Hacking energy-time entanglement-based systems with classical light

Jonathan Jogenfors, 1 Ashraf M. Elhassan, 2 Johan Ahrens, 2 Mohamed Bourennane, 2 and Jan-Åke Larsson 1, *

1Institutionen för Systemteknik, Linköpings Universitet, 581 83 Linköping, Sweden.
2Physics Department, Stockholm University, 106 91, Stockholm, Sweden.

Photonic systems based on energy-time entanglement have been proposed to test local realism using the Bell inequality. A violation of this inequality normally also certifies security of device-independent quantum key distribution, so that an attacker cannot eavesdrop or control the system. Here, we show how this security test can be circumvented in energy-time entangled systems when using standard avalanche photodetectors, allowing an attacker to compromise the system without leaving a trace.

With tailored pulses of classical light we reach Bell values up to 3.63 at 97.6% detector efficiency which is an extreme violation. This is the first demonstration of a violation-faking source that both gives tunable violation and high detector efficiency. The implications are severe: the standard Clauser-Horne-Shimony-Holt inequality cannot be used to show device-independent security for standard postselecting energy-time entanglement setups. We conclude with suggestions of improved tests and experimental setups that can re-establish device-independent security.

A Bell experiment [1] is a bipartite experiment that can be used to test for pre-existing properties that are independent of the measurement choice at each site. Formally speaking, the experiment tests if there is a “local realist” description of the experiment, that contains these pre-existing properties. Such a test can be used as the basis for security of Quantum Key Distribution [2, 3] (QKD). QKD uses a bipartite quantum system shared between two parties (Alice and Bob), that allows them to secretly share a cryptographic key. The first QKD protocol [2] (BB84) is based on quantum uncertainty [4] between non-commuting measurements, usually of photon polarization. The Ekert protocol [3] (E91) bases security on a Bell test instead of the uncertainty relation. Such a test indicates, through violation of the corresponding Bell inequality, a secure key distribution system. This requires quantum entanglement, and because of this E91 is also called entanglement-based QKD.

To properly show that an E91 cryptographic system is secure, or alternatively, that no local realist description exists of an experiment, a proper violation of the associated Bell inequality is needed. As soon as a proper violation is achieved, the inner workings of the system is not important anymore, a fact known as device-independent security [5, 6], or a loophole-free test of local realism [7]. In the security context, the size of the violation is related to the amount of key that can be securely extracted from the system. However, a proper (loophole-free) violation is difficult to achieve. For long-distance experiments, photons is the system of choice and one particularly difficult problem is to detect enough of the photon pairs; this is known as the efficiency loophole [8–10].

If the violation is not good enough, there may be a local realist description of the experiment, giving an insecure QKD system. Even worse, an attacker could control the QKD system in this case. One particular example of this occurs when using avalanche photodetectors (APD:s) which are the most commonly used detectors in commercial QKD systems: these detectors can be controlled by a process called “blinding” [11] which enables control via classical light pulses. When using photon polarization in the system, and if the efficiency is low enough in the Bell test, the quantum-mechanical prediction can be faked in such a controlled system [12, 13]. This means that the (apparent) Bell inequality violation can be faked, making a QKD system seem secure while it is not. Note that a proper (loophole-free) violation cannot be faked in this manner.

In this paper we investigate energy-time entanglement-based systems in general and the Franson interferometer [14] in particular. Traditional polarization coding is sensitive to polarization effects caused by optical fibers [15] whereas energy-time entanglement is more robust against this type of disturbance. This property has led to an increased attention to systems based on energy-time entanglement since it allows a design without moving mechanical parts which reduces complexity in practical implementations. A number of applications of energy-time entanglement, such as the QKD, quantum teleportation and quantum repeaters are described in [16].

It is already known that a proper Bell test is more demanding to achieve in energy-time-entanglement systems with postselection [17, 18], but also that certain assumptions on the properties of photons reduce the demands to the same level as for a photon-polarization-based test [19, 20]. The property in question is particle-like behavior of the photon: that it does not “jump” from one arm of an interferometer to the other. Now, clearly, classical light pulses cannot “jump” from one arm to the other, so the question arises: is it at all possible to control the output of the detectors using classical light pulses, to make them fake the quantum correlations? Below, we answer this question in the positive, give the details of such an attack, and its experimental implementation. Moreover, not only are faked quantum correlations possible to reach at a faked detector efficiency of 100%, but even the extreme predictions of nonlocal Popescu-Rohrlich boxes [21] are possible to fake at this high detector efficiency.

* jan-ake.larsson@liu.se
Such extreme predictions reaches the algebraic maximum 4 of the CHSH inequality, and would make a QKD system user suspicious; an attacker would of course not exceed the quantum bound $2\sqrt{2}$ [22].

I. BELL’S INEQUALITY AND THE FRANSON INTERFEROMETER

A Bell test of device-independent security, alternatively local realism, is always associated with a Bell inequality. The relevant part of the E91 QKD protocol up to and including the Bell test looks as follows. The general setup is a central source connected to two measurement sites, one at Alice and the other at Bob. The source prepares an entangled quantum state and distribute it to Alice and Bob who each can choose between a number of measurement settings for their devices. The output can take the values $-1$, $0$, or $+1$, denoting for example horizontal polarization, non-detection, and vertical polarization. Here we are considering a pulsed source so that there are well-defined experimental runs, and therefore also well-defined non-detection events. Alice selects a random integer $j \in \{1, 2, 3\}$ and performs the corresponding measurement $A_j$. Bob does the same with a random number $k \in \{2, 3, 4\}$ and measurement $B_k$. The quantum state and measurements are such that if $j = k$, the outcomes are highly (anti-)correlated. This preparation and measurement process is performed over and over again until enough data has been gathered.

After a measurement batch has been completed, Alice and Bob publicly announce which settings $j$ and $k$ were used (but not the corresponding outcomes!). They can then determine which measurements used the same settings $j = k$ and use the highly (anti-)correlated outcomes for key generation. The remaining outcomes corresponding to $j \neq k$ can be used for security testing, in the Bell [1]-CHSH [23] inequality

$$S_2 = |E(A_1B_2) + E(A_3B_2)| + |E(A_3B_4) - E(A_1B_4)| \leq 2,$$

where $E(A_jB_k)$ is the expected value of the product, often called “correlation” in this context. If the experimental $S_2$ is larger than 2 there is a violation, and the system is secure; there can be no local realist description of the experiment. The size of the violation is related to output key rate; the maximal quantum prediction is $2\sqrt{2}$.

However, a proper violation is difficult to achieve. There are a number ways that the test can give $S_2 > 2$ but still fail, a number of loopholes [7]. The most serious one here is the detector efficiency loophole, that non-detections or zeros are not properly taken into account. If the zeros are ignored, conditioning on detection at both sites gives the conditional correlation $E(A_jB_k|\text{coinc.})$, and a modified bound [9, 10]

$$S_{2,c} = |E(A_1B_2|\text{coinc.}) + E(A_3B_2|\text{coinc.})| + |E(A_3B_4|\text{coinc.}) - E(A_1B_4|\text{coinc.})| \leq \frac{2}{\eta} - 2.$$

The efficiency $\eta$ is the ratio of coincidences to local detections [10], and needs to be above 82% for the quantum value to give a violation. This is ignored in current experiments, almost [24–26] without exception. In the context of QKD, ignoring the zeros is allowed only if the attacker (Eve) cannot control the detectors to make no-detections depend on the local settings $j$ and $k$. Unfortunately, the commonly used APD:s can be controlled [11, 13] unless extra precautions are taken.

For this paper we have investigated a quantum device based on energy-time entanglement with postselection. While the results presented below are acquired from this particular device, the results apply to any such system. The Franson interferometer [14] is shown in Fig. 1 and is built around a source emitting time-correlated photons to both Alice and Bob. The unbalanced Mach-Zehnder interferometers have a time difference $\Delta T$ between the paths. In our pulsed setting, the time difference between a late and early source emission is $\Delta T$, giving rise to interference between the cases “early source emission, photons take the long path” and “late source emission, photons take the short path”. There will be no interference if the photons “take different paths” through the analysis stations, and those events are discarded as non-coincident in a later step.

The analysis stations have variable phase modulators and the setting choices are $\phi^A$ for measuring $A_j$ at Alice and $\phi^B$ for measuring $B_k$ at Bob. The quantum state is such that, given coincident detection, the correlation between $A_j$ and $B_k$ is high if $\phi^A_j + \phi^B_k = 0$. In the
absence of noise, the correlation between Alice’s and Bob’s outcomes will be [14]

$$E(A_j B_k | \text{coinc.}) = \cos(\phi_j^A + \phi_k^B)$$  \hspace{1cm} (3)

This again violates the CHSH inequality (1), but only if the postselection is ignored [17]. When the postselection is taken into account one arrives at the inequality (2) with $\eta = 50\%$, giving a bound of 6 which is no restriction. The question is rather if the system can be controlled by Eve, to fake the violation.

II. FAKING THE BELL INEQUALITY VIOLATION

An eavesdropper (Eve) performs the attack by replacing the source with a faked-state generator that blinds the APD:s (see Fig. 2) and makes them click at chosen instants in time. The blinding is accomplished using classical light pulses superimposed over continuous-wave (CW) illumination [11]. In normal operation, an APD reacts to even a single incoming photon. A photon that enters the detector will create an avalanche of electrical current which results in a signal, or “click”, when the current crosses a certain threshold. The avalanche current is then quenched by lowering the APD bias voltage to below the breakdown voltage, making the detector ready for another photon, and resulting in so-called Geiger mode operation. Under the influence of continuous-wave (CW) illumination, the quenching circuitry will make the current through the APD:s proportional to the power of the incoming light. This will change the behaviour of the APD into so-called linear mode, more similar to a classical photodiode. It will no longer react to single photons, nor register clicks in the usual Geiger-like way and is therefore said to be “blind”. Appropriate choice of CW illumination intensity will make the APD insensitive to single photons yet still register a click when a bright pulse of classical light is superimposed over the CW illumination [11].

What remains is to construct classical light pulses that will give clicks in the way that Eve desires, violating the Bell inequality test for the Franson interferometer. Eve uses pulses with intensity $I$ and pulse length $\tau \ll \Delta T$ intermingled with the CW light that blinds the APD:s. A single pulse emitted by the source will be split when traveling through the interferometer, resulting in two pulses in each output port with intensity $I/4$ each. If instead two pulses are emitted, separated by $\Delta T$ and with phase difference $\omega$, these two pulses will split to three. The middle pulse of the three is built up by two parts, so that the $\pm 1$ outputs show interference,

$$I^+(\phi, \omega) = I \cos^2 \left( \frac{\phi + \omega}{2} \right)$$
$$I^-(\phi, \omega) = I \sin^2 \left( \frac{\phi + \omega}{2} \right),$$  \hspace{1cm} (4)

where $\phi$ is the phase setting of the local analysis station. The chosen $\omega$ controls the $\phi$ dependence of the output. For example, if $I$ is just less than $2I_T$ and $\omega = 0$, there will be a $+1$ click for $|\phi| < \pi/2$ and a $-1$ click otherwise.

However, this is not enough to fake the Bell violation, because the detection time needs to depend on the local setting [17]. To enable this, Eve makes the source emit a group of three pulses separated by $\Delta T$, with phase difference $\omega_E$ between the first and second pulse, and $\omega_L$ between the first and second pulse, and $\omega_L$.
quantum predictions but we have here chosen to focus
on the settings used for the present Bell test: $\phi_A = 0,$ $\phi_A = \pi/2,$ $\phi_B = -\pi/4,$ and $\phi_B = -3\pi/4$ so that only $\theta$
in increments of $\pi/4.$ Eve randomly selects the hidden
variables $r$ and $\theta,$ and reads off the desired results for
the two settings at Alice. If the results are in the same
time slot, she uses two pulses, and can directly calculate
the two phase differences. The same
$r$ and $\theta$ are used to calculate the phase difference for Bob.
Repeating this procedure will produce random outcomes
(to Alice and Bob) that give exactly the quantum predictions
for the mentioned settings, violating the Bell-CHSH
inequality.

By adjusting parameters of the LHV model we can go
even further and produce Bell values up to and including
the value 4, see Fig. 3. But remember that Alice and Bob
would be very confused if their security test displayed the
value 4 as that would mean that their experiment consis-
t of unphysical and nonlocal Popescu-Rohrlich (PR)
boxes [21]. Eve would avoid this but could in principle
use this possibility to negate the effects of noise, by ad-
justing the LHV model to exceed the quantum prediction,
knowing that the noise will lower the violation to below
the quantum bound again.

III. EXPERIMENTAL RESULTS

The attack was experimentally implemented as shown
in Fig. 4. Built using standard fiber-optical components,
it is designed to meet the requirements set in the section
II. The continuous wave is produced by a CW laser while
the pulses are created by a pulsed laser. These two light
sources are combined at a fiber-optic $2 \times 2$ coupler
and then split into one beam for Alice and one for Bob.
Each of these beams is then sent into a fiber-optic $3 \times 3$
coupler (tritters) that equally divides them into three
arms. The first arm consists of a $\Delta T$ delay loop and a
phase modulator $\omega_E,$ the second arm has two $\Delta T$
delay loops and a phase modulator $\omega_M$ (so that $\omega_L = \omega_M - \omega_E$)
while the third arm performs no action. The three arms
are then combined by a second $3 \times 3$ coupler into one
Figure 4: Experimental setup of the attack on the Franson interferometer. The setup consists of a continuous-wave (CW) laser for blinding the detectors, a pulsed laser for generating the bright classical light pulses, fiber-optical couplers (C) delay loops ($\Delta T$), phase modulators ($\omega$ and $\phi$) and detectors (D).

output port that creates the output of the faked state source generator.

The source sends bright light pulses with the setting and phase difference(s) to Alice’s and Bob’s analysis stations in the Franson interferometer. Each of the two analysis stations are constructed in a similar fashion: Two fiber-optic $2 \times 2$ couplers and one delay loop $\Delta T$ and a phase modulator $\phi^A$ (Alice’s side) or $\phi^B$ (Bob’s side). Gated APD:s were used as detectors. This type of detector reduces the dark counts by raising the bias voltage above the breakdown voltage only for a short time period when an incoming photon is expected. These gated APD:s are still vulnerable to the blinding attack described above even if the details of the attack are slightly different. Since the CW power becomes unevenly distributed between detectors, the efficiency of the blinding was affected. This imbalance was avoided by installing digital variable attenuators at the output ports. In addition, optical isolators were placed in front of the detectors in order to prevent crosstalk.

Joint Alice-Bob trials were performed with the pulse amplitudes as described by eqs. (4) and (5) and depicted in Fig. 2. At the desired detector and timeslot a “click” will be forced (Fig. 2a) by constructive interference while destructive interference causes “no click” (Fig. 2b). The sampling time used was 1s and each experiment was run for at least 27s (see Fig. 5). At each point in time, the joint probabilities of Alice’s and Bob’s outcomes are computed from the detector counts and these were then used to determine the Bell value. Note that the early and late timeslots are measured in different experimental runs. The average faked Bell value is

$$S_2 = 2.5615 \pm 0.0064$$

which clearly violates the Bell bound 2. Our source has a repetition rate of of 5 kHz, and the average rate of clicks is 4.88 kHz, giving an average efficiency of 97.6%. The experimental Bell value is lower than the quantum prediction because of noise, most of which is due to unwanted clicks because pulses below the threshold are close to the threshold, and thus sensitive to small intensity variations of the lasers.
Figure 5: The faked Bell of our source is 2.5615 ± 0.0064 (solid black line) which clearly violates the CHSH inequality \( S_2 \leq 2 \). It is possible to increase the faked Bell value up to 3.6386 ± 0.0096 (dotted blue line, data for timeslots where \( p \leq r < 1/2 - p \) or \( 1/2 + p \leq r < 1 - p \)). In both cases the efficiency is 97.6\%. Each point in the diagram corresponds to the \( S_2 \) value for 1 s worth of data.

Adjusting the source to produce fake nonlocal PR boxes [21] gives a faked Bell value of

\[
S_2 = 3.6386 \pm 0.0096
\]

which is even beyond the quantum bound \( 2\sqrt{2} \). The efficiency remains at 97.6\%, and noise still lowers the value from the ideal 4. It should be noted that Eve is free to combine pulses and phases at will in order to produce any Bell value between 0 and the above value. If the noise rate of the system is known she can compensate by aiming for a higher Bell value, and letting the noise bring it back down. This allows her to reach a faked Bell value that is indistinguishable from \( 2\sqrt{2} \).

IV. COUNTERMEASURES

Our faked Bell value seemingly violates the Bell-CHSH inequality even though we are dealing with outcomes from a local realist model. The more appropriate Bell inequality (2) for conditional correlations is clearly ineffective as a test of device-independent security with energy-time entanglement that uses postselection. The bound is too high. We need to improve the security tests in such a way that they unequivocally show security, that they can give a loophole-free violation of local realism.

There are two ways to proceed: one is to use fast switching [17, 18], and the Braunstein-Caves chained Bell inequalities [8, 27] with more terms. The standard chained inequalities read

\[
S_N = |E(A_1B_2) + E(A_3B_2)| + |E(A_3B_4) + E(A_5B_4)| + \ldots + |E(A_{2N+1}B_{2N}) - E(A_1B_{2N})| \leq 2N - 2.
\]

In the Franson interferometer with fast switching (F), the chained inequalities are weakened but still produces a usable bound even after postselection on coincidence,

\[
S_{N,F} = |E(A_1B_2|\text{coinc.}) + E(A_3B_2|\text{coinc.})| + |E(A_3B_4|\text{coinc.}) + E(A_5B_4|\text{coinc.})| + \ldots + |E(A_{2N-1}B_{2N}|\text{coinc.}) - E(A_1B_{2N}|\text{coinc.})| \leq 2N - 1.
\]

This only gives the upper bound \( S_{2,F} \leq 3 \) for the Bell-CHSH value, so the standard test is not useful even with fast switching. But the quantum-mechanical prediction \( S_{N,F} = 2N \cos(\pi/2N) \) does violate this if \( N \geq 3 \), even though the violation is smaller than the standard Bell test. This re-establishes device-independent security for energy-time-entangled QKD. In practice, though, the requirements are high since the lowest acceptable visibility is 94.64\% [18].

A better solution would be to eliminate the core problem: The postselection loophole. One alternative is the use of “hugging” interferometers [28] that gives an energy-time-entangled interferometer with postselection, but without a postselection loophole. The drawback is the requirement of two fiber links each to Alice and Bob. A Bell violation has been experimentally shown [29], even with 1 km fiber length [30].

V. CONCLUSION

Bell tests are a cornerstone of quantum key distribution and is necessary for device-independent security. Device-independent Bell inequality violation must be performed with care to avoid loopholes. Time-energy-entanglement has the distinct advantage over polarization that time and energy is more easily communicated over long distances than polarization. Therefore, time-energy entanglement may be preferable as quantum resource to perform reliable key distribution.

In this paper we have shown that quantum key distribution systems based on energy-time entanglement with postselection are vulnerable to attack if the corresponding security tests use the original Bell inequality. By blinding the detectors and using an LHV model Eve lets Alice and Bob think their system violates Bell’s inequality even though she uses classical light. This lets the attacker fully control the key output and break the security of the system.
Our attack has been performed with a detector efficiency of 97.6% which is high enough to avoid the fair sampling assumption. We can compare this to Gerhardt et al. [13] where the detector efficiency was 50% when using active basis choice; that attack has an upper limit of 82.8% [9, 10], while our attack is only limited to experimental losses. In other words, the attack is possible even with perfect detection efficiency.

In addition, our attack can produce the unphysical value $S_2 = 4$ at any efficiency. It remains a fact that fast switching will restrict this largest value to 3, but our attack demonstrates the level of control an attacker can exert onto the system. In order to build a device-independent QKD system based on energy-time-entanglement the designer will either have to use stronger tests such as the Braunstein-Caves inequality, or use a system that does not contain the postselection loophole.

ACKNOWLEDGEMENTS

J.J. and J.-Å.L. were supported by CENIIT at LiU. M.B., J.A., and A.M.E. were supported by the Swedish Research Council, Ideas Plus (Polish Ministry of Science and Higher Education Grant No. IdP2011 000361) and ADOPT.

[1] J. S. Bell, Physics (Long Island City, N. Y.) 1, 195 (1964).
[2] C. H. Bennett and G. Brassard, in Proc. of the IEEE Int. Conf. on Computers, Systems, and Signal Processing, 175–179 (IEEE New York, Bangalore, India, 1984).
[3] A. K. Ekert, Phys. Rev. Lett. 67, 661 (1991).
[4] W. Heisenberg, Zeitschrift für Physik 43, 172 (1927).
[5] A. Acín, N. Gisin, and L. Masanes, Physical Review Letters 97, 120405 (2006).
[6] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. 98, 230501 (2007).
[7] J.-Å. Larsson, Journal of Physics A 47, 424003 (2014).
[8] P. Pearle, Phys. Rev. D 2, 1418 (1970).
[9] A. Garg and N. D. Mermin, Phys. Rev. D 35, 3831 (1987).
[10] J.-Å. Larsson, Phys. Rev. A 57, 3304 (1998).
[11] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photon. 4, 686 (2010).
[12] J.-Å. Larsson, Quantum Inf. Comput. 2, 434 (2002).
[13] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, V. Scarani, V. Makarov, and C. Kurtsiefer, Phys. Rev. Lett. 107, 170404 (2011).
[14] J. D. Franson, Physical Review Letters 62, 2205 (1989).
[15] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Reviews of Modern Physics 74, 145 (2002).
[16] N. Gisin and R. Thew, Nature Photonics 1, 165 (2007).
[17] S. Aerts, P. Kwiat, J.-Å. Larsson, and M. Zukowski, Physical Review Letters 83, 2872 (1999).
[18] J. Jogenfors and J.-Å. Larsson, Journal of Physics A 47, 424032 (2014).
[19] J. D. Franson, Physical Review A 61, 012105 (1999).
[20] J. D. Franson, Physical Review A 80, 032119 (2009).
[21] S. Popescu and D. Rohrlich, Foundations of Physics 24, 379 (1994).
[22] B. S. Cirel’son, Letters in Mathematical Physics 4, 93 (1980).
[23] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Phys. Rev. Lett. 23, 880 (1969).
[24] M. Giustina, A. Mech, S. Ramelow, B. Wittmann, J. Kofler, J. Beyer, A. Lita, B. Calkins, T. Gerrits, S. W. Nam, R. Ursin, and A. Zeilinger, Nature 497, 227 (2013).
[25] B. G. Christensen, K. T. McCusker, J. B. Altepeter, B. Calkins, T. Gerrits, A. E. Lita, A. Miller, L. K. Shalm, Y. Zhang, S. W. Nam, N. Brunner, C. C. W. Lim, N. Gisin, and P. G. Kwiat, Physical Review Letters 111, 130406 (2013).
[26] J.-Å. Larsson, M. Giustina, J. Kofler, B. Wittmann, R. Ursin, and S. Ramelow, Physical Review A 90, 032107 (2014).
[27] S. Braunstein and C. Caves, Annals of Physics 202, 22 (1990).
[28] A. Cabello, A. Rossi, G. Vallone, F. De Martini, and P. Mataloni, Physical Review Letters 102, 040401 (2009).
[29] G. Lima, G. Vallone, A. Chiuri, A. Cabello, and P. Mataloni, Physical Review A 81, 040101 (2010).
[30] A. Cuevas, G. Carvacho, G. Saavedra, J. Carriñen, W. a. T. Nogueira, M. Figueroa, A. Cabello, P. Mataloni, G. Lima, and G. B. Xavier, Nature Communications 4, 2871 (2013).