Insecurity of Detector-Device-Independent Quantum Key Distribution

Shihan Sajeed,[1,2,*] Anqi Huang,[1,2] Shihai Sun,[3] Feihu Xu,[4] Vadim Makarov,[1,2,5] and Marcos Curty[6]
[1]Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada
[2]Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada
[3]College of Science, National University of Defense Technology, Changsha 410073, China
[4]Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA
[5]Department of Physics and Astronomy, University of Waterloo, Waterloo, Ontario N2L 3G1 Canada
[6]Escuela de Ingeniería de Telecomunicación, Department of Signal Theory and Communications, University of Vigo, Vigo E-36310, Spain

(Received 20 July 2016; revised manuscript received 30 October 2016; published 16 December 2016)

Detector-device-independent quantum key distribution (DDI-QKD) held the promise of being robust to detector side channels, a major security loophole in quantum key distribution (QKD) implementations. In contrast to what has been claimed, however, we demonstrate that the security of DDI-QKD is not based on postselected entanglement, and we introduce various eavesdropping strategies that show that DDI-QKD is in fact insecure against detector side-channel attacks as well as against other attacks that exploit devices' imperfections of the receiver. Our attacks are valid even when the QKD apparatuses are built by the legitimate users of the system themselves, and thus, free of malicious modifications, which is a key assumption in DDI-QKD.

DOI: 10.1103/PhysRevLett.117.250505

Introduction.—Quantum key distribution (QKD), a technique to distribute a secret random bit string between two separated parties (Alice and Bob), needs to close the gap between theory and practice [1]. In theory, QKD provides information-theoretic security. In practice, however, it does not because QKD implementation devices do not typically conform to the theoretical models considered in the security proofs. As a result, any unaccounted device imperfection might constitute a side channel, which could be used by an eavesdropper (Eve) to learn the secret key without being detected [2–12].

To bridge this gap, various approaches have been proposed recently [13–17], with measurement-device-independent QKD (MDI-QKD) [17] probably being the most promising one in terms of feasibility and performance. Its security is based on postselected entanglement, and it can remove all detector side channels from QKD implementations, which is arguably their major security loophole [3–10,12]. Also, its practicality has been already confirmed both in laboratories and via field trials [18–24]. A drawback of MDI-QKD is, however, that it requires high-visibility two-photon interference between independent sources, which makes its implementation more demanding than that of conventional QKD schemes. In addition, current finite-key security bounds against general attacks [25] require larger postprocessing data block sizes than those of standard QKD, though recent proposals [26] significantly improve the performance of MDI-QKD in the finite-key regime.

To overcome these limitations, a novel approach, so-called detector-device-independent QKD (DDI-QKD), has been introduced recently [27–30]. It avoids the problem of interfering photons from independent light sources by using the concept of a single-photon Bell state measurement (BSM) [31]. As a result, its finite-key security bounds and classical postprocessing data block sizes are expected to be similar to those of prepare-and-measure QKD schemes [32]. Despite this presumed promising performance, however, the robustness of DDI-QKD against detector side-channel attacks has not been rigorously proven yet, and only partial security proofs have been introduced [27,28].

In this Letter, we show that in contrast to what has been claimed [27–30], the security of DDI-QKD *cannot* rely on the same principles as MDI-QKD, i.e., postselected entanglement. More importantly, we demonstrate that DDI-QKD is in fact vulnerable to detector side-channel attacks and to other attacks that exploit imperfections of the receiver's devices. These attacks are valid even when Alice's and Bob's state preparation processes are fully characterized and trusted, an essential assumption in DDI-QKD. Moreover, they do not require that Eve substitutes Bob's detectors with a measurement apparatus prepared by herself to leak key information to the channel [33]. That is, our attacks apply as well to the scenario where Alice and Bob build the QKD devices themselves.

MDI-QKD & DDI-QKD.—Let us start by reviewing the basic principles behind MDI-QKD and DDI-QKD. To simplify the discussion, we shall assume that Alice and Bob have at their disposal perfect single-photon sources. Note, however, that both schemes can operate as well, for instance, with phase-randomized weak coherent pulses in

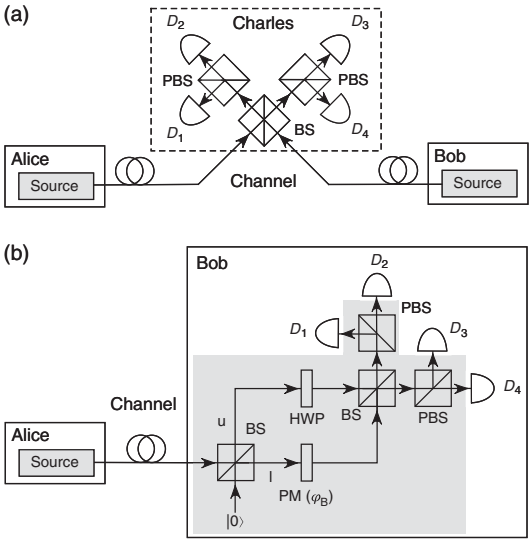

FIG. 1. Possible implementations of partially device-independent QKD with linear optics. (a) MDI-QKD [17], PBS, polarizing beam splitter; BS, 50:50 beam splitter; and D_i, with $i \in \{1, 2, 3, 4\}$, Charles' single-photon detectors. (b) DDI-QKD [28], HWP, half-wave plate, and PM, phase modulator. One single click in the detector D_1, D_2, D_3, or D_4 corresponds to a projection into the Bell state $|\Psi^+\rangle$, $|\Phi^+\rangle$, $|\Psi^-\rangle$, or $|\Phi^-\rangle$, respectively (see main text for further details). In both schemes, the gray areas denote devices that need to be characterized and trusted. Also, Alice's and Bob's laboratories need to be protected from any information leakage to the outside.

combination with decoy states [34–36], which do not prevent the attacks considered here.

An example of a possible implementation of MDI-QKD is illustrated in Fig. 1(a) [17]. Both Alice and Bob generate Bennett-Brassard 1984 (BB84) states [37] and send them to an untrusted relay, Charles. If Charles is honest, he performs a two-photon BSM that projects the incoming signals into a Bell state. In any case, Charles has to declare which of his measurements are successful together with the Bell states obtained. Alice and Bob then extract a secret key from those successful events where they used the same basis. Importantly, if Charles is honest, his BSM measurement postselects entanglement between Alice and Bob, and therefore, he is not able to learn any information about their bit values. To test whether or not Charles is honest, Alice and Bob can simply compare a randomly chosen subset of their data to see if it satisfies the expected correlations associated to the Bell states announced. That is, MDI-QKD can be seen as a time-reversed Einstein-Podolsky-Rosen QKD protocol [38]. Therefore, its security can be proven without any assumption on the behavior of Charles' measurement unit.

DDI-QKD [27–30] aims to follow the same spirit of MDI-QKD. The key idea is to replace the two-photon BSM with a 2-qubit single-photon BSM [31]. This requires that Alice and Bob use two different degrees of freedom of the single photons to encode their bit information. In so doing, one avoids the need for interfering photons from independent light sources. An example of a possible implementation is illustrated in Fig. 1(b) [28] (see also [27,29,30] for similar proposals). Here, Alice sends Bob BB84 polarization states: $(|H\rangle + e^{i\theta_A}|V\rangle)/\sqrt{2}$, where $|H\rangle$ ($|V\rangle$) denotes the Fock state of a single photon prepared in horizontal (vertical) polarization, and the phase $\theta_A \in \{0, \pi/2, \pi, 3\pi/2\}$. Bob then encodes his bit information by using the spatial degree of freedom of the incoming photons. This is done with a 50:50 beam splitter (BS) together with a phase modulator (PM) that applies a random phase $\varphi_B \in \{0, \pi/2, \pi, 3\pi/2\}$ to each input signal. Finally, Bob performs a BSM that projects each input photon into a Bell state: $|\Phi^\pm\rangle = (|H\rangle|u\rangle \pm |V\rangle|l\rangle)/\sqrt{2}$ and $|\Psi^\pm\rangle = (|H\rangle|l\rangle \pm |V\rangle|u\rangle)/\sqrt{2}$, where $|u\rangle$ ($|l\rangle$) represents the state of a photon that goes through the upper (lower) arm of the interferometer [see Fig. 1(b)]. A photon detection event (click) in only one detector D_i corresponds to a projection on a particular Bell state.

Both MDI-QKD and DDI-QKD require that Alice's and Bob's state preparation processes are characterized and trusted. This is indicated by the gray areas shown in Fig. 1. In DDI-QKD, the elements inside Bob's gray area can be regarded as his trusted transmitter (when compared to MDI-QKD). Among the trusted components there are elements, which belong to the BSM, but, importantly, the detectors D_i do not need to be trusted.

The security of DDI-QKD is not based on postselected entanglement.—At first sight, it seems that the security of DDI-QKD follows directly from that of MDI-QKD, given, of course, that the assumptions on Alice's and Bob's state preparation processes are fulfilled [27–30]. That is, it relies on the fact that the BSM postselects entanglement between Alice and Bob. A first indication that confronts this idea was given recently in [33]. There, it was shown that in contrast to MDI-QKD, DDI-QKD is actually insecure if Eve is able to replace Bob's detectors with a measurement apparatus that leaks information to the channel [33]. Although this result is important from a conceptual point of view, it violates one of the security assumptions of DDI-QKD: Bob's detectors have to be built by a trusted party (but do not need to be characterized) to avoid that they intentionally leak key information to the outside [27]. Below, we show that even in this scenario, the security of DDI-QKD cannot be based on postselected entanglement alone, unlike MDI-QKD.

For this, we will consider a slightly simplified version of the DDI-QKD scheme illustrated in Fig. 1(b). In particular, we will assume that Bob's receiver has only one active detector, say for instance, the detector D_1, while the other detectors are disabled. That is, now Bob's BSM projects the incoming photons only into the Bell state $|\Psi^+\rangle$. If the security of DDI-QKD is based on postselected entanglement, this modification should not affect its security (only

its secret key rate is reduced by a factor of four), as a projection into a single Bell state should be sufficient to guarantee security [17]. Next, we show that a blinding attack [6,8] renders DDI-QKD insecure in this situation.

In particular, suppose that Eve shines bright light onto Bob's detector D_1 to make it enter linear-mode operation [6,8]. In this mode, the detector is no longer sensitive to single-photon pulses, but it can only detect strong light. We assume that when D_1 receives a bright pulse of mean photon number μ, it always produces a click, while if the pulse's mean photon number is $\mu/2$, it never produces a click. This behavior has been experimentally confirmed in many detector types [6,8,39–44]. Once D_1 is blinded, Eve performs an intercept-resend attack on every signal sent by Alice. That is, she measures Alice's signals in one of the two BB84 bases (which Eve selects at random for each pulse), and she prepares a new signal, depending on the result obtained, that is sent to Bob. Intercept-resend attacks correspond to entanglement-breaking channels and, therefore, they cannot lead to a secure key [45]. Suppose, for instance, that the signals that Eve sends to Bob are coherent states of the form $|\sqrt{2\mu}\rangle$, with creation operator $a^\dagger = (a_H^\dagger + e^{i\phi_E} a_V^\dagger)/\sqrt{2}$. Here, $a_H^\dagger$ ($a_V^\dagger$) denotes the creation operator for horizontally (vertically) polarized photons, and the phase $\phi_E \in \{0, \pi/2, \pi, 3\pi/2\}$ depends on Eve's measurement result. More precisely, for each measured signal, Eve sends Bob a coherent state prepared in the BB84 polarization state identified by her measurement. Then, it can be shown that the state at the input ports of Bob's detectors D_i is a coherent state of the form (see Supplemental Material Sec. I [46] for details)

$$|\psi\rangle = \left|\frac{\sqrt{\mu}}{2}(e^{i\phi_E} + e^{i\varphi_B})\right\rangle_{D_1} \otimes \left|\frac{\sqrt{\mu}}{2}(1 + e^{i(\phi_E+\varphi_B)})\right\rangle_{D_2} \otimes \left|\frac{\sqrt{\mu}}{2}(e^{i\phi_E} - e^{i\varphi_B})\right\rangle_{D_3} \otimes \left|\frac{\sqrt{\mu}}{2}(1 - e^{i(\phi_E+\varphi_B)})\right\rangle_{D_4}. \quad (1)$$

This situation is illustrated in Table I, where we show the mean photon number of the incoming light to Bob's detectors for all combinations of ϕ_E and φ_B. Most importantly, from this table we can see that if D_1 is the only active detector, then Bob only obtains a click when he uses the same measurement basis as Eve, i.e., when φ_B, $\phi_E \in \{0, \pi\}$ or φ_B, $\phi_E \in \{\pi/2, 3\pi/2\}$), and $\varphi_B = \phi_E$. That is, this attack does not introduce any error. Moreover, we have that Bob and Eve select the same basis with at least $1/2$ probability. This means that the DDI-QKD scheme illustrated in Fig. 1(b) (with only one active detector) is actually insecure against the detector blinding attack for a total system loss beyond only 3 dB, just like standard QKD schemes. This confirms that the security of DDI-QKD cannot be based on postselected entanglement. The same conclusion applies as well to the DDI-QKD schemes introduced in Refs. [27,29], and [30].

TABLE I. Mean photon number of the input light to Bob's detectors as a function of the phases ϕ_E and φ_B.

(a) $\phi_E = 0$

φ_B	D_1	D_2	D_3	D_4
0	μ	μ	0	0
$\pi/2$	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$
π	0	0	μ	μ
$3\pi/2$	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$

(b) $\phi_E = \pi/2$

φ_B	D_1	D_2	D_3	D_4
0	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$
$\pi/2$	μ	0	0	μ
π	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$
$3\pi/2$	0	μ	μ	0

(c) $\phi_E = \pi$

φ_B	D_1	D_2	D_3	D_4
0	0	0	μ	μ
$\pi/2$	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$
π	μ	μ	0	0
$3\pi/2$	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$

(d) $\phi_E = 3\pi/2$

φ_B	D_1	D_2	D_3	D_4
0	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$
$\pi/2$	0	μ	μ	0
π	$\mu/2$	$\mu/2$	$\mu/2$	$\mu/2$
$3\pi/2$	μ	0	0	μ

Insecurity of DDI-QKD against detector side-channel attacks.—If Bob uses four active detectors, the detector blinding attack has one main drawback: it produces double clicks [33]. From Table I, one can already see that whenever Bob uses the same measurement basis as Eve, there are always two detectors that click. For instance, when $\varphi_B = \phi_E = 0$, the detectors D_1 and D_2 always click, similar for the other cases. This means that Alice and Bob could, in principle, try to monitor double clicks to detect the presence of Eve. So, the question is whether or not four active detectors can make DDI-QKD secure again. As we show below, the answer is no. For this, we introduce two possible eavesdropping strategies that exploit practical imperfections of Bob's detectors to avoid double clicks. See also Supplemental Material Sec. II [46] for two alternative attacks that achieve the same goal by exploiting other imperfections of Bob's linear optics network.

The first eavesdropping strategy uses the fact that single-photon detectors respond differently to the same blinding power P_B. This has been recently analyzed in Ref. [44].

There, the authors compare the response of two single-photon detectors in a commercial QKD system Clavis2 [49] to varying blinding power. They first illuminate the detectors with continuous-wave bright light of power P_B to force them enter linear-mode operation. Then they record the maximum and minimum value of the trigger pulse energy E_T for which the click probabilities are 0 and 1, respectively. The results are shown in Fig. 2(a) [44]. For a particular blinding power P_B, each point in the solid (dashed) curves shown in the figure represents the maximum (minimum) value of trigger pulse energy E_T for which the detection efficiency η_{det} is 0 (1). The blue and green colors identify the two detectors. (Note that if the energies E_T corresponding to the dashed curves are halved, the result is always below the solid curves, thus satisfying the assumption made in the previous section that pulses with mean photon number $\mu/2$ result in zero-click probability.) Next, we show how these detector characteristics could be used to avoid double clicks.

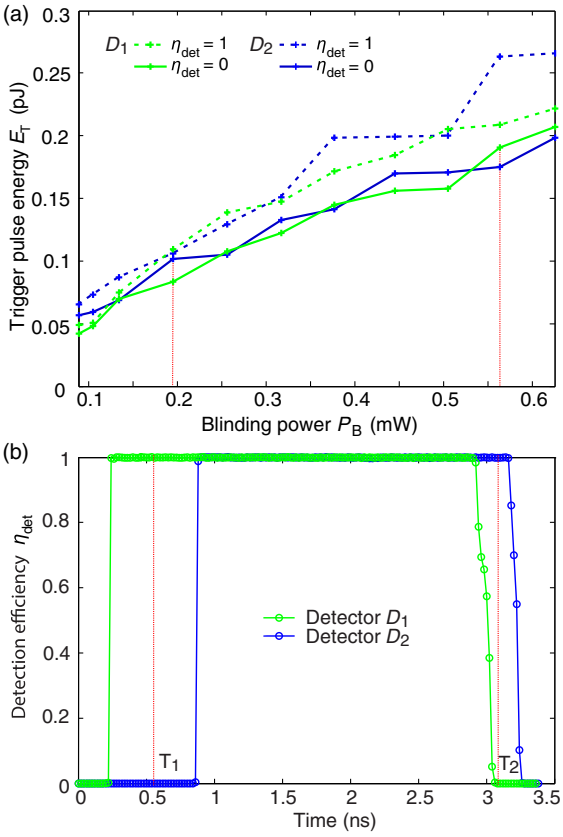

FIG. 2. Detector click probability in bright light blinded regime in commercial QKD system Clavis2. (a) Click trigger thresholds vs blinding power P_B for two different single-photon detectors D_1 and D_2. Here, for a particular blinding power P_B, each point in the solid (dashed) curves represents the maximum (minimum) value of trigger pulse energy E_T for which the detection efficiency η_{det} is 0 (1). The experimental data have been reprinted from Ref. [44]. (b) Measured detection efficiency mismatch in the time domain between two blinded single-photon detectors at $P_B = 0.32$ mW, $E_T = 0.24$ pJ, and 0.7 ns wide trigger pulse (see main text for further details).

For this, we return to the blinding attack described above against the DDI-QKD implementation illustrated in Fig. 1(b). For simplicity, let us consider again the case where $\varphi_B = \phi_E = 0$. In particular, suppose for instance that Eve wants to force a click only on detector D_1, and no click on detector D_2. Then, in order to achieve this goal, she can simply choose a combination of P_B and E_T, such that the detector D_1 (D_2) has a nonzero (zero) click probability. If the behavior of the detector D_1 (D_2) corresponds to the green (blue) curves shown in Fig. 2(a), then the values $P_B \approx 0.2$ mW and $E_T \approx 0.1$ pJ constitute an example that satisfies this criterion. Similarly, if $P_B \approx 0.56$ mW and $E_T \approx 0.19$ pJ, then Eve could make the detector D_2 (D_1) to have a nonzero click probability. Importantly, note that when Bob's basis matches that of Eve, only two out of the four detectors D_i might produce a click (see Table I). Hence, in these instances, Eve only needs to avoid double clicks between two detectors in order to remain undetected. A similar argument can be applied as well to any other value of φ_B and ϕ_E.

This attack demonstrates that if Bob's detectors are uncharacterized, as assumed in DDI-QKD, these type of schemes are indeed insecure against detector side-channel attacks. That is, Eve could learn the whole secret key without producing any error nor a double click.

A second eavesdropping strategy that also allows Eve to avoid double clicks is based on a time-shift attack [3,4] that exploits the detection efficiency mismatch between Bob's detectors. In this type of attack, Eve shifts the arrival time of each signal that she sends to Bob such that only one detector can produce a click each given time. Here, we have confirmed experimentally that this type of attack is also possible with blinded detectors. For this, we blinded two single-photon detectors from the commercial QKD system Clavis2 [49], and we measured their detection efficiency mismatch. The experimental results are shown in Fig. 2(b). We find, for instance, that whenever Bob receives a trigger pulse at the time instance T_1 (T_2), only the detector D_1 (D_2) can produce a click because this instance is outside of the response region of the detector D_2 (D_1). That is, by combining the time-shift attack with the blinding attack introduced in the previous section, Eve could again break the security of DDI-QKD without introducing errors nor double clicks.

Conclusion.—We have analyzed the security of detector-device-independent QKD, a novel scheme that promised to be robust against detector side-channel attacks. We have shown that its security is not based on postselected entanglement, as originally claimed. Most importantly, we have presented various eavesdropping attacks that demonstrate that DDI-QKD is actually vulnerable to detector side-channel attacks as well as to other side-channel attacks that exploit imperfections of Bob's receiver.

These attacks are valid even when Alice's and Bob's state preparation processes are fully characterized and trusted, and Bob's detectors are built by a trusted party and cannot be replaced with a measurement device manufactured by Eve. Alice and Bob might try to prevent these attacks by designing proper countermeasures at the detector side, just like in standard QKD schemes. In such scenarios, however, it is unclear what would be the real advantage (in terms of complexity and performance) of using DDI-QKD instead of standard QKD systems. As a final remark, let us say that the main reason for the insecurity of DDI-QKD seems to be Bob's state preparation process; while in MDI-QKD, it is assumed to be protected, in DDI-QKD it can be influenced by Eve via the signals she sends him.

This work was supported by Industry Canada, CFI, NSERC (programs Discovery, PDF, CryptoWorks21), Ontario MRI, US Office of Naval Research, National Natural Science Foundation of China (Grants No. 11304391 and 11674397), Spain MINECO, Fondo Europeo de Desarrollo Regional, FEDER (Grant No. TEC2014-54898-R), and Galician Regional Government (Programs EM2014/033, AtlantTIC). The authors thank ID Quantique for cooperation, technical assistance, and providing the QKD hardware.

S. S. and A. H. contributed equally to this work.

*shihan.sajeed@gmail.com

[1] H.-K. Lo, M. Curty, and K. Tamaki, Nat. Photonics **8**, 595 (2014).
[2] A. Vakhitov, V. Makarov, and D. R. Hjelme, J. Mod. Opt. **48**, 2023 (2001).
[3] V. Makarov, A. Anisimov, and J. Skaar, Phys. Rev. A **74**, 022313 (2006); **78**, 019905(E) (2008).
[4] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, Quantum Inf. Comput. **7**, 73 (2007).
[5] A. Lamas-Linares and C. Kurtsiefer, Opt. Express **15**, 9388 (2007).
[6] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photonics **4**, 686 (2010).
[7] F. Xu, B. Qi, and H.-K. Lo, New J. Phys. **12**, 113026 (2010).
[8] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Nat. Commun. **2**, 349 (2011).
[9] H. Weier, H. Krauss, M. Rau, M. Fürst, S. Nauerth, and H. Weinfurter, New J. Phys. **13**, 073024 (2011).
[10] P. Jouguet, S. Kunz-Jacques, and E. Diamanti, Phys. Rev. A **87**, 062313 (2013).
[11] S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, Phys. Rev. A **91**, 032326 (2015).
[12] S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, Phys. Rev. A **91**, 062301 (2015).
[13] D. Mayers and A. Yao, *Proc. 39th Annual Symposium on Foundations of Computer Science* (IEEE, Palo Alto, California, 1998), pp. 503–509.
[14] A. Acín, N. Brunner, N. Gisin, S. Massar, S. Pironio, and V. Scarani, Phys. Rev. Lett. **98**, 230501 (2007).
[15] U. Vazirani and T. Vidick, Phys. Rev. Lett. **113**, 140501 (2014).
[16] C. A. Miller and Y. Shi, *Proc. 46th Annual ACM Symposium on Theory of Computing (STOC'14)* (ACM, New York, NY, USA, 2014), pp. 417–426.
[17] H.-K. Lo, M. Curty, and B. Qi, Phys. Rev. Lett. **108**, 130503 (2012).
[18] A. Rubenok, J. A. Slater, P. Chan, I. Lucio-Martinez, and W. Tittel, Phys. Rev. Lett. **111**, 130501 (2013).
[19] T. Ferreira da Silva, D. Vitoreti, G. B. Xavier, G. C. do Amaral, G. P. Temporão, and J. P. von der Weid, Phys. Rev. A **88**, 052303 (2013).
[20] Y. Liu *et al.*, Phys. Rev. Lett. **111**, 130502 (2013).
[21] Z. Tang, Z. Liao, F. Xu, B. Qi, L. Qian, and H.-K. Lo, Phys. Rev. Lett. **112**, 190503 (2014).
[22] L. C. Comandar, M. Lucamarini, B. Fröhlich, J. F. Dynes, A. W. Sharpe, S. W.-B. Tam, Z. L. Yuan, R. V. Penty, and A. J. Shields, Nat. Photonics **10**, 312 (2016).
[23] Y.-L. Tang *et al.*, Phys. Rev. X **6**, 011024 (2016).
[24] H.-L. Yin *et al.*, Phys. Rev. Lett. **117**, 190501 (2016).
[25] M. Curty, F. Xu, W. Cui, C. C. W. Lim, K. Tamaki, and H.-K. Lo, Nat. Commun. **5**, 3732 (2014).
[26] Y.-H. Zhou, Z.-W. Yu, and X.-B. Wang, Phys. Rev. A **93**, 042324 (2016).
[27] P. González, L. Rebón, T. Ferreira da Silva, M. Figueroa, C. Saavedra, M. Curty, G. Lima, G. B. Xavier, and W. A. T. Nogueira, Phys. Rev. A **92**, 022337 (2015).
[28] C. C. W. Lim, B. Korzh, A. Martin, F. Bussières, R. Thew, and H. Zbinden, Appl. Phys. Lett. **105**, 221112 (2014).
[29] W.-F. Cao, Y.-Z. Zhen, Y.-L. Zheng, Z.-B. Chen, N.-L. Liu, K. Chen, and J.-W. Pan, arXiv:1410.2928v1 (manuscript withdrawn by authors on 23 Aug 2016 owing to the insecurity of the proposed scheme).
[30] W.-Y. Liang, M. Li, Z.-Q. Yin, W. Chen, S. Wang, X.-B. An, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **92**, 012319 (2015).
[31] Y.-H. Kim, Phys. Rev. A **67**, 040301 (2003).
[32] C. C. W. Lim, M. Curty, N. Walenta, F. Xu, and H. Zbinden, Phys. Rev. A **89**, 022307 (2014).
[33] B. Qi, Phys. Rev. A **91**, 020303 (2015).
[34] W.-Y. Hwang, Phys. Rev. Lett. **91**, 057901 (2003).
[35] H.-K. Lo, X. Ma, and K. Chen, Phys. Rev. Lett. **94**, 230504 (2005).
[36] X.-B. Wang, Phys. Rev. Lett. **94**, 230503 (2005).
[37] C. H. Bennett and G. Brassard, *Proc. IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India)* (IEEE Press, New York, 1984), pp. 175–179.
[38] E. Biham, B. Huttner, and T. Mor, Phys. Rev. A **54**, 2651 (1996).
[39] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Opt. Express **18**, 27938 (2010).
[40] C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, New J. Phys. **13**, 013043 (2011).
[41] L. Lydersen, M. K. Akhlaghi, A. H. Majedi, J. Skaar, and V. Makarov, New J. Phys. **13**, 113042 (2011).
[42] S. Sauge, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov, Opt. Express **19**, 23590 (2011).

[43] J. Jogenfors, A. M. Elhassan, J. Ahrens, M. Bourennane, and J.-A. Larsson, Sci. Adv. **1**, e1500793 (2015).
[44] A. Huang, S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, IEEE J. Quantum Electron. **52**, 8000211 (2016).
[45] M. Curty, M. Lewenstein, and N. Lütkenhaus, Phys. Rev. Lett. **92**, 217903 (2004).
[46] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.117.250505 which also includes Refs. [28,47,48] for more details about the quantum states arriving at Bob's detectors and for side-channel attacks that exploit imperfections of Bob's linear optics network.
[47] F. Xu, K. Wei, S. Sajeed, S. Kaiser, S. Sun, Z. Tang, L. Qian, V. Makarov, and H.-K. Lo, Phys. Rev. A **92**, 032305 (2015).
[48] H.-W. Li *et al.*, Phys. Rev. A **84**, 062308 (2011).
[49] Clavis2 specification sheet, http://www.idquantique.com/images/stories/PDF/clavis2-quantum-key-distribution/clavis2-specs.pdf.

Supplemental material: Insecurity of detector-device-independent quantum key distribution

Shihan Sajeed, Anqi Huang, Shihai Sun, Feihu Xu, Vadim Makarov, and Marcos Curty

I. QUANTUM STATE $|\psi\rangle$ AT THE INPUT PORTS OF BOB'S DETECTORS

In this section, we present the calculations to derive Eq. (1) in main text. To simplify the discussion, we have labeled different modes involved in the calculations in Fig. 3. Suppose that the input state in mode a is a coherent state $\left|\sqrt{2\mu}\right\rangle_a$ with creation operator $a^\dagger = (a_{\rm H}^\dagger + e^{i\phi_{\rm E}} a_{\rm V}^\dagger)/\sqrt{2}$. Also, suppose that the input signal in mode b is the vacuum state $|0\rangle_b$. Then, the output signal in modes c and d, after the action of the $50:50$ beamsplitter (BS), is given by $\left|\sqrt{\mu}\right\rangle_c \otimes \left|\sqrt{\mu}\right\rangle_d$, where $c^\dagger = (c_{\rm H}^\dagger + e^{i\phi_{\rm E}} c_{\rm V}^\dagger)/\sqrt{2}$ and $d^\dagger = (d_{\rm H}^\dagger + e^{i\phi_{\rm E}} d_{\rm V}^\dagger)/\sqrt{2}$ denote the corresponding creation operators for modes c and d. Next, we consider the phase modulator (PM) and the half-wave plate (HWP) that act on modes c and d. The former performs the unitary transformation $c^\dagger = e^{i\varphi_{\rm B}} e^\dagger$, where $e^\dagger$ is the creation operator at the output port of the PM. The HWP applies the unitary transformation $d_{\rm H}^\dagger = f_{\rm V}^\dagger$ and $d_{\rm V}^\dagger = f_{\rm H}^\dagger$, where $f_{\rm H}^\dagger$ and $f_{\rm V}^\dagger$ denote the creation operators at the output port of the HWP. This means, in particular, that the quantum state in modes e and f has the form

$$\left|\sqrt{\mu} e^{i\varphi_{\rm B}}\right\rangle_e \otimes \left|\sqrt{\mu}\right\rangle_f , \tag{2}$$

with the creation operators $e^\dagger$ and $f^\dagger$ given by $e^\dagger = (e_{\rm H}^\dagger + e^{i\phi_{\rm E}} e_{\rm V}^\dagger)/\sqrt{2}$ and $f^\dagger = (e^{i\phi_{\rm E}} f_{\rm H}^\dagger + f_{\rm V}^\dagger)/\sqrt{2}$, respectively.

Then, after applying the $50:50$ BS on modes e and f, we have that the output state in modes g and k can be expressed as

$$\exp\Big\{ \frac{\sqrt{\mu}}{2} \Big[\big(e^{i\phi_{\rm E}} - e^{i\varphi_{\rm B}}\big) g_{\rm H}^\dagger + \big(1 - e^{i(\phi_{\rm E}+\varphi_{\rm B})}\big) g_{\rm V}^\dagger + \big(e^{i\phi_{\rm E}} + e^{i\varphi_{\rm B}}\big) k_{\rm H}^\dagger + \big(1 + e^{i(\phi_{\rm E}+\varphi_{\rm B})}\big) k_{\rm V}^\dagger \Big] \Big\} |0\rangle . \tag{3}$$

Finally, if we apply the polarising beamsplitters (PBS) (which we assume reflect horizontally polarised light and let vertically polarised light pass) on modes g and k, we find that the state $|\psi\rangle$ at the input ports of Bob's detectors $\mathrm{D_i}$, with $i \in \{1,2,3,4\}$, is a tensor product of coherent states given by Eq. (1) in main text.

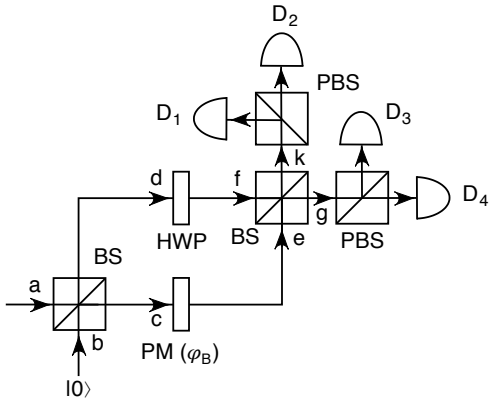

FIG. 3: Schematic representation of Bob's ddiQKD receiver [1]. The different modes a, b, c, d, e, f, g, and k correspond to those considered in the calculations of Sec. I; the receiver scheme is otherwise identical to Fig. 1(b) in main text.

II. SIDE-CHANNEL ATTACKS AGAINST BOB'S LINEAR OPTICS NETWORK

One main assumption of ddiQKD is that Bob's linear optics network (*i.e.,* the grey area within Bob's receiver in Fig. 1(b) in main text) is fully characterised and trusted. Note, however, that this does not mean that its devices need to be perfect, as this would be impossible to achieve in practice. In this section we show that Eve could also exploit various typical imperfections of Bob's linear optics to avoid double clicks when performing the blinding attack described in main text.

In particular, we will analyse two possible eavesdropping strategies in this context. In the first one, Eve uses the fact that Bob's PM $\varphi_{\rm B}$ is usually not perfect. More precisely, we study the situation where Bob's PM actually applies a phase $\varphi_{\rm B} = \bar{\varphi}_{\rm B} + \Delta_{\varphi_{\rm B}}$, where $\bar{\varphi}_{\rm B} \in \{0, \pi/2, \pi, 3\pi/2\}$ and the parameter $\Delta_{\varphi_{\rm B}}$ characterises the imperfection. In this scenario, Eve can select her phase $\phi_{\rm E} = \bar{\phi}_{\rm E} + \Delta_{\phi_{\rm E}}$, where $\bar{\phi}_{\rm E} \in \{0, \pi/2, \pi, 3\pi/2\}$ and $\Delta_{\phi_{\rm E}} > 0$ is a deviation term that Eve can select to control the detectors. According to Eq. (1) in main text, the energy at the input ports of Bob's detectors $\mathrm{D_1}$, $\mathrm{D_2}$, $\mathrm{D_3}$ and $\mathrm{D_4}$ is proportional to, respectively, $\frac{\mu}{2}[1 + \cos(\phi_{\rm E} - \varphi_{\rm B})]$, $\frac{\mu}{2}[1 + \cos(\phi_{\rm E} + \varphi_{\rm B})]$, $\frac{\mu}{2}[1 - \cos(\phi_{\rm E} - \varphi_{\rm B})]$, and $\frac{\mu}{2}[1 - \cos(\phi_{\rm E} + \varphi_{\rm B})]$. For simplicity, below we focus on the case $\bar{\varphi}_{\rm B} = \pi/2$. The other cases can be analysed similarly. We consider first the ideal scenario where $\Delta_{\varphi_{\rm B}} = 0$. The resulting normalised energies are illustrated in Fig. 4(a) as a function of $\phi_{\rm E}$. That is, as already seen in main text, when Eve's basis matches that of Bob, then two detectors receive maximum energy and, therefore, both click. If Bob and Eve use different bases then the total energy is equally distributed to all the four detectors and, given that $E_{\rm T}$ is

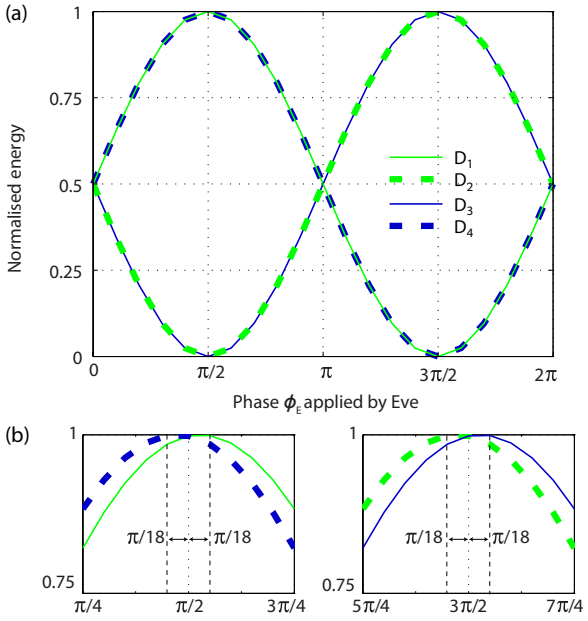

FIG. 4: Normalised energy at the input ports of Bob's detectors D_i as a function of ϕ_E, when $\bar{\varphi}_B = \pi/2$. (a) Ideal scenario with a perfect PM that has $\Delta_{\varphi_B} = 0$. (b) Example of a practical case where $\Delta_{\varphi_B} = \pi/36$ [2]. The normalised energy is defined as the energy divided by the energy of a coherent state with mean photon number μ. See text for further details.

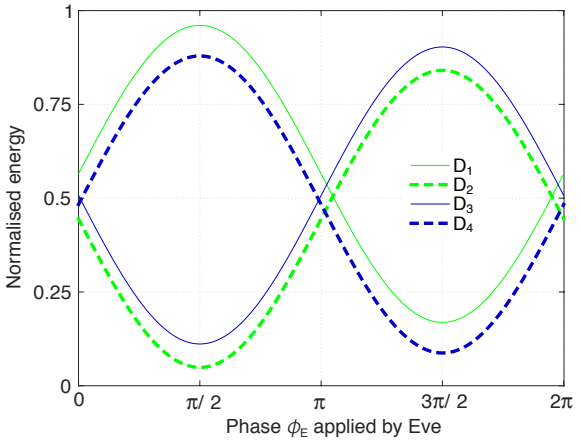

FIG. 5: Normalised energy at the input ports of Bob's detectors D_i as a function of ϕ_E, when $\varphi_B = \pi/2$. Here we assume that the splitting ratio of Bob's first (second) BS is 44 : 56 (46 : 54), and Eve's state parameter $\gamma = 0.2$. See text for further details.

chosen carefully, none of them click. Suppose now the practical scenario where Bob's state preparation is imperfect and Δ_{φ_B} is equal to say, for instance, $\pi/36$ (or 5°, which is a typical accuracy in practical systems [2]). In this situation, the energy distributions shift with respect to each other as highlighted in Fig. 4(b). If $\bar{\phi}_E = \pi/2$ and Eve selects say $\Delta_{\phi_E} = \pi/18$ ($\Delta_{\phi_E} = -\pi/18$) then the energy at the input ports of detectors D_1 and D_4 is, respectively, $E_+ \propto 0.998\mu$ and $E_- \propto 0.982\mu$ (E_- and E_+). Similarly, if $\bar{\phi}_E = 3\pi/2$ the energy at the detectors D_2 and D_3 is, respectively, E_- and E_+ (E_+ and E_-). That is, if Eve chooses carefully a suitable value of Δ_{ϕ_E} and μ such that $0.998\mu \geq \mu_{th}$ and $0.982\mu < \mu_{th}$, she can guarantee that only one detector clicks each given time, and no double-click is produced.

Finally, in the second eavesdropping strategy that we analyse we consider the situation where $\Delta_{\varphi_B} = 0$, but Eve exploits the fact that Bob's BSes are not perfect to avoid double-clicks. Although a 50 : 50 BS designed to operate at a certain wavelength (say, for example, at 1550 nm) can achieve nearly perfect splitting ratio at that wavelength, its splitting ratio can vary significantly at a different wavelength. For instance, a custom-made beamsplitter sample studied in Ref. [3] exhibited an extreme behaviour with splitting ratio of 98.6 : 1.4 (0.3 : 99.7) at 1470 nm (1290 nm). While commercial beamsplitter models may exhibit less variation, Eve in general can to some extent control the splitting ratio by simply changing the wavelength of the signals [3], and this could be used to avoid double-clicks.

In particular, suppose that Eve's signals are in a wavelength such that the splitting ratio of Bob's first (second) BS is $t_1 : 1 - t_1$ ($t_2 : 1 - t_2$). In addition, suppose that the creation operator of Eve's coherent states $|\sqrt{2\mu}\rangle$ is now given by $a^\dagger = (\sqrt{\gamma} a_H^\dagger + e^{i\phi_E}\sqrt{1-\gamma} a_V^\dagger)$, where the parameter γ is chosen by Eve. In this scenario, it can be shown that the state at the input ports of Bob's detectors D_i is a coherent state of the form

$$
\begin{aligned}
|\psi\rangle = &\left|\alpha\left(\sqrt{\hat{t}_1\hat{t}_2\hat{\gamma}}e^{i\phi_E} + \sqrt{t_1 t_2 \gamma}e^{i\varphi_B}\right)\right\rangle_{D_1} \\
&\otimes \left|\alpha\left(\sqrt{\hat{t}_1\hat{t}_2\gamma} + \sqrt{t_1 t_2 \hat{\gamma}}e^{i(\phi_E+\varphi_B)}\right)\right\rangle_{D_2} \\
&\otimes \left|\alpha\left(\sqrt{\hat{t}_1 t_2 \hat{\gamma}}e^{i\phi_E} - \sqrt{t_1\hat{t}_2\gamma}e^{i\varphi_B}\right)\right\rangle_{D_3} \\
&\otimes \left|\alpha\left(\sqrt{\hat{t}_1 t_2 \gamma} - \sqrt{t_1\hat{t}_2\hat{\gamma}}e^{i(\phi_E+\varphi_B)}\right)\right\rangle_{D_4}, \qquad (4)
\end{aligned}
$$

where $\hat{x} = 1 - x$, and $\alpha = \sqrt{2\mu}$. Note that when $t_1 = t_2 = \gamma = 1/2$ we obtain Eq. (1) in main text.

This means that, in principle, Eve might select the parameter γ and the wavelength of her signals such that the resulting splitting ratios t_1 and t_2 make the input energies at Bob's detectors asymmetric. In so doing, and following a similar argumentation to the one introduced in the previous eavesdropping strategy, Eve can guarantee that when she and Bob choose the same basis, only one detector clicks. This situation is illustrated in Fig. 5 for a particular example where $\varphi_B = \pi/2$, $t_1 = 0.44$, $t_2 = 0.46$, and $\gamma = 0.2$. In this scenario, we find that the maximum normalised energy at the input ports of Bob's detectors D_1 and D_4 when Eve selects $\phi_E = \pi/2$ is, respectively, 0.96 and 0.87. Similarly, when she chooses $\phi_E = 3\pi/2$ the

maximum normalised energy at the detectors D_3 and D_2 is, respectively, 0.9 and 0.84. Therefore, Eve can choose the energy of her signals such that only the detector D_1 (D_3) clicks when $\phi_E = \pi/2$ ($\phi_E = 3\pi/2$). That is, by changing the values of the parameters t_1, t_2, and γ, Eve can guarantee that only one detector clicks each given time.

[1] C. C. W. Lim, B. Korzh, A. Martin, F. Bussières, R. Thew, and H. Zbinden, Appl. Phys. Lett. **105**, 221112 (2014).

[2] F. Xu, K. Wei, S. Sajeed, S. Kaiser, S. Sun, Z. Tang, L. Qian, V. Makarov, and H.-K. Lo, Phys. Rev. A **92**, 032305 (2015).

[3] H.-W. Li, S. Wang, J.-Z. Huang, W. Chen, Z.-Q. Yin, F.-Y. Li, Z. Zhou, D. Liu, Y. Zhang, G.-C. Guo, W.-S. Bao, and Z.-F. Han, Phys. Rev. A **84**, 062308 (2011).