

# Randomized Ancillary Qubit Overcomes Detector-Control and Intercept-Resend Hacking of Quantum Key Distribution

Salem F. Hegazy <sup>1</sup>, Member, IEEE, Salah S. A. Obayya <sup>2</sup>, Fellow, IEEE, Fellow, OSA, and Bahaa E. A. Saleh <sup>3</sup>, Life Fellow, IEEE, Fellow, OSA

**Abstract**— Practical implementations of quantum key distribution (QKD) have been shown to be subject to various detector side-channel attacks that compromise the promised unconditional security. Most notable is a general class of attacks adopting the use of faked-state photons as in the detector-control and, more broadly, the intercept-resend attacks. In this paper, we present a simple scheme to overcome such class of attacks: A legitimate user, Bob, uses a polarization randomizer at his gateway to distort an ancillary polarization of a phase-encoded photon in a bidirectional QKD configuration. Passing through the randomizer once on the way to his partner, Alice, and again in the opposite direction, the polarization qubit of the genuine photon is immune to randomization. However, the polarization state of a photon from an intruder, Eve, to Bob is randomized and hence directed to a detector in a different path, whereupon it triggers an alert. We demonstrate theoretically and experimentally that, using commercial off-the-shelf detectors, it can be made impossible for Eve to avoid triggering the alert, no matter what faked-state of light she uses.

**Index Terms**—Detector control, faked state photons, intercept and resend attacks, optical fiber communication, plug-and-play quantum key distribution, quantum encryption.

## I. INTRODUCTION

THE unconditional security offered by quantum key distribution (QKD) relies on laws of quantum physics [1], [2], which dictate that any attempt by an adversary to know about the secret key, would inevitably introduce disturbance that alerts the legitimate parties [3], [4]. This ultimate information-theoretic security has been proved for idealized devices [4], [5], [6] and also under semi-realistic conditions [7], [8], [9]. In practice,

Manuscript received 22 March 2022; revised 15 June 2022 and 5 August 2022; accepted 7 August 2022. Date of publication 11 August 2022; date of current version 4 November 2022. This work was supported by the Academy of Scientific Research and Technology (ASRT) through JESOR Program under Grant 5283. (Corresponding author: Bahaa Saleh.)

Salem F. Hegazy is with the National Institute of Laser Enhanced Sciences, Cairo University, Giza 12613, Egypt (e-mail: salem@niles.cu.edu.eg).

Salah S. A. Obayya is with the Center for Photonics and Smart Materials, Zewail City of Science and Technology, Giza 12578, Egypt (e-mail: sobayya@zewailcity.edu.eg).

Bahaa E. A. Saleh is with the CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, FL 32816 USA (e-mail: besaleh@creol.ucf.edu).

Color versions of one or more figures in this article are available at <https://doi.org/10.1109/JLT.2022.3198108>.

Digital Object Identifier 10.1109/JLT.2022.3198108

however, real-life components of QKD systems may deviate from these idealized theoretical models, or encounter new scenarios, offering effective vulnerabilities to the adversary.

For instance, the imperfect preparation of the single-photon state may lead to leaking information about the key. This gap between theory and real-life practice allows for a plethora of source-side attacks ranging from the photon-number-splitting (PNS) attack [10], [11], the phase-remapping attack [12], [13], the wavelength-selected photon-number-splitting attack [14], and the pattern-effect attack [15], to the nonrandom-phase attacks based on unambiguous-state-discrimination [16], and laser seed control [17], [18], [19].

Compared to the source-side attacks, imperfections on the detection side are known to show much higher vulnerability to quantum hacking [20]. For example, detector imperfections such as breakdown fluorescence [21], finite ( $\sim\mu\text{s}$ ) dead time [22], nonzero dark counts, less-than-unity efficiency, and nonfixed efficiency within the gate time [23], all of which can be exploited by Eve to compromise QKD security. This leads in practice to a significant number of potential attacks such as detector fluorescence [24], faked-state [25], [26], time-shift [23], [27], time-side-channel [28], channel calibration [29], laser damage [30], [31], spatial mismatch [32], [33], detector saturation [34], and polarization shift [35] attacks. More interestingly, the single-photon detectors (SPDs) of the receiver (Bob), normally operating in the Geiger mode [36], can be turned by Eve into linear mode, which allows for various blinding and remote-control attacks [37], [38], [39], [40], [41], [42], [43], [44]. Among the detection-side attacks, the latter is widely known to be the most powerful [20], with successful demonstrations on various types of SPDs, including passively and actively quenched avalanche photodetectors (APDs) [37], [45], gated/non-gated APDs [38], [46], and superconducting nanowire single-photon detectors (SNSPDs) [47].

Since the inception of quantum encryption [1], the intercept-resend strategies have been developed through many quantum hacking paradigms. Its original version based on resending single photons was easily neutralized by QKD [3]. Employing detector imperfections, more crafty intercept-resend versions have evolved via resending faked multiphoton states either solitarily (e.g., the after-gate attack [42], the faint-after-gate attack [47], and the detector-control attack under specific laser damage [30])

or teamed with a blinding light (e.g., continuous-wave blinding attack [38], [39], sinkhole blinding attack [48], thermal blinding attack [45], [48], and pulsed illumination attack [44]).

Currently, there exist two main approaches against the intercept-resend and detector-control hacking strategies. The first is based on monitoring some detector measures, such as its photocurrent, for anomalously excessive values [49], [50], [51]. This includes also observing the detector's count rates versus random variations of either the detection efficiency [52], [53], or the attenuation in front of the detector [54]. These security patches could defeat the original attacks they were designed for, but unfortunately they fail against subsequent *ad-hoc* modified attacks [46], [55].

The second is the measurement-device-independent QKD (MDI-QKD) approach [56], which enables elimination of all detector side-channels [57], offering security regardless of the nature of the detection apparatus. However, MDI QKD builds on performing a remote Bell-state measurement, which requires high-visibility two-photon interference between independent photons from Alice's and Bob's laser sources, a practically challenging procedure.

In this paper, we present a scheme to protect practical QKD systems against various attacks based on faked-state light, including the detector-control attacks and more generally the class of intercept-resend attacks. The scheme uses phase encoding and a two-way configuration, similar to the plug-and-play configuration [58], [59], [60], which uses polarization-assisted routing through Bob's transceiver, and a Faraday mirror at Alice's site. In our scheme, however, the polarization qubit serves a different function. A photon generated at Bob's transceiver is transmitted through a polarization randomizer, which assigns it a random state of polarization, and upon reflection from the Faraday mirror it passes once more through the same randomizer, in a state orthogonal to its original state, and is directed to a specific path, whereupon the photon is detected in accordance with the phase-encoded BB84 protocol. Light pulses generated by an intruder must pass through the randomizer at the gateway to Bob's transceiver, and since they pass only once, they acquire a random state and end up in a different path, whereupon their detection triggers an alert. The randomizer is fixed during the course of the photon roundtrip and is refreshed after every cycle of photon transmission and detection. Thus, the polarization qubit serves as a carrier of a *password* that allows genuine photons to be directed to the secured detectors, while an intruder's fake photons are randomized and possibly end up at the alert detectors.

We further consider the case that Eve launches a generalized detector-control attack. To render her attack unnoticeable, she tailors the parameters of triggering pulses and blinding light in order to meet two requirements: (i) to avoid triggering alert detectors, and (ii) to be able to sometimes trigger the secured detectors in the right way. These two requirements lead us to a necessary and sufficient condition that Bob's secured and alert detectors have to satisfy. We note that commercially available detectors can violate this necessary and sufficient condition and thereby guarantee that these two requirements are impossible to meet simultaneously. We experimentally demonstrate how various faked states by Eve fail to simultaneously meet these

two requirements of unnoticeable attack. Security analysis of the system shows that for various types of attacks Eve cannot diminish the alert rate, even if she has complete control over Bob's secured detectors.

## II. QKD SCHEME

As shown in Fig. 1(a), Bob employs a single photon with two encoded qubits: a time-bin qubit communicating the *key*, and an ancillary polarization qubit serving as a *security pass* [61]. As in typical interferometric QKD systems, the photon undergoes a roundtrip from Bob to Alice, where the time-bin qubit is modulated, and sent back to Bob whereupon it is directed to two sets of detectors depending on its state of polarization. Entry into Bob's receiver is secured by a polarization randomizer applying a random transformation  $U$  (based on Haar measure) that changes every photon-roundtrip duration. Alice uses a Faraday mirror (FM) that switches the polarization qubit into an orthogonal state so that as the photon crosses the polarization randomizer in the opposite direction, the randomization is cleared. Since its state is only known to Bob, the randomizer is a secure polarization-based gateway that directs the photon to specific detectors in the receiver.

The process begins as shown in Fig. 1(a) with Bob sending single-photon pulses along path  $a$  in a polarization-path state:

$$|\psi_1\rangle = \frac{1}{\sqrt{2}}(|H\rangle + |V\rangle)|a\rangle. \quad (1)$$

This is subsequently swapped for a time-polarization state

$$|\psi_2\rangle = \frac{1}{\sqrt{2}}(|t_l\rangle + |t_s\rangle)|H\rangle, \quad (2)$$

by use of an unbalanced polarization-based Mach-Zehnder interferometer (PMZI) with a polarization controller (PC) placed in its short arm, converting the V (H) polarization into H (V) polarization.

On Alice's side, the leading time bin  $|t_s\rangle$  is encoded with a phase shift  $\phi_A$  of 0 or  $\pi$ , and  $\pi/2$  or  $3\pi/2$ . Upon reflection from the FM, the photon polarization is flipped to its orthogonal state. This compensates for the undesired polarization changes accompanying the phase modulation [62], and also for the birefringence-based polarization fluctuations along the optical fiber [63], [64]. Upon re-entry into Bob's transceiver, since  $U$  is fixed during the photon roundtrip, its effect is also cancelled out by transmission in the opposite direction. The state is now:

$$|\psi_3\rangle = \frac{1}{\sqrt{2}}(|t_l\rangle + e^{i\phi_A}|t_s\rangle)|V\rangle. \quad (3)$$

Bob's receiver is gated to select roundtrip passage via the short-long and the long-short arms of the PMZI arms. It is also configured such that with single-photon interference in the PMZI, the time-polarization state  $|\psi_3\rangle$  is swapped back to a polarization-path state

$$|\psi_4\rangle = \frac{1}{\sqrt{2}}(|H\rangle + e^{i\phi_A}|V\rangle)|b\rangle. \quad (4)$$

The photon is therefore directed to path  $b$ , which we call the *secure* path. As will be shown later, detection of a photon in path  $a$  is an indication that the system has been tampered with, and path  $a$  is therefore called the *alert* path.

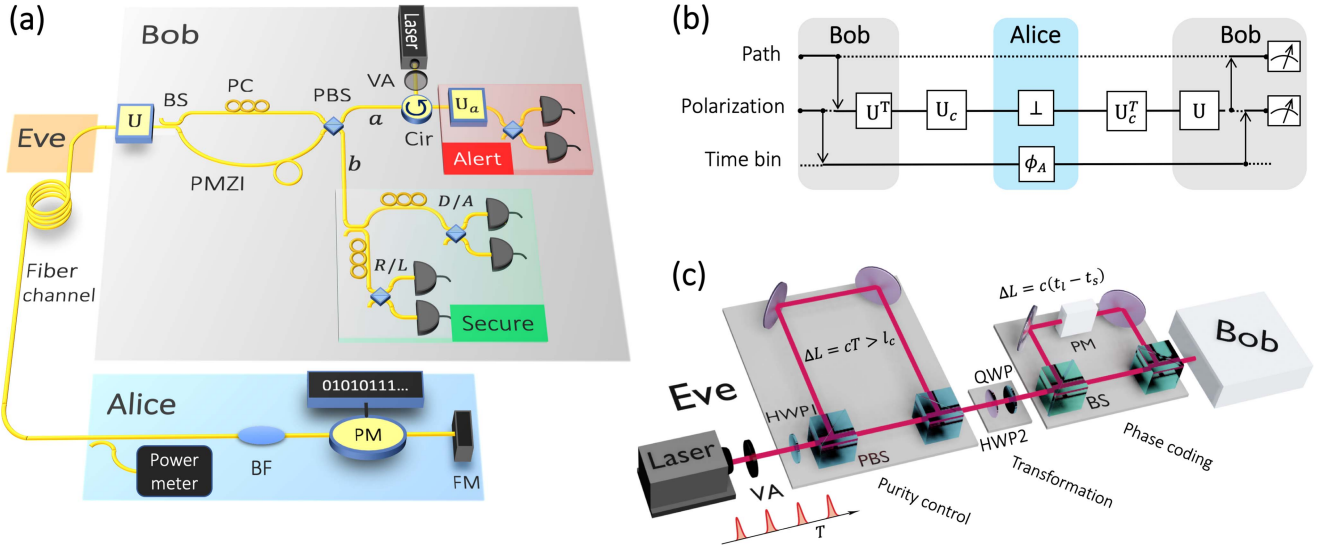


Fig. 1. (a) Optical layout of the QKD system. Bob creates single photons with time-bin (key) and polarization (ancillary) qubits. The polarization qubit is randomized by an operator  $U$ , only known to Bob. Alice's phase modulator (PM) encodes the time-bin state by a phase  $\phi_A \in \{0, \pi\}$  or  $\{\pi/2, 3\pi/2\}$ . A Faraday mirror (FM) compensates Bob's back-tracing photon for all encountered polarization variations, including the randomization  $U$ . The polarization-based Mach-Zehnder interferometer (PMZI) swaps the time-bin/polarization qubits for polarization/path qubits. Therefore, the key qubit is measured in path  $b$  in either diagonal-antidiagonal ( $D/A$ ) or right-left ( $R/L$ ) circular polarization bases. The polarization randomizer  $U$  – which may be implemented by means of high-speed electro-optic polarization controller – is active against Eve's fake photons and may direct them, without Eve's notice, to the alert detectors in path  $a$ . A click of the alert detectors in path  $a$  is a sign for Eve's intrusion. The polarization switch  $U_a$  alternates between measurements bases:  $D/A$  and  $R/L$ . BS: beam splitter; PBS: polarization beam splitter; PC: polarization controller; Cir: optical circulator; VA: variable attenuator; BF: narrow band-pass filter. (b) The timeline for the operations on qubits of the three photonic degrees of freedom, path, time bin, and polarization, during the course of a roundtrip from Bob to Alice and back along a channel  $U_c$ . The operator  $U$  describes the polarization transformation, when light enters Bob's system. In the opposite direction, it encounters a transformation  $U^T$ . (c) Optical setup demonstrating Eve's system. The half-wave plate HWPI and the following polarization-based two-path system control the purity of the polarization state via mixing orthogonal polarization components of two subsequent laser pulses. The subsequent half- and quarter-wave plates, HWP2 and QWP, alter the polarization state unitarily. The two-path system in the last stage performs the time-bin phase encoding.

After swapping the key qubit back to polarization, the BB84 measurement is performed passively in one of the conjugate bases: diagonal-antidiagonal ( $D/A$ ) or right-left ( $R/L$ ) circular polarization. The system's action on the different degrees of freedom (path, time, and polarization) of the photon during its roundtrip course is illustrated in Fig. 1(b).

In yet another measure of added security, Bob randomly directs the received photon – in a managed way – to path  $a$  instead of path  $b$  for measurement. This is accomplished by appropriate control of the polarization randomizer. This random-switching tactic unveils types of attacks that can bias triggering actions to path  $b$  such as pulsed-blinding [22], [44], [55] and wavelength-dependent attacks [69].

Alice's phase coding and Bob's gated detection require precise time synchronization between the two sides which is done via a wavelength-multiplexed classical channel carrying bright pulses. A portion of the power received by Alice is monitored to detect Trojan horse attacks [63].

Here, an ideal single-photon source is assumed for convenience. To defend against the PNS attack, Bob applies a decoy-state technique [65], [66], [67]; verifying that his produced decoy pulses encounter the same single-photon loss.

### III. RANDOMIZED ROUTING OF FAKED-STATE LIGHT

Eve's goal is to signal the detectors in the secure path  $b$  without registering a click on the detectors of the alert path  $a$ . In a typical intercept-resend strategy, Eve would measure Alice's encoded

state and then send faked-state light in a phase modulated state  $(|t_l\rangle + e^{i\phi_E}|t_s\rangle)/\sqrt{2}$ , mimicking the measured key qubit, together with a polarization qubit in a state  $\rho_p$ . Upon transmission through the PMZI, and within the detection window (centered at:  $t_s + t_l$ ), the state of Eve's photon(s) becomes

$$\begin{aligned} & \frac{1}{2} |H\rangle\langle H| (|b\rangle X + e^{i\phi_E} |a\rangle) U \rho_p U^+ \\ & \quad \times (X \langle b| + e^{-i\phi_E} \langle a|) |H\rangle\langle H| \\ & + \frac{1}{2} |V\rangle\langle V| (|a\rangle X + e^{i\phi_E} |b\rangle) U \rho_p U^+ \\ & \quad \times (X \langle a| + e^{-i\phi_E} \langle b|) |V\rangle\langle V|. \end{aligned} \quad (5)$$

The NOT operator  $X$  is due to action of the PC in the PMZI. To obtain the which-path statistics, we trace over polarization and obtain the reduced density operator of the path states

$$\begin{aligned} & p_a |a\rangle\langle a| + \cos \phi_E \langle H| U \rho_p U^+ |V\rangle |a\rangle\langle b| \\ & + \cos \phi_E \langle V| U \rho_p U^+ |H\rangle |b\rangle\langle a| + p_b |b\rangle\langle b|. \end{aligned} \quad (6)$$

The probabilities that Eve's photon(s) ends up in the alert path  $a$  is  $p_a = \langle H| U \rho_p U^+ |H\rangle$ , while that of reaching path  $b$  is  $p_b = 1 - p_a = \langle V| U \rho_p U^+ |V\rangle$ . If Eve were to know the operator  $U$ , she would be able to make  $p_a = 0$  by use of a pure state  $\rho_p = U^+ |V\rangle\langle V| U$ . Not knowing  $U$ , if she runs the conventional intercept-resend attack [1], [68] by measuring the Alice-encoded photon and re-sending a new photon prepared in accordance with the measurement outcome to Bob, then the average probability that it passes to path  $a$  is 25% (obtained

by averaging over the continuum of random realizations of  $U$  based on Haar measure, assuming ideal single-photon sources, measurements, and detection). This alert rate is on top of the normal 25% quantum bit error rate (QBER) of the BB84 key qubit.

#### IV. NECESSARY CRITERIA FOR BOB'S DETECTORS

##### A. Criteria Formulation

A more stealth intercept–resend strategy that we now investigate in more details is Eve's use of blinding light together with triggering multi-photon pulses [38], [39], [44], [45], [48]. Upon blinding, the SPD in the linear mode never clicks when the triggering pulse energy is below a threshold  $E_{never}$ , and always clicks when the energy is greater than a threshold  $E_{always}$  [44], [46]. When the energy falls between these two levels, the detector clicks with a probability between 0 and 1.

For hacking the BB84 QKD system, it is required that  $E_{always} < 2E_{never}$  so that if the trigger pulse has energy  $E_T \in [E_{always}, 2E_{never})$ , the detector will always click in the compatible basis, but will never click in the conjugate basis. Bob's detectors can then be fully controlled without elevating the QBER [38].

The reason for the potential of this hacking strategy is shown by noting that upon blinding, the alert SPDs in path  $a$  will receive double the blinding power –on average– relative to the SPDs in path  $b$  [see Fig. 1(a)]. Because  $E_{always}(I)$  and  $E_{never}(I)$  are monotonic increasing functions of the blinding power  $I$  [46], higher blinding power for the alert SPDs generally elevates their operation thresholds. As a result, one might think that the alert SPDs would be more insensitive to the triggering pulses, which could be exploited to produce an unnoticeable intrusion.

To investigate this attack further, let us consider that Eve uses triggering pulses of energy  $E_T$  carrying her measured key (time-bin) qubit together with an ancillary (polarization) state  $\rho_T$ . This is accompanied by blinding light of power  $I_B$  and polarization state  $\rho_B$ . Eve would like to optimize the attack parameters –  $E_T$ ,  $\rho_T$ ,  $I_B$ , and  $\rho_B$  – aiming to perform selective triggering of detectors in path  $b$  without registering a click in the alert SPDs in path  $a$ . In the following analysis, we show that such goal can be made impossible if Bob's SPDs are appropriately selected.

Eve's photons of the trigger pulse will be split into paths  $a$  and  $b$  with the probabilities  $p_a$  and  $p_b$ , and then split again equally between the two polarization paths of  $b$ . If the total energy of Eve's time-bin pulses is  $E_T$ , then within the gated time window there will be a portion  $\frac{1}{2}p_a E_T$  in path  $a$  (this is also the maximum energy received by any detector  $D_{ai}$ ,  $i \in \{1, 2\}$ ), and portions  $\frac{1}{4}p_b E_T$  in each arm of path  $b$  (the maximum energy received by any detector  $D_{bj}$ ,  $j \in \{1, 2, 3, 4\}$ ).

To develop a successful detector control, Eve's triggering pulse and blinding light have to satisfy concurrently the following two conditions for all possible realizations of  $U$ :

(A): The maximum trigger pulse energy that may strike an alert detector  $D_{ai}$  is less than the minimum  $E_{never}^{ai}$ , i.e.,

$$\frac{1}{2}p_{\max}E_T < E_{never}^{ai}(\min\{I_a\}), \quad (7)$$

where  $\min\{I_a\}$  is the minimum blinding power received by a detector  $D_{ai}$  and  $p_{\max}$  is the maximum value of  $p_a$  obtained over any state  $U\rho_T U^+$  (see Appendix C), which is given by

$$p_{\max} = \frac{1}{2}(1 + \sqrt{2\mathcal{P}_T - 1}), \quad (8)$$

with  $\mathcal{P}_T$  being the purity of the polarization state  $\rho_T$ .

(B): The maximum pulse energy that may strike a detector  $D_{bj}$  must be at least greater than the minimum  $E_{never}^{bj}$ , i.e.,

$$\frac{1}{4}p_{\max}E_T > E_{never}^{bj}(\min\{I_b\}), \quad (9)$$

where  $\min\{I_b\}$  is the minimum blinding power received by a detector  $D_{bj}$ , and  $p_{\max}$  is the maximum of  $p_b$  taken over any state  $U\rho_T U^+$  [same as in (8)].

Condition (A) guarantees that even if the maximum triggering-pulse energy passes to a detector  $D_{ai}$ , which is blinded with the minimum light power, this should not lead to a click. Condition (B) offers a necessary condition for detectors  $D_{bj}$  to trigger.

As shown by (7) and (9), for Eve who does not know about the transformation  $U$ , the maximum pulse energy (over all possible settings of  $U$ ) that may impinge on a detector  $D_{bj}$  is half that for an alert detector  $D_{ai}$ . Consequently, conditions (A) and (B) cannot be satisfied unless the detectors  $D_{ai}$  and  $D_{bj}$  strictly comply with the necessary and sufficient condition:

$$\frac{E_{never}^{bj}(\min\{I_b\})}{E_{never}^{ai}(\min\{I_a\})} < \frac{1}{2}, \forall i, j \quad (10)$$

Because Eve does not know the current  $U$ , she does not have the ability to reliably control the ratio of the maximum pulse energies delivered to the detectors  $D_{ai}$  and  $D_{bj}$ . Thus, Bob's setup restricts this ratio in operation to 1/2 as in (7) and (9) due to the balanced beamsplitting in path  $b$ .

Aiming to avoid the alert SPDs in path  $a$ , Eve will gain no benefit by assigning a specific time-bin state for the blinding light. We therefore assume, without loss of generality, that the blinding light is in a mixed time-bin state. For an input blinding light of power  $I_B$  and a state of polarization  $\rho_B$ , the power received by the SPDs  $D_{ai}$  and  $D_{bj}$  are, respectively,  $I_a = \frac{1}{2}r_a I_B$  and  $I_b = \frac{1}{4}r_b I_B$ , where  $r_a = \langle H|U\rho_B U^+|H\rangle$  and  $r_b = \langle V|U\rho_B U^+|V\rangle$ . The probabilities  $r_a$  and  $r_b$  are bounded over all settings of  $U$  by the same minimum value:  $\frac{1}{2}(1 - \sqrt{2\mathcal{P}_B - 1})$  (see Appendix C), where  $\mathcal{P}_B$  is the purity of the state  $\rho_B$ . It follows that:

$$\frac{\min\{I_b\}}{\min\{I_a\}} = \frac{1}{2}.$$

Note that the variations in  $U$  for each roundtrip alters the value of the blinding power illuminating the SPDs. Here, we assumed that the threshold  $E_{never}$  depends on the instantaneous blinding power. However due to the electronics of the SPD, there may be a cumulative dependence. In this case, the same 1:2 ratio is still expected due to the randomness of  $U$  along with the balanced beamsplitting in path  $b$ .

Therefore, back to (10), Eve's detector control attack can be effectively thwarted if Bob uses detectors  $D_{ai}$  and  $D_{bj}$  for which

$$\frac{E_{never}^{bj}(I/2)}{E_{never}^{ai}(I)} > \frac{1}{2}, \quad (11)$$

for any value of  $I$ . We show next that this requirement for Bob's detectors is realizable in practice.

### B. Experimental Verification of the Criteria

We demonstrate that meeting conditions (A) and (B) concurrently can be made impossible in practice by the right choice of Bob's detectors. In our demonstration, we consider an arrangement of two detectors used in the commercial QKD system Clavis2 from ID Quantique, with the values of the threshold parameters obtained from reported results of an experiment by Huang et al. [46].

Eve's source [Fig. 1(c)] consists of a pulsed laser (vertically polarized, attenuated to  $\sim 0.6$  pJ/pulse) along with the polarization purity control, unitary polarization transformation, and phase encoding (see Appendix A). The source prepares triggering multi-photon pulses with a time-bin state encoded by Eve's measured phase  $\phi_E$  and a polarization state that can be tuned to any pure or mixed state. The produced state writes:

$$\begin{aligned} & (\cos^2 \theta_1 |e\rangle\langle e| + \sin^2 \theta_1 |\bar{e}\rangle\langle \bar{e}|) \\ & \otimes \frac{1}{2} (|t_l\rangle + e^{i\phi_E} |t_s\rangle) (\langle t_l| + e^{-i\phi_E} \langle t_s|). \end{aligned} \quad (12)$$

with the polarization part be an incoherent mixture of the two arbitrary orthogonal states  $|e\rangle$  and  $|\bar{e}\rangle$  along with a pure-state time-bin part.

To assess the alert possibility, the phase  $\phi_E$  was set to zero which corresponds to Bob's detection in the basis state  $|D\rangle$  in either path  $a$  or  $b$ . Therefore, the alert possibility due to Eve's faked-state photons can be analyzed by placing the detectors:  $D_{a1}$  and  $D_{b1}$  in the alert and secure paths.

Since Eve's source is able to scan over all points of the Poincaré sphere, there is no loss of generality in fixing the randomizer  $U$  of Bob's system to a value, unknown to Eve, which we took to be

$$U = \frac{1}{\sqrt{2}} \begin{bmatrix} i & i \\ 1 & -1 \end{bmatrix}. \quad (13)$$

This matrix is equivalent to the product of Jones matrices of QWP and HWP, fixed at angles  $45^\circ$  and  $112.5^\circ$  w.r.t. the vertical axis, respectively. We used Eve's source to prepare triggering multiphoton states with purity levels:  $\mathcal{P}_T = \{1, 0.78, 0.63, 0.53, 0.5\}$ . For each purity setting, HWP2 was rotated from  $0^\circ$  to  $180^\circ$ , with the QWP fixed at  $-45^\circ$ . During the polarization sweep, the received energies of trigger pulses  $E_{a1}$  and  $E_{b1}$  that reach detectors  $D_{a1}$  and  $D_{b1}$ , respectively, were measured within the superposition time-bin window.

The threshold function  $E_{never}(I)$  is a monotonic increasing function of the blinding power  $I$  with a slightly compressive behavior [46]. The requirement in (11) can be satisfied based on this compressive behavior, and by assigning the detectors of higher sensitivity in the linear mode to the alert path [this

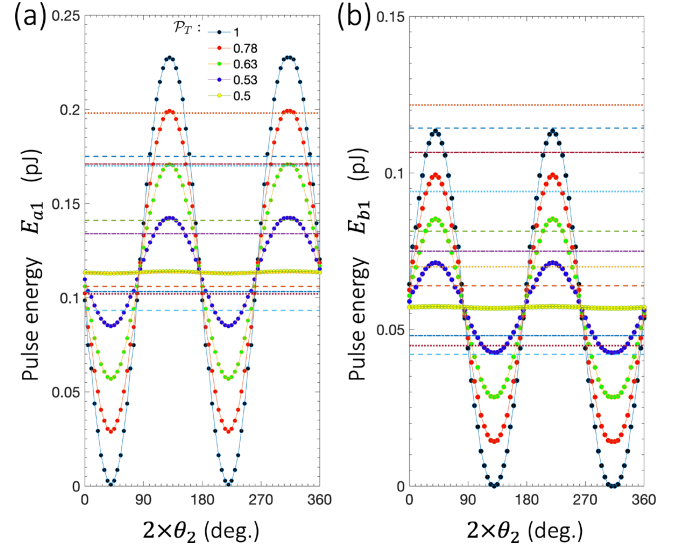


Fig. 2. Measured energies of the triggering pulses: (a)  $E_{a1}$  at alert detector  $D_{a1}$ , and (b)  $E_{b1}$  at secure detector  $D_{b1}$ ; for five purity levels of Eve's polarization state. The dashed and dotted levels are the detectors thresholds: (a)  $E_{never}^{a1}(I_B/4)$ , and (b)  $E_{never}^{b1}(I_B/8)$  at blinding powers  $I_B = \{0.72, 0.78, 0.86, 1.02, 1.09, 1.27, 1.51, 1.78, 2.02, 2.26, 2.5\}$  mW, in bottom-up order. The measurements of pulse energies were selectively performed at the superposition time window. In the measurements, the state is controlled by rotating HWP2 from  $0^\circ$  to  $180^\circ$  with the QWP fixed at  $-45^\circ$ . The measured energies was fit to sinusoidal functions of variable visibility. The measurements error is smaller than the marker size.

higher sensitivity is exhibited by the relatively lower profile of  $E_{never}(I)$ ). Therefore, based on the measurements of their thresholds (see Appendix B), we choose to assign the SPDs  $D0$  and  $D1$  of Clavis2 system, respectively, to the secure detector  $D_{b1}$  and the alert detector  $D_{a1}$ .

Because conditions (A) and (B) rely on the minimum blinding power over all possibilities of  $U$  regardless of the polarization state  $\rho_B$ , we considered an unpolarized blinding light, without loss of generality. The levels in Fig. 2(a) and Fig. 2(b) are the thresholds:  $E_{never}^{a1}(I_B/4)$  and  $E_{never}^{b1}(I_B/8)$ , respectively, taken at blinding powers:  $I_B = \{0.72, 0.78, 0.86, 1.02, 1.09, 1.27, 1.51, 1.78, 2.02, 2.26, 2.5\}$  mW, with the detector gate applied. Eve's objective is then to find out the blinding power for which the threshold in Fig. 2(a) is greater than the maximum pulse energy received by  $D_{a1}$ , and concurrently, the corresponding threshold in Fig. 2(b) is less than the maximum pulse energy received by  $D_{b1}$  (for the same purity level).

Fig. 2 shows the results. It is evident from Fig. 2 that Eve cannot meet her objective for any of these levels. Although this is not a complete polarization sweep test (i.e., not covering the entire volume of the Poincaré sphere), it is sufficient to evaluate the ability of a traceless attack. This is because it spans the entire visibility range for arbitrary (pure or mixed) polarization state. Taking into account that  $E_{never}(I)$  is a monotonic increasing function of  $I$ , it can also be verified that this cannot be possible for any other level of blinding power.

Fig. 2 shows the results for Eve's attack using pulses of fixed energy that reach Bob's system while the detectors gate

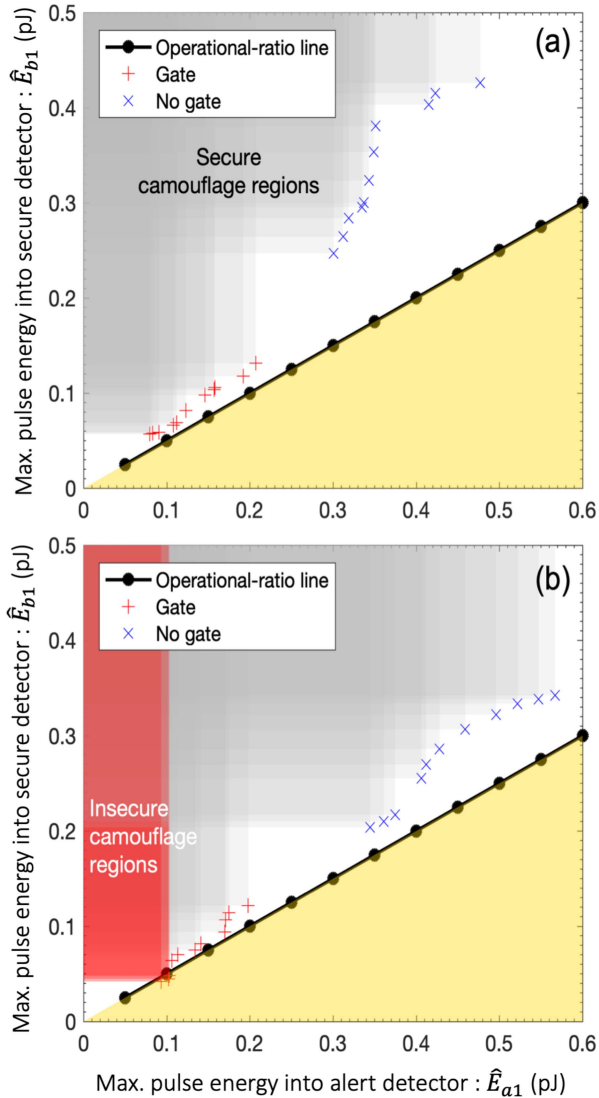


Fig. 3. Measurements of the maximum pulse energies reaching a secure-path detector:  $\hat{E}_{b1}$  and an alert-path detector:  $\hat{E}_{a1}$  showing the 1:2 ratio operational line below which the protection fails (yellow area). Intersection points of the thresholds  $E_{never}^{a1}(I_B/4)$  (vertical) and  $E_{never}^{b1}(I_B/8)$  (horizontal) are shown in the presence (+) and the absence (x) of the detector gate. Each threshold point has a camouflage region (red- and grey-colored areas) within which the maximum pulse energy delivered to  $D_{a1}$  is less than  $E_{never}^{a1}(I_B/4)$ , while the maximum pulse energy for  $D_{b1}$  is higher than  $E_{never}^{b1}(I_B/8)$ . (a) SPD  $D1$ , which is the more sensitive of the two Clavis2 detectors, is assigned to  $D_{a1}$ , while the less-sensitive  $D0$  is assigned to  $D_{b1}$  (see Appendix B). In this case all threshold points lie above the operational-ratio line, and the camouflage region does not overlap the unsafe area (marked in yellow). This renders Eve's unnoticeable attack impossible. (b) SPDs  $D1$  and  $D0$  are assigned unwisely to  $D_{b1}$  and  $D_{a1}$ , respectively. In this case, a small overlap exists between the camouflage area and the unsafe area indicating a possibility for Eve to adjust the parameters:  $E_T$ ,  $\rho_T$ ,  $I_B$ , and  $\rho_B$  and launch a successful attack.

is applied. Eve may also change the energy level of triggering pulses or launch her attack when the gate of the detector is not applied. Fig. 3 shows the results in the presence and absence of the detector gate for a span of trigger pulse energies. It depicts the operational-ratio line which specifies the strict 1:2 relation between the maximum pulse energies reachable to path- $b$  and path- $a$  detectors, as constrained by Bob's system.

Fig. 3 shows also intersection points between the threshold lines  $E_{never}^{a1}(I_B/4)$  and  $E_{never}^{b1}(I_B/8)$  (parallel to  $y$  and  $x$  axes, respectively), combining the thresholds of the two detectors at different values of total blinding power  $I_B$ . Every intersection point is associated with a camouflage region, where Eve's detector-control can be enacted tracelessly. As shown in Fig. 3(a) for a good arrangement of alert and secure detectors, all threshold points are above the operational-ratio line. This prohibits any overlap between the operational-ratio line of camouflage regions and therefore disallows unnoticeable intrusion. In this arrangement, the necessary and sufficient condition for successful intrusion in (10) is not satisfied for any threshold point. It is then impossible to avoid triggering the alert detectors, no matter what faked-state of light Eve uses.

To show how the unwise choice of Bob's alert and secure detectors may allow for unnoticeable intrusion, we considered interchanging  $D0$  and  $D1$  of Clavis2 system to be the alert detector  $D_{a1}$  and the secure detector  $D_{b1}$ , respectively. In this case, some threshold points lied under the operational-ratio line [Fig. 3(b)]. This creates a valid camouflage region (in overlap with the operational-ratio line) for Eve who can then, in principle, selectively trigger path- $b$  detectors, but not path- $a$  detectors (see Appendix B).

## V. ATTACK MODEL AND SECURITY ANALYSIS

We assume that Eve can introduce photons into Bob's receiver only through the polarization randomizer. She is acquainted with the configuration of the system, including timing and other classical information, but has no information on the specific random transformation  $U$  applied at any time. We consider a large number of quantum signals between Alice and Bob, so that all finite-size corrections required in security analysis are negligible (see, e.g., Ref. [9]).

Eve interacts identically and independently with each quantum signal. She measures the pulse encoded by Alice in one of the two bases. The outcome of Eve's measurement is described by three probabilities: 1)  $P_e^c \approx \frac{1}{2}e^{-\mu(1-F_e)\eta_e}(1 - e^{-\mu F_e\eta_e})$  is the probability that Eve's measurement is in a compatible basis and gives results in a single click in the correct detector. 2)  $P_e^w \approx \frac{1}{2}e^{-\mu F_e\eta_e}(1 - e^{-\mu(1-F_e)\eta_e})$  is the corresponding probability of a click in the wrong detector only. 3)  $P_e^{nc} \approx \frac{1}{2}e^{-\frac{\mu\eta_e}{2}}(1 - e^{-\frac{\mu\eta_e}{2}})$  is the probability that Eve's measurement is in incompatible basis and gives a click in a single detector. In these expressions,  $\mu$  is the mean number of photons per pulse,  $F_e$  is the fidelity of Eve's measurement, and  $\eta_e$  is the overall detection efficiency. We specify in the following some possible Eve's attacks.

### A. Quantum Attack

In this attack, Eve always forwards single-photon pulses to Bob. Bob performs a squashing operation whenever multiple clicks occur [70], [71], [72]; that is double clicks in different bases do not count, while if in the same basis, they give a random value [32].

In order to determine the sifted key rate and the QBER under quantum attack, we begin by writing expressions for the raw probabilities  $p_{bj}(k)$  that Bob's detector  $D_{bj}$ ,  $j = 1, 2, 3, 4$ , clicks if Eve uses the phase  $k \in \{0, \pi, \pi/2, 3\pi/2\}$  to encode her pulse. For  $k = 0$ ,

$$\begin{aligned} p_{b1}(0) &\approx c_{b1} + 1 - \exp\left(-\frac{\mu_e p_b F \eta_{b1}}{4}\right), \\ p_{b2}(0) &\approx c_{b2} + 1 - \exp\left(-\frac{\mu_e p_b (1-F) \eta_{b2}}{4}\right), \\ p_{b3(b4)}(0) &\approx c_{b3(b4)} + 1 - \exp\left(-\frac{\mu_e p_b \eta_{b3(b4)}}{8}\right), \end{aligned} \quad (14)$$

where  $c_{bj}$  is the total background rate of detector  $D_{bj}$  within the gate slot,  $F$  is the fidelity of Bob's measurement,  $\eta_{bj}$  is the overall detection efficiency of  $D_{bj}$ , and  $p_b$  is the probability that Eve's photon passes into the secure path as given in (6). Similar expressions apply for other phases  $k \in \{\pi, \pi/2, 3\pi/2\}$ .

After the squashing operation, the probability that Bob registers a click in the  $D/A$  basis, given that Eve sent a phase-encoded state for  $k = 0$  is

$$\begin{aligned} P_{DA}(0) &= [p_{b1}(0) + p_{b2}(0) - p_{b1}(0)p_{b2}(0)] \\ &\quad \times [1 - p_{b3}(0)][1 - p_{b4}(0)]. \end{aligned} \quad (15)$$

Also, after squashing, the probability that Bob registers a click on  $D_{bj}$  given that Eve sent a state coded by the phase  $k$  is  $P_{bj}(k)$ , where, for example,

$$P_{b1}(\pi) = p_{b1}(\pi)[1 - \frac{1}{2}p_{b2}(\pi)][1 - p_{b3}(\pi)][1 - p_{b4}(\pi)], \quad (16)$$

and where  $P_{DA}(k) = P_{b1}(k) + P_{b2}(k)$ .

Therefore, given that Alice's phase  $k = 0$ , the sifted key rate (in path b) is

$$\begin{aligned} R_b(0) &\approx P_e^c P_{DA}(0) + P_e^w P_{DA}(\pi) \\ &\quad + P_e^{nc} [P_{DA}(\pi/2) + P_{DA}(3\pi/2)] \\ &\quad + (1 - P_e^c - P_e^w - 2P_e^{nc})(c_{b1} + c_{b2} - c_{b1}c_{b2}). \end{aligned} \quad (17)$$

The corresponding error in Bob's measurement (in path b) is

$$\begin{aligned} E_b(0) &\approx P_e^c P_{b2}(0) + P_e^w P_{b2}(\pi) \\ &\quad + P_e^{nc} [P_{b2}(\pi/2) + P_{b2}(3\pi/2)] \\ &\quad + (1 - P_e^c - P_e^w - 2P_e^{nc})[c_{b2} - (c_{b1}c_{b2})/2]. \end{aligned} \quad (18)$$

The sifted key rates and the errors – conditioned on Alice's state with a phase  $k \in \{\pi, \pi/2, 3\pi/2\}$  – can be similarly obtained. Consequently, the total sifted key rate and the QBER under Eve's quantum attack are

$$\begin{aligned} R_b^Q &= \frac{1}{4} \sum_{k=\{0, \pi, \frac{\pi}{2}, \frac{3\pi}{2}\}} R_b(k), \\ \text{QBER}_b^Q &= \frac{1}{4R_b^Q} \sum_{k=\{0, \pi, \frac{\pi}{2}, \frac{3\pi}{2}\}} E_b(k). \end{aligned} \quad (19)$$

Bob does not apply the squashing operation on the alert detections, so that the overall alert rate is

$$\begin{aligned} R_a^Q &\approx c_{a1} + c_{a2} + 2 \\ &\quad - \frac{1}{2} \left[ \exp\left(-\frac{\mu_e p_a F \eta_{a1}}{2}\right) + \exp\left(-\frac{\mu_e p_a F \eta_{a2}}{2}\right) \right. \\ &\quad \left. + \exp\left(-\frac{\mu_e p_a \eta_{a1}}{4}\right) + \exp\left(-\frac{\mu_e p_a \eta_{a2}}{4}\right) \right]. \end{aligned} \quad (20)$$

While no obvious change appears in the sifted key rate and the QBER compared to the BB84 protocol, the presence of an alert rate, which is significantly higher than the background rate, provides an additional clear sign of Eve's attack.

### B. Blinding Attack

In this attack, Eve blinds Bob's detectors using nonpolarized light, then sends a bright pulse encoded by her measurement outcome. The bright trigger pulse has a pure polarization state. The assumption of nonpolarized blinding light is logical since it is optimal for Eve to render the blinding of all SPDs unaffected by the randomization  $U$ . We assume that Eve has complete control over Bob's measurement in the secure path so that she can limit the QBER; however, as will be shown later, this is not sufficient to limit the rate of the alert. The following analysis is presented in three cases: i) no randomization, ii) randomization, iii) randomization and switching.

i) *No randomization.* For simplicity, let us first consider the case:  $U = I$ . Aiming to trigger Bob's secure SPDs in the matched basis and avoid clicks in the unmatched one, Eve sends a trigger pulse energy  $E_T$  such that

$$2E_{never}^{bj} > \frac{1}{4}E_T > E_{never}^{bj}, \forall j. \quad (21)$$

Note that half the pulse energy  $E_T$  will pass while the gate is off (which we assume to result in no action). The lower bound in (21) assigns a threshold to enable the triggering of matched-basis detectors. The upper bound puts a limit for not triggering the ones in unmatched basis. Because an alert detector receives double the blinding power of a secure detector, and due to the compressive nature of  $E_{never}(I)$  and the higher sensitivity of alert detectors, we can infer that  $E_{never}^{ai} < 2E_{never}^{bj}, \forall i, j$ , which can be substituted into the lower bound of (21) to give

$$\frac{1}{2}E_T > E_{never}^{ai}, \forall i. \quad (22)$$

The lower bound in (21) and (22) signifies that the minimal trigger energy that enables the detectors control in the secure path will also enable triggering of the alert detectors.

ii) *Randomization.* Let us now move to the general case with a random transformation  $U$ , but without switching the paths  $a, b$ . In this case, the values of  $p_a$  are uniformly distributed between 0 and 1. Fig. 4(a) sketches the energies delivered to SPDs  $D_{ai}$  and  $D_{bj}$  in matched basis, which equal  $\frac{1}{2}p_a E_T$  and  $\frac{1}{4}(1 - p_a)E_T$ , respectively.

For simplicity, we approximate the click probability of blinded detectors by a ramp-step function as plotted in Fig. 4(b). Therefore, the alert rate on a detector  $D_{ai}$  can be obtained by

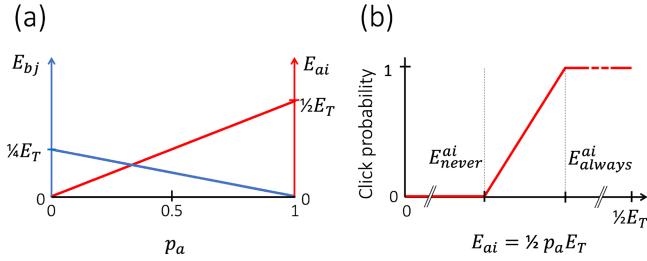


Fig. 4. (a) Portions of the trigger pulse energy  $E_T$  that strike two detectors  $D_{ai}$  and  $D_{bj}$  in matched basis plotted versus the probability  $p_a$ . (b) Ramp-step approximation of the APD click probability versus the trigger pulse energy under blinding attack. Shown is the case of alert detector  $D_{ai}$  when the bases are matched.

averaging the click probability in Fig. 4(b) over  $p_a$  as

$$R_{ai}^{Bl} \approx \frac{1}{2} \max \left\{ 1 - \frac{E_{never}^{ai} + E_{always}^{ai}}{E_T}, 0 \right\}, \quad (23)$$

where the factor  $\frac{1}{2}$  is attributed to the probability that Eve's and Bob's alert-path bases match. The trigger rate of a secure detector  $D_{bj}$  is obtained similarly by averaging its click probability over  $p_a$  as

$$R_{bj}^{Bl} \approx \max \left\{ 1 - \frac{2(E_{never}^{bj} + E_{always}^{bj})}{E_T}, 0 \right\}, \quad (24)$$

Therefore, in the absence of switching paths  $a, b$ , the total alert and secure detection rates are

$$R_a^{Bl} = \frac{1}{2} \sum_{i=1}^2 R_{ai}^{Bl}, R_{secure}^{Bl} = \frac{1}{4} \sum_{j=1}^4 R_{bj}^{Bl}. \quad (25)$$

iii) *Randomization and switching.* When Bob switches the alert and secure paths, secure-path detections are counted as alert events and vice versa. If  $R_{sw}$  is the switching rate, then the total alert and secure detection rates become

$$R_a^{Bl} = \frac{1}{2}(1 - R_{sw}) \sum_{i=1}^2 R_{ai}^{Bl} + \frac{1}{4}R_{sw} \sum_{j=1}^4 R_{bj}^{Bl},$$

$$R_{secure}^{Bl} = \frac{1}{2}R_{sw} \sum_{i=1}^2 R_{ai}^{Bl} + \frac{1}{4}(1 - R_{sw}) \sum_{j=1}^4 R_{bj}^{Bl}. \quad (26)$$

This yields the sifted key rate and QBER:

$$R_b^{Bl} = (P_e^c + P_e^w)R_{secure}^{Bl}, \text{QBER}_b^{Bl} = \frac{P_e^w}{P_e^c + P_e^w}. \quad (27)$$

Remarkably, while this complete blinding attack can keep the level of QBER unaffected by Eve's interception [as shown by (27)], it is not capable of diminishing the alert rate. Recalling that  $E_{never}^{ai} < 2E_{never}^{bj}, \forall i, j$ , Eqs. (23), (24), (25) show that the alert and secure rates are related by  $R_a^{Bl} \geq \frac{1}{2}R_{secure}^{Bl}$ . The alert rate increases proportionally with  $R_{sw}$ . For example, if  $R_{sw} = \frac{1}{2}$ , it leads to  $R_a^{Bl} = R_{secure}^{Bl}$  as given in (26).

### C. Wavelength-Dependent Blinding Attack

While narrow-band filters can be used to limit a wavelength-dependent attack, it is still possible that Eve elevates the power

values of her out-band signals to allow passage of a finite power level [69]. Such attack may target the (polarizing and non-polarizing) beam splitters and the polarization randomizer. In the former, Eve may exploit the wavelength-dependent deviation from the coupling ratio of the 3-dB coupler and the extinction ratio of the PBS. In the latter, she may exploit the dispersive nature of polarization transformers/controllers, which typically use cascaded birefringent components.

We conservatively assume that Eve can develop a wavelength-dependent blinding attack that enables the right control of secure SPDs and always avoids ticking alert SPDs (or at least keeping it below the background rate  $c_{ai}$  within the gate slot). Under these conditions, the alert and sifted key rates and QBER are

$$R_a^{W|Bl} = R_{sw},$$

$$R_b^{W|Bl} = (P_e^c + P_e^w)(1 - R_{sw}),$$

$$\text{QBER}_b^{W|Bl} = \frac{P_e^w}{P_e^c + P_e^w}. \quad (28)$$

If Bob keeps  $R_{sw} \gg c_{ai}, \forall i$ , Eve's presence will be unveiled by the alert rate.

The rates in (28) are also valid for other attack approaches that enable biasing the triggers to the secure detectors. Examples are the attacks exploiting the detector's efficiency mismatch (e.g., time-shift attacks [26], [27]) or dead time (e.g., the dead-time attack [22]). Other examples are the pulsed blinding attacks (see, e.g., Refs. [22], [44], [55]), where the linear-mode operation of the double-blinded alert detectors last for a longer period [44]; enabling to bias the triggers to secure detectors.

### D. Integrated Attacks

Eve might select one of her menu of attacks at random. If she launches a quantum attack with probability  $p_Q$ , a blinding attack with probability  $p_{Bl}$ , or a wavelength-dependent blinding attack with probability  $p_{W|Bl}$ , then the overall sifted key rate, the QBER, and the alert rate are:

$$R_b^e = p_Q R_b^Q + p_{Bl} R_b^{Bl} + p_{W|Bl} R_b^{W|Bl},$$

$$\text{QBER}_e = p_Q \text{QBER}_b^Q + (p_{Bl} + p_{W|Bl}) \frac{P_e^w}{P_e^c + P_e^w},$$

$$R_a^e = p_Q R_a^Q + p_{Bl} R_a^{Bl} + p_{W|Bl} R_a^{W|Bl}. \quad (29)$$

## VI. DISCUSSION AND CONCLUSION

We have introduced a QKD scheme that nullifies the class of practical hacking strategies exploiting faked-state light, including the detector-control attacks and more generally the intercept-resend strategies. The scheme uses a roundtrip arrangement exploiting the three optical degrees of freedom: polarization, time-bin, and path. Thanks to continuous randomization of the polarization state at the gateway to Bob's transceiver, only the genuine photon – originally created by Bob – can reliably avoid triggering the alert detectors. We have analytically proven and experimentally verified that this feature can be made unrealizable by Eve's faked-state light.

It is essential to emphasize that the randomization of the ancillary (polarization) qubit is not by itself sufficient to securely

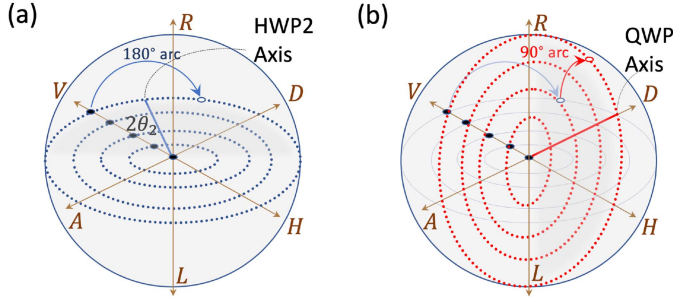


Fig. 5. Evolution of the state of polarization (SOP) of Eve's faked state on the Poincaré sphere. (a) Mixed state is generated with different purities by rotating HWP1 (black dots). Rotation of HWP2 from  $0^\circ$  to  $180^\circ$  moves the SOP to span the blue dotted circles in the plane orthogonal to the  $|R\rangle$ - $|L\rangle$  axis (b) QWP with axis at  $-45^\circ$  performs  $90^\circ$  rotation in the plane orthogonal to the  $|D\rangle$ - $|A\rangle$  axis (red dotted circles). Arbitrary transformation can be done by rotating HWP2 and QWP. Eve's transformation arrangement along with the measurement in Bob's system work in a way similar to a de Sénarmont compensator.

exchange a key without relying on the BB84 protocol to encode the key (time-bin) qubit. Without BB84, Eve could, in principle, extract the information in the time-bin qubit in a reliable manner without disturbing the single-photon state forwarded to Bob.

#### APPENDIX A EVE'S STATE PREPARATION

In order to generate light pulses with a prepared state of polarization, mimicking that potentially employed by Eve, we have used the three-stage optical system in Fig. 1(c). The first stage produces mixed-state pulses with a polarization purity set by a half-wave plate HWPI followed by a heavily unbalanced polarization-based Mach-Zehnder interferometer (PMZI). The propagation times through the two PMZI arms differ by the period of the pulsed laser, which is longer than its coherence time. The PMZI thus mixes pairs of mutually incoherent pulses of orthogonal polarization with a ratio set by the rotation angle  $\theta_1$  of HWPI.

The polarization purity is then given by

$$\mathcal{P}_T = 1 - \frac{1}{2} \sin^2 4\theta_1, \quad (30)$$

with the values  $\mathcal{P}_T = \{1, 0.78, 0.63, 0.53, 0.5\}$  used in the experiment corresponding to HWP1 angles:  $\theta_1 = \{0^\circ, 10.4^\circ, 15^\circ, 18.9^\circ, 22.5^\circ\}$ .

The second stage of the system uses a half-wave plate HWP2 (its rotation angle is  $\theta_2$ ) and a quarter-wave plate QWP to perform the unitary rotation over the Poincaré sphere (see Fig. 5) and create the polarization state in (12). After preparing the polarization state, the third stage creates phase-encoded time bin state using a Mach-Zehnder interferometer (MZI) identical to the one used by Bob.

#### APPENDIX B TRIGGERING THRESHOLDS OF "CLAVIS2" SPDs

Figs. 2 and 3 depict the threshold values:  $E_{never,0}^{gate}$ ,  $E_{never,1}^{gate}$ ,  $E_{never,0}^{no-gate}$ , and  $E_{never,1}^{no-gate}$  for the two Clavis2 detectors  $D0$  and

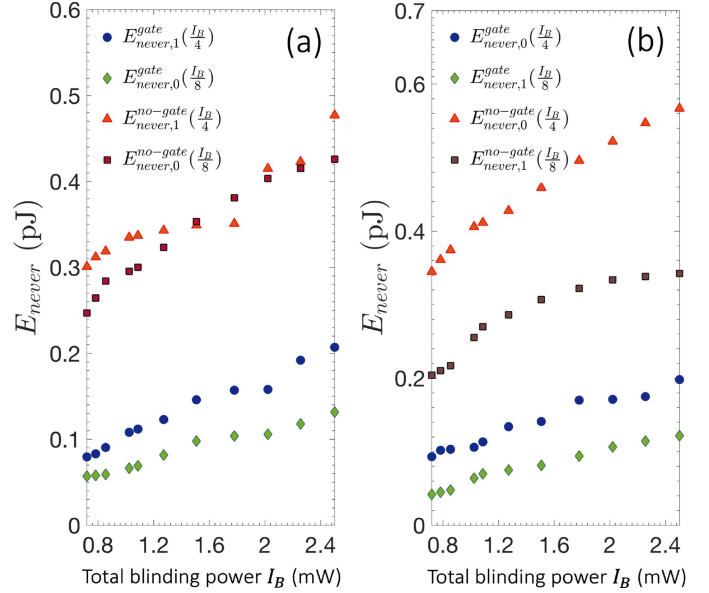


Fig. 6. Thresholds  $E_{never}$  of the two SPDs  $D0$  and  $D1$  used in the Clavis2 system in the presence and the absence of the gate, plotted versus the blinding power  $I_B$  of unpolarized light injected into Bob's system. (a)  $D0$  and  $D1$  are assigned to paths  $b$  and  $a$ , respectively. (b)  $D0$  and  $D1$  are switched to paths  $a$  and  $b$ .

$D1$  at different values of blinding power. This data was reproduced from the experimental results in [46] and supplemented by interpolations to deduce some missed points. Fig. 6(a) shows these thresholds versus the power  $I_B$  (total Eve's blinding power input to Bob's system), when the two SPDs  $D1$  and  $D0$  are inserted into the alert path  $a$  and the secured path  $b$ , respectively. Fig. 6(b) shows the other unwise alternative when  $D0$  and  $D1$  are in path  $a$  and path  $b$ , respectively. It is obvious that the condition in (10), which is necessary and sufficient for a traceless attack, is not satisfied at all points in the first case of Fig. 6(a). This verifies the security of Bob's system against Eve's detector-side attack. By contrast for the alternative arrangement, the condition (10) is satisfied at some points (particularly, the first three points of gated detection) of Fig. 6(b). This enables the overlap between the camouflage regions of these points and the operational-ratio line and allows for a traceless detector-side attack by Eve, as depicted in Fig. 3(b).

#### APPENDIX C BOUNDS FOR THE OVERLAP BETWEEN PURE AND MIXED STATES

We show here that the maximum and minimum overlaps between a pure state  $|\psi\rangle$  and a mixed state  $\rho$  after the application of an arbitrary unitary operator  $U$ , are given by

$$\begin{aligned} \max\{\langle\psi|U\rho U^+|\psi\rangle\} &= \frac{1}{2} \left(1 + \sqrt{2\mathcal{P} - 1}\right), \\ \min\{\langle\psi|U\rho U^+|\psi\rangle\} &= \frac{1}{2} \left(1 - \sqrt{2\mathcal{P} - 1}\right), \end{aligned} \quad (31)$$

where  $\mathcal{P}$  is the purity of the state  $\rho$ . Let us express the mixed state  $\rho$  as a mixture  $\lambda|v\rangle\langle v| + (1 - \lambda)|\bar{v}\rangle\langle\bar{v}|$  of two orthogonal

states  $|v\rangle$  and  $|\bar{v}\rangle$ , then, after applying  $U$ , the overlap with the state  $|\psi\rangle$  writes

$$\begin{aligned} \langle\psi|U\rho U^+|\psi\rangle &= \langle\psi|U[\lambda|v\rangle\langle v| + (1-\lambda)|\bar{v}\rangle\langle\bar{v}|]U^+|\psi\rangle \\ &= \lambda|\langle\psi|U|v\rangle|^2 + (1-\lambda)|\langle\psi|U|\bar{v}\rangle|^2. \end{aligned} \quad (32)$$

Since  $|\langle\psi|U|v\rangle|^2 = 1 - |\langle\psi|U|\bar{v}\rangle|^2$ , (32) describes the overlap as an interpolation between the two complementary probabilities  $\lambda$  and  $(1-\lambda)$ , where the interpolation weights  $|\langle\psi|U|v\rangle|^2$  and  $|\langle\psi|U|\bar{v}\rangle|^2$  vary according to  $U$ . Hence the probabilities  $\lambda$  and  $(1-\lambda)$  give the maximum and minimum overlaps between the states  $U\rho U^+$  and  $|\psi\rangle$ . The purity of  $\rho$  is  $\mathcal{P} = \text{tr}(\rho^2) = \lambda^2 + (1-\lambda)^2$ , which offers two values for  $\lambda$  based on  $\mathcal{P}$ , and leads directly to the two bounds in (31).

## REFERENCES

- [1] C. H. Bennett and G. Brassard, "Quantum cryptography: Public key distribution and coin tossing," in *Proc. IEEE Int. Conf. Comp. Syst. Signal Process*, 1984, 175–179.
- [2] A. K. Ekert, "Quantum cryptography based on bell's theorem," *Phys. Rev. Lett.*, vol. 67, pp. 661–663, 1991.
- [3] C. Bennett et al., "Reduction of quantum entropy by reversible extraction of classical information," *J. Mod. Opt.*, vol. 12, pp. 2307–2314, 1994.
- [4] P. W. Shor and J. Preskill, "Simple proof of security of the BB84 quantum key distribution protocol," *Phys. Rev. Lett.*, vol. 85, pp. 441–444, 2000.
- [5] D. Mayers, "Unconditional security in quantum cryptography," *J. ACM*, vol. 48, pp. 351–406, 2001.
- [6] R. Renner, "Security of quantum key distribution," *Int. J. Quantum Inf.*, vol. 6, pp. 1–127, 2008.
- [7] D. Gottesman, H.-K. Lo, N. Lütkenhaus, and J. Preskill, "Security of quantum key distribution with imperfect devices," *Quantum Info. Comput.*, vol. 4, pp. 325–360, 2004.
- [8] H. Inamori, N. Lütkenhaus, and D. Mayers, "Unconditional security of practical quantum key distribution," *Eur. Phys. J. D*, vol. 41, pp. 599–627, 2007.
- [9] M. Tomamichel, C. C. W. Lim, N. Gisin, and R. Renner, "Tight finite-key analysis for quantum cryptography," *Nat. Commun.*, vol. 3, 2012, Art. no. 634.
- [10] G. Brassard, N. Lütkenhaus, T. Mor, and B. C. Sanders, "Limitations on practical quantum cryptography," *Phys. Rev. Lett.*, vol. 85, pp. 1330–1333, 2000.
- [11] N. Lütkenhaus, "Security against individual attacks for realistic quantum key distribution," *Phys. Rev. A*, vol. 61, 2000, Art. no. 052304.
- [12] C.-H. F. Fung, B. Qi, K. Tamaki, and H.-K. Lo, "Phase-remapping attack in practical quantum-key-distribution systems," *Phys. Rev. A*, vol. 75, 2007, Art. no. 032314.
- [13] F. Xu, B. Qi, and H.-K. Lo, "Experimental demonstration of phase-remapping attack in a practical quantum key distribution system," *New J. Phys.*, vol. 12, 2010, Art. no. 113026.
- [14] M.-S. Jiang, S.-H. Sun, C.-Y. Li, and L.-M. Liang, "Wavelength-selected photon-number-splitting attack against plug-and-play quantum key distribution systems with decoy states," *Phys. Rev. A*, vol. 86, 2012, Art. no. 032310.
- [15] K.-I. Yoshino et al., "Quantum key distribution with an efficient countermeasure against correlated intensity fluctuations in optical pulses," *NPJ Quantum Inf.*, vol. 4, no. 1, pp. 1–8, 2018.
- [16] Y.-L. Tang et al., "Source attack of decoy-state quantum key distribution using phase information," *Phys. Rev. A*, vol. 88, 2013, Art. no. 022308.
- [17] 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*, vol. 92, 2015, Art. no. 022304.
- [18] X. L. Pang et al., "Hacking quantum key distribution via injection locking," *Phys. Rev. App.*, vol. 13, 2020, Art. no. 034008.
- [19] A. Huang, A. Navarrete, S. H. Sun, P. Chaiwongkhot, M. Curty, and V. Makarov, "Laser-seeding attack in quantum key distribution," *Phys. Rev. App.*, vol. 12, 2019, Art. no. 064043.
- [20] F. Xu, X. Ma, Q. Zhang, H. K. Lo, and J. W. Pan, "Secure quantum key distribution with realistic devices," *Rev. Mod. Phys.*, vol. 92, 2020, Art. no. 025002.
- [21] R. Newman, "Visible light from a silicon P-N junction," *Phys. Rev.*, vol. 100, 1955, Art. no. 700.
- [22] H. Weier, H. Krauss, M. Rau, M. Fürst, S. Nauwerth, and H. Weinfurter, "Quantum eavesdropping without interception: An attack exploiting the dead time of single-photon detectors," *New J. Phys.*, vol. 13, 2011, Art. no. 073024.
- [23] 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*, vol. 78, 2008, Art. no. 042333.
- [24] C. Kurtsiefer, P. Zarda, S. Mayer, and H. Weinfurter, "The breakdown flash of silicon avalanche photodiodes-back door for eavesdropper attacks," *J. Mod. Opt.*, vol. 48, pp. 2039–2047, 2001.
- [25] V. Makarov, A. Anisimov, and J. Skaar, "Effects of detector efficiency mismatch on security of quantum cryptosystems," *Phys. Rev. A*, vol. 74, 2006, Art. no. 022313.
- [26] V. Makarov and D. R. Hjelm, "Faked states attack on quantum cryptosystems," *J. Mod. Opt.*, vol. 52, pp. 691–705, 2005.
- [27] B. Qi, C.-H. F. Fung, H. K. Lo, and X. Ma, "Time-shift attack in practical quantum cryptosystems," *Quantum Info. Comput.*, vol. 7, pp. 73–82, 2007.
- [28] A. Lamas-Linares and C. Kurtsiefer, "Breaking a quantum key distribution system through a timing side channel," *Opt. Exp.*, vol. 15, pp. 9388–9393, 2007.
- [29] N. Jain et al., "Device calibration impacts security of quantum key distribution," *Phys. Rev. Lett.*, vol. 107, 2011, Art. no. 110501.
- [30] A. N. Bugge et al., "Laser damage helps the eavesdropper in quantum cryptography," *Phys. Rev. Lett.*, vol. 112, 2014, Art. no. 070503.
- [31] V. Makarov et al., "Creation of backdoors in quantum communications via laser damage," *Phys. Rev. A*, vol. 94, 2016, Art. no. 030302.
- [32] 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*, vol. 91, 2015, Art. no. 062301.
- [33] P. Chaiwongkhot et al., "Eavesdropper's ability to attack a free-space quantum-key-distribution receiver in atmospheric turbulence," *Phys. Rev. A*, vol. 99, 2019, Art. no. 062315.
- [34] H. Qin, R. Kumar, and R. Alléaume, "Quantum hacking: Saturation attack on practical continuous-variable quantum key distribution," *Phys. Rev. A*, vol. 94, 2016, Art. no. 012325.
- [35] K. Wei, W. Zhang, Y. L. Tang, L. You, and F. Xu, "Implementation security of quantum key distribution due to polarization-dependent efficiency mismatch," *Phys. Rev. A*, vol. 100, 2019, Art. no. 022325.
- [36] R. H. Hadfield, "Single-photon detectors for optical quantum information applications," *Nat. Photon.*, vol. 3, pp. 696–705, 2009.
- [37] V. Makarov, "Controlling passively quenched single photon detectors by bright light," *New J. Phys.*, vol. 11, 2009, Art. no. 065003.
- [38] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, "Hacking commercial quantum cryptography systems by tailored bright illumination," *Nat. Photon.*, vol. 4, pp. 686–689, 2010.
- [39] 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.*, vol. 2, 2011, Art. no. 349.
- [40] I. Gerhardt et al., "Experimentally faking the violation of bell's inequalities," *Phys. Rev. Lett.*, vol. 107, 2011, Art. no. 170404.
- [41] 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.*, vol. 13, 2011, Art. no. 113042.
- [42] C. Wiechers et al., "After-gate attack on a quantum cryptosystem," *New J. Phys.*, vol. 13, 2011, Art. no. 013043.
- [43] Y. J. Qian et al., "Hacking the quantum key distribution system by exploiting the avalanche-transition region of single-photon detectors," *Phys. Rev. Appl.*, vol. 10, 2018, Art. no. 064062.
- [44] Z. Wu et al., "Hacking single-photon avalanche detectors in quantum key distribution via pulse illumination," *Opt. Exp.*, vol. 28, pp. 25574–25590, 2020.
- [45] S. Sauge, L. Lydersen, A. Anisimov, J. Skaar, and V. Makarov, "Controlling an actively-quenched single photon detector with bright light," *Opt. Exp.*, vol. 19, pp. 23590–23600, 2011.
- [46] 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 Elect.*, vol. 52, no. 11, Nov. 2016, Art. no. 8000211.
- [47] L. Lydersen et al., "Superlinear threshold detectors in quantum cryptography," *Phys. Rev. A*, vol. 84, 2011, Art. no. 032320.

- [48] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, "Thermal blinding of gated detectors in quantum cryptography," *Opt. Exp.*, vol. 18, pp. 27938–27954, 2010.
- [49] Z. L. Yuan, J. F. Dynes, and A. J. Shields, "Avoiding the blinding attack in QKD," *Nat. Photon.*, vol. 4, pp. 800–801, 2010.
- [50] Z. L. Yuan, J. F. Dynes, and A. J. Shields, "Resilience of gated avalanche photodiodes against bright illumination attacks in quantum cryptography," *Appl. Phys. Lett.*, vol. 98, 2011, Art. no. 231104.
- [51] T. F. da Silva, G. B. Xavier, G. P. Temporão, and J. P. von der Weid, "Real-time monitoring of single-photon detectors against eavesdropping in quantum key distribution systems," *Opt. Exp.*, vol. 20, pp. 18911–18924, 2012.
- [52] M. Legre and G. Ribordy, "Apparatus and method for the detection of attacks taking control of the single photon detectors of a quantum cryptography apparatus by randomly changing their efficiency," U.S. Patent 10,020,937, Jul. 10, 2018.
- [53] C. C. W. Lim, N. Walenta, M. Legré, N. Gisin, and H. Zbinden, "Random variation of detector efficiency: A countermeasure against detector blinding attacks for quantum key distribution," *IEEE J. Sel. Topics Quantum Electron.*, vol. 21, no. 3, Jan. 2015, Art. no. 6601305.
- [54] Y. J. Qian et al., "Robust countermeasure against detector control attack in a practical quantum key distribution system," *Optica*, vol. 6, pp. 1178–1184, 2019.
- [55] Z. Wu et al., "Robust countermeasure against detector control attack in a practical quantum key distribution system: Comment," *Optica*, vol. 7, pp. 1391–1393, 2020.
- [56] H. K. Lo, M. Curty, and B. Qi, "Measurement-device-independent quantum key distribution," *Phys. Rev. Lett.*, vol. 108, 2012, Art. no. 130503.
- [57] S. L. Braunstein and S. Pirandola, "Side-channel-free quantum key distribution," *Phys. Rev. Lett.*, vol. 108, 2012, Art. no. 130502.
- [58] D. Stucki, N. Gisin, O. Guinnard, G. Ribordy, and H. Zbinden, "Quantum key distribution over 67 km with a plug&play system," *New J. Phys.*, vol. 4, pp. 41.1–41.8, 2002.
- [59] D. S. Bethune and W. P. Risk, "Autocompensating quantum cryptography," *New J. Phys.*, vol. 4, pp. 42.1–42.15, 2002.
- [60] C. H. Park et al., "Practical plug-and-play measurement-device-independent quantum key distribution with polarization division multiplexing," *IEEE Access*, vol. 6, pp. 58587–58593, 2018.
- [61] S. F. Hegazy and B. E. A. Saleh, "Quantum key distribution system to overcome intercept-resend and detector-control quantum hacking," U.S. Patent 63/296,711, 2022.
- [62] A. Muller, T. Herzog, B. Huttner, W. Tittel, H. Zbinden, and N. Gisin, "Plug and play" systems for quantum cryptography," *Appl. Phys. Lett.*, vol. 70, pp. 793–795, 1997.
- [63] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, "Quantum cryptography," *Rev. Mod. Phys.*, vol. 74, pp. 145–195, 2002.
- [64] D. B. Souto, J. Liñares, and X. Prieto-Blanco, "Phase auto-compensating high-dimensional quantum cryptography in elliptical-core few-mode fibres," *J. Mod. Opt.*, vol. 66, no. 9, pp. 947–957, 2019.
- [65] W. Y. Hwang, "Quantum key distribution with high loss: Toward global secure communication," *Phys. Rev. Lett.*, vol. 91, 2003, Art. no. 057901.
- [66] H. K. Lo, X. Ma, and K. Chen, "Decoy state quantum key distribution," *Phys. Rev. Lett.*, vol. 94, 2005, Art. no. 230504.
- [67] X. B. Wang, "Beating the photon-number-splitting attack in practical quantum cryptography," *Phys. Rev. Lett.*, vol. 94, 2005, Art. no. 230503.
- [68] C. H. Bennett, F. Bessette, G. Brassard, L. Salvail, and J. Smolin, "Experimental quantum cryptography," *J. Cryptol.*, vol. 5, pp. 3–28, 1992.
- [69] H.-W. Li et al., "Attacking a practical quantum-key-distribution system with wavelength-dependent beam-splitter and multiwavelength sources," *Phys. Rev. A*, vol. 84, 2011, Art. no. 062308.
- [70] N. J. Beaudry, T. Moroder, and N. Lütkenhaus, "Squashing models for optical measurements in quantum communication," *Phys. Rev. Lett.*, vol. 101, 2008, Art. no. 093601.
- [71] T. Tsurumaru and K. Tamaki, "Security proof for quantum-key-distribution systems with threshold detectors," *Phys. Rev. A*, vol. 78, 2008, Art. no. 032302.
- [72] O. Gittsovich, N. J. Beaudry, V. Narasimhachar, R. R. Alvarez, T. Moroder, and N. Lütkenhaus, "Squashing model for detectors and applications to quantum-key-distribution protocols," *Phys. Rev. A*, vol. 89, 2014, Art. no. 012325.

**Salem F. Hegazy** (Member, IEEE) was born in Hulwan, Egypt. He received the B.Sc. (Hons.), Diploma, and M.Sc. degrees in electronics and communications engineering from Cairo University, Giza, Egypt, in 2002, 2006, and 2011, respectively, and the Ph.D. degree in quantum optics and nonlinear optics under a joint-supervision scheme between CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, FL, USA, and the National Institute of Laser Enhanced Sciences (NILES), Cairo University, in 2017. He has been an Assistant Professor with the National Institute of Laser Enhanced Sciences, Cairo University, and also a Senior Research Fellow with the Centre for Photonics and Smart Materials, Zewail City of Science and Technology, Giza, Egypt. His research interests include optical quantum-information and quantum-computation processing, practical security of quantum cryptography, photonic entanglement, optical communications, and optical chaos-based systems. He was the recipient of the Egyptian State Encouragement Award in Engineering Sciences in 2019 presented to distinguished scholars under 40 years.

**Salah S. A. Obayya** (Fellow, IEEE) received the B.Sc. (Hons.) and M.Sc. degrees in electronics and communications engineering and the Ph.D. degree from Mansoura University, Mansoura, Egypt, in 1991, 1994, and 1999, respectively. From 1997 to 1999, he was with the Department of Electrical, Electronic, and Information Engineering, City University London, London, U.K., to carry out the research part of his Ph.D. under the joint supervision scheme between the City University London and Mansoura University. From 2000 to 2003, he was a Senior Research Fellow with the School of Engineering, City University London. From 2003 to 2006, he was with the School of Engineering and Design, Brunel University, London, U.K., as a Senior Lecturer, and a Reader with the School of Electronic and Electrical Engineering, University of Leeds, Leeds, U.K. Since 2008, he has been a Full Professorial Chair of photonics and led the establishment of the Nano-Photonics Research Centre, The University of Glamorgan, Pontypridd, U.K., while he was currently the Vice Chair for Academic Affairs, Zewail City of Science and Technology, Giza, Egypt. He has built an outstanding international reputation in the area of green nanophotonics with a focus on the intelligent computational modeling of modern nanophotonic devices enabling technologies for efficient generation, distribution, and utilization of sustainable energy toward low-carbon green society. He has a track record of 232 journal publications mostly in the IEEE, IET/IEE, and OSA, has authored three books, namely *Computational Photonics* (Wiley, 2010), *Computational Liquid Crystal Photonics* (Wiley, 2016), and *Computational of Photonic Sensors* (2018), which are adopted references for Graduate Programs in Photonics at international Universities, and 252 conference papers in international conference in photonics, many of these are invited talks, and keynote lectures, attracted external funding, from both industry and Research Councils, in the region of \$4 M, and supervised to successful completion 35 Ph.D. students, over 70 M.Sc./M.Res./M.Phil. students, and 15 postdoctoral research fellows.

**Bahaa E. A. Saleh** (Life Fellow, IEEE) received the Ph.D. degree from the Johns Hopkins University, Baltimore, MD, USA, in 1971. He has been Distinguished Professor with CREOL, The College of Optics and Photonics, University of Central Florida, Orlando, FL, USA, since 2009, and he was the Dean in 2009–2019. He held faculty and research positions with the University of Santa Catarina, Florianópolis, Brazil, Max Planck Institute, Germany, University of California-Berkeley, Berkeley, CA, USA, Columbia University, New York, NY, USA, University of Wisconsin-Madison, Madison, WI, USA, and Boston University, Boston, MA, USA, where he was the Chair of the ECE Department (1994–2008). He has made significant contributions in areas including coherence and statistical optics, nonlinear and quantum optics, and image science. His publications include more than 285 journal papers and three books: *Photoelectron Statistics*, (Springer, 1978), *Fundamentals of Photonics* (Wiley, 2007, with M. C. Teich), and *Introduction to Subsurface Imaging* (Cambridge, 2011). He was the Editor-in-Chief of *Journal of the Optical Society of America A* (1991–1997) and founding Editor of *OSA Advances in Optics and Photonics* (2008–2013). He is Fellow of OSA, APS, SPIE, and Guggenheim Foundation. He was the recipient of the OSA Beller Award, SPIE BACUS Award, Kuwait Prize, OSA Distinguished Service Award, and OSA Mees Medal.