Impossibility of the Counterfactual Computation for All Possible Outcomes

L. Vaidman

School of Physics and Astronomy, Raymond and Beverly Sackler Faculty of Exact Sciences, Tel-Aviv University, Tel-Aviv 69978, Israel

(Received 20 October 2006; revised manuscript received 2 January 2007; published 18 April 2007)

A recent proposal for counterfactual computation [O. Hosten *et al.*, Nature (London) **439**, 949 (2006)] is analyzed. It is argued that the method does not provide counterfactual computation for all possible outcomes. The explanation involves a novel paradoxical feature of pre- and postselected quantum particles: The particle can reach a certain location without being on the path that leads to this location.

DOI: 10.1103/PhysRevLett.98.160403 PACS numbers: 03.65.Ta, 03.67.Lx

The foundations of quantum theory, which is almost 100 years old, are still a subject of heated debate. Where is a quantum particle passing through a Mach-Zehnder interferometer? Is it present in two places simultaneously? If not, how can the interference at the output be explained? Recently, Hosten *et al.* [1] claimed that quantum mechanics allows us to build another counterintuitive device: a computer which yields the outcome of a computation without running. It is an improvement of the proposal of Jozsa [2], who used an interaction-free measurement (IFM) scheme [3] for constructing a computer which could give the outcome without running but only in the case of one particular outcome.

It was claimed [4] that the original IFM can find that there *is* an object in a particular place in an interaction-free manner, but it cannot find that the place is *empty* in the interaction-free way. Also, Jozsa and Mitchison [2,5,6] argued that a computer cannot yield all possible outcomes without running. Hosten *et al.* provided an apparent counterexample to these results. I will argue that one cannot claim that their device provides all possible outcomes without running, but the way it fails to do so yields a novel insight into quantum behavior.

The Hosten *et al.* proposal is based on "chained" Zeno effect improvement [7] of the IFM method. A simple way to describe it is as follows. We have a computer which is a device that performs a computation when a photon enters from an input port. There are only two possible outcomes of the computation: 0 and 1. If the outcome is 0, the photon passes through without any disturbance (or with a known delay which can be compensated). If the outcome is 1, the photon is absorbed by the device.

First, let us describe the original counterfactual computation (CFC) proposal [2]. Two identical optical cavities are connected by a common wall—an almost 100% reflection mirror. During the period of the oscillation in a cavity, a photon starting in the left (L) cavity evolves into a superposition of being in the left and in the right (R) cavities according to the following law:

$$\begin{aligned} |L\rangle &\rightarrow \cos\alpha|L\rangle + \sin\alpha|R\rangle, \\ |R\rangle &\rightarrow -\sin\alpha|L\rangle + \cos\alpha|R\rangle, \qquad \alpha \ll 1. \end{aligned} \tag{1}$$

The computer device is placed inside the right cavity. A single photon localized wave packet starts in cavity L, and it is left to evolve N periods of the oscillation of a single cavity. It is arranged that $N\alpha = \frac{\pi}{2}$.

Assume first that the outcome is 0. After one period, the state of the photon is $\cos\alpha|L\rangle + \sin\alpha|R\rangle$. It is easy to see that, after n periods, the state is $\cos n\alpha|L\rangle + \sin n\alpha|R\rangle$. When the number of periods equals $N = \frac{\pi}{2\alpha}$, the final state of the photon is $|R\rangle$.

Now assume that the outcome is 1. After one period, the state of the photon is $\cos\alpha|L\rangle + \sin\alpha|\mathrm{abs}1\rangle$. The state $|\mathrm{abs}1\rangle$ signifies the photon absorbed by the computer device on the first round. (The process might include macroscopic amplification, in which case the situation is usually described as a collapse to the state $|L\rangle$, with probability $\cos^2\alpha$, or collapse to the absorption of the photon, with probability $\sin^2\alpha$.) After N periods, the state is

$$\cos^N\alpha|L\rangle + \sum_{n=1}^{N} \cos^{n-1}\alpha \sin\alpha|\mathrm{abs}n\rangle. \tag{2}$$

The final step of the procedure is a measurement which tests the presence of the photon in cavity L. If we find the photon there, we know that the outcome is 1, since, if the outcome is 0, after N periods it has to be in cavity R. Also, we know that it was computed counterfactually, since computation of the outcome 1 ends by the absorption of the photon by the computing device, but the photon was detected by another detector.

If we do not find the photon in cavity L, we know that there is a high probability that the outcome is 0, but there is not any reason to claim that in this case there has been a counterfactual computation, since at every period part of the photon wave went through the computer device.

The probability to absorb the photon by the computer in the case where the outcome is 1 is $1 - \cos^{2N}\frac{\pi}{2N} \approx \pi^2/4N$, so we have constructed a CFC procedure for a single outcome with efficiency which can be arbitrarily close to 100%.

In order to perform a counterfactual computation for both outcomes, Hosten *et al.* proposed to construct three identical optical cavities using two identical almost 100%

reflection mirrors. The computer device is placed inside the third cavity; see Fig. 1.

Again, a single photon localized wave packet starts in cavity A and moves toward the almost 100% reflection mirror, which separates cavity A from cavity B. After it bounces on the mirror transmitting a localized packet of an amplitude $\sin\alpha$ to the cavity B, the mirror between cavities A and B is switched to be 100% reflective. It remains 100% reflective for N periods. During this time, the part of the photon wave in cavity A remains as it is, and the part in cavity B evolves as in the previous two-cavity case. Thus, if the outcome is 0, the state of the photon becomes

$$\cos\alpha|A\rangle + \sin\alpha|C\rangle, \tag{3}$$

and if the outcome is 1, the state of the photon becomes

$$\cos\alpha|A\rangle + \sin\alpha\left(\cos^N\alpha|B\rangle + \sum_{n=1}^{N}\cos^{n-1}\alpha\,\sin\alpha|\mathrm{abs}n\rangle\right). \tag{4}$$

At this stage, a measurement is performed which tests the presence of the photon in cavity C. The probability to find the photon in this measurement does not vanish only if the outcome is 0, but even then it is very small: $p_{\mathrm{fail}} \simeq \pi^2/4N^2$. If we find the photon, we know the outcome of the computation, but CFC fails, since the photon did pass through the computer. If the outcome is 1, we have an even smaller probability of the CFC failure, i.e., absorption of the photon by the computer device. If the photon was not absorbed, then the state of the photon in the case of outcome 0 is just $|A\rangle$, and in the case of outcome 1, the state is, up to normalization, $\cos\alpha|A\rangle + \sin\alpha\cos^N\alpha|B\rangle$. This ends the first subroutine of the computation process.

Now the mirror between cavities A and B is opened again for one bounce of the photon; i.e., it leads to the evolution described by (1), after which it is closed for N periods as before. The second run of the subroutine, as the first one, ends by the test of the presence of the photon in C. If the outcome is 0, everything is the same as in the first subroutine: The probability of the failure of the CFC is again $p_{\mathrm{fail}} \approx \pi^2/4N^2$, and if the photon is not found in C, its final state is $|A\rangle$. If the outcome is 1, we have a tiny probability for a failure, and if the photon is not absorbed by the computer, the final state of the photon is, up to normalization, $(\cos^2\alpha - \sin^2\alpha\cos^N\alpha)|A\rangle + (\sin\alpha\cos^{N+1}\alpha + \sin\alpha\cos^{2N+1}\alpha)|B\rangle$. The state is approximately equal to $\cos 2\alpha|A\rangle + \sin 2\alpha|B\rangle$. After N rounds, the state is approximately equal to $\cos N\alpha|A\rangle + \sin N\alpha|B\rangle = |B\rangle$.

Under this approximation, after N rounds of the subroutine, we end up with the state $|A\rangle$ if the outcome is 0 and the state $|B\rangle$ if the outcome is 1. The parameter α can and should be tuned a little, such that this will be an exact and not an approximate statement—except for a possibility of the failure which can be made arbitrarily small, since it is of the order of $\frac{1}{N}$. The whole procedure ends with the test in which cavity, A or B, the photon is located, and this yields the outcome of the computation: 0 or 1.

It seems that it is a counterfactual computation. Indeed, if the outcome is 1, we know that the photon was not inside the computer (otherwise, it would be absorbed), and, if the outcome is 0, we also apparently can claim that the photon was not inside the computer, because our tests of the cavity

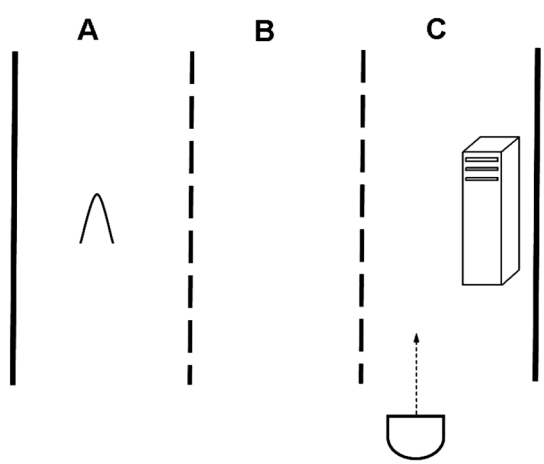

FIG. 1. Hosten *et al.* counterfactual computation method. If the outcome is 0, the photon remains in cavity A, and if the outcome is 1, the photon moves to cavity B.

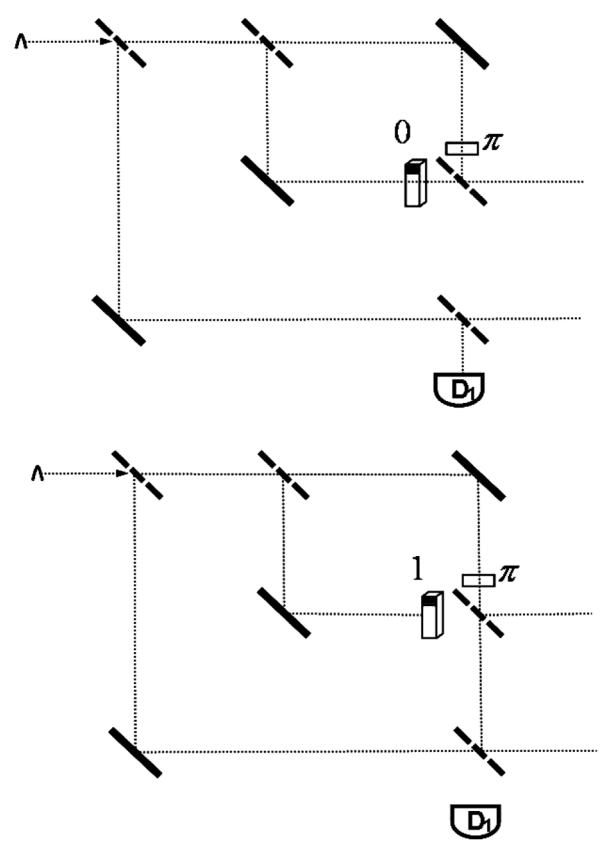

FIG. 2. The Hosten *et al.* counterfactual computation method for outcome 0. The top figure shows that if the outcome is 0, then the photon entering the inner interferometer cannot reach detector D_1, and the bottom figure shows that if the outcome is 1, then no photon at all can reach detector D_1.

C at the end of each subroutine checked every time when the photon could enter the computer, and we found that it did not.

The core of the controversy is that it is possible to make CFC when the process of computation is just the passage of the photon through the device without any change in the device and without absorption of the photon; it corresponds to the CFC of the outcome 0 in the examples above. So, following Hosten *et al.* [8], let us consider a simpler scheme. It is not optimally efficient, and it makes a decisive computation only in the case of the outcome 0.

The scheme consists of one Mach-Zehnder interferometer (MZI) nested inside another and the computer placed in one arm of the inner interferometer; see Fig. 2. The inner interferometer is tuned by a π phase shifter in such a way that if the outcome is 0, i.e., the computer is transparent, there is a destructive interference towards the output beam splitter of the large interferometer. If the outcome is 1, the photon can reach the output beam splitter of the external MZI, and it is tuned in such a way that there is a destructive interference toward detector D_1.

A possible implementation of this system is that the external MZI has two identical beam splitters which transmit two-thirds of the beam and reflect one-third. The evolution of the photon passing through such a beam splitter is

$$|H\rangle \rightarrow \sqrt{\frac{2}{3}}|H\rangle + \sqrt{\frac{1}{3}}|V\rangle, \qquad |V\rangle \rightarrow -\sqrt{\frac{1}{3}}|H\rangle + \sqrt{\frac{2}{3}}|V\rangle, \quad (5)$$

where H and V signify horizontal and vertical modes, respectively. The inner MZI has two identical half-and-half beam splitters. The evolution of the photon passing through such a beam splitter is

$$|H\rangle \rightarrow \sqrt{\frac{1}{2}}|H\rangle + \sqrt{\frac{1}{2}}|V\rangle, \qquad |V\rangle \rightarrow -\sqrt{\frac{1}{2}}|H\rangle + \sqrt{\frac{1}{2}}|V\rangle. \quad (6)$$

A simple calculation shows that, indeed, there is a destructive interference toward the vertical output mode of the inner interferometer if the computer is transparent, and there is a destructive interference toward D_1 if the lower arm of the inner interferometer is blocked by the computer.

Consider now a run of our device in which a single photon enters the interferometer and is detected by detector D_1. In this case, we get the information that the result of the computation is 0, and it is apparently a counterfactual computation. It seems that the photon has not passed through the computer, since photons passing through the inner interferometer, where the computer is located, cannot reach detector D_1.

In spite of the fact that it is a very vivid and persuasive explanation, I will argue that it is incorrect. The photon detected at outport D_1 is a quantum pre- and postselected system, and what is correct for classical systems and quantum preselected only systems might be wrong for a pre- and postselected system.

Let us consider an example of pre- and postselected systems, usually known as the "three-box paradox" [9]. A single particle is preselected in the state

$$|\psi\rangle = \frac{1}{\sqrt{3}}(|A\rangle + |B\rangle + |C\rangle) \quad (7)$$

and postselected in the state

$$\langle\phi| = \frac{1}{\sqrt{3}}(\langle A| + \langle B| - \langle C|), \quad (8)$$

where the mutually orthogonal states $|A\rangle$, $|B\rangle$, and $|C\rangle$ denote the particle being in box A, B, and C, respectively. (We use "bra" and "ket" notation to distinguish between standard, forward evolving quantum state, and backward evolving quantum state from the postselection measurement.)

The paradox is that, at the intermediate time, the particle is to be found with certainty in box A if searched there, and, at the same time, it is to be found with certainty in box B if it is searched there instead. Indeed, if a particle is not found, e.g., in box B, then its state collapses to $(1/\sqrt{2}) \times (|A\rangle + |C\rangle)$, but this is impossible since this state is orthogonal to the postselected state $\langle\phi|$.

One thing is obvious in this example: There are no grounds to say that the particle was not present at box B. Nevertheless, I can "show" that it was not present in B using similar arguments to those showing that the photon was not passing through the computer in the Hosten *et al.* example.

In fact, the Hosten *et al.* setup is an implementation of the three-box experiment. Consider three boxes on the way

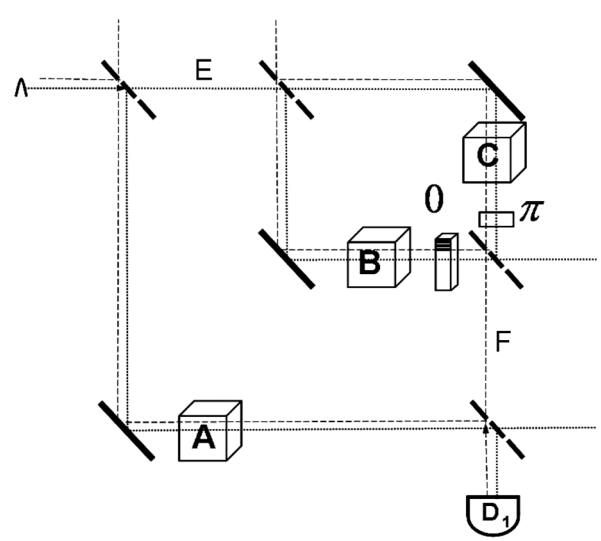

FIG. 3. Both forward and backward evolving quantum states are present in all three boxes: A, B, and C. They are described by $|\psi\rangle$ and $\langle\phi|$; see (7) and (8). However, on the way toward the inner interferometer, the backward evolving wave is not present, and on the way out, the forward evolving wave is not present.

of the photon in the three arms of the nested MZI experiment (Fig. 3). Taking into account the evolution laws (5) and (6), we see that, indeed, the quantum state of the photon in these boxes is described by $|\psi\rangle$, and, given that the computer is transparent, detection by D_1 corresponds to postselection of the state $\langle\phi|$.

According to Hosten *et al.*, box B is empty. But this argumentation is not tenable. Indeed, their argument applies also to box C, so it should be empty, too. The pre- and postselection states are symmetric under interchange of boxes A and B. Therefore, box A should be empty as well, but then where is the particle? This shows that the Hosten *et al.* argument leads to a contradiction, and, therefore, their conclusion regarding the counterfactual nature of computation is not warranted.

Let us now ask: What can we learn from experiments testing the location of the photon in the Hosten *et al.* setup? We know that the photon is to be found with certainty if searched in B. One can argue, however, that a strong nondemolition measurement of $\mathbf{P}_B$, the projection on B, changes the physical situation. Then we can perform a *weak measurement* [10] of $\mathbf{P}_B$ which requires an ensemble of N pre- and postselected particles. The weak measurement, at the limit of large N, does not change either the forward evolving state $|\psi\rangle$ or the backward evolving state $\langle\phi|$. According to the Hosten *et al.* argument, all members of the ensemble are not in B, so we should not see any effect in the measurement at B. However, the weak measurement of the projection onto the "computer" will show a different result. The outcome of the weak measurement is the *weak value*: $(\mathbf{P}_B)_w \equiv \langle\phi|\mathbf{P}_B|\psi\rangle/\langle\phi|\psi\rangle = 1$, while simultaneous weak measurements of the projections on the paths E and F, which lead to and from the inner interferometer, yield $(\mathbf{P}_E)_w = (\mathbf{P}_F)_w = 0$. This is not necessarily a gedanken experiment; Resch *et al.* performed the weak measurement of the projection in the three-box problem [11].

In the framework of these concepts we can state the following: The photon did not enter the interferometer, the photon never left the interferometer, but it was there. This is a new paradoxical feature of a pre- and postselected quantum particle.

Hosten and Kwiat [12] posed a question about the small, yet unavoidable disturbance due to weak measurements. To what extent does the weak measurement at B disturb the destructive interference in F, which is the basis of the Hosten *et al.* argument? If we perform a practical weak measurement procedure *à la* Resch *et al.*, in which each photon has its own measuring device (its transversal location), then finding a precise weak value requires strength of the interaction proportional to $1/\sqrt{N}$ and, thus, a flux through F of a number of photons. Still, the flux through F is negligible compare to the total flux, so that the destructive interference in F for every photon is almost complete. Moreover, for the conceptual issue discussed here, we can consider a gedanken experiment in which we use an external measuring device interacting very weakly with all of the photons and a rare quantum event in which all photons are postselected in D_1. In this experiment, a precise weak value of the projection on B can be obtained with interaction strength proportional to $\frac{1}{N}$, and, thus, a flux through F is much less than one photon. The operational meaning of this statement is that strong nondemolition measurements along path F have negligible probability to find even one photon during the whole process. It should be mentioned that the disturbing effect of the strong measurement at F on the weak measurement at B is not negligible at all: Although the measurement at F does not change the forward evolving quantum state at B, it nullifies the backward evolving state, causing the weak value of the projection on B to vanish.

Note that Bohmian interpretation of quantum mechanics does not exhibit such a behavior and, in fact, supports the claim of Hosten *et al.* A Bohmian particle has to "ride" on a quantum wave, but there is no quantum wave at path F, from the inner interferometer to D_1. Therefore, the (Bohmian) particle detected by D_1 did not pass through the interferometer and the computer.

Discarding the hidden variables approach, we should be able to answer any question based on the complete description of a quantum system which consists here of both forward and backward evolving quantum states. The answer cannot depend on the particular way of preparing and postselecting these states, as happens if we adopt the Hosten *et al.* approach.

It is a pleasure to thank Zhe-Xuan Gong for helpful discussions. This work has been supported in part by the European Commission under the Integrated Project Qubit Applications (QAP) funded by the IST directorate as Contract No. 015848 and by Grant No. 990/06 of the Israel Science Foundation.

[1] O. Hosten *et al.*, Nature (London) **439**, 949 (2006).
[2] R. Jozsa, in *Lecture Notes in Computer Science*, edited by C. P. Williams (Springer, London, 1998), Vol. 1509, p. 103.
[3] A. C. Elitzur and L. Vaidman, Found. Phys. **23**, 987 (1993).
[4] L. Vaidman, Found. Phys. **33**, 491 (2003).
[5] G. Mitchison and R. Jozsa, Proc. R. Soc. A **457**, 1175 (2001).
[6] G. Mitchison and R. Jozsa, quant-ph/0606092.
[7] P. G. Kwiat *et al.*, Phys. Rev. Lett. **83**, 4725 (1999).
[8] O. Hosten *et al.*, quant-ph/0607101.
[9] Y. Aharonov and L. Vaidman, J. Phys. A **24**, 2315 (1991).
[10] Y. Aharonov and L. Vaidman, Phys. Rev. A **41**, 11 (1990).
[11] K. J. Resch *et al.*, Phys. Lett. A **324**, 125 (2004).
[12] O. Hosten and P. G. Kwiat, quant-ph/0612159.