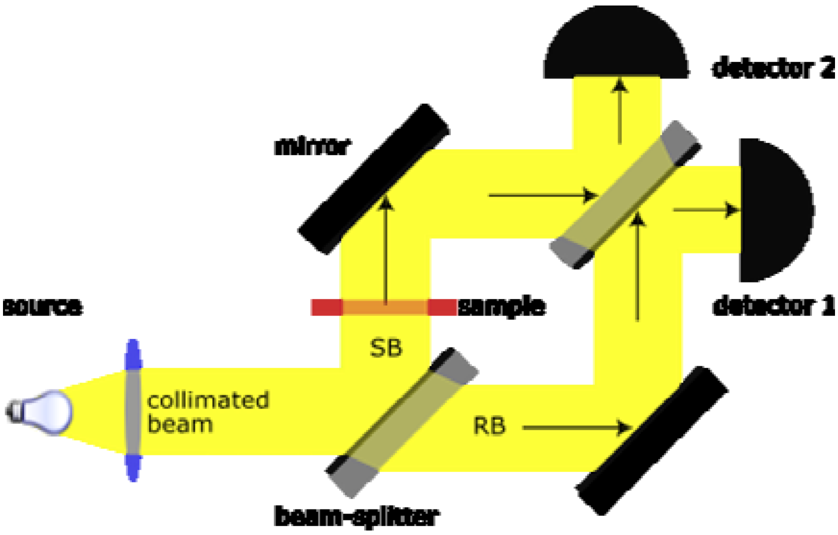
QM13: The Observability of Counterfactuals
The Elitzur-Vaidman Bomb Test, Ref.[1]
Last Update: 13/3/11

1. Counterfactuals

Suppose something could have happened, but actually did not happen. In classical physics the fact that an event *could have* happened but didn't can make no difference to any future outcome. Only those things which actually happen can influence the future evolution of the world. But in quantum mechanics it is otherwise. The potential for an event to happen can influence future outcomes even if the event does not happen. Something that could happen but actually does not is called as *counterfactual*. In quantum mechanics *counterfactuals are observable* – they have measurable consequences. The Elitzur-Vaidman bomb test provides a striking illustration of this. But first we must revise how interferometers work – specifically Mach-Zehnder interferometers.

Figure 1 The Mach-Zehnder Interferometer
RB = Reference Beam; SB = Sample Beam



The key to the behaviour of the interferometer is the phase change suffered by the reflected and transmitted waves at the mirrors. There is potential confusion over this since the phase change depends upon the type of object used as a "mirror". Here we shall just state the rules for the phase changes for two types of "mirror". The rules are derived from first principles in the Appendix.

2. Reflection/Transmission Phase Change Rules

2.1 Silvered or Half-Silvered Mirrors

These are simply conventional mirrors which consist of a flat piece of glass onto one side of which has been deposited a thin layer of silver, or similar reflective metal. The terms "silvered" and "half-silvered" refer to different thickness of metal film, the first being sufficient to prevent any transmission of light through the mirror, whereas the latter is calibrated to allow about as much transmission as reflection. The phase shift caused by such mirrors differs according to whether the incoming beam is incident on the silvered surface (the "front" of the mirror) or the glass surface (the "back" of the mirror). Note that mirrors in domestic use usually have the “back” (the glass) facing forwards, and the “front” is generally covered with something opaque to protect the silver film. The rules for the phase shifts are,

[1] A beam reflected from the front (silvered) surface undergoes a phase change of 180^o (i.e. a factor of -1);

[2] A beam reflected from the back surface undergoes no phase change due to reflection but a phase change of 2Δ due to its passage (twice) through the glass, where $\Delta = \frac{2\pi(n-1)fa}{c} = (n-1)ka$ and where $n =$ the refractive index of the glass and $a =$ the thickness of the glass;

[3] A beam transmitted through a (half-silvered) mirror undergoes a phase change of Δ.

All these phase changes are with respect to what the phase would be at the same place if propagation had involved passage through air alone (i.e., with respect to the phase $e^{i\bar{k}\cdot\bar{r}}$ of a propagating wave). The phase change rules are the same for silvered and half-silvered mirrors.

In the analysis below the phase change Δ will be assumed the same for all four mirrors. This implies that the mirrors all have the same thickness to optical precision, i.e., to an accuracy much less than the wavelength of light. In practice this is improbable and some adjustment (calibration) of the interferometer will be required prior to use to compensate for this practical limitation.

2.2 Plain Plates

Any transparent plate with a refractive index, *n*, greater than air will cause a mixture of transmission and reflection. This results from the requirement to satisfy the relevant boundary conditions at both surfaces of the plate, rather than being a result of reflection at just one surface (as for a mirror). Consequently the phase changes are different. The rule is simply,

- The reflected wave acquires a phase factor of i compared to the transmitted wave.

The transmitted wave actually picks up a phase factor relative to the incident wave, but the above rule is all that is needed to analyse the interferometer.

3. The Mach-Zehnder Interferometer – No Sample

Initially we analyse what signals we expect at the detectors 1 and 2 (Figure 1) if there is no “sample”, nor anything else put in the way of the beams (e.g., any measuring device).

Firstly let’s assume mirrors and half-silvered mirrors are used, in the orientation shown in Figure 1. The phase factors acquired by the reference beam and the sample beam at detector 2 are,

Detector 2, Mirrors: RB: $e^{i\Delta} \times -1 \times e^{i\Delta} = -e^{2i\Delta}$ SB: $-1 \times -1 \times e^{2i\Delta} = +e^{2i\Delta}$

The three terms in each expression above refer to the reflection or transmission at each mirror in sequence. The phase factors for the reference and sample beams are equal and opposite, so consequently they destructively interfere. The prediction, borne out by experiment, is that no light emerges into detector 2. If the laser brightness is turned down until there is only one photon passing through the equipment at any time, the same conclusion applies: no photons are registered by detector 2. We must hope that we get constructive interference into detector 1, or our photons will have gone missing! We do indeed get constructive interference into detector 1, thus,

Detector 1, Mirrors: RB: $e^{i\Delta} \times -1 \times -1 = +e^{i\Delta}$ SB: $-1 \times -1 \times e^{i\Delta} = +e^{i\Delta}$

Now let's examine what happens if plain plates were used instead of mirrors. In this case the phase rule is for reflection *relative to* transmission. So consider the phase changes of the sample beam with respect to the reference beam for exit into detector 2. At the first mirror the SB picks up a phase factor of i wrt the RB, and the same is true at the last (3rd) mirror, since at both these mirrors the SB is reflected but the RB is transmitted. As for the 2nd mirror, both beams are reflected so there is no relative phase change. Consequently the relative phase change between the SB and the RB at detector 2 is $(i)^2 = -1$, and we again conclude that there is destructive interference and hence no photons at detector 2.

Repeating this analysis for detector 1 the phase change of the SB relative to the RB due to the first two mirrors is identical, i.e., $i \times 1 = i$. However at the third mirror it is now the RB which is reflected and the SB which is transmitted. Hence the RB acquires a phase factor of i wrt the SB, which is equivalent to the SB acquiring a phase factor of $-i$ wrt the RB. So overall the phase change of the SB relative to the RB into detector 1 is $i \times -i = 1$, and so there is constructive interference. All the photons enter detector 1, as for the mirrors.

4. The Mach-Zehnder Interferometer With Measurements

To analyse the effect of including a measuring device in either of the beam paths, firstly let's establish some notation. Suppose the state of a photon which follows path RB into detector *j* is written $|RB\rangle_j$. Then the state of a photon following path SB into detector *j* is $|SB\rangle_j = e^{i\delta_j}|RB\rangle_j$. The preceding analysis has shown that $e^{i\delta_1} = 1$ and $e^{i\delta_2} = -1$. The state at the detectors is thus,

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|RB\rangle_j + e^{i\delta_j}|RB\rangle_j\right) \tag{1}$$

Hence the flux of photons into the detectors is proportional to,

$$\langle\psi|\psi\rangle = 1 + \cos\delta_j \tag{2}$$

thus giving constructive interference, $\langle\psi|\psi\rangle = 2$, at detector 1 and destructive interference, $\langle\psi|\psi\rangle = 0$, at detector 2.

How does this change if a measuring device is inserted into either beam path? A device to determine which path the photon takes must have two states, one of which corresponds to "the photon was detected in path RB" and the other corresponding to "the photon was detected in path SB". These states of the measuring device will be written $|M:RB\rangle$ and $|M:SB\rangle$ respectively. Now a perfect measurement must be such that if the photon is measured in path RB, i.e., if the measuring device state becomes $|M:RB\rangle$ after measurement, then the photon is definitely *not* on path SB. This means the two measurement states must be orthogonal, $\langle M:SB|M:RB\rangle = 0$. However, it is possible to envisage a poor measurement which might indicate a probability of the photon being in path RB, but with some residual possibility of being in path SB. For such an imperfect measurement we would have $\langle M:SB|M:RB\rangle \neq 0$.

The combined state of the photon and the measuring device, at either detector, is thus,

$$|\psi\rangle = \frac{1}{\sqrt{2}}\left(|RB\rangle_j |M:RB\rangle + e^{i\delta_j}|RB\rangle_j |M:SB\rangle\right) \tag{3}$$

The flux of photons into the detectors is thus,

$$\langle\psi|\psi\rangle = 1 + \Re\left\{e^{i\delta_j}\langle M:SB|M:RB\rangle\right\} \tag{4}$$

Consequently, if we have performed a perfect measurement which definitely detects which path the photon took, then, because $\langle M:SB|M:RB\rangle = 0$, we have from (4) that the photon flux is unity, $\langle\psi|\psi\rangle = 1$, into *both* detectors. All interference has been lost.

This establishes quite generally that any means of detecting which path the photon takes around the interferometer will destroy the interference (in the sense that both detectors will detect equal numbers of photons).

However, the occurrence or not of interference is not an all-or-nothing affair. An imperfect measurement, which has $0 < |\langle M:SB|M:RB\rangle| < 1$, will still leave residual interference and hence more photons into detector 1 than into detector 2.

5. The Observation of Counterfactuals: The Elitzur-Vaidman Bomb Test

Everything is now set up for this lovely example. Imagine that you have manufactured a large number of bombs. The bombs are so sensitive that the slightest interaction, say with a single photon, would make the bomb explode. The trouble is that you know that some bombs are duds but you need to identify one which is definitely not a dud.

How on earth can you do this? By the problem statement, the bombs are so sensitive that the slightest physical interaction with any given bomb will make it explode. So you need to determine if a bomb is not a dud without interacting with it at all! Of course you can easily determine which are the duds by poking all the bombs. The ones that don't explode are the duds. But unfortunately that leaves you with no live bombs.

In classical physics the problem is insoluble. But in quantum physics, amazingly, it can be solved.

A bomb is placed[1] within the RB beam path of the interferometer in such a way that a passing photon may or may not interact with it. For example, this might be accomplished by attaching a rod to the mirror at the bottom right such that there was the tiniest gap between the rod and the bomb. If the mirror is free to move when struck by a photon, the rod would then strike the bomb and set if off. (The mechanism is impractical, of course, but the principle is what matters – a truly practical device is perfectly feasible).

---

[1] Of course, you can't actually pick the bomb up and move it, or it would go off. So this is really shorthand for "a Mach-Zehnder interferometer is constructed around a given bomb". You might also be wondering what use these bombs could possibly be, since you could never move them to where you might want to destroy something. Hey, don't take this example so literally!

An active bomb has thus been made a measuring device regarding which of the paths, RB or SB, the photon takes. If it takes path RB, the bomb explodes. If the bomb does not explode either the photon took path SB or the bomb is a dud. The very sensitivity of an active bomb makes it a perfect measuring device.

Now if the bomb is a dud, then the bomb does not constitute a measuring device. A dud bomb might as well not be there. So, with a dud bomb, the photons will always emerge into detector 1, never into detector 2.

But if the bomb is *not* a dud, and assuming it does not explode, then the measurement (of the photon path being SB) destroys the interference and the photon could be detected in either detector 1 or 2. But if it is detected in detector 2 the bomb cannot be a dud! So we have successfully identified a bomb which is definitely not a dud but without it exploding. Miraculous!

The fact that the bomb *might* have gone off, but actually did not, is crucial to the bomb constituting a measuring device and hence to the identification of the unexploded bomb as not being dud. The counterfactual has had an observable consequence, namely that the bomb is now known with certainty to be live.

This curious phenomenon, and the bomb scenario described above, was originally described by Elitzur & Vaidman, Ref.[1]. Do not think that it is too theoretical to be demonstrated experimentally. On the contrary, this was done almost as soon as the effect was discovered, by Zeilinger's group in Vienna, Ref.[2].

As described, the bomb test is terribly inefficient. Of the live bombs, half are exploded, and of the remaining 50% only half of these result in a photon at detector 2 – and hence are identified definitely as live bombs. Thus, a single application of the bomb tester identifies ¼ of the live bombs. But another ¼ still remain unexploded and unidentified. Running these through the bomb tester again results in a further ¼ of this ¼ being identified as live. Hence, repeated applications identify,

$$\tfrac{1}{4} + \tfrac{1}{4} \times \tfrac{1}{4} + \tfrac{1}{4} \times \tfrac{1}{4} \times \tfrac{1}{4} \ldots = {}^{1}/_{3}$$

of the live bombs. But the remaining ${}^{2}/_{3}$ are destroyed – terribly inefficient. However, this is avoidable. In principle virtually all the live bombs can successfully be identified, as shown by Kwiat et al, Ref.[2].

6. References

[1] Elitzur A. C. and Vaidman L. (1993). Quantum mechanical interaction-free measurements. *Found. Phys.* **23**, 987-97. arxiv:hep-th/9305002

[2] P. G. Kwiat, H. Weinfurter, T. Herzog, A. Zeilinger, and M. A. Kasevich (1995). "Interaction-free Measurement". *Phys. Rev. Lett.* **74** (24): 4763.

Appendix

Derivation of the Phase Change Rules for Reflection/Transmission

A.1 Mirrors

Propagation through the glass of a mirror causes a phase change of $e^{ik'x}$ for a distance x of travel, compared with the phase change of e^{ikx} through air. The wave-numbers are related by $k' = nk$, so the phase factor relative to air propagation is $e^{i(n-1)ka} = e^{i\Delta}$, as given in the above text.

If a wave in air, e^{ikx}, meets a glass surface, in which the wave will be $Ce^{ik'x}$, the boundary conditions at the surface are that the wavefunction and its *x*-derivative be continuous. We must also account for a reflected wave, Be^{-ikx}. Hence we require, assuming the surface is at $x = 0$,

$$1 + B = C \qquad \text{and} \qquad k - Bk = k'C \tag{A.1}$$

These equations are readily solved to give,

$$B = -\left(\frac{n-1}{n+1}\right) \qquad \text{and} \qquad C = \frac{2}{n+1} \tag{A.2}$$

Since $n > 1$ it is clear that B is real and negative, corresponding to a phase change with respect to the incident wave of 180^o, a phase factor of -1. This confirms the phase change rule [1] of §2.1, i.e., reflection at the 'front' (silvered) face of a mirror causes a phase change by a factor of -1.

Now consider reflection from the "back" face. If a wave in glass meets the air boundary, the above analysis still applies except that the roles of k and k' are reversed. Hence $k' = k/n$ and hence n is replaced by $1/n$ throughout. So the reflection and transmission coefficients are now respectively,

$$B = +\left(\frac{n-1}{n+1}\right) \qquad \text{and} \qquad C = \frac{2n}{n+1} \tag{A.3}$$

The reflection coefficient is now real and positive, and so there is no phase change between the reflected and incident waves, consistent with rule [2] in §2.1. **QED**.

A.2 Plain Plate

We must now consider transmission through a finite thickness, *a*, of transparent material. The incident plus reflected waves in the region $x < 0$ are $e^{ikx} + Be^{-ikx}$. Within the plate material, $0 < x < a$, the right plus left going waves are $Ee^{ik'x} + Fe^{-ik'x}$. In the region $x > a$ the transmitted wave is Ce^{ikx}. By applying the boundary conditions (continuity of the wavefunction and its x-derivative) at both boundaries $x = 0$ and $x = a$, the coefficients B, E, F, C are found. In particular the ratio of the B and C coefficients is,

$$\frac{B}{C} = i \cdot \left(\frac{n^2 - 1}{2n}\right) \sin k'a \cdot e^{ika} \tag{A.4}$$

Now the factor of e^{ika} just accounts for the fact that the phase of the reflected wave has been reference to position $x = 0$ whereas the transmitted wave phase is referenced to $x = a$. Referencing them both to the same point, as required, leaves the phase factor from (A.4) as just i, as claimed in §2.2. **QED**.