Kasday et. al.[33]

There is not a one-to-one correspondence between Compton scattering angle and linear polarization; however, the Klein-Nishina formula

$$\frac{d\sigma}{d\Omega} \propto \frac{E}{E_0} + \frac{E_0}{E} - 2 + 4\cos^2\theta \tag{B1}$$

gives the differential cross section for a photon of initial energy E_0 and final energy E with initial and final linear polarization states separated by an angle θ. Kasday et. al. inverted this relationship to approximate initial polarizations from count rates at several scattering angles; however, this had serious flaws that left the efficiency loophole fundamentally unaddressable. Most obviously, (B1) is derived from quantum-mechanical principles; however, it can be verified empirically. The real problem is that both in quantum mechanics and in any conceivable experimental verification, the Klein-Nishina formula applies only definite elliptical polarization states, but the Bell pairs need not be in such a state or even a superposition of such states. The Bell pair's state was in fact chosen because it is not quantum-mechanically factorizable into polarization states. Thus (B1) cannot legitimately be applied to individual photons in the Bell pair; the pair must be analyzed as a single system, but the resulting quantum-mechanical predictions cannot be empirically verified. In summary, relationships between scattering angle and polarization that can be empiri-

cally verified are insufficient for indirect polarization measurements, and thus the Kasday et. al. experiment must assume the entire quantum-mechanical model of entangled photon Compton scattering.

2. Proton-Proton Scattering

While the Lamehi-Rachti and Mittig platform was closest to the Bell test's roots in spins of massive spin-1/2 particles, it also suffered from the limitations of indirect scattering measurements.[34] The setup, shown in Figure 20, produced pairs of spin-entangled protons via collision of a high-speed proton from a particle accelerator beam with an effectively stationary proton in a polyethylene target. By conservation of energy and momentum, the two now-indistinguishable protons scattered at $\pm\pi/4$ relative to the beam. Stern-Gerlach devices might seem the ideal analyzer for these proton spins, but in fact Stern-Gerlach devices are unsuitable for Bell tests on spin states of charged particles.[7] Instead, analogous to the Compton polarimeters of Kasday et. al., spin measurements were conducted indirectly via a second scattering event off of thin carbon foils.

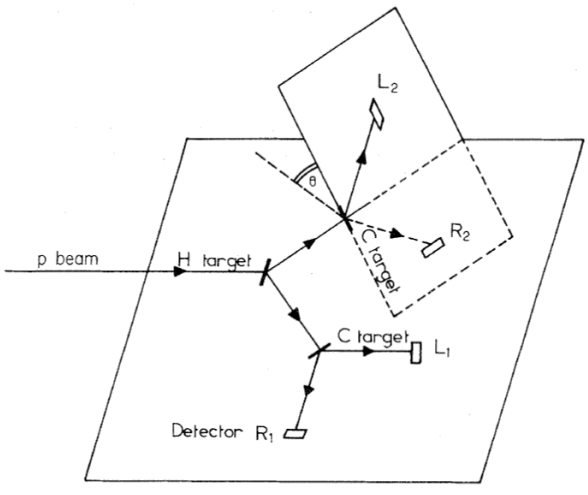


FIG. 20: Design of the Lamehi-Rachti and Mittig proton-proton scattering Bell test. A beam of protons stuck a hydrogen-rich target and scattered at $\pm\pi/2$. Both protons in each pair were then scattered off of carbon foils, with protons scattering to the left and right detected by silicon detectors L_i and R_i respectively. The relative angle θ of the planes of the detectors at the two foils was adjustable. Figure from Lamehi-Rachti and Mittig.[34]

Lamehi-Rachti and Mittig argued that this test's massive particles brought it much closer to closing the communication loophole for measurements than previous experiments. In photonic experiments, the coherence length (for atomic cascades, 0.300m; positronium annilhilation, 0.017m) was on the order of d_{AB}, requiring a supplemental assumption of photon localization to ensure spacelike separation of the measurements. By contrast, the Lamehi-Rachti protons had coherence lengths on the order of 10^{-15}m compared with $d_{AB} \approx 0.05$m. Nevertheless, Lamehi-Rachti and Mittig did not quite achieve spacelike separation of measurements due to the subluminal and non-antiparallel velocities of the protons. However, the platform relied on a modified form of (B1) to analyze proton spin indirectly from scattering angle, and thus required the same crippling supplemental assumptions.

REFERENCES

[1]A. Einstein, B. Podolsky, and N. Rosen, "Can Quantum-Mechanical Description of Physical Reality Be Considered Complete?" Physical Review **47**, 777–780 (1935).

This is the original paper well-known for introducing the EPR paradox. Although the argument is framed in terms of continuous observables (position and momentum) instead of a simpler discrete observable system, it is worth reading for the clear and explicit statement of the definitions and assumptions underlying the argument for the incompleteness of quantum mechanics.

[2]D. Bohm and Y. Aharonov, "Discussion of Experimental Proof for the Paradox of Einstein, Rosen, and Podolsky," Physical Review **108**, 1070–1076 (1957).

This theoretical paper presents a comprehensive reconstruction of the EPR authors' arguments in the more straightforward setting of discrete spin observables, discusses several contemporary possible interpretations or resolutions of the EPR paradox, and proposes a specific experiment to detect the paradox in the photons emitted by positronium annihilation.

[3]J. S. Bell, "On the problem of hidden variables in quantum mechanics," Reviews of Modern Physics **38**, 447–452 (1966).

This review dissects several prior attempts to demonstrate that no realistic theory can be consistent with quantum mechanics, most prominently John von Neumann's. Although the writing is highly technical and assumes prior knowledge of the arguments being analyzed, I found it useful in distinguishing earlier attempts from Bell's successful proof (which is not discussed here), and understanding how locality could emerge from the more arbitrary assumptions made by earlier theorists.

[4]J. S. Bell, "The Einstein Podolsky Rosen Paradox and Local HiddenVariables Theories," Physics **1**, 195–200 (1964).

This is Bell's original proof of the first Bell inequality and Bell's Theorem, and also the paper in which much of the terminology used in the Bell test literature was first adopted. Bell's central argument is fully comprehensible from this paper alone, although this version does not include stochastic local realistic theories and more succinct and effective presentations can be found elsewhere.

[5]J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test

Local Hidden-Variable Theories," Physical Review Letters **23**, 880–884 (1969).

This brief paper developed the necessary theoretical underpinnings for an experimental Bell test. This includes generalizing Bell's original inequality to an experimentally applicable form, the CHSH inequality; proposing an atomic cascade Bell test; and determining the quality of experimental setup necessary to achieve Bell inequality violations. While the paper is generally clear, its descriptions of the author's supplemental assumptions is quite abbreviated.

[6]J. F. Clauser and M. A. Horne, "Experimental consequences of objective local theories," Physical Review D **10**, 526–535 (1974).

This paper reestablishes the theoretical structure required for experimental Bell tests in the context of not just deterministic but also stochastic local realistic theories. Although the value and readability of the paper is somewhat limited by its adoption of novel and ultimately unsuccessful terminology, the authors are careful and explicit in their assumptions and derivations, among other things being the first to note the freedom-of-choice loophole (although not by that name). The authors ultimately apply their primary result, the CH inequality, to analyze the experimental results of Freedman and Clauser.

[7]J. F. Clauser and A. B. Shimony, "Bell's theorem. Experimental tests and implications," Reports on Progress in Physics **41**, 1881–1927 (1978).

Although dated, this review is a key resource. It covers all progress in the field up until 1978, both theoretical and experimental, beginning with the original EPR paper. After an exhaustive comparison of every known proof of Bell's Theorem, the paper provides first a general discussion of the necessary conditions for an experimental Bell test, and then detailed descriptions and evaluations of every significant experimental Bell test previously conducted. In certain situations, it offers more details of experimental design than the original papers themselves.

[8]P. H. Eberhard, "Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment," Physical Review A **47**, R747–R750 (1993).

This paper proves and presents theoretical and numerical models describing the conditions under which a Bell test with imperfect efficiency and nonzero background is capable of producing violations of the Bell inequalities without opening the efficiency loophole. While it is possible to follow the individual steps of the author's analysis, they often appear unmotivated, and as a result the author's methods as a whole are opaque. Also, although

its results are compelling, the paper does not do enough to distinguish theoretical and numerical conclusions.

[9]A. Aspect, "Proposed experiment to test the nonseparability of quantum mechanics," Physical Review D **14**, 1944–1951 (1976).

This paper proposes a Bell test (later conducted) in which rapid switching of measurement settings is effectively achieved with an optical switch preceding the measurement apparatus. The author carefully describes both the supplemental assumptions that this method eliminates and those that remain, and re-derives the CH inequality from first principles in the process. The author draws distinctions in terminology that were not ultimately adopted by future researchers in the field, which complicates the interpretation.

[10]A. Aspect, J. Dalibard, and G. Roger, "Experimental test of Bell's inequalities using time-varying analyzers," Physical Review Letters **49**, 1804–1807 (1982).

This paper presents the results of an atomic cascade Bell test using time-varying analyzers. The paper clearly but very concisely addresses the novel aspects of the experimental design and the results achieved.

[11]T. Scheidl, R. Ursin, J. Kofler, S. Ramelow, X.-S. Ma, T. Herbst, L. Ratschbacher, A. Fedrizzi, N. K. Langford, T. Jennewein, and A. Zeilinger, "Violation of local realism with freedom of choice," Proceedings of the National Academy of Sciences **107**, 19708–19713 (2010).

In addition to describing a Bell test that simultaneously closes the locality and freedom-of-choice loopholes, this paper addresses the limits on any Bell test's ability to close the freedom-of-choice loophole in an accessible manner. The paper's presentation of experimental methods is, by contrast, undesirably abbreviated.

[12]J.-A. Larsson, "Loopholes in Bell inequality tests of local realism," Journal of Physics A: Mathematical and Theoretical **47**, 424003 (2014).

This catalog of possible loopholes in experimental Bell tests extended beyond the scope of my research, but was useful for its in-depth discussion of the primary loopholes and the nature of superdeterminism.

[13]S. J. Freedman and J. F. Clauser, "Experimental Test of Local Hidden-Variables Theories," Physical Review Letters **28**, 938 (1972).

This is the original report of the first atomic cascade experimental Bell test. Although short, it is clear and explicit in its assumptions and methods.

[14]J. F. Clauser, "Experimental investigation of a polarization correlation anomaly," Physical Review Letters **36**, 1223–1226 (1976).

A brief report on attempts to replicate an earlier Bell test by Holt that produced results consistent with local realism and contrary to quantum mechanics. As a result of its purpose, this paper provides additional details about the experimental setup and potential sources of experimental error in atomic cascade Bell tests.

[15]E. S. Fry and R. C. Thompson, "Experimental test of local hidden-variable theories," Physical Review Letters **37**, 465–468 (1976).

This brief paper reports on an improved atomic cascade experiment using a cascade in atomic mercury. The paper clearly documents its advances over prior experiments, but the experimental design and analysis is best understood in the context of other more complete experimental papers that preceded it.

[16]A. Aspect, P. Grangier, and G. Roger, "Experimental tests of realistic local theories via Bell's theorem," Physical Review Letters **47**, 460–463 (1981).

This paper introduces the authors' basic experimental design, and in particular the novel source design, while presenting incrementally improved experimental violations of the Bell inequalities. This paper is useful in understanding future, more focused reports from the authors.

[17]A. Aspect, P. Grangier, and G. Roger, "Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities," Physical Review Letters **49**, 91–94 (1982).

This paper reports on the results of the first atomic cascade Bell test to incorporate two-channel analyzers; this represents a modification of the authors' prior design and is best understood in the context of their prior work.

[18]Y. H. Shih and C. O. Alley, "New Type of Einstein-Podolsky-Rosen-Bohm Experiment Using Pairs of Light Quanta Produced by Optical Parametric Down Conversion," Physical Review Letters **61**, 2921–2924 (1988).

This report on the first Bell test using spontaneous parametric down-conversion as a source clearly describes the setup and the derivation of the authors' theoretical description of it. Given that the Bell tests described are highly unorthodox in design (for example, a single beam splitter serves simultaneously as both analyzers and the source), this clarity is essential.

[19]P. R. Tapster, J. G. Rarity, and P. C. M. Owens, "Violation of Bell's Inequality over 4 km of Optical Fiber," Physical Review Letters **73**, 1923–1926 (1994).

This is the report of the first Franson-type experimental Bell test. The paper's descriptions of both the theory underlying the platform and the details of the experimental setup were frequently dense and confusing, and generally assumed a background in Franson interferometers.

[20]W. Tittel, J. Brendel, H. Zbinden, and N. Gisin, "Violation of Bell Inequalities by Photons More Than 10 km Apart," Physical Review Letters **81**, 3563–3566 (1998).

This is another key experimental paper based on the Franson platform; this experiment also implemented time-varying analyzers in this platform for the first time. The authors were adept at drawing parallels to earlier work in the field (including the CH inequality and Aspect et. al.'s analyzers), but truncated the development and explanation of their theoretical model, resulting in another challenging read.

[21]J. D. Franson, "Bell inequality for position and time," Physical Review Letters **62**, 2205–2208 (1989).

This proposal of a novel platform for experimental Bell tests not using discrete observables is quite mathematically dense. While the author does provide some motivation for his construction, I found it difficult to interpret.

[22]G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, "Violation of Bell's Inequality under Strict Einstein Locality Conditions," Physical Review Letters **81**, 5039–5043 (1998).

This paper reports on the first experiment to close the locality loophole for both measurement and communication. A relatively straightforward experimental design and clear writing make this paper particularly comprehensible.

[23]A. Aspect, "Bell's Inequality Test: More Ideal Than Ever," Nature **398**, 189–190 (1999).

This article is a brief, high-level overview of twentieth-century experimental Bell tests intended to be accessible to those with little background in Bell tests. It is perhaps most useful in its attempts to classify experiments into generations of Bell tests.

[24]M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Sackett, W. M. Itano, C. Monroe, and D. J. Wineland, "Experimental violation of a Bell's inequality with efficient detection," Nature **409**, 791–794 (2001).

This paper presents a novel experimental Bell test platform consisting of a pair of fixed

ions in a large trap that interact via laser pulses, and produces results that demonstrate the closure of the efficiency loophole. The paper is a bit difficult to follow without prior background in modern ion trap research methods, but valuable as a very different approach to Bell tests.

[25]B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, M. Markham, D. J. Twitchen, D. Elkouss, S. Wehner, T. H. Taminiau, and R. Hanson, "Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres," Nature **526**, 682–686 (2015).

This is the primary research paper describing the initial loophole-free experimental Bell tests at Delft.

[26]M. Giustina, M. A. M. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J. A. Larsson, C. Abellán, W. Amaya, V. Pruneri, M. W. Mitchell, J. Beyer, T. Gerrits, A. E. Lita, L. K. Shalm, S. W. Nam, T. Scheidl, R. Ursin, B. Wittmann, and A. Zeilinger, "Significant-Loophole-Free Test of Bell's Theorem with Entangled Photons," Physical Review Letters **115**, 250401 (2015).

This is the primary research paper describing the initial loophole-free experimental Bell tests at Vienna.

[27]L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, K. J. Coakley, S. D. Dyer, C. Hodge, A. E. Lita, V. B. Verma, C. Lambrocco, E. Tortorici, A. L. Migdall, Y. Zhang, D. R. Kumor, W. H. Farr, F. Marsili, M. D. Shaw, J. A. Stern, C. Abellán, W. Amaya, V. Pruneri, T. Jennewein, M. W. Mitchell, P. G. Kwiat, J. C. Bienfang, R. P. Mirin, E. Knill, and S. W. Nam, "Strong Loophole-Free Test of Local Realism," Physical Review Letters **115**, 250402 (2015).

This is the primary research paper describing the initial loophole-free experimental Bell tests at NIST. Among the three 2015 loophole-free papers, this paper is most explicit in its methods of data analysis.

[28]C. Abellán, W. Amaya, D. Mitrani, V. Pruneri, and M. W. Mitchell, "Generation of Fresh and Pure Random Numbers for Loophole-Free Bell Tests," Physical Review Letters **115**, 250403 (2015).

This paper presents an analysis of the predictability of the Abellán random number gen-

erator used by all three loophole-free tests. The paper combines empirical measures of predictability with theoretical models of the random number generator to draw conclusions about predictability under several different combinations of assumptions. While some of the distinctions drawn in the paper are technical and unintuitive, the paper as a whole is quite comprehensible.

[29]A. E. Lita, A. J. Miller, and S. W. Nam, "Counting near-infrared single-photons with 95% efficiency," Optics Express **16**, 3032 (2008).

This paper is a characterization of the properties of the authors' improved transition edge sensor, a variant of which was used in the Vienna Bell test. While many of the issues of fabrication and measurement are beyond the scope of my research, the paper also includes valuable summary information about performance and comparisons to other single-photon detectors.

[30]F. Marsili, V. B. Verma, J. A. Stern, S. Harrington, A. E. Lita, T. Gerrits, I. Vayshenker, B. Baek, M. D. Shaw, R. P. Mirin, and S. W. Nam, "Detecting single infrared photons with 93% system efficiency," Nature Photonics **7**, 210–214 (2013).

This paper is a characterization of the properties of the authors' improved superconducting nanowire single-photon detector, a variant of which was used in the NIST Bell test. While many of the issues of fabrication and measurement are beyond the scope of my research, the paper also includes valuable summary information about performance and comparisons to other single-photon detectors.

[31]A. Batalov, C. Zierl, T. Gaebel, P. Neumann, I.-Y. Chan, G. Balasubramanian, P. R. Hemmer, F. Jelezko, and J. Wrachtrup, "Temporal Coherence of Photons Emitted by Single Nitrogen-Vacancy Defect Centers in Diamond Using Optical Rabi-Oscillations," Physical Review Letters **100**, 077401 (2008).

This paper presents particular experimental results on optical properties of nitrogen vacancy centers and is not effective as an general introduction to NV centers themselves. It is referenced primarily for its superior diagram of the structure of an NV center.

[32]J. Handsteiner, A. S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H. T. Nguyen, I. Sanders, F. Steinlechner, R. Ursin, S. Wengerowsky, A. H. Guth, D. I. Kaiser, T. Scheidl, and A. Zeilinger, "Cosmic Bell Test: Measurement Settings from MilkyWay Stars," Physical Review Letters **118**, 060401 (2017).

This recent paper implementing random number selection based on distant stars in Bell tests was read primarily for key methods and results as opposed to detailed understanding of the experimental implementation.

[33]L. Kasday, J. Ullman, and C. Wu, "Einstein-Podolsky-Rosen Argument - Positron Annihilation Experiment," Bulletin of the American Physical Society **15**, 586 (1970).

This brief note in a conference proceedings announces the authors' modification of an earlier experiment on positronium annihilation to be suitable for a Bell test. However, other than declaring the results of the new experiment consistent with quantum mechanics, the note offers no additional details about the experiment's design or results; only papers by the authors' contemporaries provide this information.

[34]M. Lamehi-Rachti and W. Mittig, "Quantum mechanics and hidden variables: A test of Bell's inequality by the measurement of the spin correlation in low-energy proton-proton scattering," Physical Review D **14**, 2543–2555 (1976).

This detailed original report on an experimental Bell test on the spins of scattered protons extensively documents most aspects of the experimental apparatus, methods underlying the non-trivial calibrations involved in the experiment, and all assumptions made in the analysis.