Quantum hacking: Saturation attack on practical continuous-variable quantum key distribution

Hao Qin, Rupesh Kumar, and Romain Alléaume
LTCI, Telecom ParisTech, CNRS, 46 Rue Barrault, 75634 Paris Cedex 13, France

(Received 12 November 2015; published 15 July 2016)

We identify and study a security loophole in continuous-variable quantum key distribution (CVQKD) implementations, related to the imperfect linearity of the homodyne detector. By exploiting this loophole, we propose an active side-channel attack on the Gaussian-modulated coherent-state CVQKD protocol combining an intercept-resend attack with an induced saturation of the homodyne detection on the receiver side (Bob). We show that an attacker can bias the excess noise estimation by displacing the quadratures of the coherent states received by Bob. We propose a saturation model that matches experimental measurements on the homodyne detection and use this model to study the impact of the saturation attack on parameter estimation in CVQKD. We demonstrate that this attack can bias the excess noise estimation beyond the null key threshold for any system parameter, thus leading to a full security break. If we consider an additional criterion imposing that the channel transmission estimation should not be affected by the attack, then the saturation attack can only be launched if the attenuation on the quantum channel is sufficient, corresponding to attenuations larger than approximately 6 dB. We moreover discuss the possible countermeasures against the saturation attack and propose a countermeasure based on Gaussian postselection that can be implemented by classical postprocessing and may allow one to distill the secret key when the raw measurement data are partly saturated.

DOI: 10.1103/PhysRevA.94.012325

I. INTRODUCTION

Quantum key distribution (QKD) [1] enables two remote parties Alice and Bob to share common secure keys that are unknown to a potential eavesdropper. Unconditional security of QKD is based on the fundamental laws of quantum mechanics. Side-channel attacks nevertheless remain a crucial problem to guarantee the security of practical implementations. As a matter of fact, the models used in security proofs to describe QKD implementations may not capture all the possible deviations associated with device imperfections. This opens the possibility of attacks against QKD implementations, exploiting either passive (information leakage) or active (induced by the attacker) side channels.

In discrete-variable QKD (DVQKD), various quantum hacking strategies exploiting some implementation imperfections have been proposed and some of them have been demonstrated in experiments [2–4]. Most of the practical attacks that have been demonstrated up to now in DVQKD consist in attacks targeting the detection part of QKD systems [2,4–9] and exploit imperfection of single-photon detectors.

Continuous-variable QKD (CVQKD) is another promising approach to performing quantum key distribution. It relies on continuous modulation of the light field quadratures, which can be measured with coherent detectors such as homodyne or heterodyne detections. Continuous-variable QKD inherits several interesting features associated with the use of coherent detection instead of single-photon detectors: At the system level, CVQKD can be implemented with off-the-shelf components that are also used and optimized in modern optical communications, allowing for a convergence between quantum and classical communications [10] and also simplifying the path and the undertaking associated with photonic integration. Coherent detectors moreover act as efficient and almost single-mode filters, leading to a superior capacity for CVQKD to be wavelength multiplexed with intense classical channels over wavelength-division multiplexing networks [11].

The Gaussian-modulated coherent-state (GMCS) CVQKD protocol [12] is proven secure against collective attacks and recent works have shown progress in proving its security against arbitrary attacks [13,14]. However, similarly to DVQKD, practical CVQKD systems can face security threats linked to imperfect implementations. The validity of security proofs indeed relies on assumptions that may be violated in a practical setup, opening loopholes that may be exploited by Eve to mount attacks. For example, direct [15] or indirect [16–18] manipulation of local oscillator (LO) intensity can fully compromise the security. This imposes the monitoring of LO intensity and the use of filters to forbid wavelength-dependent LO intensity manipulations. Moreover, LO intensity fluctuations also can possibly compromise the security of practical system [19,20] and a stabilization of LO intensity is proposed to defend against such attacks [20].

In this work we have identify a loophole associated with the finite range over which coherent detectors respond nonlinearly. We have shown that it can be used to attack practical implementations of the GMCS CVQKD protocol. Instead of targeting the shot-noise calibration by manipulating the local oscillator, we propose an attack that aims at the homodyne detection located on Bob's side and more specifically at the electronics of the homodyne detection. We name our attack a saturation attack: It combines the induced saturation of the homodyne detection response with a full intercept-resend attack [21]. Based on a realistic model of the homodyne detection response and saturation behavior, we can show that the saturation attack can be used to get information about Alice-modulated input (via an intercept-resend attack, which should in theory bring the key rate to zero) while jointly manipulating the measurement results on Bob's side (taking advantage of the induced nonlinear response of the homodyne detector). For some channel and protocol parameters, the saturation attack can lead Alice and Bob to generate, at a positive rate, a key that they consider as secure, although such a key will be totally insecure due to the intercept-resend attack.

2469-9926/2016/94(1)/012325(16)

Hence the attack can lead to a full security break. Importantly, the attack is also practical and can be realistically launched against existing implementations, since all practical coherent detectors have a finite linearity domain and could be driven (if not monitored) outside this domain of linearity by displacing the mean value of the received quadratures. We however propose a countermeasure that can be implemented simply, by performing a numerical test on the measurement data. The countermeasure consists in a precalibration of the linearity domain of the homodyne detector and then application of a Gaussian postselection filter to the quadrature measurements results of Bob so that the postselected measurement results fall within the linearity domain while the postselected input data are guaranteed to be Gaussian.

This article is organized as follows. In Sec. II we present the GMCS protocol and explain how parameter estimation is performed in this protocol. In Sec. III we briefly review existing work on the practical security of CVQKD and propose the idea of the saturation attack in Sec. IV. In Sec. V we study experimentally the influence of saturation on a practical homodyne detector and propose a simple saturation model to account for it. In Sec. VI we propose a strategy to mount an active attack against the GMCS CVQKD protocol, taking advantage of induced saturation. In Sec. VII we perform numerical simulations to analyze the influence of the saturation attack on parameter estimation, in particular on channel transmission and excess noise, and then discuss the impact on the secret key rate, under two different security criteria. In Sec. VIII we discuss a possible countermeasure and present and analyze a countermeasure based on Gaussian postselection. Finally, in Sec. IX we summarize the main results of our work and discuss some perspectives.

II. GAUSSIAN-MODULATED COHERENT-STATE CONTINUOUS-VARIABLE QUANTUM KEY DISTRIBUTION

A. Protocol

In the GMCS CVQKD protocol [12], Alice encodes information on coherent states of light, which can be easily produced by a laser. The information is encoded on the quadratures X_A and P_A of coherent states, with a centered bivariate Gaussian modulation of variance $V_A N_0$. Here N_0 is the shot-noise variance that appears in the Heisenberg uncertainty relation for the noncommuting quadratures; it corresponds to the variance of the homodyne detection output when the input signal is the vacuum field. Alice sends these Gaussian-modulated coherent states, which constitute the quantum signal, to Bob through the quantum channel. On the reception side, Bob randomly chooses to measure either quadrature X or quadrature P by performing a balanced homodyne detection on the signal, using for that a strong phase reference, called a local oscillator, and switching the quadrature measurement by varying the relative phase of the LO with respect the quantum signal to be either 0 or $\pi/2$.

Keeping track of modulated quadrature data X_A (or P_A) and quadrature measurement results X_B (or P_B), Alice and Bob obtain strings of correlated classical data by repeating this process many times over successive pulses. They can then use error correction to obtain identical strings from their correlated data through reverse reconciliation [12,22] and further perform privacy amplification to obtain a secret key.

In the analysis carried out in this article, which focuses on the impact of a side channel on CVQKD, we do not consider finite-size effects [23] and we assume that all the estimations are performed in the asymptotic limit. We moreover consider the security against collective attacks to compute the secret key rate. One can moreover show that Gaussian attacks are the optimal collective attacks against the GMCS protocol in the asymptotic limit of an infinite number of signals [24,25]. Hence we can analyze the security of the protocol by considering a linear channel model with additive Gaussian noise. In this Gaussian linear model, the Alice-Bob channel is fully characterized by two parameters: the channel transmission and the excess noise. The channel transmission is related to the channel loss and can be derived from the correlation between Alice's and Bob's data. The excess noise is the variance of Bob's quadrature measurements in excess of the shot noise; it can be due to device imperfections (in particular imperfect modulation and noisy detections) or an eavesdropper's actions on the channel.

B. Parameter estimation

In order to estimate parameters from Alice's and Bob's correlated variables, the Gaussian linear model (1) with additive Gaussian noise is considered:

$$X_B = tX_A + X_N, \tag{1}$$

where $t = \sqrt{\eta T}$, with T the channel transmission and η the optical transmission through Bob's setup (including the homodyne detection's finite efficiency). On Alice's side, X_A is a Gaussian random variable centered on zero with variance V_A. Here X_N is the total noise that follows a centered normal distribution with variance $\sigma_N^2 = N_0 + \eta T\xi + v_{\text{ele}}$. This variance includes shot noise N_0, excess noise ξ, and electronic noise of Bob v_{ele}.

In this article we follow the parameter estimation procedure of Ref. [26]. We can obtain three equations relating modulated data X_A and measured data X_B to the parameter estimation:

$$V_A = \text{Var}(X_A) = \langle (X_A - \langle X_A \rangle)^2 \rangle, \tag{2}$$

$$\begin{aligned} V_B &= \text{Var}(X_B) = \langle (X_B - \langle X_B \rangle)^2 \rangle \\ &= \eta T V_A + N_0 + \eta T\xi + v_{\text{ele}}, \end{aligned} \tag{3}$$

$$\begin{aligned} \text{Cov}(X_A, X_B) &= \langle X_A X_B \rangle - \langle X_A \rangle \langle X_B \rangle \\ &= \sqrt{\eta T} V_A. \end{aligned} \tag{4}$$

Additionally, in order to measure the shot noise N_0, Bob needs to close the signal port so he can measure the variance when the input signal is in a vacuum. When there is no signal impinging on the homodyne detection, the variance of homodyne detection is used to calibrate the value of the shot noise. In this case Eq. (3) reduces to an additional equation, obtained by performing a shot-noise calibration:

$$V_{B_0} = N_0 + v_{\text{ele}}. \tag{5}$$

Note that η and v_{ele} are also calibrated values, measured before launching the protocol.

The parameters characterizing the quantum channel in the Gaussian linear model, i.e., T and ξ, can then be estimated from Eqs. (2)–(4):

$$T = \frac{\text{Cov}(X_A, X_B)^2}{\eta\,\text{Var}(X_A)^2}, \tag{6}$$

$$\xi = \frac{\text{Var}(X_B)}{\eta T} - \text{Var}(X_A) - \frac{N_0}{\eta T} - \frac{v_{\text{ele}}}{\eta T}. \tag{7}$$

Additionally, by calibrating the shot-noise variance N_0 from Eq. (5), all variances and correlations can be normalized in shot-noise units and can then be used to estimate the secret key rate.

C. Security model and achievable secret key rate

In order to estimate the secret key rate, Alice and Bob need to compute the mutual information between their data and estimate an upper bound of Eve's information. In this article parameter estimation and secret key rates will be analyzed in the context of collective attacks, in the asymptotic regime [24]. Although the security of CVQKD can be analyzed in a more general setting, we want to stress that extending our analysis to more general (and complex) security models would not qualitatively change the main finding of our article. As a matter of fact, we exhibit an explicit attack strategy, exploiting the saturation of the homodyne. As we will demonstrate, this attack leads to a complete security break against an attacker limited to collective attacks, assuming parameter estimation is performed in the asymptotic regime. By extension, our proposed attack would also lead to a complete security break under more general security models, which consist in increasing the power of the eavesdropper.

A lower bound on the secret key rate achievable against collective attack (in the asymptotic limit) for the CVQKD protocol can be expressed as $R = \beta I_{AB} - \chi_{BE}$ [27]. It is composed of two terms: I_{AB} is the mutual information between Alice and Bob, χ_{BE} is the Holevo bound of Eve's knowledge, and $\beta \in [0,1]$ is the reconciliation efficiency, related to the fact that practical error correction usually does not reach the Shannon limit (which would correspond to the case $\beta = 1$). Here I_{AB} is a decreasing function of the excess noise, while χ_{BE} is an increasing function of excess noise, hence any rise of the excess noise will lead to a decrease of the secret key rate R.

III. PRACTICAL SECURITY ISSUES: LOOPHOLES AND ATTACKS IN CVQKD

In practical CVQKD implementations [27,28], the local oscillator is transmitted publicly on the optical line between Alice and Bob, multiplexed with the quantum channel. Hence the LO can be accessed, and thus manipulated, by an attacker in practical implementations. It is important to note that the LO can in principle be generated locally on Bob's side, as demonstrated in recent proof-of-principle experiments [29–31], where the LO was phase locked with the quantum signals emitted by Alice. However, phase locking two distant lasers creates more complexity and noise and all practical CVQKD full demonstrations have so far been performed with a public LO. This opens the door to different attack strategies based on LO manipulation. An eavesdropper can, for example, modify several properties of the LO pulse, such as the intensity, the wavelength, or the pulse shape [15–20,32]. The eavesdropper can in particular bias the shot-noise calibration (5) by manipulating the LO intensity or its overlap with the quantum signal. We have indeed seen that the excess noise is expressed in shot-noise units. If the shot noise is overestimated while all the other measurements remain unchanged, the excess noise in shot-noise units will then be underestimated. As a consequence, Alice and Bob will then overestimate their secret key rate, leading to a security problem.

Most existing attacks rely on shot-noise estimation induced by different LO manipulations combined with specific strategies. An equal-amplitude attack is described in Ref. [15]. By replacing the quantum signal and the LO pulse by two squeezed states of equal amplitude, Eve can make Alice and Bob measure an excess noise estimate that is much lower than the actual shot noise. This attack may allow Eve to break the security without being detected if Bob does not monitor the LO intensity, which is strongly modified in this attack. In Ref. [32] the authors propose a strategy where Eve changes the shape of the LO pulse to introduce a delay on the clock trigger. As a consequence, the variance of the shot-noise measurement can be lowered without changing the LO power. Such a calibration attack biases the estimation of shot noise and thus the excess noise in shot-noise units. The authors propose a countermeasure based on real-time monitoring of the shot-noise method to prevent this LO manipulation loophole. In Refs. [16,17], as an extension of the equal-amplitude attack [15], a wavelength attack on a CVQKD system using heterodyne detection has been proposed. In this attack, by exploiting the wavelength-dependent property of the homodyne detection's beam splitter, Eve can bias the intensity transmissions of the LO and signal. By inserting light pulses at different wavelengths, this attack allows Eve to bias the shot-noise estimation even if it is performed in real time. This attack can be prevented by adding a wavelength filter before the beam splitter. Recently, in [18], a similar wavelength attack was proposed to compromise the practical security of the CVQKD system using homodyne detection. An improved real-time shot-noise measurement technique is also proposed to detect this attack, closing all known wavelength attack loopholes.

To summarize, the main idea of existing attacks on CVQKD consists in manipulating the local oscillator in different ways so that the eavesdropper can bias the shot-noise estimation and thus the excess noise. The threat of such attacks can be removed if Alice and Bob locally generate a LO pulse [29–31] or measure the shot noise in real time instead of relying on an offline calibration [32]. The LO is an important issue for practical security in CVQKD, but as we will demonstrate with the introduction of the saturation, it is not the only implementation loophole that should be considered in practical CVQKD.

IV. PRINCIPLE OF THE SATURATION ATTACK

Unlike the attacks aiming at the local oscillator, we introduce an attack on CVQKD that exploits the finite linearity domain of the homodyne detection response. Indeed, an implicit but nevertheless fundamental assumption in the security proofs of CVQKD is that the response of the homodyne detection is linear with respect to the input quadrature. This assumption is necessary because parameter estimation [Eqs. (2)–(4)] assumes that the quadrature measurements performed by Bob are linearly related to the optical field quadratures, in order to relate them to the parameters T and ξ of the quantum channel. However, for a practical coherent detector, such as the homodyne detection used to implement the GMCS CVQKD protocol, the linearity domain is limited. If the value of the input quadrature is too large, linearity may not be verified, leading to a saturated behavior.

From Sec. II B we can observe that, based on the Gaussian linear model (1), the parameter estimation consists in the evaluation of the covariance matrix. It is interesting to note that the different coefficients of the covariance matrix are invariant under any linear shifts of the quadratures. Indeed, the security evaluation in CVQKD relies solely on the evaluation of second-order moments of the quadrature, while the first-order moments (mean value) are not monitored. This leaves Eve the freedom to manipulate the mean value of the quadratures. Combining this observation with the existence of a finite domain of linearity for the detection, a natural attack strategy for Eve is to actively introduce a large displacement on the quadrature received by Bob in order to force the homodyne detection to operate in its saturated region. This strategy, which is the core idea of the saturation attack, enables Eve to influence Bob's measurement results and to bias parameter estimation. Importantly, unlike the attacks targeting the local oscillator, in which the shot-noise measurement is influenced, saturation attacks do not bias the shot-noise calibration but still influence the excess noise estimation.

V. SATURATION OF A HOMODYNE DETECTOR

Saturation of the homodyne detection typically occurs when the input field quadrature overpasses a threshold. This threshold depends on characteristics of the homodyne detector's electronics, such as the amplifier's linearity domain or the data acquisition (DA) card range (Fig. 1). If Bob performs quadrature measurements on input signals falling outside the detector's linearity range, the measurement statistics will be influenced by the saturation. Saturation will in particular lead to a decrease in the variance of the measurement results.

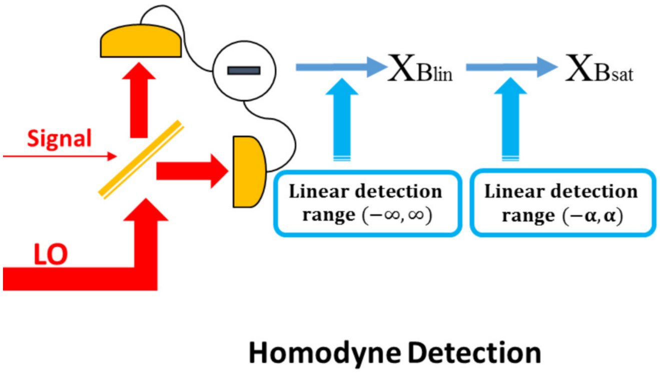
XBlin
XBsat
Signal
LO
Linear detection range $(-\infty,\infty)$
Linear detection range $(-\alpha,\alpha)$
Homodyne Detection

FIG. 1. Model for a practical homodyne detection: Its output $X_{B_{\rm sat}}$ can be seen as the ideal output $X_{B_{\rm lin}}$ on which a saturation function is applied [Eq. (8)].

A. Saturation model

The quadrature measurement performed with homodyne detection consists in the subtraction in the electronic domain of the photocurrents produced by the two photodiodes followed by an electronic front end and acquisition. The standard analysis considers that the homodyne response is linear with respect to the input quadratures. We then denote the measured quadrature by $X_{B_{\rm lin}}$ (X_B in Sec. II). However, the linear detection range of a practical homodyne detector cannot be infinite. We propose a saturation model (8) with predefined upper and lower bounds for the homodyne detection response: For quadrature input values between these two bounds, the response of homodyne detection is unaffected; otherwise it saturates to a constant value. To simplify the analysis, we have assumed in this model that the linear detection range can be described by one single parameter α intrinsic to the detector. Under this saturation model, the linear range is $[-\alpha,\alpha]$ and the measured quadrature is called $X_{B_{\rm sat}}$. The relation between $X_{B_{\rm sat}}$ and $X_{B_{\rm lin}}$ is as follows:

$$\begin{array}{rlrl} & X_{B_{\rm lin}} \geqslant & \alpha, & X_{B_{\rm sat}} = \alpha; \\ \text{if } & |X_{B_{\rm lin}}| < \alpha, & \text{then } & X_{B_{\rm sat}} = X_{B_{\rm lin}}; \\ & X_{B_{\rm lin}} \leqslant & -\alpha, & X_{B_{\rm sat}} = -\alpha. \end{array} \tag{8}$$

As expected, if $\alpha \to \infty$, the saturation model is equivalent to the standard linear model. In a typical (nonsaturated) CVQKD implementation, the value of α is large enough to ensure that field quadratures almost never overpass the saturation threshold limit. Alice and Bob can in practice guarantee the linearity by limiting the number of photons impinging on the homodyne detector to be much smaller than α^2. Since the limit α is intrinsic to the electronics of the detector, a practical way to guarantee with high probability that $\alpha \gg X_{B_{\rm lin}}$ is to lower the LO intensity so that the shot-noise value $N_0 \ll \alpha^2$. In general, input quadrature modulation variance is calibrated in shot-noise units that depend on LO intensity and Alice can choose a Gaussian modulation with $\langle X_{B_{\rm lin}} \rangle = 0$ and $\mathrm{Var}(X_{B_{\rm lin}}) \ll \alpha^2$ so that the detector does not saturate. However, as mentioned earlier, this procedure cannot cope with situations where the mean value of $X_{B_{\rm lin}}$ is strongly displaced, as will be the case in a saturation attack.

B. Experimental observation of saturation

In a practical balanced homodyne detector, the common mode rejection ratio is not infinite and the mean value of the homodyne detection in the absence of an input signal is affected by the imbalance, leading to $\langle X_{B_{0,\rm lin}} \rangle = \epsilon I_{\rm LO}$, where $I_{\rm LO}$ is the LO intensity and ϵ is the imbalance factor that is dependent on experimental imperfections such as photodiode quantum efficiency mismatch or beam-splitter imbalance.

Because of this imperfection (but in the absence of saturation), the relation between measured noise variance (in V^2) and LO intensity (in μW) usually can be written as $\mathrm{Var}(X_{B_{0,\rm lin}}) = A I_{\rm LO} + B$ [33] (we neglect the quadratic part since in our case the LO power is relatively low). Here $I_{\rm LO}$

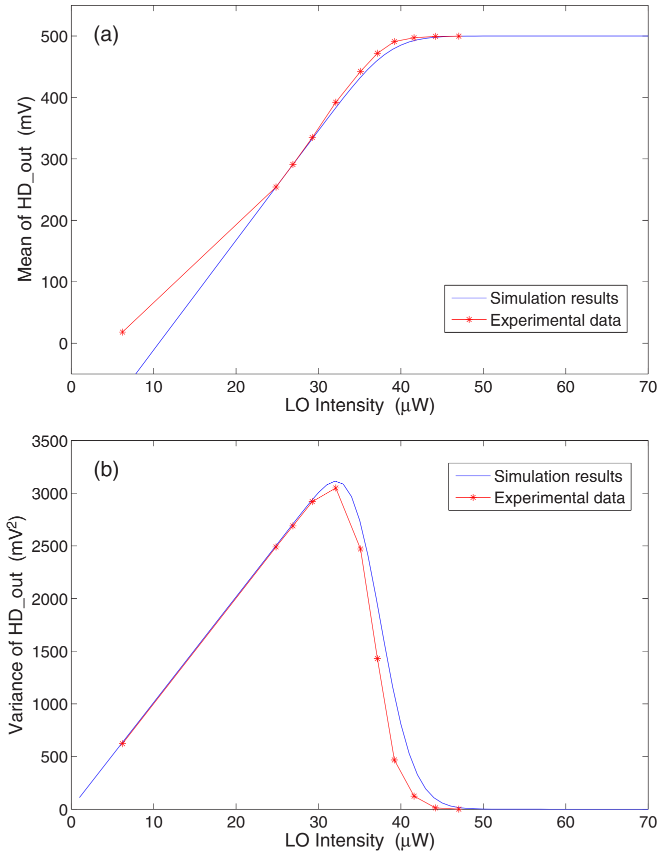

FIG. 2. Experimental characterization of the saturation behavior of a practical homodyne detection. (a) Mean of the homodyne output $\langle X_{B_0,\mathrm{sat}}\rangle$ vs LO intensity. (b) Variance of the homodyne output $\mathrm{Var}(X_{B_0,\mathrm{sat}})$ vs LO intensity.

is the LO intensity and A is linear with I_{LO} and is related to shot noise while B is independent of I_{LO} and is related to electronic noise. The values of A and B can be determined experimentally.

We have performed experimental shot-noise measurement, measuring the variance of the homodyne detection output, as a function of I_{LO}. This has revealed that the measured shot-noise variance is linear with I_{LO} in a given range and then drops when the LO intensity is above a certain value. We have analyzed this behavior with the saturation model presented in Sec. V A and compared its prediction to experimental measurements in Fig. 2. We display the measured homodyne detection output and its variance for the vacuum input signal as a function of I_{LO}. The experimental results are displayed in Fig. 2. The linear behavior can be observed when the LO intensity is below 35 μW. Due to the imbalance of homodyne detection ϵ, the mean value of the homodyne output can become large as the LO intensity increases. If these values overpass the linearity threshold (in the present case 0.5 V, due to the DA card) the homodyne detection response saturates to a constant value [Fig. 2(a)]. As a consequence, the measured shot-noise variance strongly decreases [Fig. 2(b)] when such saturation happens.

In order to check the validity of the saturation model introduced in Eq. (8), we have simulated the expected homodyne detection response with our saturation model and compared it with experimental measurements. We first determine the parameters A and B from linear fit on $\langle X_{B_{0,\mathrm{lin}}}\rangle$ and $\mathrm{Var}(X_{B_{0,\mathrm{lin}}})$ versus the LO intensity over the domain of linearity $I_{\mathrm{LO}} < 35\ \mu\mathrm{W}$. The saturation parameter α is here fixed by our DA range: $\alpha = 0.5$ V. We then apply the saturation model (8) to the variable $X_{B_{0,\mathrm{lin}}}$ to obtain $X_{B_{0,\mathrm{sat}}}$. We compute the mean $\langle X_{B_{0,\mathrm{sat}}}\rangle$ and the variance $\mathrm{Var}(X^2_{B_{0,\mathrm{sat}}})$, which result in the behavior shown in Fig. 2. For the measured shot noise under saturation, the simulation results match very well with our experimental data. This indicates that our proposed saturation model is realistic and can be further used to interpret our saturation attack.

VI. ATTACK STRATEGY

A. Intercept-resend attack

The intercept-resend attack plays an important role as one part of our saturation attack. The intercept-resend attack [21] in CVQKD is achievable with today's technologies and its security analysis has been studied in previous work [21]. In such an attack, Eve intercepts all the pulses sent by Alice on the quantum channel and measures simultaneously the quadratures X and P with the help of heterodyne detection. Eve then prepares a coherent state according to her measurement results and sends it to Bob. Under such attack the correlation between Eve's and Bob's data will be stronger than that between Alice and Bob, so Eve always has an information advantage. Due to measurement disturbance and coherent-state shot noise, the intercept-resend attack, that is, entanglement breaking, introduces two shot-noise units of excess noise. In practice, Eve's device and her action can introduce additional excess noise. A full intercept-resend attack will therefore introduce at least two shot-noise units of excess noise, which should forbid the generation of a secret key under collective attacks. This however assumes that the estimation procedure is not biased. We will see, on the contrary, that a saturation attack can bias the excess noise estimation and lead to a security break.

B. Saturation attack

The saturation attack on the GMCS CVQKD protocol is an active attack, where Eve combines a full intercept-resend attack with an induced saturation of Bob's detector. Saturation of Bob's homodyne detection is obtained by displacing the quadrature of the re-sent coherent state. The displacement value Δ is chosen by Eve but is constant for each re-sent coherent state pulse. When performing the intercept-resend attack, Eve can also choose to rescale the re-sent quadratures by a gain g.

Eve chooses the attack parameter bias of the estimated excess noise below the null key threshold (calculated under collective attack [27]) so that, according to their estimation, Alice and Bob will assume they can obtain a positive key rate and will accept to distill such a key based on parameter estimation, while no secure key can be obtained from the actual correlations since a full intercept-resend attack has been performed. We propose a visual description of our

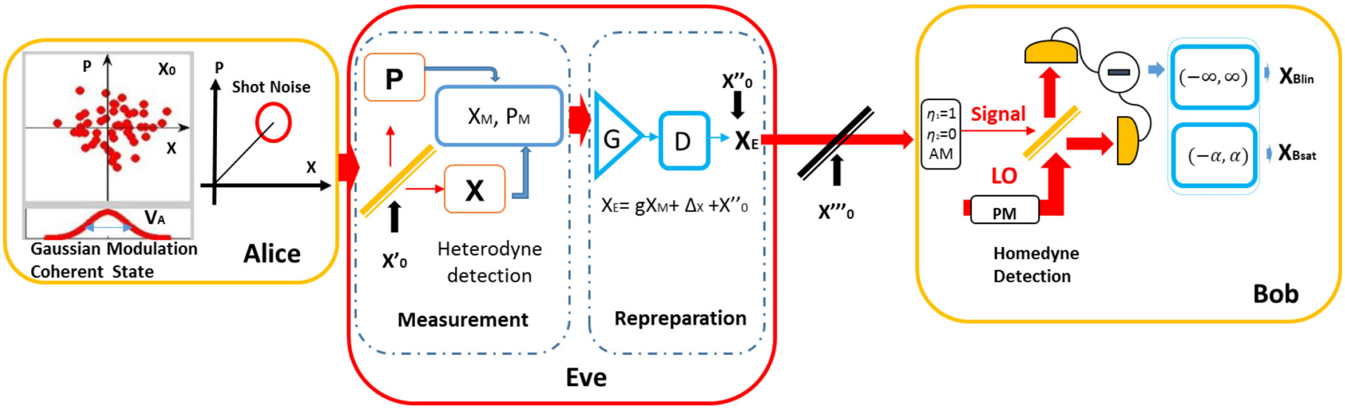

FIG. 3. General description of the GMCS CVQKD under the saturation attack: Alice, prepare the coherent state with quadratures X and P; Eve, measurement and repreparation stage; G, gain; D, displacement; Bob, perform the homodyne detection: AM, amplitude modulator; η_1 and η_2, signal transmission coefficients; PM, phase modulator; and $[-\alpha,\alpha]$, linear working range.

saturation attack in Fig. 3, in which we distinguish mainly two functional blocks: quadrature measurement and quadrature repreparation. By using heterodyne detection, Eve measures Alice's quadratures X_A and P_A simultaneously. In order to simplify our analysis, we assume that Eve's station is located at Alice's output and that the channel transmissions between the Alice-Bob and Eve-Bob channels are equal. Moreover, we assume that Alice and Bob measure their shot noise and monitor the LO intensity in real time [32], with two transmission coefficients randomly decided on Bob side ($\eta_1 = 1$ and $\eta_2 = 0$).

Eve's measurement results X_M and P_M after the heterodyne measurement are expressed as

$$X_M = \frac{1}{\sqrt{2}}(X_A + X_0 + X_0' + X_{N_{A,E}}), \tag{9}$$

$$P_M = \frac{1}{\sqrt{2}}(P_A + P_0 + P_0' + P_{N_{A,E}}), \tag{10}$$

where X_0 is a noise-term due to the coherent-state encoding of Alice while X_0' is a noise term due to a 3-dB loss in the heterodyne detection; $X_{N_{A,E}}$ is a random noise that accounts for the technical noise of Alice's preparation and Eve's measurement process with its variance $\xi_{A,E}$. In the repreparation stage, Eve prepares a coherent state with quadratures X_E and P_E according to her measurements X_M and P_M. Eve can also induce displacements Δ_X and Δ_P and an amplification g of the data X_M before optical encoding. In our further analysis, we only look at the quadrature X, but the treatment for the quadrature P is totally symmetric. The resend quadrature of Eve can be written as

$$\begin{aligned} X_E &= gX_M + \Delta_X + X_0'' && (11)\\ &= \frac{g}{\sqrt{2}}(X_A + X_0 + X_0' + X_{N_{A,E}}) + \Delta_X + X_0'', && (12)\end{aligned}$$

where X_0'' is a noise term due to the coherent-state encoding of Eve. Here X_0, X_0', and X_0'' all follow $\mathcal{N}(0,N_0)$ with their variance equal to one unit of shot noise N_0.

Introducing displacement on coherent states is experimentally achievable [34] and since Eve prepares the states, the displacement parameters Δ_X and Δ_P can be freely chosen by her. We will first consider that Eve chooses an amplification coefficient $g = \sqrt{2}$ in order to compensate for the loss from the heterodyne detection.

1. Linear detection

On Bob's side, Bob measures the quadrature sent by Eve by performing homodyne detection. We first consider that Bob uses homodyne detection whose linear detection range is infinite (Fig. 1). The measured quadrature $X_{B_{\text{lin}}}$ can be written as

$$X_{B_{\text{lin}}} = t(X_E + X_{N_{E,B}}) + \sqrt{1-t^2}X_0''' + X_{\text{ele}}. \tag{13}$$

After propagation through the lossy channel, the technical noise of Eve and Bob $X_{N_{E,B}}$ [$\text{Var}(X_{N_{E,B}}) = \xi_{E,B}$], vacuum noise $\sqrt{1-t^2}X_0'''$ [$\text{Var}(X_0''') = N_0$], and electronic noise of Bob X_{ele} [$\text{Var}(X_{\text{ele}}) = v_{\text{ele}}$] are added to the quadrature prepared by Eve (X_E). Here $t = \sqrt{\eta T}$, where T is the channel transmission between Eve and Bob and η is Bob's efficiency. The correlation between Alice and Bob and the variance of Bob can be written, respectively, as

$$\begin{aligned}\text{Cov}(X_A, X_{B_{\text{lin}}}) &= \langle X_A X_{B_{\text{lin}}}\rangle \\ &= \frac{tg}{\sqrt{2}}\langle X_A X_A\rangle + t\Delta_X\langle X_A\rangle \\ &= \frac{tg}{\sqrt{2}}\text{Var}(X_A), \end{aligned} \tag{14}$$

$$\begin{aligned}\text{Var}\big(X_{B_{\text{lin}}}\big) &= \big\langle X_{B_{\text{lin}}}^2\big\rangle - \langle X_{B_{\text{lin}}}\rangle^2 \\ &= \frac{t^2g^2}{2}[\text{Var}(X_A) + 2N_0 + \xi_{\text{sys}}] + (1-t^2)N_0 \\ &\quad + t^2N_0 + v_{\text{ele}} + t^2\Delta_X^2 - t^2\Delta_X^2 \\ &= \eta T\frac{G}{2}\text{Var}(X_A) + \eta T\frac{G}{2}(2N_0 + \xi_{\text{sys}}) \\ &\quad + N_0 + v_{\text{ele}}. \end{aligned} \tag{15}$$

In Eqs. (14) and (15), we can see that with an ideal linear detection range, the induced displacement Δ_X has no influence on the measurement results. As a matter of fact, the value of Δ_X has no impact on both correlation and variance.

Under linear detection and an intercept-resend attack with the gain $G = g^2 = 2$, the correlation (14) is not modified by Eve's action, so the estimated channel transmission is not biased ($T_{\text{lin}} = T$). Based on Eq. (7), the excess noise estimation on Alice's side is $\xi_{\text{lin}} = 2N_0 + \xi_{\text{sys}}$, where $\xi_{\text{sys}} = \xi_{A,E} + \frac{2}{G}\xi_{B,E}$. Similarly to Sec. II, we introduce the noise variable X_N, which contains all the noise added to Bob's measurement, and the variance of X_N is $\sigma_N^2 = \eta T \frac{G}{2}(2N_0 + \xi_{\text{sys}}) + N_0 + v_{\text{ele}}$.

2. *Saturated detection*

As we have seen, the linearity of homodyne detection cannot be guaranteed over an arbitrarily large detection range. A more realistic model consists in taking saturation into account, according to the saturation model described in Eq. (8): We denote by $X_{B_{\text{sat}}}$ the quadrature measured by Bob in a model taking saturation into account. Under this modified model, the quadrature measured by Bob $X_{B_{\text{sat}}} = X_{B_{\text{lin}}}$ only if $|X_{B_{\text{lin}}}| < \alpha$. Otherwise the quadrature measurement saturates to a constant value, equal to the detection limit α or $-\alpha$.

Eve can freely set the displacement values Δ_X and Δ_P so that $X_{B_{\text{lin}}}$ can partially overpass the linear range $[-\alpha,\alpha]$. In further analysis, we consider $\Delta = t\Delta_X$ as the displacement value. In order to induce a given value of Δ on the quadrature of the coherent state impinging on Bob, Eve can choose a proper Δ_X once she knows t, which typically depends on the distance. As we will see, parameter estimation affected by saturation can lead to excess noise below the null key threshold. In the next section we will show that under certain conditions of our attack strategy, Eve can manipulate the channel transmission and the excess noise estimated by Alice and Bob so that her intercept resend action can remain under cover while fully compromising the practical security of the CVQKD protocol.

C. Experimental feasibility of the saturation attack

The saturation attack is based on the ability of Eve to displace the quadrature measurement results obtained by Bob. One possible strategy to implement the attack consists, for Eve, in coherently displacing the optical quadratures of the quantum signals sent by Alice. This requires Eve to phase lock her laser with Alice's laser. This is essentially a classical phase reference sharing problem (since Alice's phase reference signal and Eve's laser are both intense) and can thus be performed in principle with arbitrary high precision, even though it could in practice be limited by the available phase-locking bandwidth. If Eve's laser is phase locked with Alice then she can control it as a coherent pump laser whose phase and intensity are also controlled by Eve, to induce an arbitrary coherent displacement on Alice quantum signals, by mixing the quantum signals with the pump on an unbalance beam splitter, with the intense pump [34].

We do not consider in this article the question of the influence of experimental limitations of Eve on the possibility to conduct the attack. We adopt instead a standard viewpoint in quantum cryptography, which is to assume that the eavesdropper has perfect hardware. Nevertheless, an important related question will be to demonstrate experimentally that a saturation attack can actually be performed in practice on a running system. This work is considered in [35].

VII. SECURITY ANALYSIS

A. Parameter estimation under the saturation attack

The channel transmission and excess noise estimation fully characterize the quantum channel of CVQKD; we thus only need to analyze the impact of saturation on these two estimated parameters. It is in particular critical to evaluate whether an induced saturation can reduce the excess noise estimation and thus open the door to severe attacks.

1. *Channel transmission estimation*

Under the saturation attack, Alice encodes X_A and Bob measures $X_{B_{\text{sat}}}$ [Eqs. (8) and (13)] and they evaluate the correlation coefficient $\text{Cov}(X_A, X_{B_{\text{sat}}})$ (calculation details can be found in Appendix A). From this correlation coefficient (A3), the estimation of the channel transmission under saturation attack $\hat{T}_{\text{sat}}$ can be expressed as

$$\hat{T}_{\text{sat}} = T\frac{G}{8}\left[1 + \text{erf}\left(\frac{\alpha - \Delta}{\sqrt{2\,\text{Var}(X_{B_{\text{lin}}})}}\right)\right]^2 \quad (\Delta > 0), \qquad (16)$$

in which erf is the error function defined in Eq. (A4) and $\text{Var}(X_{B_{\text{lin}}})$ is the variance of Bob's measurement under linear detection. As we have discussed in Sec. III, a reasonable assumption for the detector linearity limit α is that $\alpha^2 \gg \text{Var}(X_{B_{\text{lin}}})$ and $\alpha^2 \gg N_0$, so the measurement results of Bob would not be affected by saturation in the absence of displacement. This agrees with Eq. (16): If $\alpha - \Delta$ is much larger than $\sqrt{2\,\text{Var}(X_{B_{\text{lin}}})}$, then $\hat{T}_{\text{sat}} \simeq T\frac{G}{2}$, which is the estimated value under the linear model ($G = 2$ being the natural rescaling choice to compensate for the loss introduced by heterodyne detection). However, when Δ is close to α, the impact of saturation becomes important and $\hat{T}_{\text{sat}}$ becomes smaller than T. An extreme case is that when Δ is much larger than α, the error function becomes -1 and $\hat{T}_{\text{sat}} = 0$.

2. *Excess noise estimation*

From Eq. (7) the estimated excess noise depends on the variance of Bob's measurement and on the channel transmission between Alice and Bob. Under the saturation attack, these two values will both decrease. We need to evaluate these two values to see whether the induced saturation will result in a reduction of the estimated excess noise. We have already analyzed $\hat{T}_{\text{sat}}$ in the previous subsection [Eq. (16)]. With Eq. (8) we can calculate $\text{Var}(X_{B_{\text{sat}}})$ under a saturation attack (calculation details can be found in Appendix B). Based on $\hat{T}_{\text{sat}}$ and $\text{Var}(X_{B_{\text{sat}}})$, we are able to express the estimated excess noise in shot-noise units under the saturation attack

$$\begin{aligned}\frac{\hat{\xi}_{\text{sat}}}{N_0} = \frac{1}{\eta T\frac{G}{2}(1+A)^2 N_0}\Bigg[&\text{Var}(X_{B_{\text{lin}}})\left(1 + A - \frac{B^2}{\pi}\right)\\ &- 2\sqrt{\frac{2\,\text{Var}(X_{B_{\text{lin}}})}{\pi}}(\alpha - \Delta)A * B\\ &+ (\alpha - \Delta)^2(1 - A^2) - 4N_0 - 4v_{\text{ele}}\Bigg] - \frac{V_A}{N_0}, \qquad (17)\end{aligned}$$

in which A and B are given by

$$A = \mathrm{erf}\left(\frac{\alpha-\Delta}{\sqrt{2\,\mathrm{Var}(X_{B_{\mathrm{lin}}})}}\right), \quad B = \exp\left(-\frac{(\alpha-\Delta)^2}{2\,\mathrm{Var}(X_{B_{\mathrm{lin}}})}\right). \tag{18}$$

From Eq. (17) we can verify that when the value of $\alpha - \Delta$ is much larger than $\sqrt{2\,\mathrm{Var}(X_{B_{\mathrm{lin}}})}$, then $A \to 1$ and $B \to 0$, so $\hat{\xi}_{\mathrm{sat}} = \frac{\mathrm{Var}(X_B)}{\eta T} - \mathrm{Var}(X_A) - \frac{N_0}{\eta T} - \frac{v_{\mathrm{ele}}}{\eta T} = \xi_{\mathrm{lin}}$ [Eq. (7)]. It can be considered that no saturation is induced and the excess noise estimation is not affected.

3. Estimated excess noise can be made arbitrary small

We can prove, by the use of the intermediate-value theorem, that $\hat{\xi}_{\mathrm{sat}}$ can be manipulated to be any value below ξ and in particular any value below an arbitrarily small ξ.

Proposition. Under a saturation attack, for any $0 < \xi_T < \xi$, there always exists a particular value of the displacement Δ_T for which $\hat{\xi}_{\mathrm{sat}} = \xi_T$.

Proof. The $\hat{\xi}_{\mathrm{sat}}$ is a function of Δ. When $\Delta = 0$, $\hat{\xi}_{\mathrm{sat}}(0) = \xi > 0$, where $\xi = 2N_0 + \xi_{\mathrm{sys}}$ under an intercept-resend attack. When $\Delta = 2\alpha$, since we can assume that $\alpha^2 \gg \mathrm{Var}(X_{B_{\mathrm{lin}}})$, we then have

$$A = -\mathrm{erf}\left(\frac{\alpha}{\sqrt{2\,\mathrm{Var}(X_{B_{\mathrm{lin}}})}}\right) = -1$$

and $\hat{\xi}_{\mathrm{sat}}(2\alpha) \to -\infty$. Since $\hat{\xi}_{\mathrm{sat}}$ is a continuous function of Δ over the interval $[0,2\alpha)$, then for any ξ_T in $(-\infty, \xi]$ there always exists a $\Delta \in [0,2\alpha)$ so that $\hat{\xi}_{\mathrm{sat}} = \xi_T$.

B. Defining criteria of success for the saturation attack

Alice and Bob estimate the key rate based on their estimation of excess noise and channel transmission. If the excess noise is too large, it will not allow Alice and Bob to distill any secret key. A full security break consists in an attack where Eve has full knowledge of the generated key while Alice and Bob still accept this compromised key material. An intercept-resend attack is an attack strategy that leads in general to a denial of service but not to a full security break on CVQKD. On the other hand, we want to claim that the saturation attack can be used to obtain a full security break.

To clarify what we mean, we define a first criterion (level I) for a successful saturation attack, corresponding to a set of conditions to meet.

(a) The attacker Eve performs the saturation attack: an intercept-resend attack combined with displacement.

(b) Alice and Bob obtain a positive key rate from their estimated parameters $\hat{T}_{\mathrm{sat}}$ and $\hat{\xi}_{\mathrm{sat}}$.

This set of conditions corresponds to a full security break because Alice and Bob will obtain a positive key rate under the attack and thus accept key material, while this key is insecure as it can be fully obtained by Eve. Because of the proposition put forth in Sec. VII A 3, we can prove that the saturation attack can always meet the level I criterion: For any quantum channel, characterized by T and ξ, the saturation attack can cause the parameter estimation to always turn the estimated parameter to $\hat{\xi}_{\mathrm{sat}} \simeq 0$ while $0 < \hat{T}_{\mathrm{sat}} < T$. In particular, under saturation, as the estimated excess noise can be made arbitrarily close to zero, Alice and Bob will always generate some positive key rate and the level I criterion can always be met.

While the level I criterion defines conditions for a successful attack, the induced saturation can in practice strongly decrease the estimated channel transmission $\hat{T}_{\mathrm{sat}}$ [Eq. (16)]. This can be a problem in practice since Alice and Bob usually have a good *a priori* estimate of the channel transmission based on their knowledge of the channel length and of the fiber attenuation coefficient. In addition, channel loss are usually calibrated before any new optical device (such as a QKD system) is installed. If the measured channel transmission is much lower than the expected value for a given link distance, Alice and Bob can reasonably be suspicious and they may decide to reject the key. This motivates us to introduce additional conditions to the list and to define a level II criterion for a successful saturation attack.

(a) The attacker Eve performs the saturation attack (an intercept-resend attack on each pulse combined with displacement of each re-sent pulse).

(b) Maintain the channel transmission estimation unaffected ($\hat{T}_{\mathrm{sat}} = T$).

(c) Alice and Bob obtain a positive key rate from their estimated parameters $\hat{T}_{\mathrm{sat}}$ and $\hat{\xi}_{\mathrm{sat}}$.

The strategy for Eve, in order to meet this level II criterion, will be to adjust not only the displacement Δ, but also the gain g, in the saturation attack.

C. Analysis and simulation results

We will formalize two strategies for Eve and study numerically whether they can be used to meet the two criteria for the success of the saturation attack, respectively. We use Eqs. (16) and (17) to perform numerical evaluation of $\hat{T}_{\mathrm{sat}}$ and $\hat{\xi}_{\mathrm{sat}}$ in order to study the impact of the saturation attack.

1. Assumptions used in the numerical simulations

We have performed numerical simulations of the estimated excess noise $\hat{\xi}_{\mathrm{sat}}$, the estimated channel transmission $\hat{T}_{\mathrm{sat}}$, and the secret key rate under a collective attack. We have chosen our simulation parameters in order to match typical parameters that can be achieved and chosen for the operation of an experimental CVQKD system.

(i) First is deployment over a dark fiber, with the quantum channel wavelength in the C band and the fiber attenuation coefficient $a = 0.2$ dB/km.

(ii) Next is the total optical transmission (including homodyne detection finite efficiency) of Bob: $\eta = 0.55$.

(iii) Then we choose the linear detection limit of Bob's homodyne detection: $\alpha = 20\sqrt{N_0}$.

(iv) The variance of the electronic noise is chosen as $v_{\mathrm{ele}} = 0.015N_0$, i.e., a result that can typically be achieved with a 10-MHz bandwidth homodyne detection and system clock rate of 1 MHz [28].

(v) We have chosen a conservative value $\xi_{\mathrm{sys}} = 0.1$ for the system's excess noise (equivalent to the excess noise at the input) in our simulations. This value is relatively high compared to some experimental results demonstrated in CVQKD [11,28] but it has been encountered in [21], when performing the experimental demonstration of the intercept-resend attack. Adopting a pessimistic value for system excess noise is conservative and will not weaken our predictions concerning the power of the saturation attack on practical systems.

(vi) In a practical CVQKD deployment, the value of Alice variance modulation V_A depends on the link distance. This is in particular due to finite reconciliation efficiency in practice. To achieve a high reconciliation efficiency in practical CVQKD ($\beta = 0.95$), optimized error correction codes work at a fixed signal-to-noise ratio (SNR). Therefore, Alice needs to optimize her modulation variance with respect to the distance in order to work at a given SNR. We have followed the procedure described in Ref. [22] to choose Alice's variance with respect to distance in our numerical simulations.

2. *Attack strategy I: Meeting level I criteria by varying* Δ

Let us define strategy I more precisely.

(a) Eve implements the saturation attack as described in Sec. VI B.

(b) Eve chooses a fixed gain value $G = g^2 = 2$ in order to compensate for the loss due to heterodyne detection.

(c) By choosing the value of Δ, Eve biases the excess noise estimation of Alice and Bob below the null key threshold, so Alice and Bob can obtain a positive key rate.

The key idea of this strategy is that, for a given distance with the knowledge of Var$(X_{B_{\text{lin}}})$, Eve can manipulate $\hat{\xi}_{\text{sat}}$ by changing Δ. More importantly, Eve needs to manipulate the excess noise evaluation so that $\hat{\xi}_{\text{sat}}$ falls below the null key threshold but remains positive, to meet the level I success criterion. Here $\hat{\xi}_{\text{sat}}$ is a function of Δ [Eq. (16)]; the behavior of $\hat{\xi}_{\text{sat}}$ versus Δ is shown in Fig. 4(a). Under the linear model, the total estimated excess noise under a full intercept-resend attack is $\hat{\xi}_{\text{lin}} = \xi = 2.1$, including 0.1 technical noise [black curves in Fig. 4(a)]. With such an excessive noise, no key rate can be established by Alice and Bob. However, $\hat{\xi}_{\text{sat}}$ can be manipulated by changing the value of Δ. In Fig. 4(a), for long distance (i.e., above 20 km) $\hat{\xi}_{\text{sat}}$ always decreases when Δ increases. In particular, when Δ is close to α, $\hat{\xi}_{\text{sat}}$ is significantly reduced, which agrees with the analysis in Sec. VII A. For short distance (i.e., below 15 km), when Δ increases, $\hat{\xi}_{\text{sat}}$ first increases and then decreases, but $\hat{\xi}_{\text{sat}}$ can still become arbitrarily small when Δ is large enough. Importantly, from Fig. 4(a) we can see that Eve can obtain an arbitrarily small value of $\hat{\xi}_{\text{sat}}$ by manipulating Δ at any distance, which agrees with the analysis in Sec. VII A 2.

A drawback of this strategy I of saturation attack is that the estimated channel transmission can be strongly reduced, i.e., we can have $\hat{T}_{\text{sat}} \ll T$ [Eq. (16)]. In Fig. 4(b) we plot the estimated channel transmission on a logarithmic scale versus distance, in which the black curve is the estimated channel transmission T versus distance in the absence of an attack and the other curves are the estimated channel transmission $\hat{T}_{\text{sat}}$ under the saturation attack. We can see that the estimated channel transmission can be strongly reduced in comparison to the actual transmission in the absence of an attack. This is especially so if Δ is large, which will be the case, as we can see in Fig. 4(a) for short links, where it is necessary to use a large value of Δ to effectively reduce the excess noise estimation and meet criterion I. Hence, even though the attack strategy I can always be mounted, it may lead to using a large displacement value Δ (typically or even beyond the saturation limit α set to 20 in our simulations). This will strongly reduce the effective transmission of the channel $\hat{T}_{\text{sat}}$ and therefore the achievable secret key rate.

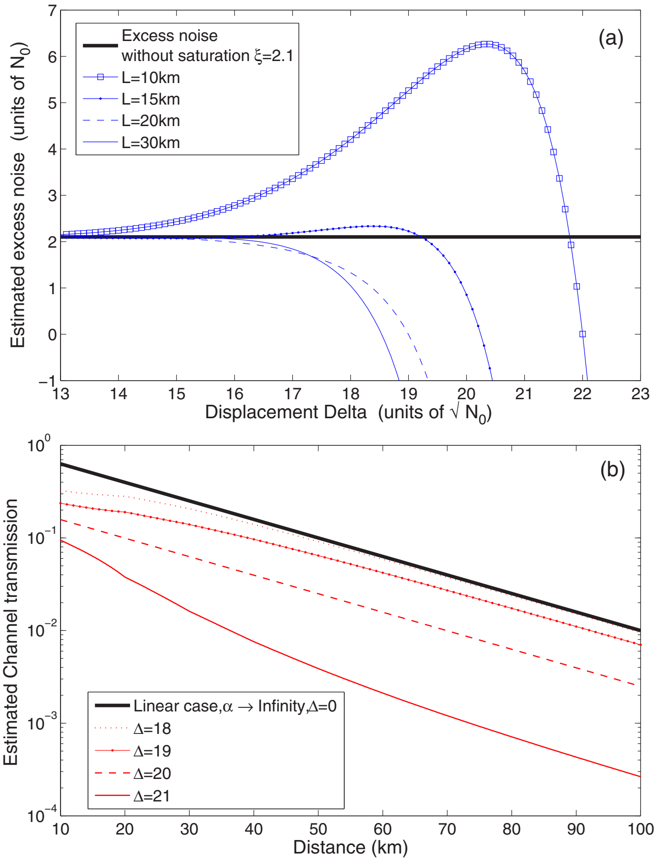

FIG. 4. Simulations of estimated excess noise and channel transmission under attack strategy I. See Sec. VII C 1 for the simulation parameters. (a) Estimated excess noise $\hat{\xi}_{\text{sat}}$ versus Δ for different distances. (b) Estimated transmission $\hat{T}_{\text{sat}}$ versus distance for different Δ.

3. *Attack strategy II : Meeting level II criteria by varying* Δ *and* g

As we have just discussed, inducing the saturation of the homodyne detection (through the displacement Δ) can lower the correlation between Alice's and Bob's data, which will result in a decrease of the estimated channel transmission $\hat{T}_{\text{sat}}$ [Fig. 4(b)] and thus also of the achievable secret key rate by performing the GMCS CVQKD protocol over the channel. However, as already stated in Sec. VII B, there are many practical cases where Alice and Bob may monitor, or at least perform, some kind of consistency check on the estimated transmission and could therefore identify a problem if the estimated transmission, which becomes $\hat{T}_{\text{sat}}$ under the attack, is significantly smaller than the value of T they expect.

This motivates us to define a second attack strategy, capable of meeting the level II criterion: Perform the saturation attack and obtain a positive key rate while leaving the estimation of the channel transmission unchanged. A level II criterion clearly could not have been achieved by solely varying the displacement Δ applied on the coherent states. However, the intercept-resend attack that is part of the saturation attack leaves an additional degree of freedom: Eve can rescale the value of the re-sent quadratures (classical result obtained after heterodyne detection) by a gain g that she can also freely choose.

To meet the success criterion II, we will study a second strategy is similar to strategy I except for the second step, where the gain g will be set according to Eq. (19), in which $\text{Var}(X_{B_{\text{lin}}})$ is given by Eq. (15). As a matter of fact, if Eq. (19) is verified, then $\langle X_A X_{B_{\text{sat}}}\rangle = \frac{1}{2}\langle X_A^2\rangle t$ and we will thus have $\hat{T}_{\text{sat}} = T$, which guarantees that the channel transmission estimation for Alice and Bob is not biased,

$$\frac{2\sqrt{2}}{g} - 1 = \text{erf}\left(\frac{\alpha - \Delta}{\sqrt{\text{Var}(X_{B_{\text{lin}}})}}\right). \tag{19}$$

Now g can be considered as a function of Δ, as displayed in Fig. 5(a). Furthermore, in order to see whether we can meet criterion II and have a full security break with this choice of g we still need to analyze the estimated of excess noise and secret key rate. By taking the g solutions of Eq. (19) into account, the behavior of $\hat{\xi}_{\text{sat}}$ versus Δ is shown in Fig. 5(b).

We can see that if the distance is longer than 30 km it is always possible to reduce $\hat{\xi}_{\text{sat}}$ close to zero by choosing a value of Δ close to α and thus to have an attack meeting criterion II. On the other hand, if the distance is smaller than approximately 30 km, it will not be possible to meet the attack success criterion II and to jointly maintain the estimate of the channel transmission unchanged and have a positive key rate while launching the saturation attack. Thus the capacity to launch a successful saturation attack under success criterion II is dependent on the distance, as can be seen in Fig. 5(b).

We also need to study the condition for $\hat{\xi}_{\text{sat}} < \xi_{\text{null}}$ as we previously did in Sec. VII C 2. However, the analysis is now simpler, since the estimated channel transmission is not biased and the null key threshold does not depend on the attack parameter Δ and only varies with distance. In Fig. 6(a) we enlarge the scale of Fig. 5(b) and compare the estimated excess noise to the null key threshold for different distances. As we can see, when the distance reaches 31 km, the condition $\hat{\xi}_{\text{sat}} < \xi_{\text{null}}$ can only be satisfied with a choice of $\Delta \simeq 19.5$ and level II criterion conditions cannot be met for smaller distances.

We also estimate the secret key rate of Alice and Bob versus distance [Fig. 6(b)]. A set of parameters Δ and g can always be found to meet success criterion II as long as the distance is large enough (larger than 31 km with our simulation

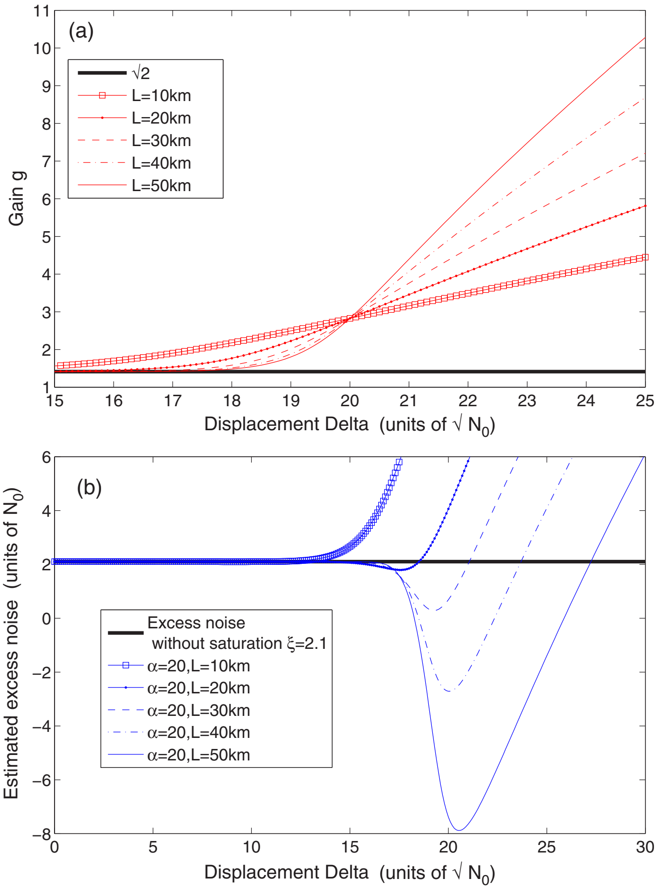

FIG. 5. Simulations under attack strategy II, step (a). See Sec. VII C 1 for simulation parameters. (a) Gain g verifying Eq. (19) as a function of Δ. (b) Estimate excess noise as a function of Δ for different link lengths.

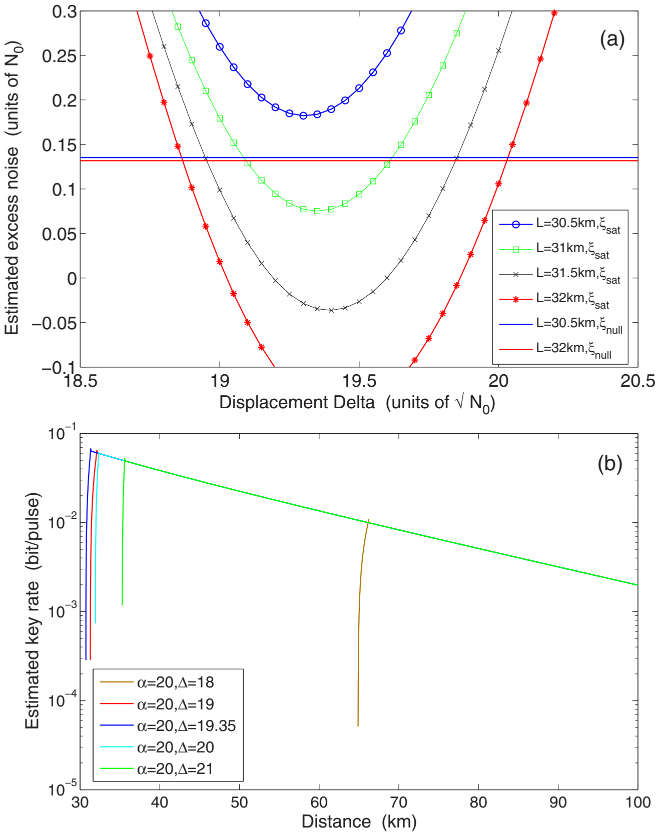

FIG. 6. Simulations under saturation attack strategy II, step (b). See Sec. VII C 1 for the simulation parameters. (a) Excess noise versus Δ and distance. Solid lines with symbols show the estimate excess noise $\hat{\xi}_{\text{sat}}$ and dashed lines the null key threshold ξ_{null}. (b) Key rate versus distance for different values of Δ. No attack is possible for links shorter than 31 km under criteria II (see the text).

parameters, as detailed in Sec. VII C 1). Since the estimated channel transmission T is unchanged, the estimated key rate will be identical to the key rate in the absence of an attack. Hence, reaching strategy II, although it cannot be launched on short channels (high transmission), is a more powerful and more convincing strategy.

VIII. COUNTERMEASURES AGAINST THE SATURATION ATTACK

The vulnerability to the saturation attack, studied in previous sections, is related to the fact that the first moment (mean value) of the measured quadratures are by default not monitored in a CVQKD protocol and can therefore be freely manipulated by an attacker, opening a practical security loophole. The essence of a countermeasure against the saturation attack will therefore consist in adding some test procedure to the CVQKD protocol in order to rule out the possibility that the detector saturates, i.e., that some input optical state has a quadrature X_{in} larger than α, characterizing the linear range of the detection. We present what could be this test procedure, also called countermeasures against saturation. They range from postselection tests, which can be implemented without any modification of CVQKD hardware, to more structural modifications of the protocols, requiring extra hardware. Importantly, we first recall that most countermeasures rely on some preexisting calibration of the detector.

A. Prerequisite: Calibration and characterization of the homodyne detection linear range

The scope of this article is restricted to prepare and measure CVQKD, where the detector can be considered trusted, i.e., not influenced by the eavesdropper. In this context, Alice and Bob, the legitimate users, can rely on a (trusted) calibration of the detector and in particular on a characterization of the detector linear range. The homodyne detection is a phase-sensitive device that transforms an input optical state of quadrature X_{in}, measured with respect to the phase reference (local oscillator), into a measured voltage V_{out}. For an unsaturated homodyne detection, the relation between X_{in} and V_{out} is linear, with a linear gain that depends on several parameters such as the local oscillator amplitude, the optical loss, the mode matching, and the electronic gain of the electronics (a transimpedance circuit is commonly used to perform low-noise measurement of the small differential photocurrent associated with X_{in}). All these parameters are not easy to measure independently, but we can calibrate them globally by measuring (offline, for example before launching the CVQKD protocol) the variance of the homodyne output voltage V_{out} when the input is in a vacuum, possibly for different values of the local oscillator power. This corresponds to measuring the shot-noise variance (in voltage) $N_{0,V}$. Quadrature measurements X_{out} are then usually expressed as the square root of shot-noise units, which means that they are renormalized, based on shot-noise calibration, $X_{\text{out}} \equiv V_{\text{out}}/\sqrt{N_{0,V}}$, where X_{out}, expressed as the square root of shot-noise units, corresponds to the quadrature measurement we want to perform with our homodyne detection. If the input optical state is also expressed as the square root of shot-noise units then we have, after calibration, $X_{\text{in}} = X_{\text{out}}$, provided the detection is linear.

However, as illustrated in Fig. 2, a realistic detector saturates and in practice the linearity between X_{in} and X_{out} can only be guaranteed over a finite range. We have called α the linearity bound such that $\forall |X_{\text{in}}| < \alpha$, $X_{\text{out}} = X_{\text{in}}$. Here α is a characteristic of the homodyne detector for a given local oscillator power. In practice, the local oscillator power should be chosen not too large such that the saturation limit is much larger than the shot noise, i.e., if α is expressed in shot-noise units, we want to have $\alpha^2 \gg 1$. Inversely, the local oscillator power should be chosen not too small so that the variance of the electronic noise, in shot-noise units, is much smaller that the shot-noise variance $\upsilon_{\text{elec}} \ll 1$.

B. From an intuitive but faulty countermeasure to an efficient countermeasure based on radical postselection

A simple (but faulty) countermeasure would consist in postselecting quadrature measurement results provided they fall in a confidence interval where the homodyne detection is known to be linear, i.e., typically if they fall within an interval of the form $[-(\alpha - \beta),\alpha - \beta]$. Here β can be seen as a confidence margin, with $0 < \beta < \alpha$. We can however see that such a countermeasure would trigger new problems: It would give Eve the possibility to influence which data are postselected and which are not, just by controlling the displacement value. The postselected data would not be Gaussian and no security could be guaranteed. This observation has motivated the development of the countermeasure based on Gaussian postselection that we will detail at the end of this section.

A more radical countermeasure is however possible, which consists in discarding measurement data blocks if *any* of the measured data fall out of the confidence interval $[-(\alpha - \beta),\alpha - \beta]$. Quadrature measurements and parameter estimation are in practice realized over blocks of large size in order to limit finite-size effects (for example, 10^8 in [28]). If the detector has been properly calibrated and if its characteristics are stable over time, then a countermeasure based on what we will call radical postselection will guarantee by construction that, provided it passes the test, a data block has been acquired by a detector operating in a linear regime. The drawback of this radical postselection procedure is that it might lead to the removal of some or possibly almost all data blocks if the confidence interval is not large enough (and/or not centered) with respect to the impinging optical quadrature variance. This countermeasure will efficiently protect against saturation attacks and is easily implementable. Provided the distribution of the input quadrature X_{in} is centered, the confidence interval should typically be larger than six standard deviations, i.e., $\alpha - \beta > 6\sqrt{\text{Var}(X_{\text{in}})}$, so the probability (per measurement event) to have a saturation is below 2×10^{-9} and hence the probability to discard "good" data blocks remains relatively small. On the other hand, if the linearity domain of the homodyne detection is not large enough, then this countermeasure might strongly affect the effective key rate, even in the absence of any attack, which has motivated us to propose a more refined countermeasure.

C. Gaussian postselection

We now propose a refined version of the radical postselection. This method retains the important advantage of being implementable at the software level by modifying the postprocessing stage. It can moreover cope with detector saturation in a more gentle way than throwing away the entire data block as soon as a saturated measurement is detected, as is the case with radical postselection.

The method is based on performing a Gaussian postselection of the measurement data. The key idea of such a method is to extract a set of (almost) Gaussian-distributed data among the raw measurement data and to adjust the parameters of the postselection so that postselected data fall almost certainly within the (calibrated) linearity domain of the homodyne detector. Calling $g(x)$ the probability distribution of data points after postselection, the goal of the postselection procedure is to choose the parameters of the non-normalized Gaussian filter $g(x)$ (mean value μ_g and variance σ_g^2) under the following constraints:

(i) Here g is a postselection function, which implies that there are fewer points after than before postselection: $0 \leqslant g(x) \leqslant f(x)\, \forall x \in [-\alpha,\alpha]$.

(ii) Postselected data should be almost Gaussian, i.e., the support of $g(x)$ should be almost contained in the linearity domain: $\int_{-\alpha}^{\alpha} g(x)dx \simeq \int_{-\infty}^{\infty} g(x)dx$.

(iii) The number of postselected points $N' \equiv \int_{-\alpha}^{\alpha} g(x)dx$ should be maximized under the two previous constraints.

Performing the Gaussian postselection first consists in binning the measured data (size N). Calling $f(x)$ the normalized distribution distribution of the raw data (quadrature measurement data) and considering bins centered on measurement result x and of width δx, there are approximately $Nf(x)\delta x$ raw data points falling within a bin. The Gaussian postselection consists in randomly selecting a fraction $g(x)/f(x)$ of those points falling within the bin and throwing away the others. After this procedure, applied to each bin, the postselected data will have a probability distribution close to $g(x)$ (provided we have large enough data blocks and use small enough bins so that finite-size effects remain small).

In order to illustrate how this Gaussian postselection method could be realized in practice, we have simulated a CVQKD experiment affected by saturation. The results are displayed in Fig. 7. The total number of points is $N = 10^6$. We have assumed a lossy channel between A and B, a displacement Δ at B, and a homodyne detection affected by saturation, as described in Sec. V A. We have used the following numerical values: channel distance 25 km, Alice variance $V_A = 11.58N_0$, and displacement $\Delta = 19.2\sqrt{N_0}$. The blue dots correspond to measurement results with a perfect homodyne detection (no saturation) while the red dots correspond to the results for the realistic detection, affected by saturation, with a linearity limit characterized by $\alpha = 20\sqrt{N_0}$. The green dots correspond to the Gaussian postselected data, applying the procedure detailed above. In order to find the parameter of the Gaussian filter $g(x)$ we first removed the data points falling outside the linearity domain $[-\alpha,\alpha]$ and then built the histogram of X_{B_i} with bin size $\delta x = 0.1\sqrt{N_0}$. We could then estimate the probability distribution $f(x)$. The essential remaining step was to choose the parameters (variance, mean value, and amplitude A) of the Gaussian filter

$$g(x) \equiv \frac{A}{\sqrt{2\pi}\sigma_g} \exp\left(-\frac{(x-\mu_g)^2}{2\sigma_g^2}\right). \tag{20}$$

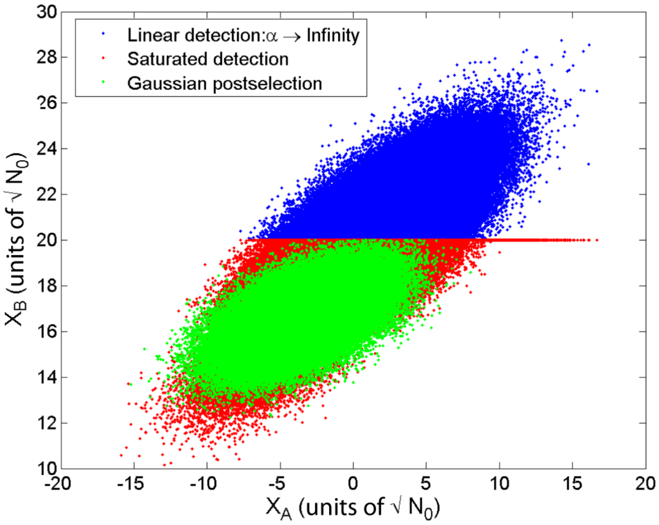

FIG. 7. Simulation of Gaussian postselection. Blue dots show the simulated experimental data with an infinite linear detection limit, $\Delta = 19.2\sqrt{N_0}$, $L = 20$ km, $V_A = 11.58N_0$, and data number $N = 10^6$. Red dots show the simulated experimental data with a linear detection limit $\alpha = 20\sqrt{N_0}$; other parameters are the same as the blue ones. Green dots show the Gaussian postselected data among the red dots, with the Gaussian postselection parameter, $\sigma_g^2 = 2.5N_0$, $\mu_g = 16.55N_0$, and the number of postselected points $N' = 15.37\%N$. For other simulation parameters see Sec. VII C 1.

We have optimized numerically these parameters to maximize the number of postselected points N' under constraints (i) and (ii). We have obtained a total number of postselected points $N' = 15.37\%N$ (green dots in Fig. 7), while guaranteeing that the $L2$ distance between the normalized postselected data distribution and a perfect Gaussian distribution is below 10^{-3}.

The Gaussian postselection is more complex to implement than the radical postselection, but has the advantage of allowing one to generate some key, even in the presence of moderate saturation. The postselection also guarantees that postselected points fall within the linearity domain of the detector (therefore guaranteeing linearity of the detector on these postselected data points). Moreover, the Gaussian postselection also guarantees that the distribution of the input data, after postselection, is Gaussian and thus that the postselected data still implement the GG02 protocol, although with different channel parameters T' and ξ'. As a consequence, provided T' and ξ' are compatible with secure key generation, some key can be distilled from the postselected data, free from any security threat and attacks exploiting detector saturation.

D. Countermeasures relying on existing techniques, necessitating some additional hardware

The two (radical and Gaussian) postselection measures discussed above can be implemented without any additional hardware, which constitutes an important practical advantage. For completeness we also discuss here other ways to counterattack related to saturation, which, however, all involve some changes not only of the protocol, but also of the hardware.

As proposed and implemented in [36], varying randomly the attenuation, on the signal port, at the input of Bob, allows one to test the linearity of the homodyne output with respect to the signal input. The saturation attack exploits in particular the fact that the shot-noise variance is calibrated in the absence of a signal (attenuation $\eta = 0$) while the total noise variance is measured with no attenuation ($\eta = 1$) and with saturation (when the attack is launched). As a consequence, the excess noise is underestimated. Such an attack cannot however be performed if the total noise variance is estimated with more than two attenuation values and in particular is the linearity of the total noise with the attenuation checked. This is precisely the countermeasure proposed in [36], which constitutes an effective parade against the saturation attack. The drawback of this countermeasure is, however, that it requires another hardware element, causing additional complexity, and also that it leads to attenuation of the signal and thus a reduction of the key rate.

Measurement-device-independent (MDI) CVQKD can be used to perform QKD with untrusted detectors [37]. It could therefore be used in particular to perform CVQKD securely with practical homodyne detectors, subject of saturation. This would, however, be at the expense of a significant increase of the experimental complexity (in particular phase locking of two distant lasers) and also at the expense of performance since only low to moderate losses can be tolerated in MDI CVQKD [38].

Finally, one could notice that the saturation attack requires one to strongly displace the mean value of the quadrature signal. The signals impinging on the homodyne detection must therefore have a high energy. The method proposed in [39], in another context, could then be used as a countermeasure: It consists in upper bounding, with an auxiliary homodyne detector (sometimes called watchdog in other contexts [4]), the energy of the impinging signals. One limitation of this method is, however, that it is in general not possible to predict in which mode an attacker will try to send energy in order to saturate the detector. It can thus be very difficult to design in practice a watchdog capable of detecting any attempts to saturate the detector.

IX. CONCLUSION

We have studied quantitatively how the saturation attack can be used to compromise the security of practical CVQKD systems. The main finding of our study is that the excess noise can be actively reduced by displacing the quadrature's mean value of the coherent states received by Bob and that this effect can compromise the security of the Gaussian-modulated coherent state CVQKD protocol, operated with practical detectors whose linearity response can only be guaranteed over a restricted domain of quadrature values.

We have proposed an explicit attack, called a saturation attack, that combines displacement with a full intercept-resend attack. We have performed numerical simulations that show the feasibility of our attack under realistic experimental conditions. The saturation attack consists in strongly displacing the quadrature mean values to induce saturation of Bob's detector. Our attack is achievable with current technology and may impact the security of all CVQKD implementations since any practical detectors is subject to saturation. An experimental demonstration of this attack is the topic left for future work.

While all previous attacks on CVQKD had focused on local oscillator manipulation and biasing excess noise evaluation, our attack has no influence on the local oscillator and thus cannot be ruled out by generating the local oscillator locally [29–31]. It is therefore important to propose practical solutions against this attack and we have presented in detail two effective countermeasures based on postselection that can be implemented without requiring any modification at the hardware level. This work illustrates the importance of putting under great scrutiny the hypothesis under which the security proof can be derived, but also illustrates that secure and yet still practical QKD implementations are within reach.

ACKNOWLEDGMENTS

R.A. thanks Frédéric Grosshans for stimulating discussions about countermeasures. This research was supported by the French National Research Agency, through the FREQUENCY (Grant No. ANR-09-BLAN-0410), and by the European Union through the Project Q-CERT (Grant No. FP7-PEOPLE-2009-IAPP) and the ANR Project Quantum-WDM (Grant No. ANR-12-EMMA-0034). H.Q. acknowledges support from the Institut Mines-Télécom, Paris, France.

APPENDIX A: CALCULATION OF THE CORRELATION UNDER THE SATURATION ATTACK

In order to clearly show the calculation, we consider y_{sat}, y, x, and z as the notation $X_{B_{\text{sat}}}$, $X_{B_{\text{lin}}}$, X_A, and X_N respectively. We use $X_{B_{\text{sat}}}$ [Eq. (8)] to calculate the correlation $\text{Cov}(X_A,X_{B_{\text{sat}}})$ under the saturation attack. We assume here $\alpha \gg 1$ and consider $\Delta \geqslant 0$, while the analysis of $\Delta \leqslant 0$ is similar. The saturation model can be considered as

$$\begin{aligned}
&y_{\text{sat}} = \alpha, \quad t\frac{g}{\sqrt{2}}x + z + \Delta \geqslant \alpha,\\
&y_{\text{sat}} = t\frac{g}{\sqrt{2}}x + z + \Delta, \quad \left|t\frac{g}{\sqrt{2}}x + z + \Delta\right|\\
&\quad < \alpha \quad (\alpha \gg 1,\ \Delta \geqslant 0),\\
&y_{\text{sat}} = -\alpha, \quad t\frac{g}{\sqrt{2}}x + z + \Delta \leqslant -\alpha,
\end{aligned} \tag{A1}$$

where $x \sim \mathcal{N}(0,\sigma_x^2)$ and $z \sim \mathcal{N}(0,\sigma_z^2)$ are both centered Gaussian variables with probability density functions $p_X(x)$ and $p_Z(z)$, respectively,

$$p_X(x) = \frac{e^{-x^2/2\sigma_x^2}}{\sqrt{2\pi}\sigma_x}, \quad p_Z(z) = \frac{e^{-z^2/2\sigma_z^2}}{\sqrt{2\pi}\sigma_z}, \tag{A2}$$

in which $\sigma_x^2 = \mathrm{Var}(X_A)$ and $\sigma_z^2 = N_0 + \eta T \xi + v_{\mathrm{ele}}$. By knowing $p_X(x)$ and $p_Z(z)$, we can calculate $\mathrm{Cov}(x, y_{\mathrm{sat}})$ with a double integral of x and z in the domain D_{xz}. The domain D_{xz} is defined in Eq. (A1): $-\alpha < \frac{tg}{\sqrt{2}}x + z + \Delta < \alpha$, $\frac{tg}{\sqrt{2}}x + z + \Delta \leqslant -\alpha$, and $\frac{tg}{\sqrt{2}}x + z + \Delta \geqslant \alpha$. A long but straightforward calculation of $\mathrm{Cov}(x, y_{\mathrm{sat}})$ is presented as follows:

$$\begin{aligned}\mathrm{Cov}(X_A, X_{B_{\mathrm{sat}}}) &= \langle x y_{\mathrm{sat}}\rangle - \langle x\rangle\langle y_{\mathrm{sat}}\rangle = \langle x y_{\mathrm{sat}}\rangle = \iint_{D_{xz}} x y p_X(x) p_Z(z) dx\, dz \\ &= \iint_{-\alpha < (tg/\sqrt{2})x + z + \Delta < \alpha} \left(\frac{tg}{\sqrt{2}}x^2 + x\Delta + xz\right) p_X(x) p_Z(z) dx\, dz \\ &\quad + \iint_{(tg/\sqrt{2})x + z + \Delta \leqslant -\alpha} -\alpha x p_X(x) p_Z(z) dx\, dz + \iint_{(tg/\sqrt{2})x + z + \Delta \geqslant \alpha} \alpha x p_X(x) p_Z(z) dx\, dz, \end{aligned} \tag{A3}$$

$$\begin{aligned}\mathrm{Cov}(X_A, X_{B_{\mathrm{sat}}}) &= \frac{1}{2\pi\sigma_x\sigma_z}\int_{-\infty}^{\infty}\left(\frac{tg}{\sqrt{2}}x^2 + x\Delta\right)e^{-x^2/\sigma_x^2}dx\int_{-\alpha-\Delta-(tg/\sqrt{2})x}^{\alpha-\Delta-(tg/\sqrt{2})x} e^{-z^2/2\sigma_z^2}dz \\ &= \frac{1}{2\pi\sigma_x\sigma_z}\int_{-\infty}^{\infty}\left(\frac{tg}{\sqrt{2}}x^2 + x\Delta\right)e^{-x^2/\sigma_x^2}\sqrt{\frac{\pi}{2}}\sigma_z\left[\mathrm{erf}\left(\frac{\alpha + \Delta + \frac{tg}{\sqrt{2}}x}{\sqrt{2}\sigma_z}\right) + \mathrm{erf}\left(\frac{\alpha - \Delta - \frac{tg}{\sqrt{2}}x}{\sqrt{2}\sigma_z}\right)\right]dx \\ &= \frac{1}{2\pi\sigma_x}\sqrt{\frac{\pi}{2}}\left[\frac{tg}{\sqrt{2}}\int_{-\infty}^{\infty} x^2 e^{-x^2/2\sigma_x^2}dx + \Delta\int_{-\infty}^{\infty}\mathrm{erf}\left(\frac{\alpha - \Delta - \frac{tg}{\sqrt{2}}x}{\sqrt{2}\sigma_z}\right)x e^{-x^2/2\sigma_x}dx\right] \\ &= \frac{tg}{2\sqrt{2}\pi\sigma_x}\sqrt{\frac{\pi}{2}}\sqrt{2\pi}\sigma_x^3 + \frac{tg}{2\sqrt{2}\pi\sigma_x}\sqrt{\frac{\pi}{2}}\sqrt{2\pi}\sigma_x^3\mathrm{erf}\left(\frac{\alpha - \Delta}{\sqrt{2(\sigma_z^2 + \frac{t^2g^2}{2}\sigma_x^2)}}\right) \\ &= \frac{tg}{2\sqrt{2}}\sigma_x^2\left[1 + \mathrm{erf}(\frac{\alpha - \Delta}{\sqrt{2(\sigma_z^2 + \frac{t^2g^2}{2}\sigma_x^2)}})\right]. \end{aligned} \tag{A4}$$

Thus we can conclude that

$$\mathrm{Cov}(X_A, X_{B_{\mathrm{sat}}}) = \frac{tg}{2\sqrt{2}}\langle X_A^2\rangle\left[1 + \mathrm{erf}\left(\frac{\alpha - \Delta}{\sqrt{2\,\mathrm{Var}(X_{B_{\mathrm{lin}}})}}\right)\right], \tag{A5}$$

in which the error function $\mathrm{erf}(x)$ is defined as

$$\mathrm{erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt \tag{A6}$$

and we have used the integral formulas of $\mathrm{erf}(x)$ provided in [40]. In Eq. (A4), $\mathrm{Var}(X_{B_{\mathrm{lin}}}) = \sigma_z^2 + \frac{t^2g^2}{2}\sigma_x^2$ is the variance of Bob with no saturation. In this calculation, the integrals of the odd functions with symmetric bounds $(-\infty, \infty)$ are equal to zero.

APPENDIX B: CALCULATION OF THE VARIANCE OF BOB UNDER THE SATURATION ATTACK

In order to calculate the variance of Bob under the saturation attack, we use the step function $\theta(x)$, which is defined as

$$\theta(x) = \begin{cases} 1, & x \in [0, \infty) \\ 0, & x \in (-\infty, 0]. \end{cases} \tag{B1}$$

With Eq. (B1) we can transform Eq. (A1) into

$$\begin{aligned} y_{\mathrm{sat}} &= y\theta(y + \Delta + \alpha)\theta(-y - \Delta + \alpha) + \alpha[1 - \theta(y + \Delta + \alpha)\theta(-y - \Delta + \alpha)] \\ &\approx \alpha + (y + \Delta - \alpha)\theta(-y - \Delta + \alpha) = \alpha + (y - \varepsilon)\theta(-y + \varepsilon), \end{aligned} \tag{B2}$$

in which

$$\varepsilon = \alpha - \Delta \quad (\alpha > 0,\ \Delta \geqslant 0), \tag{B3}$$

$$y = t\frac{g}{\sqrt{2}}x + z. \tag{B4}$$

Since x and z are both Gaussian variables, y is also a Gaussian variable [$y \sim \mathcal{N}(0, \sigma_y^2)$], with its probability function $p_Y(y) = \frac{e^{-y^2/2\sigma_y^2}}{\sqrt{2\pi}\sigma_y}$ and $\sigma_y^2 = \mathrm{Var}(X_{B_{\mathrm{lin}}})$ is the variance of Bob under linear detection. In order to estimate $\mathrm{Var}(X_{B_{\mathrm{sat}}}) = \mathrm{Var}(y_{\mathrm{sat}}) = \langle y_{\mathrm{sat}}^2\rangle -$

$\langle y_{\text{sat}}\rangle^2$, we need to calculate $\langle y_{\text{sat}}\rangle$ and $\langle y_{\text{sat}}^2\rangle$, respectively,

$$\langle y_{\text{sat}}\rangle = \alpha + \langle (y-\varepsilon)\theta(-y+\varepsilon)\rangle = \alpha + C, \tag{B5}$$

$$\langle y_{\text{sat}}^2\rangle = \langle \alpha^2 + 2\alpha(y-\varepsilon)\theta(-y+\varepsilon) + (y-\varepsilon)^2\theta(-y+\varepsilon)\rangle \tag{B6}$$

$$= \alpha^2 - 2\alpha C + D, \tag{B7}$$

in which C and D are equal to $\langle (y-\varepsilon)\theta(-y+\varepsilon)\rangle$ and $\langle (y-\varepsilon)^2\theta(-y+\varepsilon)\rangle$ and can be calculated as follows:

$$C = \int_{-\infty}^{\infty} p_Y(y)(y-\varepsilon)\theta(-y+\varepsilon)dy \tag{B8}$$

$$= \int_{-\infty}^{\infty} p_Y(y'+\varepsilon)y'\theta(-y')dy' = \int_{-\infty}^{0} p_Y(y'+\varepsilon)y'dy' \tag{B9}$$

$$= -\left[\frac{\sigma_y}{\sqrt{2\pi}}e^{-\varepsilon^2/2\sigma_y^2} + \frac{\varepsilon}{2} + \frac{\varepsilon}{2}\text{erf}\left(\frac{\varepsilon}{\sqrt{2}\sigma_y}\right)\right], \tag{B10}$$

$$D = \langle (y-\varepsilon)^2\theta(-y+\varepsilon)\rangle = \int_{-\infty}^{\infty} p_Y(y)(y-\varepsilon)^2\theta(-y+\varepsilon)dy \tag{B11}$$

$$= \int_{-\infty}^{\infty} p_Y(y'+\varepsilon)y'^2\theta(-y')dy' = \int_{-\infty}^{0} p_Y(y'+\varepsilon)y'^2dy' \tag{B12}$$

$$= \frac{\varepsilon\sigma_y}{\sqrt{2\pi}}e^{-\varepsilon^2/2\sigma_y^2} + \frac{\varepsilon^2+\sigma_y^2}{2}\left[1 + \text{erf}\left(\frac{\varepsilon}{\sqrt{2}\sigma_y}\right)\right]. \tag{B13}$$

We have used $y' = y - \varepsilon$ in the calculations of C and D. Provided with C and D, we can calculate $\text{Var}(y_{\text{sat}})$:

$$\begin{aligned}\text{Var}(y_{\text{sat}}) &= \left\langle y_{\text{sat}}^2\right\rangle - \langle y_{\text{sat}}\rangle^2\\ &= \alpha^2 - 2\alpha C + D - (\alpha + C)^2 = D - C^2\\ &= \sigma_y^2\left[\frac{1+\text{erf}(\frac{\varepsilon}{\sqrt{2}\sigma_y})}{2} - \frac{e^{-\varepsilon^2/\sigma_y^2}}{2\pi}\right] - \frac{\varepsilon\sigma_y}{\sqrt{2\pi}}\text{erf}\left(\frac{\varepsilon}{\sqrt{2}\sigma_y}\right)e^{-\varepsilon^2/2\sigma_y^2} + \frac{\varepsilon^2}{4}\left[1 - \text{erf}^2\left(\frac{\varepsilon}{\sqrt{2}\sigma_y}\right)\right]\\ &= \text{Var}(X_{B_{\text{lin}}})\left(\frac{1+A}{2} - \frac{B^2}{2\pi}\right) - (\alpha - \Delta)\sqrt{\frac{\text{Var}(X_{B_{\text{lin}}})}{2\pi}}A * B + \frac{(\alpha-\Delta)^2}{4}(1 - A^2),\end{aligned} \tag{B14}$$

in which

$$A = \text{erf}\left(\frac{\alpha - \Delta}{\sqrt{2\,\text{Var}(X_{B_{\text{lin}}})}}\right), \quad B = e^{-(\alpha-\Delta)^2/2\,\text{Var}(X_{B_{\text{lin}}})}. \tag{B15}$$

[1] V. Scarani, H. Bechmann-Pasquinucci, N. J. Cerf, M. Dušek, N. Lutkenhaus, and M. Peev, Rev. Mod. Phys. **81**, 1301 (2009).

[2] F.-H. Xu, B. Qi, and H.-K. Lo, New J. Phys. **12**, 113026 (2010).

[3] L. Lydersen, C. Wiechers, C. Wittmann, D. Elser, J. Skaar, and V. Makarov, Nat. Photon. **4**, 686 (2010).

[4] I. Gerhardt, Q. Liu, A. Lamas-Linares, J. Skaar, C. Kurtsiefer, and V. Makarov, Nat. Commun. **2**, 349 (2011).

[5] B. Qi, C.-H. F. Fung, H.-K. Lo, and X. Ma, Quantum Inf. Comput. **7**, 7382 (2007).

[6] Y. Zhao, C.-H. F. Fung, B. Qi, C. Chen, and H.-K. Lo, Phys. Rev. A. **78**, 042333 (2008).

[7] C. Wiechers, L. Lydersen, C. Wittmann, D. Elser, J. Skaar, C. Marquardt, V. Makarov, and G. Leuchs, New J. Phys. **13**, 013043 (2011).

[8] H. Weier, H. Krauss, M. Rau, M. Füerst, S. Nauerth, and H. Weinfurter, New J. Phys. **13**, 073024 (2011).

[9] V. Makarov, New J. Phys. **11**, 065003 (2009)

[10] S. Kleis and C. G. Schaeffer, *Proceedings of the ITG Symposium on Photonic Networks, Leipzig*, 2014 (VDE, Berlin, 2014), pp. 1–5.

[11] R. Kumar, H. Qin, and R. Alléaume, New J. Phys. **17**, 043027 (2015).

[12] F. Grosshans, G. van Assche, J. Wenger, R. Brouri, N. J. Cerf, and P. Grangier, Nature (London) **421**, 238 (2003).

[13] A. Leverrier, R. García-Patrón, R. Renner, and N. J. Cerf, Phys. Rev. Lett. **110**, 030502 (2013).

[14] R. Renner and J. I. Cirac, Phys. Rev. Lett. **102**, 110504 (2009).

[15] H. Häseler, T. Moroder, and N. Lütkenhaus, Phys. Rev. A **77**, 032303 (2008).

[16] J.-Z. Huang, C. Weedbrook, Z.-Q. Yin, S. Wang, H.-W. Li, W. Chen, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **87**, 062329 (2013).

[17] X.-C. Ma, S.-H. Sun, M.-S. Jiang, and L.-M. Liang, Phys. Rev. A **87**, 052309 (2013).

[18] J.-Z. Huang, S. Kunz-Jacques, P. Jouguet, C. Weedbrook, Z.-Q. Yin, S. Wang, W. Chen, G.-C. Guo, and Z.-F. Han, Phys. Rev. A **89**, 032304 (2014).

[19] X.-C. Ma, S.-H. Sun, M.-S. Jiang, and L.-M. Liang, Phys. Rev. A **88**, 022339 (2013).

[20] X.-C. Ma, S.-H. Sun, M.-S. Jiang, M. Gui, Y.-L. Zhou, and L.-M. Liang, Phys. Rev. A **89**, 032310 (2014).

[21] J. Lodewyck, T. Debuisschert, R. Garcia-Patron, R. Tualle-Brouri, N. J. Cerf, and P. Grangier, Phys. Rev. Lett. **98**, 030503 (2007).

[22] P. Jouguet, S. Kunz-Jacques, and A. Leverrier, Phys. Rev. A **84**, 062317 (2011).

[23] A. Leverrier, F. Grosshans, and P. Grangier, Phys. Rev. A **81**, 062343 (2010).

[24] R. García-Patrón and N. J. Cerf, Phys. Rev. Lett. **97**, 190503 (2006).

[25] M. Navascués, F. Grosshans, and A. Acín, Phys. Rev. Lett. **97**, 190502 (2006).

[26] P. Jouguet, S. Kunz-Jacques, E. Diamanti, and A. Leverrier, Phys. Rev. A **86**, 032309 (2012).

[27] J. Lodewyck, M. Bloch, R. García-Patrón, S. Fossier, E. Karpov, E. Diamanti, T. Debuisschert, N. J. Cerf, R. Tualle-Brouri, S. W. McLaughlin, and P. Grangier, Phys. Rev. A **76**, 042305 (2007).

[28] P. Jouguet, S. Kunz-Jacques, A. Leverrier, P. Grangier, and E. Diamanti, Nat. Photon. **7**, 378 (2013).

[29] B. Qi, P. Lougovski, R. Pooser, W. Grice, and M. Bobrek, Phys. Rev. X **5**, 041009 (2015).

[30] D. B. S. Soh, C. Brif, P. J. Coles, N. Lütkenhaus, R. M. Camacho, J. Urayama, and M. Sarovar, Phys. Rev. X **5**, 041010 (2015).

[31] D. Huang, P. Huang, D. Lin, C. Wang, and G. Zeng, Opt. Lett. **40**, 3695 (2015).

[32] P. Jouguet, S. Kunz-Jacques, and E. Diamanti, Phys. Rev. A **87**, 062313 (2013).

[33] Y. Chi, B. Qi, W. Zhu, L. Qian, H.-K. Lo, S.-H. Youn, A. I. Lvovsky, and L. Tian, New J. Phys. **13**, 013003 (2011).

[34] M. G. Paris, Phys. Lett. A **217**, 78 (1996).

[35] R. Kumar, H. Qin, and R. Alléaume (unpublished).

[36] S. Kunz-Jacques and P. Jouguet, Phys. Rev. A **91**, 022307 (2015).

[37] S. Pirandola, C. Ottaviani, G. Spedalieri, C. Weedbrook, S. L. Braunstein, S. Lloyd, T. Gehring, C. S. Jacobsen, and U. L. Andersen, Nat. Photon. **9**, 397 (2015).

[38] C. Ottaviani, G. Spedalieri, S. L. Braunstein, and S. Pirandola, Phys. Rev. A **91**, 022320 (2015).

[39] F. Furrer, Phys. Rev. A **90**, 042325 (2014).

[40] E. W. Ng and M. Geller, J. Res. Natl. Bur. Stand. Sect. B **73**, 1 (1969).