Device Calibration Impacts Security of Quantum Key Distribution

Nitin Jain,[1,2,*] Christoffer Wittmann,[1,2] Lars Lydersen,[3,4] Carlos Wiechers,[1,2,5] Dominique Elser,[1,2]
Christoph Marquardt,[1,2] Vadim Makarov,[3,4] and Gerd Leuchs[1,2]

[1]*Max Planck Institute for the Science of Light, Günther-Scharowsky-Straße 1, Bau 24, 91058 Erlangen, Germany*
[2]*Institut für Optik, Information und Photonik, University of Erlangen-Nuremberg, Staudtstraße 7/B2, 91058 Erlangen, Germany*
[3]*Department of Electronics and Telecommunications, Norwegian University of Science and Technology, NO-7491 Trondheim, Norway*
[4]*University Graduate Center, NO-2027 Kjeller, Norway*
[5]*Departamento de Física, Campus León, Universidad de Guanajuato, Lomas del Bosque 103, Fraccionamiento Lomas del Campestre, 37150, León, Gto, México*

(Received 24 March 2011; published 9 September 2011)

Characterizing the physical channel and calibrating the cryptosystem hardware are prerequisites for establishing a quantum channel for quantum key distribution (QKD). Moreover, an inappropriately implemented calibration routine can open a fatal security loophole. We propose and experimentally demonstrate a method to induce a large temporal detector efficiency mismatch in a commercial QKD system by deceiving a channel length calibration routine. We then devise an optimal and realistic strategy using faked states to break the security of the cryptosystem. A fix for this loophole is also suggested.

DOI: 10.1103/PhysRevLett.107.110501 PACS numbers: 03.67.Dd, 03.67.Ac, 03.67.Hk, 42.50.Ex

Quantum key distribution (QKD) offers unconditionally secure communication as eavesdropping disturbs the transmitted quantum states, which in principle leads to the discovery of the eavesdropper Eve [1]. However, practical QKD implementations may suffer from technological and protocol-operational imperfections that Eve could exploit in order to remain concealed [2,3].

Until now, a variety of eavesdropping strategies have utilized differences between the theoretical model and the practical implementation, arising from (technical) imperfections or deficiencies of the components. Ranging from photon number splitting [4] and Trojan horse [5], to leakage of information in a side channel [6], time shifting [7], and phase remapping [8], several attacks have been proposed and experimentally demonstrated. Recently, proof-of-principle attacks [9–11] based on the concept of faked states [12] have been presented. Eve targets imperfections of avalanche photodiode (APD) based single-photon detectors [13] that allow her to control them remotely.

Another important aspect of QKD security not yet investigated, however, is the calibration of the devices. A QKD protocol requires a classical and a quantum channel; while the former must be authenticated, the latter is merely required to preserve certain properties of the quantum signals [2,14]. The establishment of the quantum channel remains an implicit assumption in security proofs: channel characterization (e.g., channel length) and calibration of the cryptosystem hardware, especially the steps involving two-party communication, have not yet been taken into account. As we show, the calibration of the QKD devices must be carefully implemented, otherwise it is prone to hacks that may strengthen existing (or create new) eavesdropping opportunities for Eve.

In this Letter, we propose and experimentally demonstrate the hacking of a vital calibration sequence during the establishment of the quantum channel in the commercial QKD system model Clavis2 from ID Quantique [15]. Eve induces a parameter mismatch [16] between the detectors that can break the security of the QKD system. Specifically, she causes a temporal separation of the order of 450 ps of the detection efficiencies by deceiving the detection system, shown in Fig. 1. This allows her to control Bob's detection outcomes using time, a parameter already shown to be instrumental in applying a time-shift attack [7]. Alternatively, she could launch a faked-state attack (FSA) [16] for which we calculate the quantum bit error rate (QBER) under realistic conditions. Since FSA is an intercept-resend attack, Eve has full

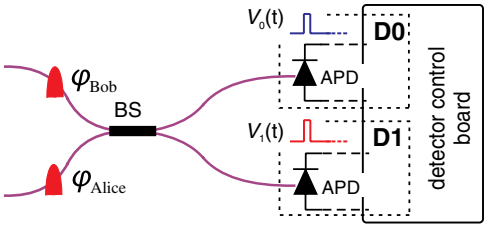

FIG. 1 (color online). Typical detection system in a Mach-Zehnder interferometer based QKD implementation. The bit and basis choices of Alice and Bob (phases φ_{Alice} and φ_{Bob}) determine the interference result at the 50:50 beam splitter (BS), deciding in turn which of the two detectors $D0$ or $D1$ would click. It is thus crucial that $D0$ and $D1$ are indistinguishable to the outside world (i.e., Eve). If gated mode APDs are employed, the detector control board ensures that the activation of $D0$ and $D1$ [via voltage pulses $V_0(t)$ and $V_1(t)$] happens almost simultaneously, to nullify any existing temporal efficiency mismatch.

information-theoretic knowledge about the key as long as Alice and Bob accept the QBER at the given channel transmission T and do not abort key generation [17]. Constricting our FSA to match the raw key rate expected by Bob and Alice, i.e., maintaining T at nearly the exact preattack level, we find that the security of the system is fully compromised. Our hack has wide implications: most practical QKD schemes based on gated APDs, in both plug-and-play and one-way configurations [18–20], need to perform channel characterization and hardware calibration regularly. A careful implementation of these steps is required to avoid leaving inadvertent back doors for Eve.

The optical setup of Clavis2 is based on the plug-and-play QKD scheme [15,18]. An asymmetric Mach-Zehnder interferometer operates in a double pass over the quantum channel by using a Faraday mirror; see Fig. 2(a) without Eve. The interference of the paths taken by two pulses traveling from Bob to Alice and back is determined by their relative phase modulation ($\varphi_{\mathrm{Bob}} - \varphi_{\mathrm{Alice}}$), and forms the principle for encoding the key. Any birefringence effects of the quantum channel are passively compensated. As a prerequisite to the key exchange, Clavis2 calibrates its detectors in time via a sequence named line length measurement (LLM). Bob emits a pair of *bright* pulses and applies a series of detector gates around an initial estimate of their return. The timing of the gates is electronically scanned (while monitoring detector clicks) to refine the estimation of the channel length and relative delay between the time of arrival of the pulses at $D0$ and $D1$. Alice keeps her phase modulator (PM) switched off, while Bob applies a uniform phase of $\pi/2$ to one of the incoming pulses. Therefore, both detectors are equally illuminated and their

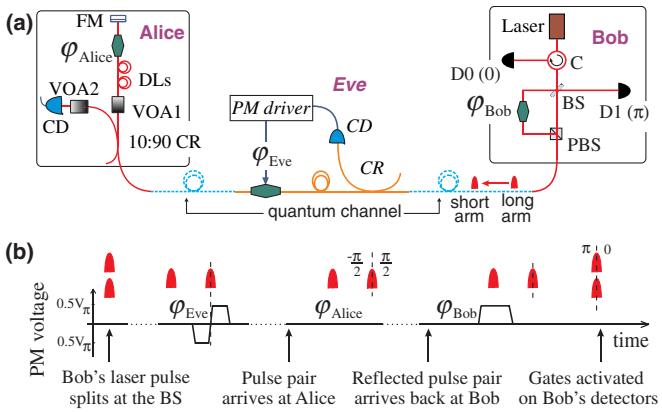

FIG. 2 (color online). Manipulation of the calibration routine. (a) Simplified version of Alice and Bob devices and Eve (in italic) gearing for the hack. FM, Faraday mirror; CD, classical photodiode; DLs, delay loops; VOA, variable optical attenuator; CR, coupler; BS, 50:50 beam splitter; PBS, polarizing beam splitter; C, optical circulator. The hexagonal-shaped objects are phase modulators (PMs); φ_X, where X is Bob, Alice, or Eve, represents the applied modulation. (b) Timeline for a cycle of the hacked LLM. V_π, PM voltage for a π phase shift.

detection efficiencies, denoted by $\eta_0(t)$ and $\eta_1(t)$, can be resolved in time. Any existing mismatch can thus be minimized by changing the gate-activation times (see Fig. 1).

However, the calibration routine does not always succeed; as reported in [7], a high detector efficiency mismatch (DEM) is sometimes observed after a normal run of LLM. For example, we have noticed a temporal mismatch as high as 400 ps in Clavis2. This physical limitation of the system—arising due to fast and uncontrollable fluctuations in the quantum channel or electromagnetic interference in the detection circuits—is the vulnerability that the time-shift attack exploits. However, the attack has some limitations: it is applicable only when the temporal mismatch happens to exceed a certain threshold value, which is merely 4% of all the instances [7]. Also, Eve can neither control the mismatch (as it occurs probabilistically) nor extract its value (as it is not revealed publicly).

We exploit a weakness of the calibration routine to induce a large and deterministic DEM without needing to extract any information from Bob. As depicted in Fig. 2(a), Eve installs her equipment in the quantum channel such that the laser pulse pair coming out of Bob's short and long arm passes through her PM. Eve's modulation pattern is such that a rising edge in the PM voltage flips the phase in the second (long arm) optical pulse from $-\pi/2$ to $\pi/2$, as shown in Fig. 2(b). As a result of this hack, when the pulse pair interferes at Bob's 50:50 beam splitter, the two temporal halves have a relative phase difference ($\varphi_{\mathrm{Bob}} - \varphi_{\mathrm{Eve}}$) of π and 0, respectively. This implies that photons from the first (second) half of the interfering pulses yield clicks in $D1$ ($D0$) deterministically. As the LLM localizes the detection efficiency peak corresponding to the optical power peak, an *artificial* temporal displacement in the detector efficiencies is induced. An inverse displacement can be obtained by simply inverting the polarity of Eve's phase modulation.

In the Supplemental Material [21], we describe a proof-of-principle experiment to deceive the calibration routine. With this setup, we record the temporal separation Δ_{01}, i.e., the difference between the delays for electronically gating $D0$ and $D1$, for several runs of LLM. Relative to the statistics from the normal runs (denoted by $\Delta_{01}^{\text{no Eve}}$), the hacked runs yield an average shift, $\Delta_{01}^{\text{Eve}} - \Delta_{01}^{\text{no Eve}} = 459$ ps with a standard deviation of 105 ps. Figure 3 shows the detection efficiencies $\eta_0(t)$ and $\eta_1(t)$ (measurement method explained in [21]) for the normal and hacked cases. It also provides a quantitative comparison between the usual and induced mismatch. Note that a larger mismatch can be obtained by modifying the shape of laser pulses coming from Bob.

After inducing this substantial efficiency mismatch, Eve can use an intercept-resend strategy employing "faked states" [12] to impose her will upon Bob (and Alice). Compared to her intercepted measurements, she prepares the opposite bit value in the opposite basis and sends it with

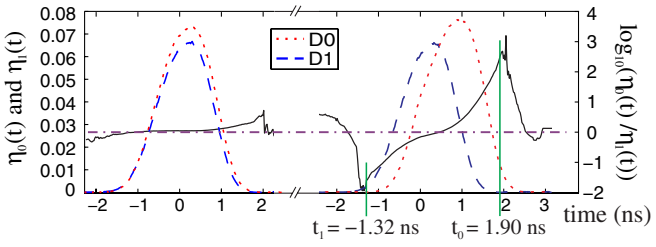

FIG. 3 (color online). Induced temporal mismatch. Efficiencies $\eta_0(t)$ (dotted line) and $\eta_1(t)$ (dashed line) from normal LLMs, on the left, and after Eve's hack that induced a separation of 459 ps, on the right. The logarithm of their ratio, quantifying the degree of mismatch (solid line), is at least an order of magnitude higher in the flanks after Eve's hack: the dash-dotted line indicates zero mismatch. To eavesdrop successfully, Eve times the arrival of "appropriately bright" faked states at $t = t_0$ or t_1 in Bob.

such a timing that the detection of the opposite bit value is suppressed due to negligible detection efficiency. As an example, assume that Eve measures bit 0 in the Z basis [in a phase-coded scheme, measuring in Z (X) basis $\Leftrightarrow$ applying $\varphi = 0(\pi/2)$]. Then, she resends bit 1 in the X basis, timed to be detected at $t = t_0$ (see Fig. 3), where $D1$ is almost blind. Using the numerical data on the induced mismatch, Eq. (3) from [16] yields a QBER $<0.5\%$ if the FSA is launched at times t_0 and t_1 where the efficiency mismatch is high.

However, it can be observed that the detection probabilities for $D0$ and $D1$ are quite low in this case. A considerable decrease in the rate of detection events in Bob could ensue an alarm. Also, the (relatively increased) dark counts would add significantly to the QBER. In fact, Eve needs to *match* the channel transmission T that Alice and Bob expect, without exceeding the QBER threshold at which they abort key generation [17]. Experimentally, we have found that the QBER abort threshold depends on the channel loss seen by Clavis2; for an optical loss of 1–6 dB (that should correspond to $0.79 > T > 0.25$), it lies between 5.94% and 8.26%.

Eve solves these problems by increasing the mean photon number of her faked states. To evaluate her QBER, we elaborate the approach of [16] by generalizing Table I of Ref. [16]. Our attack strategy, carefully accounting for all the involved factors, is summarized in Table I. For instance, in the first row we replace the probability of detection $\eta_0(t_0)/2$ by $1 - \exp[-\mu_0\eta_0(t_0)/2]$ for a coherent-state pulse of mean photon number μ_0 impinging on Bob's detectors at time t_0. Including the effect of the dark counts into this expression, Bob's probability to register 0 becomes $\mathbf{q}_0 = d_0 + (1 - d_0)\{1 - \exp[-\mu_0\eta_0(t_0)/2]\}$, where d_0 is the dark count probability in detector $D0$. A row for double clicks, i.e., simultaneous detection events in $D0$ and $D1$, is added for every (resent) state.

Because of the FSA, the $D0/1$ click probability at time t no longer depends solely upon $\eta_{0/1}(t)$. Summing over all the states sent by Alice (by extending Table I), the total detection probabilities in $D0$ and $D1$ when the attack is launched at specific times t_0 and t_1 are

$$p_0(\mu_0, \mu_1) = 0.75 + 0.25d - 0.25(1 - d) \times (e^{-0.5\mu_0\eta_{00}} + e^{-0.5\mu_1\eta_{01}} + e^{-\mu_1\eta_{01}}), \quad (1)$$

$$p_1(\mu_0, \mu_1) = 0.75 + 0.25d - 0.25(1 - d) \times (e^{-0.5\mu_0\eta_{10}} + e^{-0.5\mu_1\eta_{11}} + e^{-\mu_0\eta_{10}}). \quad (2)$$

TABLE I. Faked-state attack, given that Alice prepared bit 0 in the Z basis and that Bob measured in the Z basis (only matching basis at Alice and Bob remains after sifting). The first column contains the basis chosen by Eve and her measurement result. The second column shows parameters of the faked state resent by Eve: basis, bit, mean photon number, timing. The third column shows Bob's measurement result; $0 \cap 1$ denotes a double click. The last column shows the corresponding click probabilities (ignoring possible superlinearity effect in gated detectors [22]). Note that the first result ($\rightarrow$ Eve $\equiv Z, 0$) is twice as likely to occur as the other two.

$\rightarrow$ Eve	Eve $\rightarrow$	Bob's result	Detection probability
$Z, 0$	$X, 1, \mu_0, t_0$	0	$\mathbf{q}_0 = d_0 + (1 - d_0)\{1 - \exp[-\mu_0\eta_0(t_0)/2]\}$
		1	$\mathbf{q}_1 = d_1 + (1 - d_1)\{1 - \exp[-\mu_0\eta_1(t_0)/2]\}$
		$0 \cap 1$	$\mathbf{q}_0\mathbf{q}_1$
		Loss	$1 - (\mathbf{q}_0 + \mathbf{q}_1 - \mathbf{q}_0\mathbf{q}_1)$
$X, 0$	$Z, 1, \mu_0, t_0$	0	$\mathbf{r}_0 = d_0$
		1	$\mathbf{r}_1 = d_1 + (1 - d_1)\{1 - \exp[-\mu_0\eta_1(t_0)]\}$
		$0 \cap 1$	$\mathbf{r}_0\mathbf{r}_1$
		Loss	$1 - (\mathbf{r}_0 + \mathbf{r}_1 - \mathbf{r}_0\mathbf{r}_1)$
$X, 1$	$Z, 0, \mu_1, t_1$	0	$\mathbf{s}_0 = d_0 + (1 - d_0)\{1 - \exp[-\mu_1\eta_0(t_1)]\}$
		1	$\mathbf{s}_1 = d_1$
		$0 \cap 1$	$\mathbf{s}_0\mathbf{s}_1$
		Loss	$1 - (\mathbf{s}_0 + \mathbf{s}_1 - \mathbf{s}_0\mathbf{s}_1)$

Here $\eta_{jk} = \eta_j(t_k)$ with $j, k \in \{0, 1\}$ and $d = \text{mean}(d_0, d_1)$ are used to simplify the expressions. Similarly, one can compute the expression for $p_{0\cap 1}$, the total double-click probability. Eve's error probability, the arrival probability of the optical signals in Bob, and the QBER are

$$p_{\text{error}}(\mu_0, \mu_1) = 0.75 + 0.25d - 0.5p_{0\cap 1} - 0.125 \times (1 - d)(e^{-\mu_0\eta_{10}} + 2e^{-0.5\mu_0\eta_{10}} + e^{-\mu_1\eta_{01}} + 2e^{-0.5\mu_1\eta_{01}}), \quad (3)$$

$$p_{\text{arrive}}(\mu_0, \mu_1) = p_0 + p_1 - p_{0\cap 1}, \quad (4)$$

$$\text{QBER}\,(\mu_0, \mu_1) = p_{\text{error}}(\mu_0, \mu_1)/p_{\text{arrive}}(\mu_0, \mu_1). \quad (5)$$

Here double clicks are assumed to be assigned a random bit value by Bob [23], causing an error in half the cases.

If Alice and Bob are connected back to back (channel transmission $T \approx 1$), the click probabilities in Bob should be slightly less than half of the peak values in Fig. 3. This is because of optical losses ($\gtrsim 3$ dB) in Bob's apparatus. Eve's constraints can now be formalized as follows: starting in the vicinity of $p_0 = 0.038$ and $p_1 = 0.032$, not only does she have to match Bob's expected detection rate for any given $T < 1$, but she also has to keep the resultant QBER below the threshold at which Clavis2 aborts the key exchange. We assume Eve detects photons at Alice's exit using a perfect apparatus and resends perfectly aligned faked states.

Substituting $t_1 = -1.32$ ns, $t_0 = 1.90$ ns (marked in Fig. 3) and $d = 2.4 \times 10^{-4}$ in Eqs. (1)–(5), Eve collects tuples $[p_0, p_1, \text{QBER}]$ by varying μ_0 and μ_1 in a suitable range. Out of all tuples that feature the same detection probabilities (arising from different combinations of μ_0 and μ_1), Eve chooses the one having the lowest QBER. A contour plot in Fig. 4 displays this minimized error $\min_{\mu_0,\mu_1}\text{QBER}[(\mu_0, \mu_1)|(p_0, p_1)]$. The thick shaded line shows that for $T > 0.25$, Eve not only maintains the detection rates within 5% of Bob's expected values, but also keeps the QBER below 7% [24], thus breaking the security of the system. Note that the simulation assumes a lossless Eve, but in principle she can cover loss from her realistic detection apparatus by increasing μ_0 and μ_1 further and/or including t_0 and t_1 in the minimization.

To counter this hack, Bob should randomly apply a phase of 0 or π (instead of $\pi/2$ uniformly) while performing LLM. This modification is implementable in software and has already been proposed to ID Quantique. More generally, a method to shield QKD systems from attacks that exploit DEM is described in Ref. [25].

In conclusion, we report a proof-of-principle experiment to induce a large detector efficiency mismatch in a commercial QKD system by deceiving a vital calibration routine. An optimized faked-state attack on such a compromised system would not alarm Alice and Bob as it would introduce a QBER $<7\%$ for a large range of expected channel transmissions. Thus, the overall security of the system is broken. With initiatives for standardizing QKD [26] under way, we believe this report is timely and shall facilitate elevating the security of practical QKD systems.

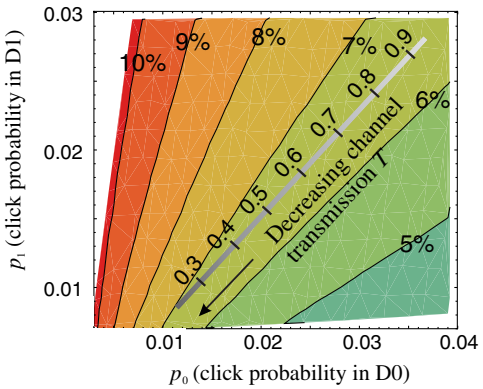

FIG. 4 (color online). Minimum QBER versus click probabilities in $D0$ and $D1$. Eve minimizes the error with a suitable choice of the mean photon number of the faked states (for this plot, $1 < \mu_0 < 100$ and $21 < \mu_1 < 120$ at Bob's detectors). The thick shaded line indicates Bob's detection probabilities. The QBER introduced by Eve stays below 7% for $T \gtrsim 0.25$.

We thank M. Legré from ID Quantique and N. Lütkenhaus for helpful discussions and Q. Liu, L. Meier, and A. Käppel for technical assistance. This work was supported by the Research Council of Norway (Grant No. 180439/V30), and DAADppp mobility program financed by NFR (Project No. 199854) and DAAD (Project No. 50727598). C. Wiechers acknowledges support from FONCICYT Project No. 94142.

*nitin.jain@mpl.mpg.de

[1] C. H. Bennett and G. Brassard, in *Proceedings of the IEEE International Conference on Computers, Systems, and Signal Processing* (IEEE, New York, 1984), pp. 175–179; P. W. Shor and J. Preskill, Phys. Rev. Lett. **85**, 441 (2000) and references therein.
[2] V. Scarani *et al.*, Rev. Mod. Phys. **81**, 1301 (2009).
[3] V. Scarani and C. Kurtsiefer, arXiv:0906.4547.
[4] B. Huttner *et al.*, Phys. Rev. A **51**, 1863 (1995); N. Lütkenhaus and M. Jahma, New J. Phys. **4**, 44 (2002).
[5] N. Gisin *et al.*, Phys. Rev. A **73**, 022320 (2006); A. Vakhitov, V. Makarov, and D. R. Hjelme, J. Mod. Opt. **48**, 2023 (2001).
[6] A. Lamas-Linares and C. Kurtsiefer, Opt. Express **15**, 9388 (2007); S. Nauerth *et al.*, New J. Phys. **11**, 065001 (2009).
[7] Y. Zhao *et al.*, Phys. Rev. A **78**, 042333 (2008).
[8] C.-H. F. Fung *et al.*, Phys. Rev. A **75**, 032314 (2007); F. Xu *et al.*, New J. Phys. **12**, 113026 (2010).
[9] L. Lydersen *et al.*, Nat. Photon. **4**, 686 (2010).

[10] L. Lydersen *et al.*, Opt. Express **18**, 27938 (2010); C. Wiechers *et al.*, New J. Phys. **13**, 013043 (2011).
[11] I. Gerhardt *et al.*, Nature Commun. **2**, 349 (2011).
[12] V. Makarov and D. R. Hjelme, J. Mod. Opt. **52**, 691 (2005).
[13] V. Makarov, New J. Phys. **11**, 065003 (2009).
[14] N. Gisin *et al.*, Rev. Mod. Phys. **74**, 145 (2002).
[15] Data sheet of Clavis2, http://www.idquantique.com.
[16] V. Makarov, A. Anisimov, and J. Skaar, Phys. Rev. A **74**, 022313 (2006).
[17] D. Gottesman *et al.*, Quantum Inf. Comput. **4**, 325 (2004); H.-K. Lo, X. Ma, and K. Chen, Phys. Rev. Lett. **94**, 230504 (2005).
[18] D. Stucki *et al.*, New J. Phys. **4**, 41 (2002).
[19] D. S. Bethune and W. P. Risk, IEEE J. Quantum Electron. **36**, 340 (2000); M. Bourennane *et al.*, Opt. Express **4**, 383 (1999).
[20] Z. Yuan and A. Shields, Opt. Express **13**, 660 (2005).
[21] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.107.110501 for experimental details.
[22] L. Lydersen *et al.*, arXiv:1106.2119 [Phys. Rev. A (to be published)].
[23] N. Lütkenhaus, Phys. Rev. A **59**, 3301 (1999).
[24] The QBER can be reduced even further if Bob checks only the *overall* detection probability $p_0 + p_1$.
[25] L. Lydersen, V. Makarov, and J. Skaar, Phys. Rev. A **83**, 032306 (2011).
[26] "Quantum key distribution (QKD); Security proofs," European Telecommunications Standards Institute, ETSI GS QKD 005 V1.1.1; T. Länger and G. Lenhart, New J. Phys. **11**, 055051 (2009).