A practical Trojan Horse for Bell-inequality-based quantum cryptography

Jan-Åke Larsson

Quantum Information and Computation, Vol. 2, No. 6 (2002) 434–442

A PRACTICAL TROJAN HORSE FOR BELL-INEQUALITY-BASED QUANTUM CRYPTOGRAPHY

JAN-ÅKE LARSSON*
MaPhySto, Department of Mathematical Sciences, University of Aarhus, Ny Munkegade, DK-8000 Aarhus C, Denmark

Received January 10, 2002
Revised August 20, 2002

Quantum Cryptography, or more accurately, Quantum Key Distribution (QKD) is based on using an unconditionally secure "quantum channel" to share a secret key among two users. A manufacturer of QKD devices could, intentionally or not, use a (semi-)classical channel instead of the quantum channel, which would remove the supposedly unconditional security. One example is the BB84 protocol, where the quantum channel can be implemented in polarization of single photons. Here, use of several photons instead of one to encode each bit of the key provides a similar but insecure system. For protocols based on violation of a Bell inequality (e.g., the Ekert protocol) the situation is somewhat different. While the possibility is mentioned by some authors, it is generally thought that an implementation of a (semi-)classical channel will differ significantly from that of a quantum channel. Here, a counterexample will be given using an identical physical setup as is used in photon-polarization Ekert QKD. Since the physical implementation is identical, a manufacturer may include this modification as a Trojan Horse in manufactured systems, to be activated at will by an eavesdropper. Thus, the old truth of cryptography still holds: you have to trust the manufacturer of your cryptographic device. Even when you do violate the Bell inequality.

Keywords: Quantum cryptography, Trojan horse, Ekert protocol, Bell inequality
Communicated by: H-K Lo & H Zbinden

1. Introduction

QKD is an application of quantum techniques proposed to give a way of securely sharing a secret key between two or more users. The appeal of this technique is that the security is based on laws of nature rather than number-theoretical calculations that seem, and perhaps are, intractable. Progress is quite rapid in this field [1], experimental implementations make the necessary technological advances available [2, 3] while theoretical analysis provides better insight into security questions under more and more general conditions [4, 5, 6]. And recently (spring 2002), commercial QKD systems have become available. Let us suppose that Alice and Bob have bought such a commercial QKD system, intending to use it to share a crypto key secretly from Eve. Now, it is possible that they have bought a system which claims to be and behaves similarly to a QKD system but actually is a (semi-)classical key distribution

*Present address: Dept. of Math., Linköping University, SE-581 83 Linköping, Sweden.
E-mail: jalar@mai.liu.se

system. It would then not provide any security of the sort inferred from a proper QKD system. Let us look at an example.

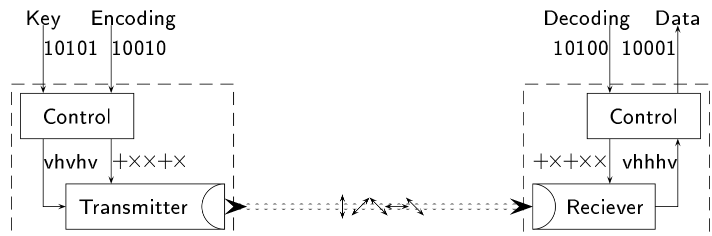

Fig. 1. An example of Quantum Key Distribution: BB84. Alice generates two random bit-sequences, one as the basis of the secret key, and another as the encoding for the secret key in the quantum channel. Alice encodes each bit of the key, here in the polarization state of a photon, following a certain scheme: If the corresponding encoding-bit is 0, Alice uses 0 = horizontal, and 1 = vertical, whereas if the encoding-bit is 1, Alice uses the same encoding in a 45° rotated frame. The photons are transmitted to Bob who uses a third random bit-sequence to decode the bit-sequence with the same scheme as Alice. This means that Bob will have used the same de/encoding as Alice only half of the time, but by communicating the *settings* used to each other via an unjammable classical channel, they can establish for which bits they used the same settings. These bits can then be used as cryptographic key.

In the BB84 protocol [7], security is based on transmitting the key from Alice to Bob through a quantum channel in such a way that attempts to eavesdrop is directly detectable as an increased noise-level in the transmitted key (see Figure 1). The security is here provided by the quantum-mechanical nature of the photon. Eve cannot use a beamsplitter to tap off part of the signal for herself (the so-called "Beamsplitter attack"); a single photon cannot be split in such a manner that Bob doesn't notice [8]. Eve cannot either faithfully copy the polarization state of the photon, due to the no-cloning theorem [9]. Lastly, because of the encoding scheme described above, Eve cannot measure the polarization and then retransmit another photon to Bob (the "Intercept-resend attack"). If she tries, there will be an increased error rate in the data received by Bob. Alice and Bob will detect this by sacrificing a portion of the key to estimate the error rate. Then, if the key is not too noisy, Alice and Bob can use classical privacy amplification [10, 11]. Otherwise, they will be forced to abandon the QKD as being compromised.

This provides good security given that the quantum channel really is quantum. By constructing the device used by Alice to send several photons for each key bit, the key can (somewhat simplified) be successfully extracted by Eve using the Beamsplitter attack. This will not introduce noise in Bob's data, and thus, there is no security provided by such a system [12]. Many devices used in practical implementations does not use single photons but weak laser pulses for which less than one photon is present in each pulse on the average, and present research indicates that these devices provide the expected security only under certain restrictions [13].

Here, we will concentrate on another system, less discussed in this context: the Ekert protocol [14] (see Figure 2). Security is here provided by the quantum-mechanical properties of entangled photon pairs. In this protocol there is less probability that the settings are the same, but there is a compensation since one does not need to sacrifice key bits for eavesdropper detection. One uses the left-over bits instead, whose correlation provides a violation of a Bell

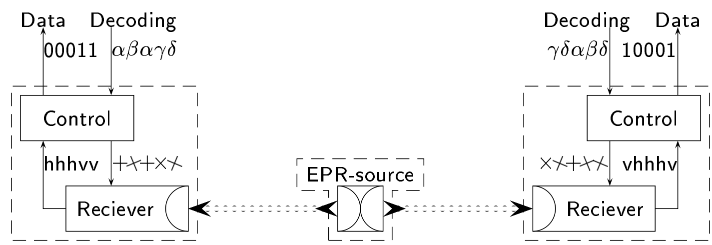

Fig. 2. The Ekert protocol: The quantum channel contains a bidirectional source and a random key is generated at detection at Alice and Bob. In the photon-polarization case, the protocol is as follows: the source sends one photon to Alice and one to Bob in an entangled state, an EPR-Bohm pair [16, 17], which is such that equally oriented polarization measurements yield identical results. Both receivers have, in this case, four settings ($\alpha = 0°$, $\beta = 22.5°$, $\gamma = 45°$, $\delta = 67.5°$) and Alice an Bob use random independent settings at their respective sites. Again, by communicating on an unjammable classical channel, they can establish for which bits the settings were the same. These bits provide the key.

inequality [15]. The correlation is

$$E(\phi,\varphi) = \frac{N_{\text{same}}(\phi,\varphi) - N_{\text{different}}(\phi,\varphi)}{N_{\text{same}}(\phi,\varphi) + N_{\text{different}}(\phi,\varphi)}, \tag{1}$$

where ϕ and φ are the settings at Alice and Bob, respectively, and N_{same} and $N_{\text{different}}$ are the number of photon pairs for which the results are the same and different, respectively. It is known that an EPR-Bohm pair [16, 17] yields correlations that violate the CHSH form of the Bell inequality [18]:

$$|E(\alpha,\beta) + E(\gamma,\beta)| + |E(\gamma,\delta) - E(\alpha,\delta)| \leq 2. \tag{2}$$

If the difference between Alice's and Bob's setting is 22.5° (as in $E(\alpha,\beta)$, $E(\gamma,\beta)$ and $E(\gamma,\delta)$), the correlation between the results is $1/\sqrt{2} \approx 0.7071$, and if the difference is 67.5° (as in $E(\alpha,\delta)$), the correlation is $-1/\sqrt{2}$. This yields a left-hand side of $2\sqrt{2}$ in (2), i.e., a violation of the CHSH inequality. An attempt to eavesdrop will establish EPR-elements of reality, and by the local random choice of settings at the detection sites, the resulting correlation must obey the inequality (2). Thus, the eavesdropper's presence can be detected by checking the violation [14]. If it is sufficient, classical privacy amplification can be used, but in the worst case, the QKD must be abandoned as being compromised.

2. A Trojan Horse

It would seem that violation of the Bell inequality ensures that a quantum channel really is used, since by the violation, there *can be no pre-existing key that can be eavesdropped upon.* Unfortunately, this is not entirely true. There are situations which allow for a pre-existing key and random local choices of the settings *while still violating the Bell inequality* (formally). This is related to the so-called detector-efficiency loophole discussed by many authors (e.g. [1, 19, 20, 21] and references therein). In what follows, a "cheat" will be outlined, using the same physical setup as in the true Ekert QKD sketched above, only changing the "Control" boxes of Figure 2. The reason for using the word "cheat" is that it is unlikely that a manufacturer would make a device according to the below specification while believing that it is a true QKD system (but see the tempting "Really cheap KD" in Appendix A).

The first task will be to choose the (semi-)classical channel to be used instead of the polarization state of each photon. Let us use the timing of events as our "hidden variable"[a] (this will be especially simple if we use a pulsed variant of the common parametric down-conversion source [22]). At the start of the protocol, the devices are initiated so that they are in sync, say by a bright light-pulse from the source. After this initiation, let us establish a numbering of time slots, say 8000 short time slots (repeated if necessary). The controlling electronics is now changed so that instead of using the polarization measurement result, the received time slot is mapped into the translation tables in Figure 3 to obtain a measurement result.

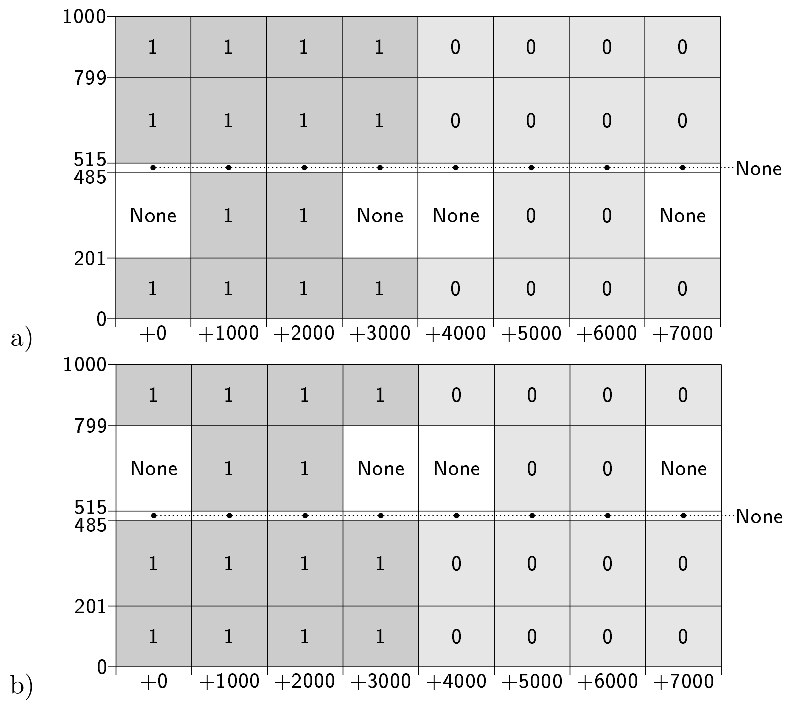

Fig. 3. Translation tables from time slot number to measurement result to be used for the $\alpha = 0°$setting. Each $22.5°$ shift corresponds to a shift of 1000 in the numbering of the time slots (wrapped if it exceeds 8000). a) Translation at Alice's detector. For example, a photon arriving at Alice in time slot 2223 would give the result "1" at the setting $\alpha = 0°$. Had the setting been $\beta = 22.5°$, the controlling electronics would have shifted the slot to $2223 + 1000 = 3223$, and there would have been no detection reported. Similarly, there would have been no detection at $\gamma = 45°$, and a "0" at $\delta = 67.5°$. b) Translation at Bob's detector.

Some received photons will register as "None" in the table, and will not be reported as bits in the output data, so the resulting efficiency of the detector array will be lowered somewhat. If the photon pairs are evenly distributed over the 8000 (repeated) slots, the overall rate of

[a]The recent claim made by Hess and Phillip [23, 24, 25], that time dependence has been neglected in the Bell theorem(s) may seem similar to this "cheat". However, their claim is stronger and more controversial, since they claim their model works with ideal detectors. Their construction is quite different than the one used here, and by all evidence their model is non-local. In other words, it cannot be used for the present task; it cannot violate the Bell inequality while being separated into the three "boxes": Source, Alice's detector, and Bob's detector, unless the settings are communicated *before* the measurements.

reported detections will be decreased somewhat. Of the 8000 slots, there are $4*201+12*485$ for which there is a result reported, so

$$P\big(\text{ bit reported} \;\big|\; \text{single detection }\big) = \frac{4*201+12*485}{8000} = 0.828. \tag{3}$$

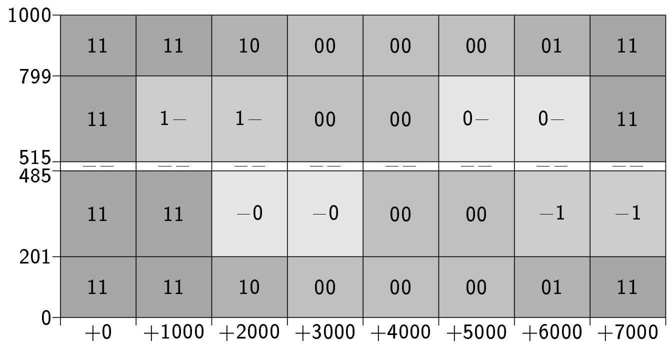

Fig. 4. An example of a setup where Alice has chosen the $\beta = 22.5°$setting and Bob has chosen the $\gamma = 45°$setting. For example, a photon arriving at the slot 2810 will give a 1 at Alice and a 0 at Bob, while a pair arriving in the slot 5560 will give a 0 at Alice and no result at Bob.

In Figure 4, we can see an example of simultaneous results obtained by Alice and Bob for a certain setting at each side. Clearly, for a pair to be reported, the photon pair needs to be detected in a slot where results are reported on both sides. In a similar fashion as above, there are $8*201+8*485$ such slots, and thus,

$$P\big(\text{ pair reported} \;\big|\; \text{pair detection }\big) = \frac{8*201+8*485}{8000} = 0.686. \tag{4}$$

An important property of these probabilities of single photon detection and pair detection is that they do not vary with the settings. Also, these events are approximately statistically independent since $0.828^2 \approx 0.686$; the approximation improves with a larger table. In a realistic situation, if the efficiency is 5% [26] the modified efficiency would be roughly 4%, which would not cause great concern in a commercial system. The reader may recognize (3) as being just below the efficiency bound for the CHSH inequality.

The obtained results are always the same if the settings are equal. A knowledgeable user would be concerned with the absence of noise in this particular implementation, but it is easy to introduce an appropriate amount of artificial noise in the translation tables of Figure 3, or use the modified scheme discussed below. If the "orientations" differ by $\pm 22.5°$ (e.g., as in $E(\alpha,\beta)$), the correlation is

$$E(\alpha,\beta) = \frac{(4*201+8*485)-4*201}{8*201+8*485} = \frac{485}{686} \approx 0.7070. \tag{5}$$

The "error" relative to the true QM value $1/\sqrt{2}$ is approximately $1.1*10^{-4}$, which can be "improved" by using a larger table. A similar calculation yields $E(\alpha,\delta) \approx -0.7070$, and these correlations will yield a violation of the CHSH inequality. Even when there are EPR-elements of reality (there are here!) and the settings are chosen locally and at random.

While this is not a quantum system, the obtained statistics mimics one: the Ekert protocol will give the users *an insecure key* while *the Bell inequality will be violated*. An eavesdropper needs not, in this case, access the quantum channel since simple knowledge of the translation tables together with the public communication between Alice and Bob will give Eve all she needs to obtain the key.

One obvious shortcoming of this simple setup is that the results will show a strong correlation to the time slot number in which the result was obtained, and detection or no-detection will show a weak correlation. Also, the devices do not need to be oriented manually when the system is first set up. In real QKD, at least in this implementation, the bits correspond to some planar polarization state of a photon. The cheat may now be slightly modified to use polarization for one of the four settings, letting the result keep or invert the translation table, depending on if the result was “horizontal” or “vertical”. This would require manual orientation of the detectors, and would altogether remove the result correlation to the time slot number. In addition, it would also require Eve to do a simplified intercept-resend on the quantum channel.

3. Discussion

It is possible to continue the discussion of these and other tests and ways to counter them, but our space is limited. More importantly, such tests may not reveal any misbehavior from the receiving devices. This is because the physical implementation described above is identical to the one used in true QKD, only changing the mapping from measurement results to output data. The devices may operate in “QKD mode” when purchased and provide a true QKD system, but e.g. when fed a certain initiation pulse, change into “Trojan mode”. The users does not notice this, since the data they obtain follow the same statistics (even the lowered efficiency of the Trojan mode can be masked by a permanent artificial lowering of the efficiency in QKD mode). An eavesdropper may now “tap” the key distribution scheme by inserting appropriate devices that emit this special initiation pulse when the normal initiation pulse is received from the source. Having switched the receptors of Alice and Bob into Trojan mode, the eavesdropper can listen to their conversation unhindered. And devices like this can be changed from QKD mode to Trojan mode at will by the eavesdropper.

In principle, a Trojan of this type can be identified but that would require reverse-engineering the receiving devices. A complete dissemination of the internals, including the content of any ROM or flash RAM present, will show whether the system is a true QKD implementation or if it does contain a Trojan of the above type. It is not enough to examine the device superficially, since the Trojan is such that only the internals (the program) of the “Control” is changed (see Figure 2). This will unfortunately be difficult for a normal user. Furthermore, one-chip devices, normally intended to make the devices cheaper, will make it difficult even for a specialist. Of course, a successful extraction of the key while still violating the Bell inequality would be direct proof of a Trojan, but one would need to activate the Trojan first, and this is not easy since the activation (if it exists) is unknown to Alice and Bob.

Thus, to determine whether a system is true QKD or not will require advanced testing, and users of commercial QKD devices cannot generally be expected to have access to the technology needed. Indeed, if they did, they would be able to build their own QKD system rather than buying one. It would be possible to defer the tests to a certifying authority, but in the present world, perhaps this is not a great improvement: a potential user would then need to trust the certifying authority rather than the manufacturer.

It would naturally be desirable to close the detector-efficiency loophole in QKD implementations. The recent result by Rowe *et al* [19] shows promise, but is not yet usable in QKD setups. Furthermore, some present implementations use Franson interferometry [1, 27] in which the quantum channel is established using time-energy entanglement instead of the polarization entanglement used here. And there is a local realistic model of the Franson setup even in the ideal case [28]. In other words, the model works even at 100% efficiency, and it would be simple to use this this model to create a Trojan for the corresponding QKD scheme, along the lines indicated here.

In the case of BB84 QKD, it has been suggested that a self-checking source can be constructed via a certain experimental setup and a Bell inequality test [29]. The considerations here should be taken into account when testing and using such a source. For instance, one may consider buying the source and the testing equipment from different manufacturers (but this would be only slightly better than buying both from one manufacturer).

True QKD does have many good properties, and the Trojan described does not remove any of these good properties. However, users must make sure that the system they intend to use really is a true QKD system. Ultimately, the old truth of cryptography still holds: you need to trust the manufacturer of your cryptographic device. Even when you do violate the Bell inequality.

Acknowledgments

This work was supported by the Wenner-Gren foundation and the Royal Swedish Academy of Science. MaPhySto, The Centre for Mathematical Physics and Stochastics, is funded by the Danish National Research Foundation.

References

1. Nicolas Gisin, Grégorie Ribordy, Wolfgang Tittel, and Hugo Zbinden (2002), *Quantum cryptography*, Rev. Mod. Phys. 74:145–195.
2. Grégoire Ribordy, Jürgen Brendel, Jean-Daniel Gautier, Nicolas Gisin, and Hugo Zbinden (2001), *Long-distance entanglement-based quantum key distribution*, Phys. Rev. A 63:012309–1–12.
3. W. T. Buttler, R. J. Hughes, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. G. Peterson (2000), *Daylight quantum key distribution over 1.6 km*, Phys. Rev. Lett. 84:5652–5655.
4. D. Mayers (1996), *Quantum key distribution and string oblivious transfer in noisy channels*, in *Advances in cryptology — Proceedings of Crypto '96*, Springer, pp. 343–357.
5. Peter W. Shor and John Preskill (2000), *Simple proof of security of the BB84 quantum key distribution protocol*, Phys. Rev. Lett. 85:441–444.
6. D. S. Naik, C. G. Peterson, A. G. White, A. J. Berglund, and P. G. Kwiat (2000), *Entangled state quantum cryptography: Eavesdropping on the Ekert protocol*, Phys. Rev. Lett. 84:4733–4736.
7. C. H. Bennett and G. Brassard (1984), *Quantum cryptography: Public key distribution and coin tossing*, in *Proc. of the IEEE Int. Conf. on Computers, Systems, and Signal Processing, Bangalore, India*, IEEE (New York), pp. 175–179.
8. John F. Clauser (1974), *Experimental distinction between the quantum and classical field-theoretic predictions for the photoelectric effect*, Phys. Rev. D 9:853–860.
9. W. K. Wooters and W. H. Zurek (1984), *A single quantum cannot be cloned*, Nature 299:802–803.
10. C. H. Bennett, G. Brassard, and J.-M. Robert (1986), *How to reduce your enemy's information*,

in *Advances in Cryptology — Proceedings of Crypto '85* (Lecture Notes in Computer Science, vol. 218), Springer (Berlin), pp. 468–476.
11. C. H. Bennett, G. Brassard, C. Crepeau, and U. M. Maurer (1995), *Generalized privacy amplification*, IEEE Transactions on Information Theory, 41:1915–1923
12. C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin (1992), *Experimental quantum cryptography*, J. Crypto. 5:3–28.
13. Gilles Brassard, Norbert Lütkenhaus, Tal Mor, and Barry C. Sanders (2000), *Limitations on practical quantum cryptography*, Phys. Rev. Lett. 85:1330–1333.
14. Artur K. Ekert (1991), *Quantum cryptography based on Bell's theorem*, Phys. Rev. Lett. 67:661–663.
15. J. S. Bell (1964), *On the Einstein-Podolsky-Rosen paradox*, Physics 1:195–200.
16. A. Einstein, B. Podolsky, and N. Rosen (1935), *Can quantum-mechanical description of physical reality be considered complete?*, Phys. Rev. 47:777–780.
17. David Bohm (1951) *Quantum Theory*, Prentice-Hall (New York).
18. John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt (1969), *Proposed experiment to test local hidden-variable theories*, Phys. Rev. Lett. 23:880–884.
19. M. A. Rowe, D. Kielpinski, V. Meyer, C. A. Scakett, W. M. Itano, C. Monroe, and D. J. Wineland (2001), *Experimental violation of a Bell's inequality with efficient detection*, Nature 409:791–794.
20. Jan-Åke Larsson (1998), *The Bell inequality and detector inefficiency*, Phys. Rev. A 57:3304–3308.
21. Jan-Åke Larsson (1999), *Modeling the singlet state with local variables*, Phys. Lett. A 256:245–252.
22. P. G. Kwiat, K. Mattle, H. Weinfurter, A. Zeilinger, A. V. Sergienko, and Y. Shih (1995), *New high-intensity source of polarization-entangled photon pairs*, Phys. Rev. Lett. 75:4337–4341.
23. K. Hess and W. Phillip (2001), *A possible loophole in the theorem of Bell*, Proc. Nat. Acad. Sci. USA 98:14224–14227.
24. K. Hess and W. Phillip (2001), *Bell's theorem and the problem of decidability between the views of Einstein and Bohr*, Proc. Nat. Acad. Sci. USA 98:14228–14233.
25. K. Hess and W. Phillip (2002), *Exclusion of time in the theorem of Bell*, Europhys. Lett. 57:775–781.
26. Gregor Weihs, Thomas Jennewein, Christoph Simon, Harald Weinfurter, and Anton Zeilinger (1998), *Violation of Bell's inequality under strict Einstein locality conditions*, Phys. Rev. Lett. 81:5039–5043.
27. J. D. Franson (1991), *Violations of a simple inequality for classical fields*, Phys. Rev. Lett. 67:290–293.
28. Sven Aerts, Paul Kwiat, Jan-Åke Larsson, and Marek Żukowski (1999), *Two-photon Franson-type experiments and local realism*, Phys. Rev. Lett. 83:2872–2875.
29. Dominic Mayers and Andrew Yao (1998), *Quantum cryptography with imperfect apparatus*, in *Proceedings 39th Annual Symposium on Foundations of Computer Science*, IEEE Comput. Soc. (Los Alamitos), pp. 503–509.

Appendix A: Really Cheap KD

A really cheap key-distribution system that pretends to be an Ekert QKD system can be built out of the following components:

Source: Two pulse transmitters of the type commonly used in fiber-optic communications, controlled to transmit pulses simultaneously at random moments in time. In addition, some provision should be made for the initial time sync.

Receivers: Suitable receiving devices, again of the type commonly used in fiber-optic communications, controlled by electronics that use the translation tables of Figure 3 after having established the time sync.

This system will be cheap, fast, and efficient, compared to a single-photon system. This is because it can be built out of existing standard components, the speed is only limited by the specifications of the components, and the output efficiency is that of the translation tables. The system will yield a key and violate the Bell inequality. For a manufacturer, the above construction is tempting. Violation is all that is needed for secure QKD, is it not? *No. This is not a QKD system.* Do not build/buy a system like this. It is cheap, but you get what you pay for: *no security.*