Research Article | Vol. 26, No. 16 | 6 Aug 2018 | OPTICS EXPRESS 21020

Optics EXPRESS

Eavesdropping and countermeasures for backflash side channel in quantum cryptography

PAULO VINICIUS PEREIRA PINHEIRO,[1,2,3,9] POOMPONG CHAIWONGKHOT,[3,4,*] SHIHAN SAJEED,[3,5,6] ROLF T. HORN,[3] JEAN-PHILIPPE BOURGOIN,[3,4] THOMAS JENNEWEIN,[3,4,7] NORBERT LÜTKENHAUS,[3,4] AND VADIM MAKAROV[8,4]

[1]*Engineering Department, Paraíso Faculty of Ceará, 63010-465, Juazeiro do Norte, Ceará, Brazil*
[2]*Department of Teleinformatic Engineering, Federal University of Ceará, C.P. 6007, Campus do Pici, Fortaleza, Brazil*
[3]*Institute for Quantum Computing, University of Waterloo, Waterloo, ON, N2L 3G1, Canada*
[4]*Department of Physics and Astronomy, University of Waterloo, Waterloo, ON, N2L 3G1, Canada*
[5]*Department of Electrical and Computer Engineering, University of Toronto, Toronto, ON, M5S 3G4, Canada*
[6]*Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, N2L 3G1, Canada*
[7]*Quantum Information Science Program, Canadian Institute for Advanced Research, Toronto, ON, Canada*
[8]*Russian Quantum Center and MISIS University, Moscow, Russia*
[9]*paulovpp@gmail.com*
[*]*poompong.ch@gmail.com*

Abstract: Quantum key distribution (QKD) promises information theoretic secure key as long as the device performs as assumed in the theoretical model. One of the assumptions is an absence of information leakage about individual photon detection outcomes of the receiver unit. Here we investigate the information leakage from a QKD receiver due to photon emission caused by detection events in single-photon detectors (backflash). We test commercial silicon avalanche photodiodes and a photomultiplier tube, and find that the former emit backflashes. We study the spectral, timing and polarization characteristics of these backflash photons. We experimentally demonstrate on a free-space QKD receiver that an eavesdropper can distinguish which detector has clicked inside it, and thus acquire secret information. A set of countermeasures both in theory and on the physical devices are discussed.

1. Introduction

Quantum key distribution (QKD) is one of the most developed branches of quantum communications. QKD offers protocol security in the sense that the QKD protocols [1–3] can be proven secure with a composable security definition under the model assumptions about Alice and Bob's devices, but without assumptions about the adversaries capabilities [4–7]. Nowadays, a large number of experimental systems are available [8–14]. However, the issue of implementation security is still a matter of concern as the security proof of the protocols make model assumptions. By definition, there is a gap between the behaviour of actual devices and their model. This leads to possible side channels exploitable by an eavesdropper Eve [15–29]. These side-channels may compromise the security of a QKD implementation if they are not taken into account. It is important to monitor potential side-channels and to design countermeasures to minimize their impact. With side-channels controlled in a best-practice approach, QKD will then show an implementation security that secures the generated key against future technological and algorithmic advances. The protocol security proof is an important component of that claim, even

#330517 https://doi.org/10.1364/OE.26.021020
Journal © 2018 Received 30 Apr 2018; revised 22 Jun 2018; accepted 28 Jun 2018; published 31 Jul 2018

Optics EXPRESS

when the implementation security claim is not a mathematical proof in itself. Different aspects of QKD systems have been exploited, including but not limited to timing [30], information leakage via Trojan-horse attack [15, 29, 31, 32], pulse-energy-monitoring system [24, 26], device calibration [33], source flaws [34], laser seeding [35], and laser damage [26]. However, most of the reported attacks exploited detectors [16, 19, 21, 25–28], making them the most vulnerable part of the system.

Among the exploitable vulnerabilities of the detectors such as efficiency mismatch [16, 25, 30], detector control [19, 21, 28, 36–39], and wavelength dependency [40], one has attracted considerably less attention: the backflash emission [41–48]. It has been known for a long time [49] that a reverse biased p-n junctions in a silicon avalanche photodiode in Geiger mode emits light upon the detection of a photon. Chynoweth and McKay [41] reported a detailed study of the phenomenon and predicted that the light emission originates due to the recombination of the energetic electrons and holes in the avalanche breakdown region. Subsequently, several other papers stated distinct possible causes for the phenomenon and quantified this emission [42–48]. In 2001, Kurtsiefer and his coworkers [50] raised the question: can this emission from the detectors employed in practical quantum communication systems affect the security? The outcome of their study suggested that the backflash photons might leak information about the detection to Eve, though the leakage of information was not quantified. Recently, a study about the backflash in InGaAs/InP avalanche photodiodes (APDs) was done [51]. The latter also suggests the possibility that Eve could measure state of backflash photons and learn about detection in the receiver without causing errors in the key.

The quantum state of the backflash photons is not expected to be correlated to that of the photon that triggered the effect. However, and unfortunately from a security point of view, the backflash photons may pass through other security critical components of Bob's receiver and carry out information about the state of those components back to the channel. For example, in polarization-based QKD with a passive basis-choice scheme, backflash photons from the horizontal (vertical) detectors will come out into the channel horizontally (vertically) polarized when they pass different arms of polarization beam-splitters (PBSes). In this case, Eve can measure the polarization of the backflash photons and predict with high probability which detector they originated from, thus compromising the security. Another possible method of distinguishing backflash photons from different channels is monitoring the difference in time delay of backflash photons from each channel. However, for the device studied in this article, preliminary tests have shown that the difference in time delay of backflash between channels is not sufficiently distinguishable to be used to determine the source of backflash. Thus, we do not investigate the latter method here.

The Article is organized as follows. In Section 2, we characterize backflash emission probability from APD and photomultiplier tube (PMT) instead of InGaAs/InP studied in Ref. [51]. Furthermore, in Section 3, we characterize backflash photons from a free-space polarization encoding receiver, and use that information to demonstrate a practical attack on the receiver. We also quantify the information leakage to Eve in this attack scheme. In Section 4, we introduce a countermeasure for this attack that reduces reverse transmission efficiency of the receiver from the detectors to channel to reduce information leakage. We also introduce a characterization procedure and modify the key rate equation to take into account the remaining information leakage. We conclude in Section 5.

2. Characterization of backflash emission

In order to study the effect of backflash photon emission during the avalanche breakdown, a series of experiments are conducted on two different types of detectors. The first device tested is a Si-APD detector module (Excelitas SPCM-AQRH-12-FC) with a circular active area of 180 μm and peak photon detection efficiency of 0.7 at 700 nm [52]. The second device tested is a PMT

Optics EXPRESS

(Hamamatsu H7422P-40), which has a GaAsP photocathode, with 5 mm diameter and a peak photon detection efficiency of 0.4 at 580 nm [53]. Both are thermoelectrically cooled.

2.1. Si avalanche photodiode

The first step in quantifying the information leakage is to find the probability of backflash P_b, i.e., the probability that a detection (click) leads to emission of at least one photon that leaks out of the detector. To find the value of P_b, we perform a measurement using the setup in Fig. 1(a). Two identical APD modules, one marked as device under test (DUT) and another marked as single-photon counting module (SPCM), are connected by a 2 m long 105 μm core diameter multimode fiber (Thorlabs M43L01). Click coincidences between them are recorded by a time interval analyzer (Stanford Research Systems SR620). In this setup, we record clicks caused by dark counts in the DUT. We record until the total clicks in DUT reach $N = 10^6$, and plot the histograms of coincidence clicks between the two detectors in Fig. 2. The right-most peak represents the backflash photons from DUT coupled through the fiber and detected by SPCM, which occur $\approx$ 10 ns after the detections in DUT owing to the optical delay. We have added a 40 ns electrical delay so that the coincidence click appears at a delay of $\approx$ 50 ns in the plot. This also allows us to see the backflash from SPCM recorded by DUT, which is the left-most peak having a similar shape but time-inversed. The shape of the coincidence peak roughly matches that of the current flowing through the APD I_{APD}, which we have measured using a small resistor added at the APD's cathode and a wideband differential oscilloscope probe. We divide the histogram into three regions. Region I shows rapid increase in coincidence counts that resembles the exponential increase of the number of avalanche electrons flowing through the APD. Region II shows decay in the coincidence counts resembling the decrease of avalanche electrons owing to the voltage across the APD dropping as its capacitance discharges. Region III is where the voltage across the APD is further lowered below breakdown by the quenching circuit. At that time the photon emission drops to near zero. The rough match between the current shape and the photon emission suggests that the backflash photons originate from the electric current across the APD during the avalanche.

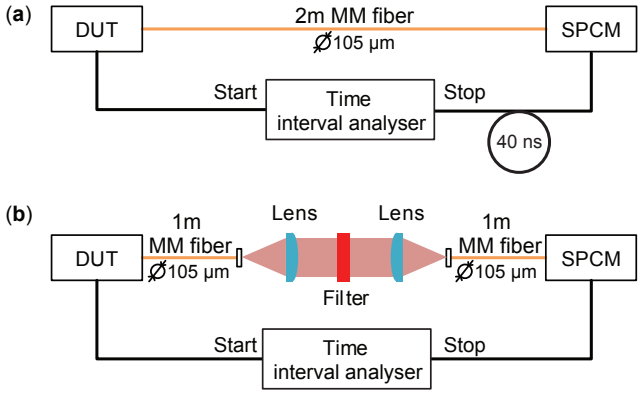

Fig. 1. Setup for measuring probability of backflash emission. (a) Two identical APDs are connected with a 2 m long multimode (MM) fiber causing 10 ns optical delay between the two detectors. An electronic delay line of 40 ns is added so that the backflash photons from SPCM could also be recorded. (b) To perform spectral analysis, a free-space interference narrowpass filter is added to the setup. The filter represents one often used at the entrance of a practical QKD receiver.

Optics EXPRESS

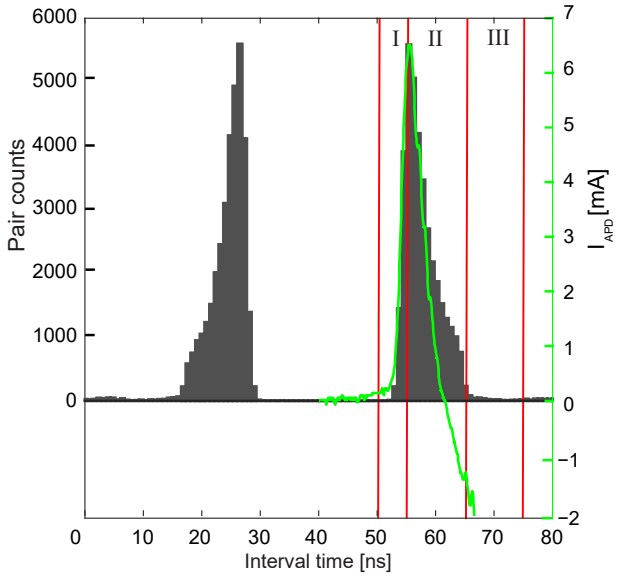

Fig. 2. Histogram of time-intervals (dark grey) measured from the coincident clicks from the setup in Fig. 1. The peak on the right is backflash from DUT detected by SPCM. Regions I, II, and III of the histogram represent different stages of detector operation cycle. The shape of histogram resembles the APD current I_{APD} (green line, measured separately). The current shape is not exact owing to a finite common-mode rejection ratio of the differential probe used to measure I_{APD}. The apparent abrupt drop of current at the border between regions II and III is common-mode interference from the quenching circuit that lowers the bias voltage and thus ends the avalanche. This coincides with a drop of photon emission almost to zero. The peak on the left is backflash from SPCM detected by DUT.

We count coincident clicks C within the right-hand peak. Here, we take into account channel transmission efficiency $T = 0.97$, and average detection efficiency of the SPCM in 500–900 nm spectral band $\eta = 0.6$ [52]. Since the SPCM can only detect photons efficiently in this narrow spectral band, our measurement provides only a lower bound estimate of $P_b \gtrsim C/(\eta T N)$. We note that this and subsequent calculations of backflash probability are approximate in the case where $P_b \ll 1$. For this specific setup, there are 37643 coincident detections, corresponding to $P_b \gtrsim 0.065$. Furthermore, we have measured the electrical charge flowing through the APD per avalanche, by monitoring the current consumption from the high-voltage bias source. We have found that the APD under test passes on average $n_{\mathrm{e}^-} = 2.7 \times 10^8$ electrons through the APD per avalanche. The probability of backflash photon emission per avalanche electron $P_{\mathrm{e}^-} \gtrsim P_b/n_{\mathrm{e}^-} = 2.4 \times 10^{-10}$. We remark that a detector circuit that reduces n_{e^-} would be expected to have lower backflash.

While the wideband measurement above is imprecise, many free-space QKD setups employ a narrowband spectral filter at Bob's entrance, in order to cut background light entering Bob [54–58]. The same filter would restrict the backflash emission to the narrow band that can be measured much more precisely in our setup. We have added a free-space narrowpass filter with center wavelength of 808 nm and bandwidth of 3 nm [see Fig. 1(b)], in order to mimic spectral filter inside a practical QKD receiver [58]. We have repeated the counting process and found 2306 coincident detections. At this specific wavelength, the SPCM has detection efficiency of 0.62 [52]. The coupling efficiency of the channel in this setup is $T = 0.83$. The probability of at least one

Optics EXPRESS

backflash photon leaking through this filter is $P_b^{\text{filter}} = 4.5 \times 10^{-3}$. The spectral filter indeed reduces the emission significantly, which reduces the information leakage as we prove later in Section 4.

We have performed another measurement to characterize the spectral distribution of the backflash photons, using a sensitive spectrum analyzer (Acton Spectrapro 2750). Unfortunately, we could not fully calibrate the spectrum analyzer for this specific setup, and the result is only qualitative. The measurement indicates that the backflash emission is broadband, spanning continuously from 550 nm to >1000 nm with a gentle peak around 900 nm (see Appendix A). This broadband characteristic leads to the possibility of including a narrow bandpass spectral filter in the system. The filter limits the wavelength range in which the backflash probability needs to be characterised, reduces the backflash emission from Bob, and thus reduces the information leakage.

2.2. Photomultiplier tube

Photomultiplier tube (PMT) is another type of detector widely used for its larger sensitive area and moderate dark count rate [59]. We have replaced DUT in Fig. 1 (a) with a PMT unit. Since the dark count rate of the PMT is low, additional weak laser pulses have been coupled to the active area of PMT to induce clicks. After recording 10^6 counts in the PMT, we have found fewer than 100 coincidences for both the fiber and free-space setups. This coincidence level is close to the dark count level of the SPCM, implying that the probability of backflash in PMT is negligible within the spectral range of our measurement.

3. Eavesdropping experiment

In this section, we experimentally quantify Eve's ability to identify which detector the backflash photons originated from, by measuring the backflash photon's polarization state. Bob's receiver used in this test is an integrated receiver built by INO (National Optics Institute of Canada) designed for a free-space passive polarization encoding QKD system running at 785 nm. Fig. 3 shows its optical scheme. The receiver consists of a pinhole to prevent spatial mode attack [25], coupling lens to focus incoming beam into optical fibers, and an integrated optics module. The latter consists of a beamsplitter (BS) to passively select the basis of measurement and PBSes in each basis to discriminate the four polarizations of the incoming photons: horizontal (H), vertical (V), diagonal (D), and antidiagonal (A). Next, we characterize the backflash emission as a possible side channel.

3.1. Reverse loss and extinction ratio

As the photons back-propagate through the setup, they experience the reverse loss of the receiver, i.e., the loss from originating detector to the channel input. This could reduce probability that backflash photon leaks into the channel. The setup shown in Fig. 4 is used to estimate this loss. An 808 nm laser (wavelength close to the operating wavelength of the receiver) is connected to the receiver's output multimode fiber, one channel at a time. The laser power at the end of receiver's fiber is $P_1 = 40\ \mu$W. We adjust the polarization controller PC to maximize throughput power, providing an upper bound of the reverse transmission. We then measure laser power P_2 emitted at the front of the receiver module, between the focusing lens and receiver's pinhole in 4. The reverse transmission efficiency of the receiver for the optimum polarization is then $T_b = P_2/P_1$. We have measured the average reverse transmission efficiency over all four channels of this receiver $T_b \approx 0.091$ (the individual values lie in the range 0.088 to 0.094). Assuming backflash photons are randomly polarized, their transmission should be approximately half of this upper bound.

Next, we demonstrate Eve's ability to distinguish the originating channel of backflash photon. For that, we measure polarization extinction ratio of the reverse emitted beam from the receiver.

Optics EXPRESS

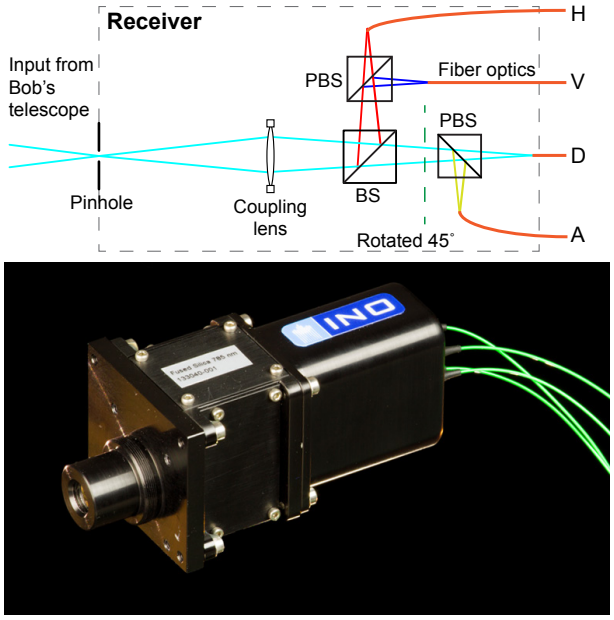

Fig. 3. Receiver designed by INO working as a passive basis choice polarization analyzer at 785 nm. Top: the important optical components consists of a pinhole, coupling lens, beamsplitter (BS), and polarizing beamsplitters (PBSes). Bottom: photo of the receiver. Four multimode fibers lead to the four detectors (not shown).

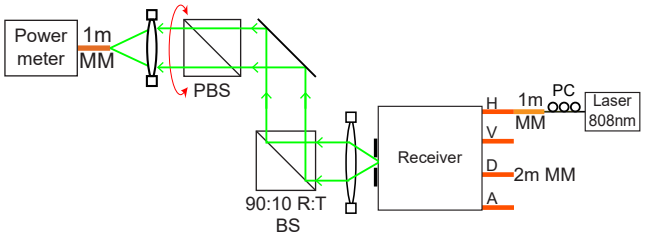

Fig. 4. Setup for measurement of the reverse propagation loss and polarization extinction ratio. An 808 nm laser is connected to each of the output channels of the receiver, one at a time. A 90:10 reflection:transmission (R:T) ratio beamsplitter diverts the reverse propagating beam to the measurement unit. The latter consists of a fiber-coupled optical power meter, and a rotating PBS to measure power and polarization extinction ratio of the reverse propagation beam. A polarization controller PC is used to maximize throughput power from each receiver channel.

Optics EXPRESS

Table 1. **Reverse propagating extinction ratio measurement of Bob's setup. The photons from H and V channel could be distinguished with high probability. The measured extinction ratios of A and D channels are low, presumably owing to polarization becoming elliptical at reflections in the measurement unit.**

Output channel	max		min		Extinction ratio
	Angle (deg)	Power (μW)	Angle (deg)	Power (μW)	
H	3	25.0	91	0.15	167
V	94	19.8	1	0.03	660
D	315	20.7	223	1.94	10.7
A	49	23.5	141	3.69	6.4

In Fig. 4, a 90:10 reflection:transmission (R:T) ratio beamsplitter is added to divert the outgoing beam from the receiver to a measurement unit consisting of a PBS and a fiber-coupled optical power meter. This additional setup has throughput efficiency $T_e = 0.60$. For each receiver channel input, we rotate the PBS to find a pair of angles that results in maximum and minimum power at the power meter. The optimal angles for each channel and respective extinction ratios are shown in Table 1. The drastically lower extinction ratio in D and A polarization is likely a result of polarization distortion caused by Fresnel effect on the dielectric mirror and the 90:10 BS used by Eve. These reflective surfaces were aligned at a certain angle along the axis corresponding to V polarization. This alignment distorted the diagonal polarization of the reflected beam, by inducing a phase difference between its H and V polarization components. In real eavesdropping, Eve can correct this polarization distortion using a phase compensator or waveplate. She can also split the incoming backflash photons into two PBSes oriented at the angles that yield the highest extinction ratios in both bases. This should allow her to distinguish the photons from all four channels with high probability.

3.2. *Timing of backflash photons through the receiver*

The previous experiment suggests that by measuring the polarization of the backflash photons, Eve could estimate which detector they originated from. However, in real life scenario, Eve's detection might not solely be from the backflash photons; it can be a result of stray light in the channel, reflection of Alice's signal from Bob's optical components or dark counts in Eve's detector – all unwanted noise. To avoid those unwanted signals, Eve needs to synchronize her measurement apparatus with Alice and Bob's signal pulses, and activate her detector at a specific time when the backflash photons are expected to arrive. The synchronization can be done by monitoring Alice's and Bob's signals prior to the eavesdropping. This section demonstrates a practical setup to measure timing characteristics of the backflash photons.

The experimental setup is shown in Fig. 5. A train of 3 *ns* wide laser pulses with 200 *ns* period is sent to Bob's receiver to simulate signals from Alice. The detector used as DUT in Section 2.1 is connected to one channel of the receiver at a time. A time interval analyzer (TIA) is used to record the coincidence time between the signal sent by Alice and Eve's SPCM clicks. In Fig. 6, we plot two histograms of the coincidence time from the APD in H channel. The green histogram is the coincidence time when DUT is powered off. Thus the detections in Eve resulted from reflections from the receiver's optical components. The positions of the peaks correspond to optical delay between reflective components in the setup and Eve's SPCM. The leftmost peak is a result of backreflection off the free-space optics at the front of Bob, such as his lenses and BS. The next peak matches the time delay from fiber splices in the receiver's fiber, indicated by short bars in Fig. 5. The third peak is the backreflection from the APD (in H channel only, as the

Optics EXPRESS

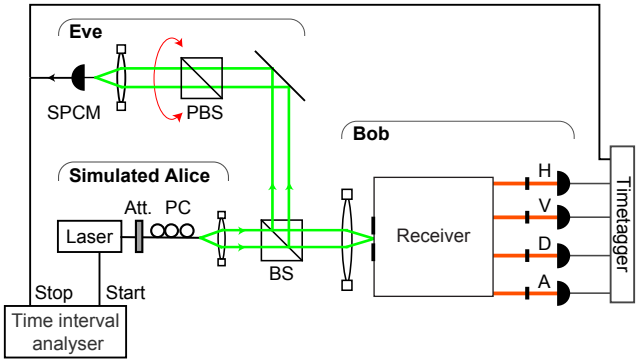

Fig. 5. Eavesdropping setup for timing characterization and proof-of-principle attack. The 90:10 R:T BS diverts photons from Bob to Eve's detector. Eve's setup consists of a PBS that can be rotated to find the optimal angle for Eve to distinguish the source of backflash photon. The time interval analyser (TIA) is used to find the time delay of the backflash photon in the channel. The timetagging unit records coincidence time between Bob's and Eve's detections in the proof-of-principle attack.

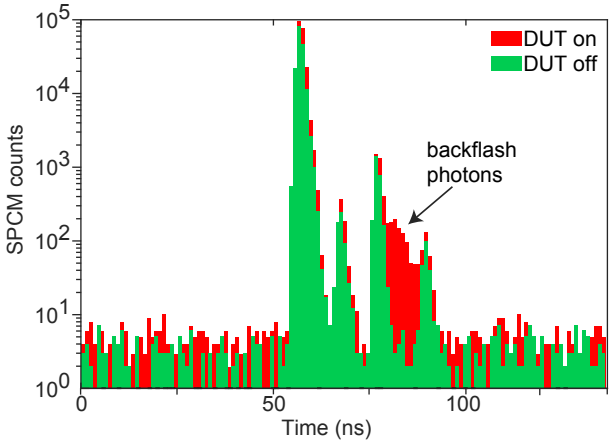

Fig. 6. Histogram of time intervals between emitting Alice's laser pulse and detection in Eve's SPCM. The histogram with DUT powered on (red) has an area of coincidence peak well above the level when DUT is powered off (green). The timing of this area matches the optical time delay between Eve's receiver and DUT, indicating backflash emission. The other peaks are optical reflections in the setup (see text for details).

fiber in the other channels has been terminated with matching gel that eliminates backreflections). The time delay of the right-most peak matches the round-trip of triple reflection between the APD and fiber splice. The red histogram is the coincidence time when DUT is powered on. Extra counts due to backflash photons can clearly be seen at 80–87 *ns*. The time delay matches optical delay between DUT and Eve's SPCM. Since the coincident counts of backflash events are $\approx$ 1.5 orders of magnitude higher than the back-reflection and noise level, the probability of Eve registering back-reflected pulses within this time window is small. Similar result could be seen when connecting the DUT to V, D, and A channels.

Optics EXPRESS

3.3. Proof-of-principle eavesdropping demonstration

We next emphasize the threat of this attack by demonstrating Eve's performance using a practical setup, shown in Fig. 5. In this experiment, we demonstrate Eve's ability to distinguish backflash emissions in one basis, between H and V channels. We only consider those photons that are coupled back to the optical channel and thus could carry information to Eve. We first repeat the alignment procedure as described in Section 3.1 by sending laser beam through the receiver's fibers, and rotating the PBS in Eve to find two optimal angles where the detection rate from the laser sent through Bob's H channel is maximum but V channel is minimum, and vise versa. Bob is then equipped with four powered-on APDs, one at each channel of the receiver, as in a real QKD setup. As seen in Section 3.2, Eve needs to register the coincidence counts within a specific time window to filter out back-reflection events. For that, we replace TIA with a timetagger (Dotfast Consulting 78-ps resolution 8-channel module) set to register the events where Eve's detector clicks within 25–30 ns after Bob's detection, which matches the time delay between Bob's and Eve's detectors. A train of 3 ns wide laser pulses with 200 ns period are sent to Bob to simulate QKD signal pulses from Alice. For each orientation of Eve's PBS, we count the number of detections in Bob and coincidence count in Eve over 10 s. We record the ratio of coincidence events $R_{ij} = E_{ij}/B_i$, where B_i is the number of clicks in Bob's ith detector, and E_{ij} is the number of Eve's coincident clicks with Bob's ith detector when she sets her PBS angle to maximise clicks from Bob's channel j. For example, R_{HH} represents probability of a click in Bob's H channel causing a coincident click while Eve aligns her PBS to measure signal from H channel, i.e., the probability that Eve gets a correct detection.

The probability of Eve gaining information (about H channel detection) is the chance of getting correct detection (R_{HH}) less the chance that she gets a wrong detection (R_{HV}). Note that the backflash probability P_b and reverse transmission efficiency T_b are already accounted in these coincidence ratios. Our measurements show that, for Bob's H detection, $R_{HH} = 5.00 \times 10^{-3}$ and $R_{HV} = 1.45 \times 10^{-3}$, causing information leakage of 3.5×10^{-3}. For Bob's V detection, $R_{VV} = 5.69 \times 10^{-3}$ and $R_{VH} = 3.66 \times 10^{-3}$, causing information leakage of 2.0×10^{-3}. From the calibration measurements we have expected the information leakage to be less than $\eta T_e T_b P_b/2 = 1.1 \times 10^{-3}$, which poorly matches the leakage observed in the eavesdropping experiment. We could not explain this discrepancy.

This result shows that Eve could learn a fraction of Bob's detections by monitoring the backflash photons. On the one hand, the information leakage is small, and we don't have the spectral filter in Bob in this experiment. On the other hand, our Eve's setup is not an optimal one for the attack. Proper countermeasures both in physical implementation and in post-processing step need to be considered.

4. Countermeasure

In this section, we discuss about possible countermeasures for attacks exploiting backflash photons. For physical implementation, using PMT can eliminate the possibility of generating backflash photons (although this conclusion is subject to the limitations of our measurement in Section 2.2). Another possible countermeasure is using measurement-device-independent QKD (MDI-QKD) [60, 61], in which the detection outcomes are public, thus Eve gains no new information from the backflash. However implementation of MDI-QKD in free-space is challenging [62–64]. If a non-MDI-QKD system uses APDs, the information leakage could be limited by decreasing reverse transmission efficiency T_b either by adding narrow-band spectral filter as shown in Section 2, or an optical isolator. These measures could reduce but not eliminate the leakage of information. The remaining leakage needs to be taken into account when calculating the required shortening of the key during privacy amplification.

The following procedure could be employed. Bob follows the procedure in Section 2.1 to find the APD's probability of backflash P_b and receiver's reverse transmission efficiency T_b. This T_b

Optics EXPRESS

includes all optical isolators and filters added to the receiver to limit the information leakage. If Bob does not include a narrow-pass filter, these parameters need to be characterized in a very wide spectral range, because typical free-space optics and air are transparent in a wide spectral band. This wide spectral characterization will be challenging. However, if a band-pass filter is used, it is sufficient to characterize the parameters over its spectral pass-band. From the result in Section 3.1, it is reasonable to assume that in the worst case, with ideal equipment, Eve could distinguish the origin of backflash photons with certainty. The information leakage to Eve is then $P_E = P_b T_b$. In other words, a fraction P_E of Bob's detections is tagged by Eve without disturbing the quantum state or inducing error. Then the privacy amplification for QKD with tagged signal [5, 65] can be used to take care of the information leakage.

As an example, let us consider the key rate equation for the Bennett-Brassard 1984 (BB84) protocol in QKD system with single-photon signals. Under the backflash attack, the secret key rate per signal sent by Alice becomes

$$l \geq AP_{det}(1 - h(\frac{e}{A})) - leak_{EC}, \tag{1}$$

where P_{det} is the probability of detection per signal, e is the error rate, $h(x) = -x \log x - (1 - x) \log(1 - x)$ is the binary Shannon entropy, and $leak_{EC}$ is the portion of key disclosed during error correction. The correction term $A = (P_{det} - P_E)/P_{det}$, where P_E is the information leakage calculated in the characterization step above.

The theoretical analysis in this paper considers only the worst-case scenario where Eve has the ability to collect and distinguish all backflash photons and map them to the raw key in Alice and Bob. This analysis also provides only the lower bound on the secret key rate, which could be improved by more careful analysis.

5. Conclusion

We have quantified the backflash emission of photons from APD-based single-photon detectors, and verified that these photons can be used by an eavesdropper to learn about the key in QKD systems. We have found that, for a system without spectral filter, at least 0.065 of the clicks in actively-quenched Si detector module result in backflash. This probability is reduced by a factor of 14 when a narrowband spectral filter is added, suggesting the latter is an efficient countermeasure. For PMT the backflash emission is negligible within the sensitivity of our measurement. Our experiment with a real polarization-encoding QKD receiver shows that Eve can distinguish polarization of backflash photons with near certainty. The proof-of-principle attack shows that Eve could learn 2.0×10^{-3} fraction of raw key using our today's imperfect setup. The information leakage may be higher for an ideal Eve. To close this loophole, we discuss a procedure to characterize the system and quantify Eve's information, then modify the key rate equation to take care of the information leakage due to backflash emission. We hope that our study will contribute to the development of certification and standardization of practical QKD against side-channels.

A. Spectral distribution measurement

Figure 7 shows the spectral distribution of backflash emission measured with a sensitive spectrum analyzer (Acton Spectrapro 2750). Due to difficulties we have encountered in spectrometer calibration, this measurement has a large margin of error comparing with the narrow-band filter measurement at a specific wavelength. Thus, we omit this result from the main Article. Even so, this measurement shows some important characteristics of backflash emission. The backflash emission is broadband, spanning continuously across our range of measurement from 550 to 1000 nm with a gentle peak around 900 nm. This suggest the possibility of having

backflash emission beyond our range of measurement. This emphasizes the necessity of adding the narrow-band filter to ease the characterization process and limit the information leakage.

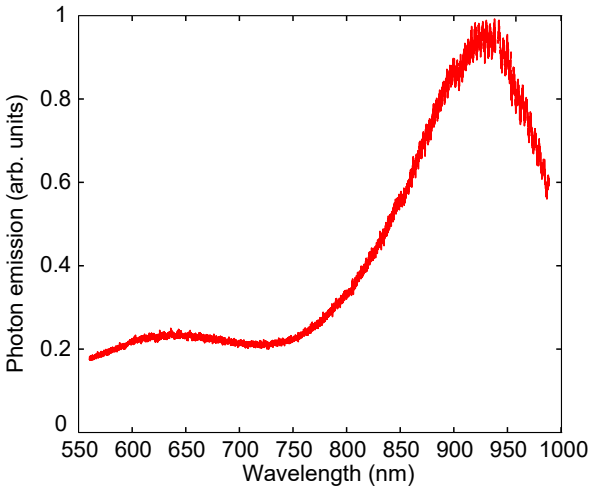

Fig. 7. Spectral distribution of backflash.

Funding

US Office of Naval Research (ONR) (N00014-13-1-0579); Industry Canada; CFI; Ontario MRI; Natural Sciences and Engineering Research Council of Canada (NSERC) (341495); the Canadian Space Agency (CSA).

Acknowledgments

This work was supported by the US Office of Naval Research, Industry Canada, CFI, Ontario MRI, NSERC (programs Discovery and CryptoWorks21), and the Canadian Space Agency. P.V.P.P. was also supported by a Brazilian CAPES scholarship (project no. 014633/2013-02). P.C. was also supported by a Thai DPST scholarship. J.-P.B. was also supported by FED DEV.

Author contributions: P.V.P.P. and P.C. contributed equally to this study. P.V.P.P. conducted measurements on the individual detectors and integrated receiver. P.C. conducted the eavesdropping demonstration experiment and theoretical analysis. S.S., R.T.H., and J.-P.B. assisted in setting up the experiments. N.L. supervised the theoretical analysis. V.M., T.J., and N.L. supervised the study. P.C. and P.V.P.P. wrote the article with input from all authors.

References

1. C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE International Conference on Computers, Systems, and Signal Processing (Bangalore, India),* (IEEE Press, 1984), pp. 175–179.
2. A. K. Ekert, "Quantum cryptography based on Bell's theorem," Phys. Rev. Lett. **67**, 661–663 (1991).
3. C. H. Bennett and D. P. DiVincenzo, "Quantum information and computation," Nature **404**, 247–255 (2000).
4. H.-K. Lo and H. F. Chau, "Unconditional security of quantum key distribution over arbitrarily long distances," Science **283**, 2050–2056 (1999).
5. N. Lütkenhaus, "Security against individual attacks for realistic quantum key distribution," Phys. Rev. A **61**, 052304 (2000).
6. P. W. Shor and J. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol," Phys. Rev. Lett. **85**, 441–444 (2000).
7. R. Renner, N. Gisin, and B. Kraus, "Information-theoretic security proof for quantum-key-distribution protocols," Phys. Rev. A **72**, 012332 (2005).

Optics EXPRESS

8. M. P. Peloso, I. Gerhardt, C. Ho, A. Lamas-Linares, and C. Kurtsiefer, "Daylight operation of a free space, entanglement-based quantum key distribution system," New J. Phys. **11**, 045007 (2009).
9. M. Peev, C. Pacher, R. Alléaume, C. Barreiro, J. Bouda, W. Boxleitner, T. Debuisschert, E. Diamanti, M. Dianati, J. F. Dynes, S. Fasel, S. Fossier, M. Fürst, J.-D. Gautier, O. Gay, N. Gisin, P. Grangier, A. Happe, Y. Hasani, M. Hentschel, H. Hübel, G. Humer, T. Länger, M. Legré, R. Lieger, J. Lodewyck, T. Lorünser, N. Lütkenhaus, A. Marhold, T. Matyus, O. Maurhart, L. Monat, S. Nauerth, J.-B. Page, A. Poppe, E. Querasser, G. Ribordy, S. Robyr, L. Salvail, A. W. Sharpe, A. J. Shields, D. Stucki, M. Suda, C. Tamas, T. Themel, R. T. Thew, Y. Thoma, A. Treiber, P. Trinkler, R. Tualle-Brouri, F. Vannel, N. Walenta, H. Weier, H. Weinfurter, I. Wimberger, Z. L. Yuan, H. Zbinden, and A. Zeilinger, "The SECOQC quantum key distribution network in Vienna," New J. Phys. **11**, 075001 (2009).
10. S. Nauerth, F. Moll, M. Rau, C. Fuchs, J. Horwath, S. Frick, and H. Weinfurter, "Air-to-ground quantum communication," Nat. Photonics **7**, 382 (2013).
11. Clavis2 specification sheet, http://marketing.idquantique.com/acton/attachment/11868/f-00a0/1/-/-/-/-/Clavis%20QKD%20Datasheet.pdf, visited 13 March 2018.
12. P. V. P. Pinheiro and R. V. Ramos, "Two-layer quantum key distribution," Quantum Inf. Process. **14**, 2111–2124 (2015).
13. T. Gehring, V. Händchen, J. Duhme, F. Furrer, T. Franz, C. Pacher, R. F. Werner, and R. Schnabel, "Implementation of continuous-variable quantum key distribution with composable and one-sided-device-independent security against coherent attacks," Nat. Commun. **6**, 8795 (2015).
14. S.-K. Liao, W.-Q. Cai, W.-Y. Liu, L. Zhang, Y. Li, J.-G. Ren, J. Yin, Q. Shen, Y. Cao, Z.-P. Li, F.-Z. Li, X.-W. Chen, L.-H. Sun, J.-J. Jia, J.-C. Wu, X.-J. Jiang, J.-F. Wang, Y.-M. Huang, Q. Wang, Y.-L. Zhou, L. Deng, T. Xi, L. Ma, T. Hu, Q. Zhang, Y.-A. Chen, N.-L. Liu, X.-B. Wang, Z.-C. Zhu, C.-Y. Lu, R. Shu, C.-Z. Peng, J.-Y. Wang, and J.-W. Pan, "Satellite-to-ground quantum key distribution," Nature **549**, 43 (2017).
15. A. Vakhitov, V. Makarov, and D. R. Hjelme, "Large pulse attack as a method of conventional optical eavesdropping in quantum cryptography," J. Mod. Opt. **48**, 2023–2038 (2001).
16. V. Makarov, A. Anisimov, and J. Skaar, "Effects of detector efficiency mismatch on security of quantum cryptosystems," Phys. Rev. A **74**, 022313 (2006). Erratum ibid. **78**, 019905 (2008).
17. B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," Quantum Inf. Comput. **7**, 73–82 (2007).
18. A. Lamas-Linares and C. Kurtsiefer, "Breaking a quantum key distribution system through a timing side channel," Opt. Express **15**, 9388–9393 (2007).
19. L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, "Hacking commercial quantum cryptography systems by tailored bright illumination," Nat. Photonics **4**, 686–689 (2010).
20. F. Xu, B. Qi, and H.-K. Lo, "Experimental demonstration of phase-remapping attack in a practical quantum key distribution system," New J. Phys. **12**, 113026 (2010).
21. I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, "Full-field implementation of a perfect eavesdropper on a quantum cryptography system," Nat. Commun. **2**, 349 (2011).
22. H. Weier, H. Krauss, M. Rau, M. Fürst, S. Nauerth, and H. Weinfurter, "Quantum eavesdropping without interception: an attack exploiting the dead time of single-photon detectors," New J. Phys. **13**, 073024 (2011).
23. P. Jouguet, S. Kunz-Jacques, and E. Diamanti, "Preventing calibration attacks on the local oscillator in continuous-variable quantum key distribution," Phys. Rev. A **87**, 062313 (2013).
24. S. Sajeed, I. Radchenko, S. Kaiser, J.-P. Bourgoin, A. Pappa, L. Monat, M. Legré, and V. Makarov, "Attacks exploiting deviation of mean photon number in quantum key distribution and coin tossing," Phys. Rev. A **91**, 032326 (2015).
25. S. Sajeed, P. Chaiwongkhot, J.-P. Bourgoin, T. Jennewein, N. Lütkenhaus, and V. Makarov, "Security loophole in free-space quantum key distribution due to spatial-mode detector-efficiency mismatch," Phys. Rev. A **91**, 062301 (2015).
26. V. Makarov, J.-P. Bourgoin, P. Chaiwongkhot, M. Gagné, T. Jennewein, S. Kaiser, R. Kashyap, M. Legré, C. Minshull, and S. Sajeed, "Creation of backdoors in quantum communications via laser damage," Phys. Rev. A **94**, 030302 (2016).
27. A. Huang, S. Sajeed, P. Chaiwongkhot, M. Soucarros, M. Legré, and V. Makarov, "Testing random-detector-efficiency countermeasure in a commercial system reveals a breakable unrealistic assumption," IEEE J. Quantum Electron. **52**, 8000211 (2016).
28. S. Sajeed, A. Huang, S. Sun, F. Xu, V. Makarov, and M. Curty, "Insecurity of detector-device-independent quantum key distribution," Phys. Rev. Lett. **117**, 250505 (2016).
29. S. Sajeed, C. Minshull, N. Jain, and V. Makarov, "Invisible Trojan-horse attack," Sci. Rep. **7**, 8403 (2017).
30. Y. Zhao, C.-H. F. Fung, B. Qi, C. Chen, and H.-K. Lo, "Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems," Phys. Rev. A **78**, 042333 (2008).
31. N. Gisin, S. Fasel, B. Kraus, H. Zbinden, and G. Ribordy, "Trojan-horse attacks on quantum-key-distribution systems," Phys. Rev. A **73**, 022320 (2006).
32. N. Jain, E. Anisimova, I. Khan, V. Makarov, C. Marquardt, and G. Leuchs, "Trojan-horse attacks threaten the security of practical quantum cryptography," New J. Phys. **16**, 123030 (2014).
33. N. Jain, C. Wittmann, L. Lydersen, C. Wiechers, D. Elser, C. Marquardt, V. Makarov, and G. Leuchs, "Device calibration impacts security of quantum key distribution," Phys. Rev. Lett. **107**, 110501 (2011).
34. F. Xu, K. Wei, S. Sajeed, S. Kaiser, S. Sun, Z. Tang, L. Qian, V. Makarov, and H.-K. Lo, "Experimental quantum key

Optics EXPRESS

distribution with source flaws," Phys. Rev. A **92**, 032305 (2015).
35. S.-H. Sun, F. Xu, M.-S. Jiang, X.-C. Ma, H.-K. Lo, and L.-M. Liang, "Effect of source tampering in the security of quantum cryptography," Phys. Rev. A **92**, 022304 (2015).
36. L. Lydersen and J. Skaar, "Security of quantum key distribution with bit and basis dependent detector flaws," Quantum Inf. Comput. **10**, 60–76 (2010).
37. L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, "Thermal blinding of gated detectors in quantum cryptography," Opt. Express **18**, 27938–27954 (2010).
38. L. Lydersen, J. Skaar, and V. Makarov, "Tailored bright illumination attack on distributed-phase-reference protocols," J. Mod. Opt. **58**, 680–685 (2011).
39. L. Lydersen, M. K. Akhlaghi, A. H. Majedi, J. Skaar, and V. Makarov, "Controlling a superconducting nanowire single-photon detector using tailored bright illumination," New J. Phys. **13**, 113042 (2011).
40. 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, "Attacking a practical quantum-key-distribution system with wavelength-dependent beam-splitter and multiwavelength sources," Phys. Rev. A **84**, 062308 (2011).
41. A. G. Chynoweth and K. G. McKay, "Photon emission from avalanche breakdown in silicon," Phys. Rev. **102**, 369–376 (1956).
42. M. Waldschmidt and S. Wittig, "Backscattering and bremsstrahlung of electrons in a silicon detector," Nucl. Instr. Meth. **64**, 189–191 (1968).
43. P. A. Childs and W. Eccleston, "Impact ionization induced minority carrier injection by avalanching p-n junctions," J. Appl. Phys. **55**, 4304–4308 (1984).
44. D. K. Gautam, W. S. Khokle, and K. B. Garg, "Photon emission from reverse-biased silicon P-N junctions," Solid-State Electron. **31**, 219–222 (1988).
45. A. Lacaita, S. Cova, A. Spinelli, and F. Zappa, "Photon-assisted avalanche spreading in reach-throught photodiodes," Appl. Phys. Lett. **62**, 606–608 (1993).
46. N. Akil, S. E. Kerns, D. V. Kerns Jr., A. Hoffmann, and J.-P. Charles, "Photon generation by silicon diodes in avalanche breakdown," Appl. Phys. Lett. **73**, 871–872 (1998).
47. A. Pacelli, A. S. Spinelli, and A. L. Lacaita, "Impact ionization in silicon: A microscopic view," J. Appl. Phys. **83**, 4760–4764 (1998).
48. T. Huang, J. Shao, X. Wang, L. Xiao, and S. Jia, "Photon emission characteristics of avalanche photodiodes," Opt. Eng. **44**, 074001 (2005).
49. R. Newman, "Visible light from a silicon p–n junction," Phys. Rev. **100**, 700–703 (1955).
50. C. Kurtsiefer, P. Zarda, S. Mayer, and H. Weinfurter, "The breakdown flash of silicon avalanche photodiodes—back door for eavesdropper attacks?" J. Mod. Opt. **48**, 2039–2047 (2001).
51. A. Meda, I. P. Degiovanni, A. Tosi, Z. Yuan, G. Brida, and M. Genovese, "Quantifying backflash radiation to prevent zero-error attacks in quantum key distribution," Light. Sci. Appl. **6**, e16261 (2016).
52. SPCM-AQRH single photon counting module data sheet, http://www.excelitas.com/Downloads/DTS_SPCM-AQRH.pdf, visited 13 March 2018.
53. Photosensor modules H7422 series, https://www.hamamatsu.com/resources/pdf/etd/m-h7422e.pdf, visited 13 March 2018.
54. W. T. Buttler, R. J. Hughes, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. G. Peterson, "Daylight quantum key sistribution over 1.6 km," Phys. Rev. Lett. **84**, 5652–5655 (2000).
55. R. J. Hughes, J. E. Nordholt, D. Derkacs, and C. G. Peterson, "Practical free-space quantum key distribution over 10 km in daylight and at night," New J. Phys. **4**, 43 (2002).
56. C. Kurtsiefer, P. Zarda, M. Halder, H. Weinfurter, P. M. Gorman, P. R. Tapster, and J. G. Rarity, "Quantum cryptography: A step towards global key distribution," Nature **419**, 450–450 (2002).
57. C. Kurtsiefer, P. Zarda, M. Halder, P. M. Gorman, P. R. Tapster, J. G. Rarity, and H. Weinfurter, "Long distance free space quantum cryptography," Proc. SPIE **4917**, 25–31 (2002).
58. J.-P. Bourgoin, E. Meyer-Scott, B. L. Higgins, B. Helou, C. Erven, H. Hübel, B. Kumar, D. Hudson, I. D'Souza, R. Girard, R. Laflamme, and T. Jennewein, "A comprehensive design and performance analysis of low Earth orbit satellite quantum communication," New J. Phys. **15**, 023006 (2013).
59. R. H. Hadfield, "Single-photon detectors for optical quantum information applications," Nat. Photonics **3**, 696–705 (2009).
60. H.-K. Lo, M. Curty, and B. Qi, "Measurement-device-independent quantum key distribution," Phys. Rev. Lett. **108**, 130503 (2012).
61. Y. Liu, T.-Y. Chen, L.-J. Wang, H. Liang, G.-L. Shentu, J. Wang, K. Cui, H.-L. Yin, N.-L. Liu, L. Li, X. Ma, J. S. Pelc, M. M. Fejer, C.-Z. Peng, Q. Zhang, and J.-W. Pan, "Experimental measurement-device-independent quantum key distribution," Phys. Rev. Lett. **111**, 130502 (2013).
62. B. Qi, H. K. Lo, C. C. W. Lim, G. Siopsis, E. A. Chitambar, R. Pooser, P. G. Evans, and W. Grice, "Free-space reconfigurable quantum key distribution network," in *2015 IEEE International Conference on Space Optical Systems and Applications (ICSOS),* (2015), pp. 1–6.
63. W. Le, Z. Sheng-Mei, G. Long-Yan, and C. Wei-Wen, "Free-space measurement-device-independent quantum-key-distribution protocol using decoy states with orbital angular momentum," Chin. Phys. B **24**, 120307 (2015).
64. E. Diamanti, H.-K. Lo, B. Qi, and Z. Yuan, "Practical challenges in quantum key distribution," NPJ Quantum Inf. **2 (2016).**
65. D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, "Security of quantum key distribution with imperfect devices," Quantum Inf. Comput. **4**, 325–360 (2004).