Entanglement

Wesley C. Campbell

February 15, 2018

1 Uncertainty in photon counting experiments

In the next few experiments, you are going to be using a single-photon counter to measure count rates. It is important for the interpretation of your results that you understand how to calculate the uncertainty in your measurement of the count rate. In this section, we are going to show that if an experiment records μ detector clicks in a time T, the measured count rate is given by $R_{\text{meas.}} = (\mu \pm \sqrt{\mu})/T$. That uncertainty, $\sqrt{\mu}$, is often called *shot noise* and is so fundamental to experimental physics that we're going to devote some serious time to it here. We will begin with some simple probability theory.

1.1 D&D

As with most concepts in probability, I find it is helpful to start first with an example before talking about the mathematical details of the general case. Consider, for a moment, that we have one of those 20-sided dies from Dungeons and Dragons and we're interested in calculating the probability that we roll exactly two 8s in three tries. We can approach the solution to this question in many ways, but it will be instructive in this case for us to start by calculating the probability of getting a particular, 3-trial sequence of success (s, meaning we rolled an 8) and failure (f, meaning we rolled something that was not an 8), say (s, s, f).

The probability of getting the first result is clearly $\frac{1}{20}$, so the probability of getting (s, s) is $\left(\frac{1}{20}\right)^2$. We then multiply this by the probability of failure in the last roll, which is $\frac{19}{20}$ to give the probability of the exact sequence (s, s, f),

$$\mathcal{P}\left[(s, s, f)\right] = \left(\frac{1}{20}\right)^2 \frac{19}{20}. \tag{1}$$

We now simply multiply this result by the number of sequences that contain exactly two successes, which is 3, to get the overall probability of rolling exactly two 8s in three trials,

$$\mathcal{P}\left[N_s = 2\right] = 3\left(\frac{1}{20}\right)^2 \frac{19}{20} \approx 0.007. \tag{2}$$

1.2 The binomial distribution

It is not difficult to see how to generalize our D&D example to a general case. If the probability of a success is denoted p, the probability of failure is $q = 1 - p$. If we consider n trials and wish to know the probability that exactly ν of them resulted in success, we can rewrite Eq. 1 as $p^\nu q^{n-\nu}$. The number of different ways to get ν successes in n trials is n choose ν, so we have

$$\mathcal{P}\left[\nu \text{ successes in } n \text{ trials}\right] = \binom{n}{\nu} p^\nu q^{n-\nu} \equiv B_{n,p}(\nu). \tag{3}$$

Equation 3 is called the *binomial distribution*, and you will recognize the binomial coefficient[1]

$$\binom{n}{\nu} \equiv \frac{n!}{\nu!(n-\nu)!}. \tag{4}$$

The binomial distribution $B_{n,p}(\nu)$ tells us the probability of getting exactly ν successes in n trials (given the probability of success in a single trial is p). For a given number of trials n, and given the success probability per trial p, the average number of successes will be

$$\bar{\nu} = np. \tag{5}$$

More interesting than this, which is probably fairly intuitive, is the expected variance in the number of measured successes each time n trials are performed. The standard deviation of the number of successes is

$$\sigma_\nu = \sqrt{np(1-p)}. \tag{6}$$

Applying this to our D&D example, we predict that the probability of rolling exactly two 8s is $B_{3,1/20}(2) \approx 0.007$, as we found before, the expected number of 8s in any set of 3 rolls is $3/20 = 0.15$, and the set-to-set fluctuations in the number of 8s per set of 3 trials will have a standard deviation of $\sigma_2 \approx 0.38$.

1.3 Application to photon counting

Now we're ready to get back to physics. Say we have a single-photon counting detector (with no number resolution), and we're tasked with measuring the average count rate of some source of photons. What I mean by a *single-photon counting detector* is a device that "clicks" once whenever it registers that it has been hit by a photon. You should think of the output as being digital, as opposed to analog – more light means a faster rate of clicks, as opposed to an increasing analog current or voltage or whatever. If two photons are absorbed by the detector simultaneously, you still just get one click, and it looks just like a single-photon click.

---

[1] You may recall the *binomial expansion*, $(p+q)^n = \sum_{\nu=0}^{n} \binom{n}{\nu} p^\nu q^{n-\nu}$, which we could rewrite as $(p+q)^n = \sum_{\nu=0}^{n} B_{n,p}(\nu)$.

Now, for the single-photon detectors you will use in this course, there is a certain amount of *dead time* (T_d) after a photon is detected during which the device recovers and any photons that arrive during this time will not cause a click. So if you want to use this type of detector to accurately determine the average count rate, you have to make sure that the count rate is much less than T_d^{-1}. Even if you don't know T_d, you can check that you're in this *linear regime* by doing things like blocking half the light and making sure the count rate goes down by a factor of 2.

To put this in the language we developed in the previous sections, we have stipulated that if p represents the probability of registering a click in a time bin of width T_d, we have $p \ll 1$. Keep in mind that we do not need to know T_d or p in order to experimentally ensure that this is the case.

So, given this definition of p as the probability of registering a click in a single time bin of width T_d, what would the probability be of getting exactly 1 click in a time interval that is two dead times long, $2 \times T_\mathrm{d}$? Well, this will be the product of

1. the probability of getting a click in the first time bin, p, times
2. the probability of not getting a click in the second time bin, q, times
3. the number of ways to get one click in two bins, 2

which gives us

$$2 \times p \times q. \tag{7}$$

Generalizing this, we can ask what is the probability of getting ν clicks in n bins of width T_d? By analogy with our development above, it will be given by the binomial distribution

$$B_{n,p}(\nu) = \binom{n}{\nu} p^\nu q^{n-\nu}. \tag{8}$$

In fact, to take this one step futher, it is very frequently the case that $n \gg 1$ and $p \ll 1$. In this limit, the binomial distribution is indistinguishable from a Poisson distribution,

$$P_\mu(\nu) \equiv \mathrm{e}^{-\mu} \frac{\mu^\nu}{\nu!}, \tag{9}$$

which is characterized entirely by its mean number of "successful" outcomes μ. (It is also worth noting that in this same limit, the binomial and Poisson distributions are indistinguishable from a Gaussian distribution with mean μ and standard deviation $\sigma = \sqrt{\mu}$.)

Now, this may not look incredibly useful for a physicist who may not know what T_d is, because she also will not know what n or p are. However, if she does know that the detector is in the linear regime, that is enough to apply these results, because that was the only real stipulation we made.

For instance, if she measures μ clicks in a time T, she can then ask what is the uncertainty in the measured click rate $R_\mathrm{best} = \mu/T$? Well, given her knowledge of the arguments we just made, her best guess for the parent distribution that gave rise to this result μ is that it's a binomial distribution with an average number of successes $\bar{\nu} = \mu$. So she can already say that $np = \mu$. She can then look at the standard deviation of such a distribution, Eq. 6, and apply the approximation $p \ll 1$ to infer that the uncertainty is given by

$$\sigma \approx \sqrt{np} = \sqrt{\mu}. \tag{10}$$

This result is the main point of all of this analysis. Given a measurement that has recorded μ successes, the uncertainty in the measurement of the mean will be $\sqrt{\mu}$. Our researcher can now properly replort her measurement result for the count rate,

$$R = \left(\frac{\mu}{T} \pm \frac{\sqrt{\mu}}{T}\right). \tag{11}$$

Exercise 1.1 *Given a photon counting experiment with a mean count rate of μ/T, predict (with a quantified uncertainty) the count rate that would be found in an otherwise identical experiment lasting some other time $T_2 \neq T$.*

This is a task that must be done carefully in order to get the uncertainty correct. I say this because the mean can be done by inspection since we clearly expect the same mean count rate, $R_{\text{best}} = \mu/T$. However, the fractional uncertainty in the rate will not be the same as that shown in Eq. 11 unless $T_2 = T$. You can see this intuitively by imagining that T_2 is really really long compared to T. The measurement of the rate done over the T_2 timescale amounts to a ton of averaging – the fluctuations in the inferred rate will decrease as more and more data are accumulated.

The important step that you have to be sure to do when calculating shot noise is to make sure that the uncertainty in the number of counts *is the square root of the mean* number of counts, *not the count* rate. *The mean number of counts expected in time T_2 is $\left(\frac{\mu}{T}\right) \times T_2$, so the uncertainty in the number of counts expected is $\sqrt{\mu T_2/T}$, which gives us the desired expression for the rate,*

$$R_2 = \left(\frac{\mu}{T} \pm \sqrt{\frac{\mu T_2}{T}} \times \frac{1}{T_2}\right) = \left(\frac{\mu}{T} \pm \sqrt{\frac{\mu}{T T_2}}\right). \tag{12}$$

Exercise 1.2 *A movie producer wants to use a poll to decide if more people like movies about Godzilla or Gamera. The poll results in 35 people saying they like movies about Godzilla and 28 said they like movies about Gamera. The movie producer concludes that people are 25% more likely to watch movies about Godzilla than Gamera. Is this conclusion supported by the poll results?*

Well, to figure this out, we can estimate the uncertainties in these measurements that are trying to determine the "true" mean value for each quantity. A Poisson distribution with a mean of 35 *will have a standard deviation of* $\sqrt{35}$, *and we see that the measured number of people who like Godzilla movies is* 35 ± 6. *Likewise, the measured number of people who like Gamera movies is* 28 ± 5. *So the poll results are perfectly consistent with no net preference one way or another, and (arguably) even with the average person being more likely to prefer Gamera movies than Godzilla movies. The poll is therefore not consistent with the quoted 25% preference, and is clearly cannot be used to make a claim of any preference either way.*

2 Polarization-entangled photons

2.1 Polarization measurements

Last week, we discussed the process of Spontaneous Parametric Down-Conversion (SPDC) and how Type I phase-matched SPDC in the experimental apparatus you will be using can create photons in the state

$$|\psi\rangle = \tfrac{1}{\sqrt{2}}(|H_1H_2\rangle + \mathrm{e}^{i\phi}|V_1V_2\rangle). \tag{13}$$

We haven't yet specified what the phase ϕ is, but as long as it is fixed (as opposed to a random variable that changes unpredictably for each photon pair), the state Eq. 13 is *entangled*. By *entangled*, we mean that it cannot be factored into a product of a state that describes photon 1 and a state that describes photon 2. The best description of the state of the system (i.e. the description that gives us the best agreement between predictions and measurements) is one that requires we treat the photons as being part of a joint state such as Eq. 13 that cannot be factored into states of the individual photons.

In order to verify this claim, we will need some single-photon polarization measurements that return yes/no answers to some specific polarization questions we will motivate in the following sections. In particular, we need the measurement outcomes to be either $+1$ or -1 for each photon we measure.

One example of a measurement we will need is one that returns $+1$ if the photon is H-polarized and -1 if it is V-polarized. This can be accomplished with a polarizing beamsplitter that has a single-photon detector at each output. If the polarizing beamsplitter transmits $|H\rangle$ into detector H and reflects $|V\rangle$ into detector V, we can assign a click from the H detector a value of $+1$ and a click on the V detector a value of -1. (Crucially, for single photons, there will never be two clicks). The operator that describes this measurement (applied, for instance, to photon 1) is given by

$$\hat{\mathbf{A}}(0) \equiv |H_1\rangle\langle H_1| - |V_1\rangle\langle V_1|. \tag{14}$$

(For reasons that will be explained in the next section, we will be associating measurements on photon 1 with the letter "A" and measurements of photon 2 with the letter "B.")

Now let's see what happens when we rotate this same polarization analyzer, that is, we will rotate the whole thing including the detectors about the axis of propagation of the input light. If the rotation is through an angle α, we already developed the formalism to see that the new operator is given by

$$\hat{\mathbf{A}}(\alpha) = \mathbf{R}(\alpha)\hat{\mathbf{A}}(0)\mathbf{R}^{-1}(\alpha) \tag{15}$$

where $\mathbf{R}(\theta)$ is the same rotation matrix we used when we talked about polarization in week 1:

$$\mathbf{R}(\theta) = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}. \tag{16}$$

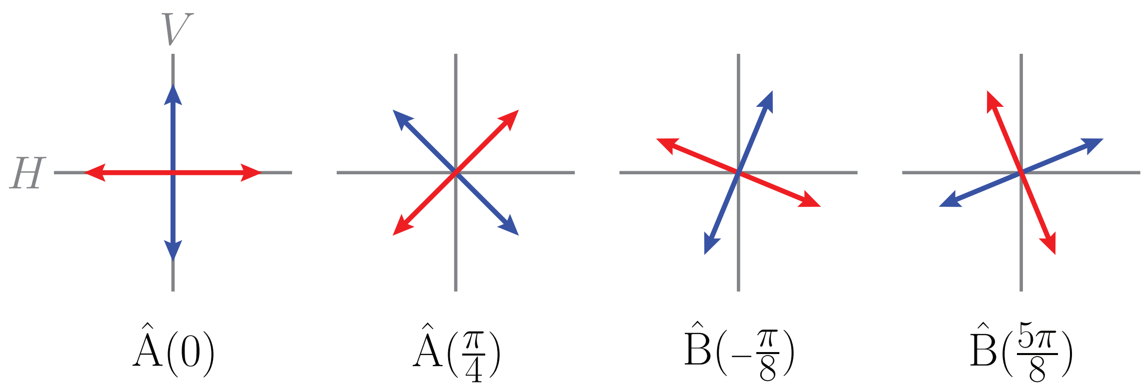

Figure 1: Spatial depiction of the basis for each of the four measurements of linear polarization that will be needed for Bell's inequality. Red corresponds to the direction yielding a $+1$ result, blue corresponds to a -1 result. We recognize $\hat{\mathbf{A}}(0)$ and $\hat{\mathbf{A}}(\frac{\pi}{4})$ as measurements in the $H-V$ and $D-A$ (Diagonal-Antidiagonal) bases.

Since the matrix form of Eq. 14 is

$$\hat{\mathbf{A}}(0) = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{17}$$

we have

$$\hat{\mathbf{A}}(\alpha) = \begin{pmatrix} \cos 2\alpha & \sin 2\alpha \\ \sin 2\alpha & -\cos 2\alpha \end{pmatrix}. \tag{18}$$

3 Bell's inequality

We will introduce four measurements for linearly-polarized photons that will return ± 1 for each measurement instance (each single photon). The averages of these measurements over many repeated measurement instances will be given by the expectation values of the following four operators:

$$\begin{aligned}
\hat{\mathbf{A}}(0) &\equiv |H_1\rangle\langle H_1| - |V_1\rangle\langle V_1| && (19)\\
\hat{\mathbf{A}}(\tfrac{\pi}{4}) &\equiv |V_1\rangle\langle H_1| + |H_1\rangle\langle V_1| && (20)\\
\hat{\mathbf{B}}(-\tfrac{\pi}{8}) &\equiv \tfrac{1}{\sqrt{2}}\Big(|H_2\rangle\langle H_2| - |H_2\rangle\langle V_2| - |V_2\rangle\langle H_2| - |V_2\rangle\langle V_2|\Big) && (21)\\
\hat{\mathbf{B}}(\tfrac{5\pi}{8}) &\equiv \tfrac{1}{\sqrt{2}}\Big(-|H_2\rangle\langle H_2| - |H_2\rangle\langle V_2| - |V_2\rangle\langle H_2| + |V_2\rangle\langle V_2|\Big). && (22)
\end{aligned}$$

As we discussed, these are all really just rotated versions of a polarizing beam splitter (PBS) with two detectors. The point of the "A" and "B" notation is that the first two operators will be implemented by a character named Alice to measure photon 1, while the second two

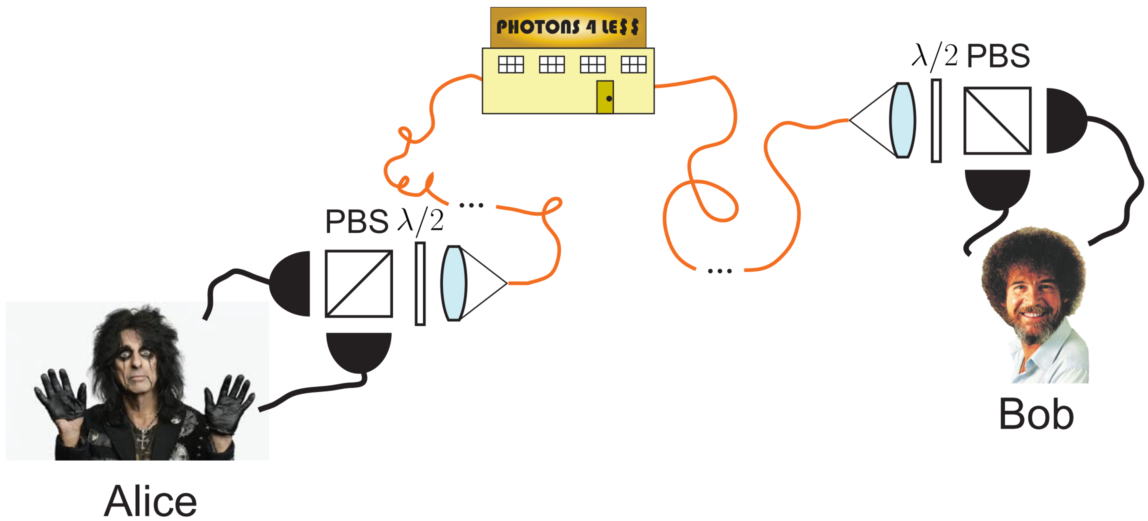

Figure 2: Pairs of photons are sent via single-mode, PM fiber to Alice and Bob, who are separated from each other by a large distance. Each of them has a polarization measurement apparatus, and they record the results of their measurements.

will be operated by a character named Bob on photon 2, as shown in Fig. 2. Of course, instead of actually rotating the PBS and detectors, they use a half wave plate to rotate the input light by $-\theta$ instead, which produces the same result with less experimental hassle.

The $+1$ and -1 eigenstates of $\hat{\mathbf{A}}(0)$ and $\hat{\mathbf{A}}(\frac{\pi}{4})$ are $|H\rangle$ and $|V\rangle$; $|D\rangle$ and $|A\rangle$. For $\hat{\mathbf{B}}(-\frac{\pi}{8})$ and $\hat{\mathbf{B}}(\frac{5\pi}{8})$, they're given by

$$\begin{pmatrix} \cos(-\frac{\pi}{8}) \\ \sin(-\frac{\pi}{8}) \end{pmatrix} \text{ and } \begin{pmatrix} \sin(-\frac{\pi}{8}) \\ \cos(-\frac{\pi}{8}) \end{pmatrix}; \begin{pmatrix} \cos(\frac{5\pi}{8}) \\ \sin(\frac{5\pi}{8}) \end{pmatrix}, \text{ and } \begin{pmatrix} \sin(\frac{5\pi}{8}) \\ \cos(\frac{5\pi}{8}) \end{pmatrix}. \tag{23}$$

3.1 Local hidden variables

What we are interested in finding out is whether or not we can devise an experiment that can tell the difference between photon states that are classically correlated and those that require the nonlocality of quantum mechanics to describe.

First, let's discuss what is meant by "classically correlated." I would claim that if I can create a system that produces the same measurement results as whatever quantum system we're considering without needing quantum mechanics, that system shows only classical correlations, and it would not be out-of-bounds to even say the system itself *is classical* in the sense that it has not been proved that quantum mechanics is necessary to describe it.

For instance, consider for a moment a black box with two fiber connections that can produce pairs of photons where the photons sent out of the two ports are either both vertically polarized or both horizontally polarized when they leave the device. Maybe there is a little homunculus in the box who flips coins or something – the point is that the polarization of the photons are determined before they leave the box, according to a classical probability distribution, say, 50% $V_1 V_2$ and the same for $H_1 H_2$. This could even be classical bursts of light for all we care – a large number of photons get sent in a burst, but the measurement result in the $H-V$ basis is guaranteed to give either $+1$ or -1.

If we send the outputs of this source to Alice and Bob and they measure each of them in the $H-V$ basis (i.e. they both measure with $\theta = 0$ above) *their results will show perfect correlation!* Every time Alice measures $+1$, Bob will also measure $+1$ and vice versa. But this is clearly not a very "quantum" system – classical mechanics certainly allows for a homunculus in a box to flip a coin and make short bursts of light that are H polarized, then for him to switch and make a short burst that is V polarized and so forth. Nothing had to travel faster than the speed of light to allow for spooky "communication" between Alice and Bob's experiments, nothing here had to be nonlocal (*i.e.* acting at a distance).

As a more elaborate example, let's consider the possibility that instead of sending Alice and Bob a series of photons, I send them a series of pieces of paper with their measurement outcomes written on it ($+1$ or -1) for each of their two choices for θ ($\theta_{\rm A} = 0, \pi/4$ for Alice and $\theta_{\rm B} = -\pi/8, 5\pi/5$ for Bob from Eq. 19-22). Alice and Bob make a random choice of which setting they will have for their detector for each measurement instance and then open the envelope and read off their results. Bob's instructions can even be seen on Alice's paper and vice versa, so both experimentalists may get a letter in the mail that looks something like the following:

"Photon #53"			
Alice		Bob	
$\theta_{\rm A}$	Outcome	$\theta_{\rm B}$	Outcome
0	$+1$	$-\frac{\pi}{8}$	$+1$
$\frac{\pi}{4}$	-1	$\frac{5\pi}{8}$	$+1$

We could think of Alice's part of this table as the result of some function $a(\theta_{\rm A}, \lambda_{53})$ of the choice Alice made for $\theta_{\rm A}$ and some other parameter called λ_{53} that rides along with her "particle" (and likewise for $b(\theta_{\rm B}, \lambda_{53})$ for Bob). λ is a so-called *local hidden variable*, and each photon can carry a copy if need be. Furthermore, to ensure that $a(\theta_{\rm A}, \lambda)$ mimics quantum mechanics with all it's fluctuations and predictions, there is a probability distribution of different λ from which experimental values of λ are drawn that makes sure that $a(\theta_{\rm A}, \lambda) = \langle \psi_1 | \hat{\mathbf{A}}(\theta_{\rm A}) | \psi_1 \rangle$ for single-photon wavefunctions $|\psi_1\rangle$. For single photons and single-photon states, each measurement instance may produce a different answer due to the different λ for each case (since this is what we see in real experiments), and the average will give the same answer as quantum mechanics.

The crucial thing here is that these outcomes are independent of the other experimentalist's apparatus setting; Alice's outcomes can't depend upon Bob's choice for $\theta_{\rm B}$ and vice versa, because Bob's choice of $\theta_{\rm B}$ and the recording of his measurement result are separated from Alice's by a spacelike interval. If Bob changes his choice for $\theta_{\rm B}$, this has no impact on Alice's photon's hidden variable λ since it is assumed local.

Note once again that this whole business can be carried by manifestly classical objects such as pieces of paper, or watermelons with measurement outcomes written on them with a sharpie. If the local hidden variable needs to mimic the behavior of Hamiltonian evolution in transit or something, the watermelon may need a little magnetometer in it and a touch screen or whatever, but it is restricted to change deterministically based on local (in the special relativity sense of "within a timelike interval") conditions.

The only reasons we call the local variable "hidden" is that we haven't found a piece of photon anatomy that can necessarily do this yet; in quantum mechanics, we don't consider that the results of measurements of Alice's two choices for θ_A can be knowable ahead of time since those operators don't commute. If a photon's polarization is known to be either pure H or pure V, a measurement in the $D-A$ basis should give 50/50 results. (Another reason it's good to call it "hidden" is that we probably want to make sure Alice and Bob don't "look" before choosing their measurement basis).

3.2 Bell's observable

The question we are trying to answer is, in a sense, whether the measurement of part of a quantum system at spacetime location A can depend upon something that happens at spacetime location B that is separated from A by a spacelike interval. In other words, is there a measurable difference between photons coming from some classical (local hidden variable) source and a truly quantum state such as:

$$|\psi_s\rangle = \frac{|H_1 H_2\rangle - |V_1 V_2\rangle}{\sqrt{2}}, \tag{24}$$

and if so, what is the experimental result?

After all, the EPR paper argues fairly convincingly that the predictions of quantum mechanics lead to some fairly uncomfortable interpretations. If there is a way to empirically test whether quantum nonlocality is actually needed to predict the outcomes of even a single experiment, it will therefore be of great interest. As physicists, we are used to the idea of treating a state like $|\psi_s\rangle$ as if it actually *exists* and maybe even evolves as the photons are traveling, but perhaps all of this stuff is actually predetermined by local hidden variables that potentially also evolve in time (albeit *deterministically*).

The key to uncovering this experimental test is an insight that is attributed to John Bell. In the language of our particular system, he decided to consider the following observable:

$$\hat{\mathbf{S}} \;\equiv\; \hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(-\tfrac{\pi}{8}) \;+\; \hat{\mathbf{A}}(0)\hat{\mathbf{B}}(-\tfrac{\pi}{8}) \;+\; \hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(\tfrac{5\pi}{8}) \;-\; \hat{\mathbf{A}}(0)\hat{\mathbf{B}}(\tfrac{5\pi}{8}). \tag{25}$$

This operator $\hat{\mathbf{S}}$ is Bell's observable[2], and the trick to the whole business is somehow contained in that one minus sign at the end. Each of the four terms in this expression is the product of an observable of particle 1 and an observable of particle 2. We will show that the expected value of this observable is different for a world governed by a local hidden variable theory and one governed by quantum mechanics.

[2]This is actually a variant on it called a CHSH inequality, but the basic idea is the same.

3.3 Measurement outcome for a local hidden variable theory

The local hidden variable theory analog of a state like Eq. 24 will have a variable that describes the outcome of the four measurements appearing in Eq. 25. Each photon will have a copy of this λ, and we presume that Alice and Bob are sufficiently far apart and sufficiently synchronized in their measurement timing that the locality of these hidden variables guarantees that Alice's measurement outcomes or choices of measurement bases can't influence the hidden variable encoding measurement outcomes on Bob's end, in accordance with special relativity.

Let's denote by q, r, s, and t (all taking values ± 1) the possible single measurement instance outcomes of $\hat{\mathbf{A}}(0)$, $\hat{\mathbf{A}}(\frac{\pi}{4})$, $\hat{\mathbf{B}}(-\frac{\pi}{8})$, and $\hat{\mathbf{B}}(\frac{5\pi}{8})$, respectively. For a given local hidden variable λ, the outcomes are a function of λ, but in a series of measurements, we certainly permit λ to be different for each one. Assuming enough statistics can be gathered in our experiments, we can model the various outcomes in terms of their probability (or frequency) of occurrence, which theoretically would be some average involving $a(\theta_{\mathrm{A}}, \lambda)$ and $b(\theta_{\mathrm{B}}, \lambda)$ over the distribution governing λ. We will call the probability of a particular outcome (averaged over the distribution governing λ) $p(q, r, s, t)$. Probability normalization guarantees that the sum of the probabilities of all possible outcomes must be 1:

$$\sum_{q,r,s,t=\pm 1} p(q, r, s, t) = 1. \tag{26}$$

This probability $p(q, r, s, t)$ is a little hard to interpret, particularly since readers who are used to quantum mechanics will have a hard time imagining a probability like $p(q, r)$ where those two outcomes q and r correspond to results of measurements in bases that don't commute. However, for a local hidden variable theory, we will presume that λ can be used to compute definite outcomes for both possible measurements. Each photon pair in our scenario will have sixteen different possibilities for the (q, r, s, t) argument of this probability, and if enough statistics can be gathered experimentally, one could figure out what the sixteen different $p(q, r, s, t)$ are. One could say that each individual experimental instance is a faithful measurement of a single combination of (q, r, s, t) that is pulled from a distribution of the sixteen possibilities that is governed by the $p(q, r, s, t)$s.

The expectation value of an operator in this local hidden variable formalism will simply be the sum of the product of each possible measurement outcome with the probability of that outcome. We will denote this expectation value with $E\left\{\right\}$. For example,

$$E\left\{\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(-\tfrac{\pi}{8})\right\} = \sum_{q,r,s,t=\pm 1} p(q, r, s, t) \times rs \tag{27}$$

We therefore have

$$\begin{aligned} E\left\{\hat{\mathbf{S}}\right\} &= \sum_{q,r,s,t=\pm 1} p(q, r, s, t) \times (rs + qs + rt - qt) \qquad (28) \\ &= \sum_{q,r,s,t=\pm 1} p(q, r, s, t) \times rs + \sum_{q,r,s,t=\pm 1} p(q, r, s, t) \times qs \end{aligned}$$

$$+ \sum_{q,r,s,t=\pm1} p(q,r,s,t) \times rt - \sum_{q,r,s,t=\pm1} p(q,r,s,t) \times qt \tag{29}$$

$$= E\left\{\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(-\tfrac{\pi}{8})\right\} + E\left\{\hat{\mathbf{A}}(0)\hat{\mathbf{B}}(-\tfrac{\pi}{8})\right\} + E\left\{\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(\tfrac{5\pi}{8})\right\} - E\left\{\hat{\mathbf{A}}(0)\hat{\mathbf{B}}(\tfrac{5\pi}{8})\right\} \tag{30}$$

Looking at Eq. 28, we see that we can factor it as follows

$$\sum_{q,r,s,t=\pm1} p(q,r,s,t) \times (rs + qs + rt - qt) = \sum_{q,r,s,t=\pm1} p(q,r,s,t) \times \Big((r+q)s + (r-q)t\Big). \tag{31}$$

This factorization is possible because we have declared that the measurement of Bob's photon is determined by its local hidden variable, which does not have time to "learn" about Alice's choice of measurement basis. The probability that Bob measures s cannot depend upon whether Alice chose to measure in the $\theta_{\mathrm{A}} = 0$ or $\pi/4$ basis, so the probability that Bob measures s when Alice chooses $\theta_{\mathrm{A}} = 0$, when we sum over Alice's possible outcomes r has to be the same as the probability he measures s when Alice chooses $\theta_{\mathrm{A}} = \pi/4$ once sum over Alice's possible outcomes q.

However, for all possible values of r and q, either $(r+q) = 0$ or $(r-q) = 0$, so $(r+q)s + (r-q)t$ can never be larger than 2. Using Eq. 26, we have

$$\sum_{q,r,s,t=\pm1} p(q,r,s,t)\Big((r+q)s + (r-q)t\Big) \leq 2 \tag{32}$$

which gives us Bell's inequality:

$$E\left\{\hat{\mathbf{S}}\right\} = E\left\{\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(-\tfrac{\pi}{8})\right\} + E\left\{\hat{\mathbf{A}}(0)\hat{\mathbf{B}}(-\tfrac{\pi}{8})\right\} + E\left\{\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(\tfrac{5\pi}{8})\right\} - E\left\{\hat{\mathbf{A}}(0)\hat{\mathbf{B}}(\tfrac{5\pi}{8})\right\} \leq 2. \tag{33}$$

A local hidden variable theory cannot produce results whose expected value for this thing $\hat{\mathbf{S}}$ exceeds 2.

3.4 Measurement outcome for quantum mechanics

We can also calculate the expectation value of $\hat{\mathbf{S}}$ for the quantum state in Eq. 24:

$$\begin{aligned}
\left\langle\hat{\mathbf{S}}\right\rangle &= \langle\psi_s|\hat{\mathbf{S}}|\psi_s\rangle \\
&= \langle\psi_s|\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(-\tfrac{\pi}{8})|\psi_s\rangle + \langle\psi_s|\hat{\mathbf{A}}(0)\hat{\mathbf{B}}(-\tfrac{\pi}{8})|\psi_s\rangle + \langle\psi_s|\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(\tfrac{5\pi}{8})|\psi_s\rangle - \langle\psi_s|\hat{\mathbf{A}}(0)\hat{\mathbf{B}}(\tfrac{5\pi}{8})|\psi_s\rangle \\
&= \left(\frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right) + \left(\frac{1}{\sqrt{2}}\right) - \left(-\frac{1}{\sqrt{2}}\right) \\
&= 2\sqrt{2}.
\end{aligned} \tag{34}$$

Here we have a clear difference between the measurement result predicted by a local hidden variable theory and that predicted by quantum mechanics. This is what physics is all about! Maybe the universe works one way, maybe it works another, and here we have an actual, feasible test that will be able to rule out one of the two theories. And, as you are probably aware, experiments clearly rule out local hidden variables. It is possible to violate Bell's inequality to many, many σ.

3.5 Discussion

So, what do these experimental results prove? Well, first of all, that no classical system (in the sense we used the term earlier) can mimic these results. The homunculus in the box can be sending papers with detailed measurement results, little robots programmed however he wants, or can even be sending people to try to fake out the results. As long as the measurement basis choice that Alice and Bob do is truly random and that these choices are unknown by the person/paper/robot on the other end, Bell's inequality can't be violated without an entangled quantum state.

The key thing that quantum mechanics has that is missing in the classical system is *nonlocality.* Alice and Bob's measurement results will be correlated with one another in a way that suggests that their photons "knew" something about the measurement made by the other person. Something about Bob's measurement of his photon restricts Alice's possible results and vice versa. This does not, however, mean that Alice can *cause* something to happen at Bob's location, and this lack of causally connected spacelike intervals is often cited as the reason that this does not violate special relativity.

3.6 Residual entanglement

In a real experiment, it is somewhat challenging to create the state $|\psi_s\rangle$ (Eq. 24). In particular, one has to be very careful that there is no part of the state that would make it possible in principle to learn whether the photons are polarized in the horizontal direction or the vertical one.

The reason for this can be seen without needing recourse to a density matrix formalism. Let us assume for a moment that there is some other degree of freedom of the system that contains information about whether the photons are in $|H_1H_2\rangle$ or $|V_1V_2\rangle$. Examples of such things include

- Time bins t_H and t_V. The H photons may be delayed compared to the V photons by, for instance, dispersion or birefringence in the SPDC crystal. The polarization could then be inferred by measuring the arrival time of the photons at some location in space.

- Spatial modes. The H photons may be emitted into a spatial mode that is distinguishable from the mode for V photons due to a difference in waist location z_0 or $\mathbf{k}$-vector or any other aspect of the spatial mode of the electromagnetic field associated with the emitted light. This is why the fiber coupling of the SPDC light into single-mode fibers in the experiment is done in a way that is specifically designed to couple equally well for the first and second crystals.

- Photon frequencies leading to relative delay time tagging. Type I SPDC can create pairs of down-converted photons that are not the same color. This, by itself, would not matter, since this would be just as true for the two H photons as the two V photons. However, if the dispersion of the doubling crystals is large enough to cause a relative delay time between the two photons (due to their different colors), photons created in the first crystal will get twice as much delay time added between them as photons created in the second crystal. A pair of photons that are spaced far apart in time would be more likely to have originated in the first (upstream) crystal, and vice versa for tight spacing in time.

Any property of the photon state that could reveal information about which polarization the two photons have without actually measuring the polarization is a potential source of residual entanglement, which will look just like dephasing between $|H_1H_2\rangle$ and $|V_1V_2\rangle$.

For instance, if there is an arrival time difference between H photons (which arrive at time t_H after the pump photon is sent into the nonlinear crystals) and V photons (which arrive arrive at time t_V), the actual state that is created will be

$$|\psi_s'\rangle = \frac{|H_1, H_2, t_H\rangle - |V_1, V_2, t_V\rangle}{\sqrt{2}}. \tag{35}$$

If we then compute the expected value of one of the parts of Bell's observable, we have

$$\begin{aligned}
\langle\psi_s'|\hat{\mathbf{A}}(\tfrac{\pi}{4})\hat{\mathbf{B}}(-\tfrac{\pi}{8})|\psi_s'\rangle &= \tfrac{1}{\sqrt{2}}\Big(\langle H_1, H_2, t_H| - \langle V_1, V_2, t_V|\Big)\Big(|V_1\rangle\langle H_1| + |H_1\rangle\langle V_1|\Big) \\
&\quad \otimes\tfrac{1}{\sqrt{2}}\Big(|H_2\rangle\langle H_2| - |H_2\rangle\langle V_2| - |V_2\rangle\langle H_2| - |V_2\rangle\langle V_2|\Big) \\
&\quad \times\tfrac{1}{\sqrt{2}}\Big(|H_1, H_2, t_H\rangle - |V_1, V_2, t_V\rangle\Big) \\
&= \tfrac{1}{2\sqrt{2}}\Big(\langle V_1, H_2, t_H| - \langle H_1, V_2, t_V|\Big) \\
&\quad \times\Big(|H_1, H_2, t_H\rangle + |V_1, H_2, t_V\rangle - |H_1, V_2, t_H\rangle + |V_1, V_2, t_V\rangle\Big) \qquad (36) \\
&= \tfrac{1}{\sqrt{2}}\Big(\langle t_H|t_V\rangle + \langle t_V|t_H\rangle\Big). \qquad (37)
\end{aligned}$$

So we see that if the two states associated with these two delay times don't have any appreciable overlap with each other, the interference is erased and this expectation value is zero. Importantly, the state overlap $\langle t_H|t_V\rangle$ has nothing to do with the detector response or anything about this actually being *observed*. It is enough to say that this could be measured *in principle* to see that the interference is destroyed. Residual entanglement with other degrees of freedom therefore has to be eliminated to access the two-body coherence in the state (24).

Last, for the specific case we considered (different delay times), what determines $\langle t_H|t_V\rangle$? Surely, there is a small value for the differential delay that results in no discernible difference between these states. The answer is that it's *the length of the photons*. If the pump laser has some linewidth $\Delta\omega_\mathrm{p}$, we can expect its phase coherence to last a time $\tau_\mathrm{p} = 1/\Delta\omega_\mathrm{p}$. This is the length of time over which we believe the pump photons can be described by a pure state, and the down converted photons should inherit the phase properties of the pump. We therefore expect the red photons to have a length (in time) of τ_p, so as long as the differential delay between H and V is significantly shorter than this, learning the delay time associated with a particular photon doesn't provide very much information about the polarization, and the interference will survive in tact.