

Upper bounds on key rates in device-independent quantum key distribution based on convex-combination attacks

Karol Łukanowski^{1,2}, Maria Balanzó-Juandó³, Máté Farkas^{4,3}, Antonio Acín^{3,5}, and Jan Kołodyński¹

¹Centre for Quantum Optical Technologies, Centre of New Technologies, University of Warsaw, Banacha 2c, 02-097 Warszawa, Poland

²Faculty of Physics, University of Warsaw, Pasteura 5, 02-093 Warszawa, Poland

³ICFO – Institut de Ciències Fòniques, The Barcelona Institute of Science and Technology, 08860 Castelldefels, Spain

⁴Department of Mathematics, University of York, Heslington, York, YO10 5DD, United Kingdom

⁵ICREA-Institució Catalana de Recerca i Estudis Avançats, Lluís Companys 23, 08010 Barcelona, Spain

The device-independent framework constitutes the most pragmatic approach to quantum protocols that does not put any trust in their implementations. It requires all claims, about e.g. security, to be made at the level of the final classical data in hands of the end-users. This imposes a great challenge for determining attainable key rates in *device-independent quantum key distribution* (DIQKD), but also opens the door for consideration of eavesdropping attacks that stem from the possibility of a given data being just generated by a malicious third-party. In this work, we explore this path and present the *convex-combination attack* as an efficient, easy-to-use technique for upper-bounding DIQKD key rates. It allows verifying the accuracy of lower bounds on key rates for state-of-the-art protocols, whether involving one-way or two-way communication. In particular, we demonstrate with its help that the currently predicted constraints on the robustness of DIQKD protocols to experimental imperfections, such as the finite visibility or detection efficiency, are already very close to the ultimate tolerable thresholds.

1 Introduction

Device-independent quantum key distribution (DIQKD) is the strongest form of quantum cryptographic protocols [1, 2]. It does not require the honest users to make any assumptions about the inner workings of devices at their hands and, hence, opens doors to assuring security without putting any trust into the manufacturer providing a given key-distribution system. As long as the parties can assure the classical data they generate during the protocol does not leak out without control—an assumption at the foundation of any cryptographic protocol [3]—

Karol Łukanowski: k.lukanowski@cent.uw.edu.pl

Jan Kołodyński: j.kolodynski@cent.uw.edu.pl

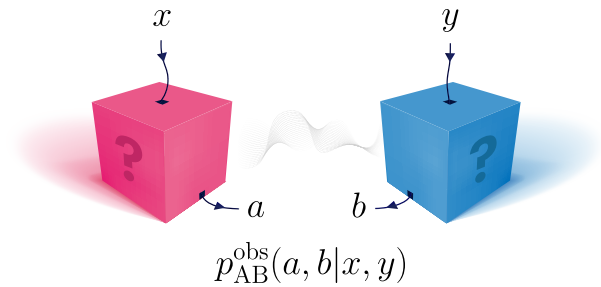


Figure 1: **Device-independent view of a QKD protocol.** Each honest user, Alice (or Bob), ignores the inner workings of her (his) device and treats it as a “black box”, here red (or blue), which in each round of the protocol takes the *setting* x (y) as an input and outputs the *outcome* a (b). As a result, by publicly revealing some of the generated data, the parties can verify to be sharing devices whose behaviour is described by a particular *correlation* $p_{AB}^{\text{obs}}(a, b|x, y)$ —the only property to be trusted when validating the security of the protocol.

then by revealing some of the data and verifying it to exhibit non-local correlations [4], they can extract a cryptographic key whose security is guaranteed by the correctness of quantum theory [1, 2], or even just the no-signalling paradigm [5–7]. As this makes DIQKD immune to all quantum implementation flaws, these cannot be exploited anymore to perform any hacking attack [8–11].

Quantum key distribution (QKD) protocols are based on the setting in which two distant honest parties, Alice and Bob, aim at sharing a cryptographic key, while assuring it to be unknown to any potential eavesdropper. In QKD, this can be achieved by distributing entangled quantum states between Alice and Bob in each round of the protocol, during which they then measure their corresponding part of the state [12]. Within the DIQKD framework, however, the “black-box” approach depicted schematically in Fig. 1 is pursued. From the perspective of the users, they are just provided with devices that allow to vary the type of measurement being implemented in each round, the measurement *setting*, whose *outcome* is then outputted by the device. Still, by revealing some of the results between each other, Alice and Bob

can verify what is the probability distribution—the *correlation*—describing the operation of their devices (boxes), i.e. specifying the probabilities with which the outcomes occur for each of the chosen setting. Note that, for the sake of simplicity, we adopt here a terminology in terms of the observed correlation that is meaningful in a scenario consisting of independent and identically distributed (i.i.d.) realisations of the experiment. In a general security proof, one should consider the estimated frequencies of all the observed events for a given finite number of rounds.

Crucially, within the DI paradigm the users do not assume anything about the origin of the correlation, apart from the fact that it must be compliant with the laws of quantum mechanics. Nonetheless, if the observed correlation violates a Bell inequality [4], Alice and Bob can estimate the information that any potential eavesdropper may have about their recorded outcomes—the *raw data*—opening up the possibility for the parties to extract a secure key. Although establishing performance when only finite amount of data is available is important for real-life implementations [13–15], the first step is always to verify whether the asymptotic key rate—which we refer to here as just the (DIQKD) *key rate*—can be even positive, i.e. the number of secret bits being finally shared by Alice and Bob divided by the number of protocol rounds employed (the size of the raw data), in the limit of the latter going to infinity.

The task of estimating key rates and proving the security of DIQKD protocols constitutes a great challenge, being a subject of intensive theoretical research. In the *one-way* scenario, in which the parties can publicly communicate only in one direction when distilling the key from the raw data, say from Alice to Bob, the amount of secrecy in any of Alice’s outcomes can be quantified by its corresponding von Neumann entropy conditioned on an eavesdropper’s quantum side information, $H(A|E)$. Thanks to recent developments, such a statement is now crucially true not only when considering *collective attacks* [16, 17], but also when allowing for the most powerful *coherent attacks* [13, 18, 19]. Still, the challenge is to compute (or at least lower-bound) $H(A|E)$ for a given non-local correlation being shared, in order to determine (lower-bound) the corresponding DIQKD key rate. The first approaches (c.f. [1, 2]) have succeeded in providing lower bounds based on the violation of the Clauser–Horne–Shimony–Holt (CHSH) Bell inequality [20], while the more recent works generalised these to include biased CHSH inequalities [21, 22], accounting also for *noisy preprocessing* of the raw data [22–24]. Another valid approach is based on lower-bounding $H(A|E)$ via the min-entropy [25], which apart from being again relatable to the violation of CHSH [26], can be accurately lower-bounded by resorting to numerical convex-programming methods [27–30], based on a convergent hierarchy of relaxations [31, 32]. This

has been lately done also for DIQKD protocols involving *random postselection* of the raw data [33]—a procedure performed jointly by the users for which, however, the security beyond i.i.d. attacks [34] has not been proven so far. Moreover, a convergent hierarchy has been recently proposed for the conditional von Neumann entropy, $H(A|E)$, itself [35, 36], see also [37].

In parallel, complementary methods of upper-bounding the DIQKD key rates have been proposed [38–41]. These were put forward, however, with general aim in mind of dealing with *all* potential DIQKD protocols that may involve even *two-way* communication between the parties—e.g. advantage distillation of the raw data [42]—so that the upper bounds may then serve as ultimate benchmarks beyond which no DIQKD protocol can venture. In this work, we follow this path but focus instead on attacks that an eavesdropper may adapt for a *particular* DIQKD scenario. As a result, the upper bounds on key rates we obtain account for the special features of the DIQKD protocol considered, e.g.: whether it involves one-way or two-way communication, the type of preprocessing used, or the postselection stage.

In particular, our approach is to propose concrete strategies a malicious third-party can play when distributing and controlling the devices, so that the data in hands of the users is consistently recovered, but some information about it—which can be explicitly quantified—remains in possession of the eavesdropper, Eve. We consider *individual attacks* [12] that yield a tripartite (Alice, Bob and Eve) classical (i.i.d.) model that in case of one-way scenarios describes a broadcast (wiretap) channel [43], while in the two-way case allows for unconstrained public discussion [44, 45]. The key rate attained between Alice and Bob within such a model constitutes then an upper bound on the DIQKD key rate associated with their shared data. Moreover, as allowing for more powerful eavesdroppers (collective or coherent attacks) can only decrease the attainable key rate, such an upper bound remains valid beyond individual attacks. In particular, in case it vanishes, it is assured that no secure key can be distilled by the honest parties from a given data set.

In our work, we focus on a class of individual attacks—dubbed *convex-combination* (CC) attacks—in which Eve randomly alternates between distributing either devices that yield stronger non-local correlations than the ones exhibited by the data being shared, or devices that yield classical (local) correlations with Eve possessing full knowledge about the output data of both Alice and Bob. We show that, while correctly reproducing the shared data on average, the CC attack can be optimised by means of linear programming to maximise the probability of the local correlation being shared. As a result, it provides a direct method of upper-bounding the key rates

that turns out to be very effective in predicting the zero-key regions, in which no DIQKD is possible. Although this is not our motivation here, let us note that in the non-zero regions our technique could also be merged with the other methods [38–40] to determine the tightest overall upper bounds on the key rate [46].

In contrast to our previous work [41], in which we have focused on the two-way scenario in order to find regimes in which the CC attack precludes any key to be extracted despite the shared correlation exhibiting non-locality, here we study the limits the attack imposes on imperfections within the correlations being shared—given the DIQKD protocol (incl. any processing of the raw data) followed by the parties.

In particular, we use the CC attack to determine thresholds on experimentally motivated parameters: *visibility* and *detection efficiency* (level of losses) beyond which DIQKD cannot be made possible both in the one-way and two-way scenarios—no matter how well one improves current techniques of lower-bounding the exact DIQKD key rates [35–37, 47]. To our knowledge, at the time of preparing this manuscript, the best known values of tolerable detection efficiency above which fully secure one-way DIQKD becomes feasible read $\eta \geq 80.00\%$ [35] and $\eta \geq 80.26\%$ [47], while the CC attack allows us to verify that without changing the structure of these protocols, the thresholds could at most be improved to 79.04% and 79.15%, respectively. While noisy pre-processing of the shared data is crucial for the parties to reach the above tolerable efficiencies [35, 47], it simultaneously makes the CC attack more efficient, so that it provides very tight lower bounds. In general, the CC-based value diminishes to 75% when optimising over all forms of (one-way [17]) pre-processing potentially employed by the parties. On the other hand, when the eavesdropper can be assumed to perform at most collective attacks, one may further allow the parties to publicly perform random postselection before one-way communication. In such a scenario, the CC-based threshold decreases further to $\approx 66.6\%$ —the fundamental value imposed by non-locality of the shared correlation [48]—what is consistent with the smallest up-to-date known tolerable efficiency of $\eta \geq 68.5\%$ established with random postselection [33]. Finally, by considering the recent protocol of Ref. [49], we demonstrate that the CC attack can be easily applied also to DIQKD schemes motivated by Bell violations with more than two measurement settings and outcomes.

Our results make us believe that the CC attack constitutes a useful tool that, not only allows to easily verify whether there is much room for improvement of the state-of-the-art estimates [22–24, 26–28, 33, 35, 37, 47, 49] of the key rate for a given DIQKD protocol, but also can be very helpful in seeking ways to modify the protocol in order to improve

its robustness to imperfections. Finally, it is also useful to benchmark the lower bounds on Eve’s entropy obtained through the existing hierarchies and understand how much can be gained by increasing the level in the hierarchy.

The manuscript is structured as follows. In Sec. 2, we discuss the formulation of standard DIQKD protocols under individual attacks and, in particular, the upper bounds on one-way and two-way key rates such attacks yield. We then introduce the CC attack as a special case of an individual attack in Sec. 3, including its geometric formulation and optimisation via a linear program. In Sec. 4, we describe the noise models of finite visibility and detection efficiency that we will use to benchmark the robustness of current DIQKD schemes by means of the CC attack. In particular, in Sec. 5, we firstly apply the CC attack to both one-way and two-way protocols that rely on non-local correlations arising from maximally entangled states, discussing in detail the construction and its consequences. We then move onto one-way protocols involving partially entangled states in Sec. 5.4, which currently provide the state-of-the-art key rates and robustness to noise. In Sec. 5.5, we further demonstrate that the CC attack can be straightforwardly applied also to scenarios in which the parties employ more than two measurement settings and outcomes. Finally, we conclude our findings in Sec. 6.

2 DIQKD under individual attacks

2.1 Standard DIQKD protocols

In a DIQKD protocol two parties, Alice and Bob, have access to a bipartite quantum state, $\rho_{AB} \in \mathcal{B}(\mathcal{H}_A \otimes \mathcal{H}_B)$, defined on the tensor product of their corresponding Hilbert spaces. The protocol consists of several rounds, in each of which Alice and Bob choose a particular quantum measurement to measure their part of a fresh copy of ρ_{AB} . In particular, Alice chooses her measurement according to a random variable X , whose instance x labels the (measurement) *setting* selected out of $|X| = m_A$ possibilities. Similarly, Bob chooses his measurement according to $Y = y$ with $|Y| = m_B$. The (measurement) *outcome* a (b) recorded then by Alice (Bob) corresponds to an instance of the random variable A (B) that we assume, without loss of generality, to take the same number of $|A| = n_A$ ($|B| = n_B$) values for any setting, x (y), chosen.

According to quantum theory, each of the m_A (m_B) measurements of Alice (Bob) is described by a positive-operator-valued measure $\{M_a^x\}_{a=1}^{n_A}$ ($\{N_b^y\}_{b=1}^{n_B}$), so that the *correlation* shared by the parties generally reads

$$p_{AB}^{\text{obs}}(a, b|x, y) = \text{Tr}\{\rho_{AB} (M_a^x \otimes N_b^y)\}, \quad (1)$$

specifying the probability of obtaining the outcomes

a and b , given that the measurements x and y were selected. We say that p_{AB}^{obs} in Eq. (1) is the *observed* correlation within the $m_A m_B n_A n_B$ -scenario [50].

After each protocol round, Alice and Bob store both the observed measurement outcomes, a and b , as well as the measurement settings, x and y , they employed. The records constitute then the *raw data*, out of which Alice and Bob distil a secret key with help of public communication, so that at the end of the procedure they aim at holding identical strings that appear perfectly random to any third party. In this work, we focus on estimating the asymptotic *key rate*, i.e. the length of such secret strings divided by the overall number of protocol rounds, in the limit of the latter going to infinity.

We consider here *standard* DIQKD protocols, i.e. ones in which both parties announce publicly the measurement settings employed in each round [41]. Although we primarily focus on protocols that further use the outcomes of a pre-agreed fixed pair of settings to extract the key, let us emphasise already that the CC attack, which is our main interest, can be applied to any scenario by following step-by-step every stage of a given protocol within the attack, e.g. see [41] for its application to the scheme of [15] involving multiple key-settings. In standard protocols, Alice and Bob firstly record strings of outcomes and settings over sufficiently many protocol rounds. Since individually they only have access to the marginal distributions, they publicly reveal part of their data in order to estimate the full correlation $p_{AB}^{\text{obs}}(a, b|x, y)$. This part of the dataset is then discarded. They also reveal the settings for the remaining dataset, and keep those outcomes that correspond to a pre-agreed *key setting* pair, (x^*, y^*) , distributed according to $p_{AB}^{\text{obs}}(a, b|x^*, y^*)$. If estimation shows that the error probability is low enough, they extract the final key from this dataset, using either two-way or one-way public communication schemes known as *privacy amplification* and *error correction* [43–45]—and abort the protocol otherwise.

2.2 Individual attacks

In this work, we consider *individual attacks* [12] of the eavesdropper, Eve, in which her register at the end of each protocol round corresponds to a random variable, E , being somehow correlated with the outcomes of Alice and Bob, determined by A and B , respectively. As E may take as many values as required, it may, for example, consist of doubles (ordered pairs), i.e. $e = (\tilde{a}, \tilde{b})$ where $|E| = n_A n_B$, and \tilde{a} and \tilde{b} stand for Eve’s guesses of Alice’s and Bob’s outcomes, respectively. In such a case, the situation in which Eve knows perfectly both the outcomes corresponds simply to the (tripartite) correlation $p_{ABE}(a, b, e = (\tilde{a}, \tilde{b})) = \delta_{a\tilde{a}}\delta_{b\tilde{b}}/(n_A n_B)$ with $\delta_{\alpha\beta}$ denoting the Kronecker delta function. Note that,

generalising naturally Eq. (1) to p_{ABE} , such attacks “force” Eve to measure her part of now a tripartite state ρ_{ABE} in the same way at the end of each protocol round [12], and exclude the possibility of her possessing a quantum memory [51].

As a result, each round of the protocol and, hence, the protocol on the whole, is then completely described by a tripartite correlation incorporating also the eavesdropper:

$$p_{ABE}(a, b, e|x, y) \quad (2)$$

$$\text{s.t. } \forall_{a,b,x,y} : \sum_e p_{ABE}(a, b, e|x, y) = p_{AB}^{\text{obs}}(a, b|x, y),$$

where the above constraint assures that a given quantum correlation observed by Alice and Bob is indeed recovered on average, despite the presence of Eve.

In general, in order to consistently define the attack one should specify the form of the correlation (2), in particular, its quantum origin, i.e. the state being shared between all three parties and the measurements they perform [12]. However, for our purposes we consider individual attacks in which the strategy of Eve is to simply distribute different boxes—bipartite correlations shared by Alice and Bob (known to her and labelled by λ)—in each protocol round, so that Eq. (2) takes the form:

$$p_{ABE}(a, b, e|x, y)$$

$$= \sum_{\lambda} q(\lambda) p(e|\lambda) p_{AB}(a, b|x, y, \lambda) \quad (3)$$

$$\text{s.t. } \forall_{x,y} : \sum_{\lambda} q(\lambda) p_{AB}(a, b|x, y, \lambda) = p_{AB}^{\text{obs}}(a, b|x, y).$$

Such an attack is then specified by the probabilities, $q(\lambda)$, of Eve distributing each bipartite correlation $p_{AB}(a, b|x, y, \lambda)$, each of which must be decomposable as in Eq. (1) to be consistent with quantum theory, and her knowledge $p(e|\lambda)$ about the outcomes of Alice and Bob for each of these correlations.

Note that this individual attack can be implemented by Eve via sharing the same tripartite state in each measurement round and measuring her part of the state, producing the outcome e . Importantly, this can be done such that e preserves all information about λ that can later be used by Eve to post-process the variable e . This will be important in standard DIQKD protocols, in which Alice and Bob at some point reveal their measurement settings for each round, and this information—together with λ —can be used by Eve to improve her guess on Alice’s and Bob’s outcomes. The knowledge of λ practically means that Eve always knows which term in the convex decomposition of p_{AB}^{obs} in Eq. (3) is used, whereas Alice and Bob have access only to the average distribution p_{AB}^{obs} . For an explicit construction of the tripartite state and the measurements of Alice, Bob and Eve, see App. A.

2.3 Upper bounds on one-way key rates

Formally, the key rate for all QKD and, hence, also DIQKD protocols assisted by one-way communication (say from Alice to Bob) includes a maximisation over all *preprocessing* maps (performed then by Alice), i.e. [17]:

$$r_{1\text{-way}}(A \rightarrow B) := \max_{p_{A'|A}, p_{M|A'}} r_{1\text{-way}}(A \rightarrow B|A', M), \quad (4)$$

where A and B are the outcome variables of Alice and Bob when they both select the key settings, x^* and y^* , respectively. The mapping $A \rightarrow A'$ described by the stochastic map $p_{A'|A}$ is applied by Alice on her outcome, followed by $A' \rightarrow M$ (described by $p_{M|A'}$), whose output is then sent to Bob over a public channel¹. In contrast, all the operations performed by Bob on his outcomes can be ignored, as this would lead to an underestimation of the rate due to an overestimation of the fraction of bits required to perform the error correction [52].

For any given preprocessing strategy, however, the one-way key rate may be generally lower-bounded by the so-called *Devetak-Winter* (DW) rate [16], which is valid for all *collective attacks* (more powerful than individual [12]) that Eve may perform, i.e. [17]:

$$r_{1\text{-way}}(A \rightarrow B|A', M) \geq r_{\text{DW}} := H(A'|E, M) - H(A'|B, M), \quad (5)$$

where $H(A'|E, M)$ is the von Neumann entropy conditioned on the information possessed by the most general *quantum* eavesdropper—denoted here by a roman letter E to explicitly distinguish quantum side-information from random variables signified throughout the text by italic characters—while $H(A'|B, M)$ is the conditional (Shannon) entropy between Alice's and Bob's outcomes for the key settings, both conditioned also on the classical data M revealed by Alice during the preprocessing stage.

The DW rate (5) can be intuitively understood as the difference between the contributions attributed to privacy amplification (PA) and error correction (EC). In particular, the PA-term, $H(A'|E, M)$, represents the fraction of bits that are at least available to Alice after she compresses her bit-string sufficiently to ensure that it is no longer correlated anyhow with any eavesdropper. The EC-term, $H(A'|B, M)$, denotes instead the fraction of bits that she must still publicly communicate to Bob for him to correct his bit-string to be perfectly matching the one of hers. However, note that the latter is fully determined by the correlation $p_{AB}^{\text{obs}}(a, b|x^*, y^*)$ shared by Alice and Bob (and the chosen preprocessing) and, hence, is actually unaffected by the presence of any eavesdropper.

Strikingly, within the DIQKD framework the inequality in Eq. (5) has been shown via the entropy

¹Without loss of generality, the ranges of M and A' can be set to be finite [43, 45].

accumulation theorem (EAT) [13] to hold for the most powerful quantum eavesdroppers, i.e. all *coherent attacks* [12], as long the data announced publicly in a given round is independent from the device outputs generated in preceding rounds [53, Section 6.1]. This is true, in particular, for a large family of DIQKD protocols where the publicly disclosed data is restricted to (random) device inputs, whereas the preprocessing performed by Alice is limited to some stochastic mapping $A \rightarrow A'$, in which case the DW rate (5) just reads $r_{\text{DW}} = H(A'|E) - H(A'|B)$ [13, 18, 19]. This applies, for instance, to scenarios when A and B constitute dichotomous variables, while *noisy preprocessing* of the raw data is included [54]. In particular, Alice applies then a symmetric bit-flip map² $A \rightarrow A'$ to introduce extra randomness (errors) and make her outputs less correlated with Eve by an amount larger than the one required for them to be corrected during the EC stage, so that the DW rate goes up overall [17]—as recently demonstrated also within the context of DIQKD [22–24].

On the other hand, by considering any particular *individual attack* and fixing the preprocessing strategy, we may construct an upper bound on the one-way rate (DIQKD or not) that is valid for all one-way protocols employing this strategy [43, 45]:

$$r_{1\text{-way}}(A \rightarrow B|A', M) \leq H(A'|E, M) - H(A'|B, M), \quad (6)$$

which, in contrast to Eq. (5), assumes a *classical* eavesdropper, i.e. Eq. (6) is completely determined by the tripartite distribution (2) for the key settings, $p_{ABE}(a, b, e|x^*, y^*)$, and, in particular, its marginals p_{AE} and $p_{AB} = p_{AB}^{\text{obs}}$ specifying the PA- and EC-terms, respectively. Note that this upper bound remains valid also when stronger attacks are considered, as these may only decrease the rate. Furthermore, by maximising the upper bound (6) (typically by numerical heuristic methods) over all preprocessing strategies, i.e. maps $p_{A'|A}$ and $p_{M|A'}$, we obtain an upper bound on the key rate that is universally valid for one-way protocols (4).

2.4 Upper bounds on two-way key rates

When it comes to two-way protocols within the DIQKD framework, i.e. the setting in which Alice and Bob are allowed to perform unconstrained public discussion [44, 45], lower bounds on the corresponding two-way key rates have been established only when constraining the power of Eve to collective attacks and the communication between Alice and Bob to the so-called *advantage distillation* protocol [42]. On the other hand, universal upper bounds on the two-way DIQKD rates have been recently proposed [38–

²With some probability $0 < p < 1$ of flipping '0' onto '1', and symmetrically '1' onto '0'.

41] that base on, e.g., measures of reduced entanglement [39] or the CHSH-inequality violation [40].

Here, following the approach based on individual attacks described above and our previous work [41], we consider upper bounds on the two-way DIQKD key rate constructed with help of intrinsic information [55, 56]—originally employed when considering non-signalling eavesdroppers [6, 7]. In particular, for any individual attack of Eve described by the tripartite correlation (2), the following upper bound on the two-way key rate generally holds [55, 56]:

$$r_{2\text{-way}}(A \leftrightarrow B) \leq I(A:B \downarrow E), \quad (7)$$

where the *intrinsic information*,

$$I(A:B \downarrow E) := \min_{p_{F|E}} I(A:B|F), \quad (8)$$

is defined as the conditional mutual information evaluated on the tripartite correlation

$$p_{ABF}(a, b, f|x^*, y^*) = \sum_e p_{F|E}(f|e) p_{ABE}(a, b, e|x^*, y^*), \quad (9)$$

which is further minimised over all potential mappings $E \rightarrow F$ that Eve can perform on her variable E ³.

In general, the computation of Eq. (8) may require heuristic methods, as the minimisations over mappings $E \rightarrow F$ constitutes a non-convex optimisation problem. However, any map $p_{F|E}$ provides a valid upper bound on two-way rate, since $I(A:B \downarrow E) \leq I(A:B|F)$. Moreover, let us note that the conditional mutual information $I(A:B|F)$ and, hence, the intrinsic information (8) is a monotonic decreasing function under stochastic maps applied on either A or B . Thus, as the two-way rate $r_{2\text{-way}}(A \leftrightarrow B)$ by definition involves maximisation over all stochastic maps that the parties may apply on their bits (supplemented by any two-way communication), the r.h.s. in Eq. (7) correctly incorporates already such a maximisation in its form and is thus a universal upper bound—in stark contrast to the upper bound (6) on the one-way rate, which similarly to the DW rate (5) may increase under preprocessing. In fact, by applying any transformations $A \rightarrow A'$ and $B \rightarrow B'$ on the r.h.s. of Eq. (7), we obtain a valid upper bound on the two-way rate in protocols with fixed preprocessing that may only be smaller than the universal bound, i.e.,

$$r_{2\text{-way}}(A' \leftrightarrow B') \leq I(A':B' \downarrow E) \leq I(A:B \downarrow E), \quad (10)$$

where the preprocessed A' and B' are now the random variables initially available to the parties.

³Without loss of generality, F can be taken to have the same number of outcomes as E [56].

3 The convex-combination attack

We consider a subclass of individual attacks taking the form (3), in particular, the *convex-combination (CC) attacks* introduced by us in [41], being inspired by the considerations of [6, 7] in which, however, Eve is allowed to possess even stronger than quantum, but still non-signalling, correlations with the raw data.

In short, within the CC attack, Eve mimics the ‘observed’ non-local correlation (pair of boxes) being shared between Alice and Bob, $p_{AB}^{\text{obs}}(a, b|x, y)$, by distributing interchangeably ‘local’ (exhibiting a local-hidden-variable model [4]) and ‘non-local’ correlations, in such a way that on average the ‘observed’ correlation is recovered and the attack proceeds unnoticed by the parties. Here we are interested in protocols involving two parties, but such a strategy may be analogously generalised to scenarios in which more parties are involved [57].

An (overpessimistic⁴) assumption is then made, restricting Eve to possess no knowledge about the outcomes of the honest parties whenever she distributes any ‘non-local’ correlation. This contrasts strongly the case of distributing ‘local’ correlations, for each of which Eve can be shown to possess full knowledge about all the outcomes. Motivated by this difference, it is further assumed within the CC attack that it is best for Eve to maximise the overall probability of using local boxes. As a result, once the ‘non-local’ boxes to be used by Eve are specified, the optimal ‘local’ correlation to be distributed most frequently by her can always be found by means of linear programming.

In what follows, we first provide a geometrical interpretation of the CC attack, in order then to describe its optimisation in terms of a linear program, which we subsequently employ in Sec. 5 to find the tightest upper bounds on the DIQKD key rates that the CC attack can provide.

3.1 Geometric formulation of the CC attack

As stated above, the CC attack constitutes an example of the individual attack described by Eq. (3). In particular, in its simplest form, Eve distributes either a local or a non-local correlation, denoted by $p_{AB}^{\mathcal{L}}$ or $p_{AB}^{\mathcal{NL}}$, respectively, such that the tripartite correlation (2) reads:

$$p_{ABE}(a, b, e|x, y) = q^{\mathcal{L}} p_{AB}^{\mathcal{L}}(a, b|x, y) \delta_{e,(a,b)} + q^{\mathcal{NL}} p_{AB}^{\mathcal{NL}}(a, b|x, y) \delta_{e,?}, \quad (11)$$

which corresponds to setting $\lambda = \{0, 1\}$ in Eq. (3) to a binary variable, whose outcome heralds that either a local or a non-local correlation is distributed by Eve, with probabilities $p(\lambda = 0) = q^{\mathcal{L}}$ and $p(\lambda =$

⁴If Eve possesses any information about the outcomes also in the ‘non-local’ rounds, this may only improve the CC attack further—yield even tighter upper bounds on the key rates.

1) = $q^{\mathcal{NL}} = 1 - q^{\mathcal{L}}$, respectively. Moreover, Eve knows and controls which boxes are being used in each protocol round, so whenever $p_{AB}^{\mathcal{L}}$ is distributed she has perfect knowledge and $p(e|\lambda = 0) = \delta_{e,(a,b)}$ in Eq. (3), i.e. her outcome is perfectly correlated with the outcomes of Alice and Bob; while in case $p_{AB}^{\mathcal{NL}}$ is used $p(e|\lambda = 1) = \delta_{e,?}$ in Eq. (3), i.e. she registers a special extra outcome “?” giving her no knowledge about the outcomes of the honest parties. Note that for simplicity, we collected all the ‘local’ terms in $\lambda = 0$. In practice, every local correlation can be decomposed as a convex combination of deterministic correlations. The $\lambda = 0$ term contains this convex decomposition, and as stated earlier, Eve knows exactly which term in the convex decomposition is being used. Hence, once the inputs of Alice and Bob are announced, Eve knows their outcomes exactly, which explains the $p(e|\lambda) = \delta_{e,(a,b)}$ term for each $\lambda = 0$ case.

Recall from Eq. (3) that for such an individual attack to be valid the actual correlation observed by the parties, $p_{AB}^{\text{obs}}(a, b|x, y)$, must be recovered on average. Evaluating the relevant marginal of Eq. (11), this corresponds to the following constraint:

$$q^{\mathcal{L}} p_{AB}^{\mathcal{L}}(a, b|x, y) + q^{\mathcal{NL}} p_{AB}^{\mathcal{NL}}(a, b|x, y) = p_{AB}^{\text{obs}}(a, b|x, y). \quad (12)$$

Therefore, “reversing” the above construction, any *convex combination*—hence, the name of the attack—of a local and a non-local quantum correlation satisfying Eq. (12) can be used to construct a valid CC attack defined by the tripartite classical correlation (11). Still, the best choice of the convex decomposition (12) may strongly depend on the setting in which the CC attack is applied.

As depicted geometrically in Fig. 2, for our purposes of constructing upper bound DIQKD key rates, we assume that it is best for Eve to maximise the probability with which she distributes the local correlation, i.e. $q^{\mathcal{L}}$ in Eq. (12) that we refer to as the *local weight*. A particular correlation being shared, p_{AB}^{obs} , constitutes a point within the probability space that we mark in Fig. 2. Then, maximising $q^{\mathcal{L}}$ corresponds to finding two other points collinear with it: $p_{AB}^{\mathcal{L}}$ contained within the local set \mathcal{L} , and $p_{AB}^{\mathcal{NL}}$ outside of \mathcal{L} but within the quantum set \mathcal{Q} ; such that the ratio of distances of $p_{AB}^{\mathcal{L}}$ to p_{AB}^{obs} and $p_{AB}^{\mathcal{NL}}$ to p_{AB}^{obs} is minimised, see Fig. 2.

In the above argumentation, we have stated that Eve perfectly knows all the outcomes whenever she distributes a local correlation to Alice and Bob. This follows from the fact that, because the local set \mathcal{L} forms a convex polytope in the probability space [4], any local correlation $p_{AB}^{\mathcal{L}}$ can always be decomposed into the extremal points of the polytope, i.e. $p_{AB}^{\mathcal{L}} = \sum_i \mu_i p_{AB}^{\mathcal{L},(i)}$ where it is actually the extremal local correlations $p_{AB}^{\mathcal{L},(i)}$ that are distributed by Eve in every protocol round, each with probability $\mu_i q^{\mathcal{L}}$. Now, as any such $p_{AB}^{\mathcal{L},(i)}$ corresponds to a deterministic strategy

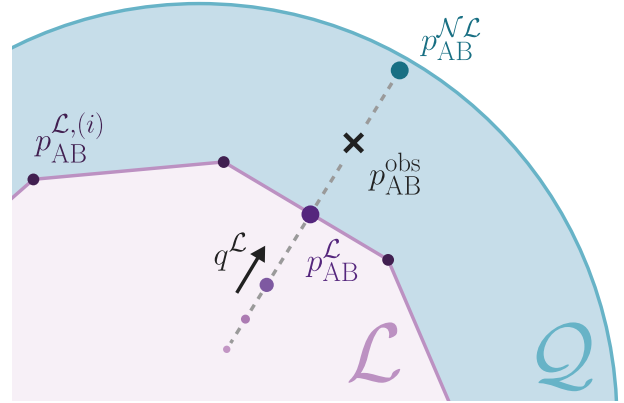


Figure 2: Geometric formulation of the CC attack. The correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ observed by Alice and Bob resides in the quantum set \mathcal{Q} of the correlation space. In the CC attack, Eve decomposes p_{AB}^{obs} into a local correlation $p_{AB}^{\mathcal{L}}$, contained within the local set \mathcal{L} , and a nonlocal, quantum correlation $p_{AB}^{\mathcal{NL}}$, contained within \mathcal{Q} but outside of \mathcal{L} . She performs the decomposition under the constraint that the convex combination of $p_{AB}^{\mathcal{L}}$ and $p_{AB}^{\mathcal{NL}}$, with weights $q^{\mathcal{L}}$ and $1 - q^{\mathcal{L}}$ respectively, reproduces p_{AB}^{obs} on average. At the same time she strives to maximise the local weight $q^{\mathcal{L}}$, which, after fixing $p_{AB}^{\mathcal{NL}}$, corresponds to moving $p_{AB}^{\mathcal{L}}$ in the correlation space along the line connecting $p_{AB}^{\mathcal{NL}}$ and p_{AB}^{obs} in the direction of p_{AB}^{obs} . This finally results in $p_{AB}^{\mathcal{L}}$ lying at the boundary of \mathcal{L} . Moreover, because \mathcal{L} is a convex polytope, $p_{AB}^{\mathcal{L}}$ can be further decomposed into a convex combination of the vertices of \mathcal{L} , corresponding to deterministic correlations $p_{AB}^{\mathcal{L},(i)}$. This allows Eve to possess perfect knowledge of the outcomes whenever distributing a local correlation to Alice and Bob, as she may then equivalently distribute deterministic strategies with predetermined outputs.

with predetermined outputs [4], by tracking which extremal local correlation she uses in every round, Eve is able to perfectly infer the outcomes of both Alice and Bob—for whom it still appears that $p_{AB}^{\mathcal{L}}$ is being shared (on average).

Moreover, as displayed in Fig. 2 within the geometric construction, the maximisation of the local weight, $q^{\mathcal{L}}$, leads to an optimal local correlation $p_{AB}^{\mathcal{L}}$ lying at the border of the local polytope \mathcal{L} . As a consequence, the optimal $p_{AB}^{\mathcal{L}}$ must always belong to one of the facets of \mathcal{L} . This means that not only one may perform such maximisation by solving a linear program, as we now show, but also the facet at which the optimal $p_{AB}^{\mathcal{L}}$ lies can be unambiguously determined. Although identifying facets of the local set may be hard in general [58], once the Bell inequality associated with a particular facet is identified, one can in principle determine the corresponding expression for the local weight analytically, and hence provide an analytic solution to the problem.

In what follows, we succeed in doing so when the observed correlation p_{AB}^{obs} of Alice and Bob arises from a maximally entangled state being shared by them, so that the non-local correlation $p_{AB}^{\mathcal{NL}}$ used in the attack by Eve corresponds to Tsirelson boxes [59], i.e. the

non-local correlation violating maximally the CHSH inequality [4]. Moreover, when analysing the robustness of DIQKD to finite detection efficiency (losses), which requires the use of partially entangled states by the honest users [48], we also obtain analytic results in the limit of the shared state approaching its product form—a feature typically required to reach the highest robustness to losses [48].

3.2 Optimisation of the CC attack via a linear program

For brevity, we drop within this section the subscript $p \equiv p_{AB}$, as all the probabilities refer here to bipartite correlations shared by the honest users—which Eve distributes within the CC attack—unless specified otherwise.

The extremal correlations defining the local polytope, \mathcal{L} in Fig. 2, correspond to deterministic strategies of assigning particular outcomes, a and b , for each combination of measurement settings, x and y [4]. Hence, for a given $m_A m_B n_A n_B$ -scenario considered, there exists $n_A^{m_A} n_B^{m_B}$ such extremal points. We shall label by the vector $\mathbf{p}^\mathcal{L} = (p_i^\mathcal{L})_i$ the set of all such extremal local correlations, and by $\mathbf{q}^\mathcal{L} = (q_i^\mathcal{L})_i$ the vector of the corresponding probabilities that Eve assigns to each of them within the CC attack. On the other hand, we assume the average non-local correlation that she distributes to be a mixture of pre-chosen non-local quantum correlations forming a vector $\mathbf{p}^{\mathcal{NL}} = (p_j^{\mathcal{NL}})_j$, each of which is distributed by Eve with the corresponding probability from the vector $\mathbf{q}^{\mathcal{NL}} = (q_j^{\mathcal{NL}})_j$. Finally, let us recall that for the attack to succeed Eve must reproduce on average the true correlation observed by Alice and Bob, i.e. $p^{\text{obs}} \equiv p_{AB}^{\text{obs}}(a, b|x, y)$ of Eq. (1), for *all* the measurement settings x and y .

In order to optimise the CC attack, Eve seeks a probability vector $\mathbf{q} = \mathbf{q}^\mathcal{L} \oplus \mathbf{q}^{\mathcal{NL}}$ such that the local correlations are distributed as frequently as possible. This corresponds to solving the following linear program [60], which maximises the overall probability of sending any local boxes:

$$\begin{aligned} \mathbf{q}_{\text{CC}}[\mathbf{p}^{\mathcal{NL}}, p^{\text{obs}}] &= \underset{\mathbf{q}}{\text{argmax}} \sum_i q_i^\mathcal{L} & (13) \\ \text{s.t.} \quad & \mathbf{q}^\mathcal{L} \cdot \mathbf{p}^\mathcal{L} + \mathbf{q}^{\mathcal{NL}} \cdot \mathbf{p}^{\mathcal{NL}} = p^{\text{obs}}, \\ & \sum_i q_i^\mathcal{L} + \sum_j q_j^{\mathcal{NL}} = 1, \\ & \forall_{i,j}: \quad 0 \leq q_i^\mathcal{L}, q_j^{\mathcal{NL}} \leq 1, \end{aligned}$$

where the first constraint is just the generalisation of Eq. (12) enforcing Eve to distribute on average the observed correlation, while the other constraints ensure \mathbf{q} to constitute a valid probability vector. Note that the set of extremal local correlations, $\mathbf{p}^\mathcal{L}$, is not an input to the linear program. Rather, it is a predetermined collection defined by the considered scenario

and remains fixed for all programs computed within that scenario, for different choices of $\mathbf{p}^{\mathcal{NL}}$ and p^{obs} .

The above construction requires to specify multiple local correlations, $\mathbf{p}^\mathcal{L}$ (extremal points of the local polytope), and for generality we have also allowed for multiple non-local boxes, $\mathbf{p}^{\mathcal{NL}}$. However, by defining now the effective local correlation as the average $p^\mathcal{L} := \mathbf{q}^\mathcal{L} \cdot \mathbf{p}^\mathcal{L}$, and similarly $p^{\mathcal{NL}} := \mathbf{q}^{\mathcal{NL}} \cdot \mathbf{p}^{\mathcal{NL}}$ for the non-local case, we always recover the binary setting described in the previous section and Fig. 2. In particular, the resulting CC attack is completely specified by the tripartite correlation (11), where now the local weight reads $q^\mathcal{L} := \sum_i q_i^\mathcal{L}$ (and similarly $q^{\mathcal{NL}} := \sum_i q_i^{\mathcal{NL}}$ as the non-local weight), while one must substitute the solution of Eq. (13), $\mathbf{q}_{\text{CC}} = \mathbf{q}_{\text{CC}}^\mathcal{L} \oplus \mathbf{q}_{\text{CC}}^{\mathcal{NL}}$, for the corresponding local and non-local probability vectors.

4 Robustness of non-local correlations

It should be clear from the previous section and the geometric picture that for the CC attack to be applicable the observed correlation, p_{AB}^{obs} in Fig. 2, cannot lie at the border of the quantum set \mathcal{Q} , in which case $q^\mathcal{L}$ is necessarily zero. This, however, never happens in real-life implementations due to the inevitable noise perturbing the desired correlation and forcing it to be decomposable in the form of a convex combination depicted in Fig. 2. The two models applicable to experimental realisations [14, 30, 61], commonly used to verify robustness of DIQKD protocols [1, 2, 22–24, 42], are the scenarios of finite visibility and finite detection efficiency that we summarise below.

4.1 Finite visibility

Although within the DI framework we are restricted to perform the analysis at the level of correlations, the noise models associated with particular implementations are typically defined assuming certain form of quantum states and measurements employed. The finite *visibility*, in particular, is associated with the probability $V \in [0, 1]$ with which Alice and Bob succeed in sharing the intended bipartite state ρ_{AB} , while with probability $1 - V$ it is the maximally mixed state that is rather distributed. As a result, the actual state they share becomes

$$\rho_{AB}(V) := V \rho_{AB} + \frac{1 - V}{d_A d_B} \mathbb{1}_{d_A d_B}, \quad (14)$$

where $d_A = \dim \mathcal{H}_A$ and $d_B = \dim \mathcal{H}_B$.

However, given that Alice and Bob perform projective (von Neumann) measurements for which $n_A = d_A$ and $n_B = d_B$, we may then write their observed correlation (1) as

$$p_{AB}^{\text{obs}}(a, b|x, y) = V Q_{AB}(a, b|x, y) + \frac{1 - V}{n_A n_B}, \quad (15)$$

where by Q_{AB} we denote the ideal correlation shared at $V = 1$. Hence, the finite visibility model is then equivalent to the uniform noise being admixed with all the $n_A n_B$ outcomes occurring with equal probability, so that at $V = 0$ a uniformly random distribution of the outcomes is always observed by the parties, i.e. independently of the measurement settings chosen.

4.2 Finite detection efficiency

The second model of finite detection efficiency is attributed to the problem of photonic losses in optical implementations of DIQKD [14, 30, 61]. At the level of the shared correlation, this results in both Alice and Bob failing to detect any signal with probability $\bar{\eta} := 1 - \eta$, where $\eta \in [0, 1]$ is the *detection efficiency* parameter. Such a non-detection event constitutes then an additional outcome for any of the measurement used, a ‘no-click’, that we denote by \emptyset . For example, if the original outputs of the parties’ devices are binary, $a, b \in \{0, 1\}$, after the inclusion of non-detection events (which changes the scenario to have $n_A = n_B = 3$ outcomes) the observed correlation can be expressed in a convenient table format as⁵

$$p_{AB}^{\text{obs}}(a, b|x, y) = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 Q_{00}^{xy} & \eta^2 Q_{01}^{xy} & \eta \bar{\eta} Q_0^x \\ 1 & \eta^2 Q_{10}^{xy} & \eta^2 Q_{11}^{xy} & \eta \bar{\eta} Q_1^x \\ \hline \emptyset & \bar{\eta} \eta Q_0^y & \bar{\eta} \eta Q_1^y & \bar{\eta}^2 \end{array} \quad (16)$$

where by $Q_{ab}^{xy} := Q_{AB}(a, b|x, y)$ we denote again the ideal quantum correlation observed by the users for $\eta = 1$, with its marginals of Alice and Bob reading $Q_a^x = \sum_b Q_{ab}^{xy}$ and $Q_b^y = \sum_a Q_{ab}^{xy}$, respectively.

Note, that if one wanted to consider the effect of imperfect visibility and finite detection efficiency at the same time, it suffices to substitute $p_{AB}^{\text{obs}}(a, b|x, y)$ from Eq. (15) for Q_{ab}^{xy} in the correlation (16).

Moreover, it is worth noting that in some protocols, *binning* the ‘no-click’ outcomes is considered for security enhancement, which in the table notation corresponds to aggregating the rows and columns for the ‘no-click’ events with the other proper outcomes. An example is provided by the CHSH protocol, which we analyze in the subsequent section.

5 Applications to DIQKD protocols

In the following, we apply the CC attack to derive upper bounds on the one-way and two-way key rates in noisy scenarios, i.e. as functions of detection efficiency η and visibility V , for a range of DIQKD protocols. Most importantly, as a result, we determine *critical* visibilities V_{crit} and detection efficiencies η_{crit} , below which our upper bounds on the key rates become negative and preclude a secure experimental realisation of

⁵Dealing with more than two outcomes leads just to more rows and columns in Eq. (16).

a given protocol. As these critical values signify then lower bounds on minimal robustness parameters that the protocol can tolerate—below these values there exists an explicit attack, the CC attack, that invalidates the security—by comparing them with the ones obtained from the state-of-the-art security proofs, one can judge how much room there exists for potential improvement of the latter.

In order to apply the CC attack and upper-bound the key rate in a noisy one-way (6) or two-way (7) DIQKD protocol, one must first specify the correlation $Q_{AB}(a, b|x, y)$ that would be shared by the parties in the absence of imperfections. The true noisy correlation being observed, $p_{AB}^{\text{obs}}(a, b|x, y)$, is then decomposed within the attack into the local and non-local parts. Although the local contribution is determined via the linear program (13), the eavesdropper must specify in advance the set of nonlocal correlations \mathbf{p}^{NL} to be used within the convex decomposition. In this work, we choose \mathbf{p}^{NL} to consist of only one correlation, namely, the noiseless $Q_{AB}(a, b|x, y)$. We find this choice to be optimal for our purposes by heuristic methods, however, we leave it open whether the upper bounds on key rates derived under this choice can be further improved by performing a rigorous optimisation of \mathbf{p}^{NL} .

In this section, we consider the application of the CC attack to particular DIQKD protocols, which exhibit state-of-the-art robustness to noise. In particular, we summarise the experimentally relevant bounds on critical visibilities and detection efficiencies below which the protocols become vulnerable to the CC attack and thus insecure. As an example, an explicit derivation of the CC-based upper bound on the key rate and the resulting lower bounds on tolerable noise levels are presented for the CHSH-based protocol with deterministic binning of the non-detection events, while similar derivations applicable to the other protocols considered are relegated to the [Appendices](#).

5.1 Protocols based on the CHSH violation

Within the canonical CHSH-based protocol [1, 2], the parties strive to obtain correlations maximally violating the CHSH inequality [20]. The value of the CHSH violation may then be used to construct a lower bound on the DW rate (5) [1, 2, 22], which, if the violation is high enough, may be positive and thus certify the possibility of distilling a secure cryptographic key. For this to be possible, Alice uses two binary-outcome measurements, labelled by her input $x \in \{0, 1\}$, while Bob uses three labelled by the inputs $y \in \{0, 1, 2\}$, corresponding to a scenario with $m_A = 2$ and $m_B = 3$. In each round they select their inputs randomly, and only the rounds with $x, y \in \{0, 1\}$ are used to estimate the CHSH violation, whereas only the rounds with $(x^*, y^*) := (0, 2)$, i.e. with the *key settings* chosen, are

used to distil the key. This formally constitutes the 2322-scenario⁶, also when finite visibility ($V < 1$ in Eq. (15)) is accounted for, while in case of imperfect detection ($\eta < 1$ in Eq. (16)) it becomes the 2333-scenario with the third extra outcome corresponding to the ‘no-click’ event observed by any of the parties.

Within the canonical protocol [1, 2] the security relies on the CHSH-scenario with binary outcomes. Therefore, in case of imperfect detection and the 2333-scenario, the standard technique used in security proofs is to have the parties bin the third ‘no-click’ outputs, i.e. assign them to one of the two ‘proper’ measurement outcomes (0 or 1), for the inputs $x, y \in \{0, 1\}$ used to estimate the CHSH-violation, with $x^* = 0$ being also used by Alice to distil the key. Although the construction of the CC attack may be performed for any given correlation, it must include all the steps conducted within the protocol being considered, in particular, also the binning procedure.

In what follows, we assume the typical choice of binning [22, 23, 35, 42], i.e. the *deterministic* assignment of all the ‘no-clicks’ \emptyset to one of ‘proper’ outcomes, say 0, by each party. Nonetheless, within **Appendices** we consider the option of not binning at all, as well as other binning strategies, which in combination with any preprocessing applied by Alice on her outcome for $x^* = 0$, i.e. $p_{A'|A}$ in Eq. (4), correspond to just instances of stochastic maps that may be further optimised over to determine a preprocessing-independent upper bound on the one-way key rate.

5.1.1 Generating the observed non-local correlations.

Here, we study a family of protocols inspired by the original CHSH construction and adopt the convention of [22], in which the ideal correlation $Q_{AB}(a, b|x, y)$ available to the parties should be understood as the one obtained by them when sharing a *partially entangled state* of two qubits:

$$|\psi_\theta\rangle := \cos\left(\frac{\theta}{2}\right)|00\rangle + \sin\left(\frac{\theta}{2}\right)|11\rangle, \quad (17)$$

with the measurements of Alice and Bob, M_a^x and M_b^y in Eq. (1), corresponding to eigenstate projectors of the dichotomic observables A_x and B_y , respectively: $A_0 = B_2 = \sigma_z$ for the key settings (x^*, y^*) , while A_1 and $B_{0/1}$ are chosen to maximise the CHSH-type functional:

$$S_{\text{det}}(\eta, \theta) := \eta^2 \langle B_0(A_0 + A_1) + B_1(A_0 - A_1) \rangle + 2\eta\bar{\eta} \langle A_0 + B_0 \rangle + 2\bar{\eta}^2, \quad (18)$$

which accounts already for the finite detection efficiency, $\eta < 1$ in Eq. (16), and assumes deterministic

⁶Recall that by $m_A m_B n_A n_B$ we denote a scenario with m_A (m_B) measurement settings and n_A (n_B) possible outcomes for each measurement setting on Alice’s (Bob’s) side.

binning of the ‘no-click’ events. Note that, due to the linearity of the expression (18), the above choice of measurements remains optimal when also the finite visibility, $V < 1$ in Eq. (14), is considered⁷.

5.2 One-way CHSH protocols involving maximally entangled states

5.2.1 Finite detection efficiency

Firstly, we sketch the calculation of the upper bound on the one-way key rate (6) based on the CC attack (see Apps. C and D for a more detailed derivation) for the above CHSH-based protocol with finite detection efficiency η and deterministic binning. For protocols in which Alice does not announce publicly any variable M and bins her key setting outcome A deterministically, the bound (6) reads

$$r_{1\text{-way,det}}(A \rightarrow B|A') \leq H(A'|E)_{\text{det}} - H(A'|B)_{\text{det}}, \quad (19)$$

where the binary variable A' is obtained by transforming the ternary outcome A of Alice’s measurement with $x^* = 0$ by the stochastic map

$$\mathcal{S}_{\text{det}} := \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad (20)$$

responsible for binning the ‘no-click’ events (last column) deterministically onto the ‘0’ outcome. For instance, the marginal probability of Alice then reads

$$p_{A'}^{\text{obs}}(a'|x^*) = \sum_a \mathcal{S}_{\text{det}}(a'|a) p_A^{\text{obs}}(a|x^*), \quad (21)$$

describing now the distribution of A' rather than A . Although it is the full correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ that determines the value of the local weight $q^{\mathcal{L}}$ within the CC attack (see Eq. (13)), the key is distilled only from the (x^*, y^*) -rounds. As a consequence, both entropies in Eq. (19) are computed for the key settings and we may drop for convenience the conditioning on (x^*, y^*) in all the following expressions, so that, e.g., $Q_{ab} := Q_{ab}^{x^*y^*}$, $Q_a^A := Q_a^{x^*}$, $Q_b^B := Q_b^{y^*}$, or $p_{AB}^{\text{obs}}(a, b) := p_{AB}^{\text{obs}}(a, b|x^*, y^*)$.

We calculate first the EC-term $H(A'|B)_{\text{det}}$ in Eq. (19), which depends solely on the correlation being observed. After applying the stochastic map (20) on the outcome of Alice in Eq. (16), we obtain the resulting shared correlation as

$$p_{A'B}^{\text{obs}}(a', b) = \begin{array}{c|ccc} a' \backslash b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 Q_{00} & \eta^2 Q_{01} & \bar{\eta}\eta Q_0^A \\ & +\eta\bar{\eta} Q_0^B & +\eta\bar{\eta} Q_1^B & +\eta^2 \\ \hline 1 & \eta^2 Q_{10} & \eta^2 Q_{11} & \eta\bar{\eta} Q_1^A \end{array}, \quad (22)$$

whose first row is obtained by summing the first and third rows in Eq. (16). Now, the conditional probability distribution of Alice is obtained by dividing the

⁷With $\langle X \rangle = \text{Tr}\{\psi_\theta X\}$ in Eq. (18) being then replaced by $\langle X \rangle = \text{Tr}\left\{ \left[V\psi_\theta + (1-V)\frac{\mathbb{1}}{4} \right] X \right\}$ according to Eq. (14).

columns in Eq. (22) by the corresponding marginal probabilities of Bob, $(p_B^{\text{obs}}(0), p_B^{\text{obs}}(1), p_B^{\text{obs}}(\emptyset)) = (\eta Q_0^B, \eta Q_1^B, \bar{\eta})$ obtained by summing the columns in Eq. (22), i.e.:⁸

$$p_{A'|B}^{\text{obs}}(a'|b) = \frac{p_{A'B}^{\text{obs}}(a', b)}{p_B^{\text{obs}}(b)}$$

$a' \setminus b$	0	1	\emptyset
0	$\eta \frac{Q_{00}^B}{Q_0^B} + \bar{\eta}$	$\eta \frac{Q_{01}^B}{Q_1^B} + \bar{\eta}$	$\eta Q_0^A + \bar{\eta}$
1	$\eta \frac{Q_{10}^B}{Q_0^B}$	$\eta \frac{Q_{11}^B}{Q_1^B}$	ηQ_1^A

(23)

As a result, we may directly compute the relevant conditional entropy as

$$H(A'|B)_{\text{det}} = \sum_b p_B^{\text{obs}}(b) H(A'|B=b) \quad (24)$$

$$= \eta Q_0^B h \left[\frac{\eta Q_{10}^B}{Q_0^B} \right] + \eta Q_1^B h \left[\frac{\eta Q_{11}^B}{Q_1^B} \right] + \bar{\eta} h[\eta Q_1^A],$$

where $h[x] := -x \log_2 x - (1-x) \log_2(1-x)$ is the binary entropy function.

On the contrary, in order to determine the PA-term $H(A'|E)_{\text{det}}$ in (19), one needs to find the value of $q^{\mathcal{L}}$ for the specified correlation. This can be done by means of linear programming, but often also analytically, as discussed in Sec. 3. Let us also recall that whenever Eve distributes a local correlation within the CC attack, she possesses full knowledge about the outcome of Alice, A , and hence of A' , since it is obtained via a deterministic transformation. Therefore, the PA term is completely determined by the non-local rounds in which the noiseless correlation $Q(a, b|x, y)$ is distributed, so that

$$H(A'|E)_{\text{det}} = (1 - q^{\mathcal{L}}) h[Q_0^A] \quad (25)$$

corresponds to the entropy of the marginal Q_a^A multiplied by the probability of a round being nonlocal.

Finally, we obtain an upper bound on the key rate as a function of detection efficiency η by subtracting the EC-term (24) from the PA-term (25), i.e.:

$$r_{1\text{-way, det}}(A \rightarrow B|A') \leq (1 - q^{\mathcal{L}}) h[Q_0^A] - \eta Q_0^B h \left[\frac{\eta Q_{10}^B}{Q_0^B} \right] - \eta Q_1^B h \left[\frac{\eta Q_{11}^B}{Q_1^B} \right] - \bar{\eta} h[\eta Q_1^A], \quad (26)$$

which applies for any correlation (16), given the deterministic binning of no-clicks and one-way communication in the protocol.

However, recall that the CC-based upper bound (26) requires the local weight $q^{\mathcal{L}}$ to be determined

⁸Note that although it leads to an easy calculation of the formula for the conditional entropy $H(A'|B)$, this division, and hence Eq. (23), is incorrect for the extremal values $\eta = 0$ and $\eta = 1$ due to dividing by 0. In these edge cases, to calculate $H(A'|B)$ one should rather use the chain rule $H(A'|B) = H(A', B) - H(B)$ on the joint probability distribution (22), which ultimately yields the same result as in Eq. (24). This applies as well to other instances of conditional entropy calculated in the paper.

for a particular p_{AB}^{obs} . We present first the solution when Alice and Bob share a maximally entangled Bell state, $|\Phi^+\rangle$, i.e. set $\theta = \pi/2$ in Eq. (17), which yields $Q_{ab} = \delta_{ab}/2$ and $Q_a^A = Q_b^B = 1/2$ for the key measurements $A_0 = B_2 = \sigma_z$. The other measurements maximising the expression (18) turn out then to be the standard CHSH-optimal observables, i.e.

$$A_1 = \sigma_x, \quad B_{0/1} = \frac{1}{\sqrt{2}} (\sigma_z \pm \sigma_x). \quad (27)$$

We show in App. B.2 how to determine then the maximal local weight analytically, which reads

$$q^{\mathcal{L}} = (1-\eta) \left(1 + \left(3 + 2\sqrt{2} \right) \eta \right) \quad \text{for } \eta > \eta_{\text{loc}} \quad (28)$$

where $\eta_{\text{loc}} := 2(\sqrt{2} - 1) \approx 82.8\%$ is the detection efficiency below which the resulting correlation (16) becomes local [48, 62] and, thus, disallows any DIQKD to be possible. Finally, we can write Eq. (26) as

$$r_{1\text{-way, det}}(A \rightarrow B|A') \leq \left(3 + 2\sqrt{2} \right) \eta^2 - 2 \left(1 + \sqrt{2} \right) \eta - \frac{\eta}{2} h[\eta] - \bar{\eta} h \left[\frac{\eta}{2} \right], \quad (29)$$

which becomes negative below $\eta_{\text{crit}} \approx 89.16\%$. This formally demonstrates that for detection efficiencies $\eta_{\text{loc}} \leq \eta \leq \eta_{\text{crit}}$, no positive key is possible despite the correlation (16) being non-local [41]. We include $\eta_{\text{crit}} \approx 89.16\%$ in Tab. 1 (see the penultimate column for the 2333-scenario) presenting it against the best-known efficiency threshold, $\eta_{\text{DW}}^{\uparrow} \approx 90.78\%$, above which the DW rate (5) is assured to be positive [22]. Hence, it follows that the true⁹ DW-threshold fulfils $89.16\% \leq \eta_{\text{DW}} \leq 90.78\%$, with the CC attack leaving less than 2% for the improvement of $\eta_{\text{DW}}^{\uparrow}$ by devising stronger lower bounds on the DW rate.

Note that the above efficiency window applies when considering the most general eavesdropping attacks. On one hand, the DW rate (5) is valid for coherent attacks despite the deterministic binning, as the EAT still holds, see the discussion above Eq. (6). On the other, as we consider a particular attack, by improving its strength η_{crit} can only be increased.

In Fig. 3a) we explicitly compare the upper bound (29) with the analytic lower bound on the DW rate (5) established in Ref. [22] as a function of η . It can be seen that the CC-based upper bound remains relatively tight in the whole region of positive key rates, with the maximal difference between the two bounds never exceeding 0.15 of a bit per round.

5.2.2 Finite visibility

From the above analysis, it is now straightforward to determine the CC-based upper bound on the key rate if instead the finite visibility ($V < 1$) is considered.

⁹Given deterministic binning of ‘no-click’ events.

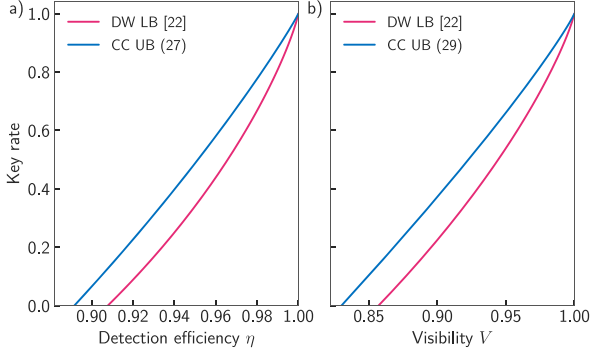


Figure 3: **CC-based upper bounds vs lower bounds on the DW rate (5) for the CHSH-based protocols involving maximally entangled states**, as a function of (a): the detection efficiency η in the 2333 scenario; and (b): the visibility V in the 2322 scenario. *Blue lines* correspond to the analytic CC-based upper bounds on key rates (29) and (31), whereas *red lines* are the corresponding lower bounds on the DW rate (5) derived in Ref. [22]. The points at which the curves cross the zero in (a) and (b) are the critical detection efficiencies and visibilities cited for the 2333- and 2322-scenarios, respectively, in the ‘none’ column of Tab. 1.

It is obtained by letting $\eta = 1$ within the EC-term (24) and replacing therein the noiseless correlation Q_{ab} with $P_{ab} := VQ_{ab} + (1 - V)/4$ in accordance with Eq. (15). On the other hand, the PA term (25) is left intact, as Eve again distributes the noiseless correlation Q_{ab} within the non-local rounds of our CC attack. Hence, focusing again on Q_{ab} maximally violating the CHSH inequality, we have $P_{ab} = (V/2)\delta_{ab} + (1 - V)/4$ and $P_a^A = P_b^B = 1/2$ for (x^*, y^*) , while measurements (27) remain optimal for the other settings. The complete correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ of Eq. (15) constrains then the maximal local weight to (see [41] and App. B.1):

$$q^{\mathcal{L}} = (1 - V)/(1 - 1/\sqrt{2}) \quad \text{for } V > 1/\sqrt{2}, \quad (30)$$

with $1/\sqrt{2}$ being the well-known locality threshold for Werner states. As a result, we obtain for this 2322-scenario¹⁰ the following CC-based upper bound:

$$r_{1\text{-way}}(A \rightarrow B) \leq V(2 + \sqrt{2}) - h\left[\frac{1 + V}{2}\right] - \sqrt{2} - 1, \quad (31)$$

which ceases to be positive at $V_{\text{crit}} \approx 83.00\%$. Hence, given the best-known value of V above which the DW rate is positive, $V_{\text{DW}}^{\uparrow} \approx 85.70\%$ [22], the CC attack constrains narrowly any further improvement of this threshold to about 2.7%, i.e. $83.00\% \leq V_{\text{DW}} \leq 85.70\%$, see the first row of column ‘none’ in Tab. 1. Again, the above window in which the true visibility threshold lies is valid for coherent attacks of the eavesdropper, by the same arguments as discussed in the finite efficiency case.

¹⁰For finite visibility ($V \leq 1$) but perfect detection efficiency ($\eta = 1$), no binning strategy is required due to outcomes of measurements remaining binary.

Furthermore, we explicitly compare in Fig. 3b) the upper bound (31) with the corresponding analytic lower bound on the DW rate (5) found in Ref. [22] as a function of V . Similarly to the case of finite detection efficiency and Eq. (29), we observe that the CC-based upper bound remains relatively tight in the whole region of positive key rates, and the difference between the two bounds again does not exceed 0.15 of a bit per round.

5.2.3 22- scenarios

We further repeat the above derivation for the less favourable situation in which Bob, similarly to Alice, uses the same measurement setting to distil the key as for the CHSH violation. This yields then 2233- and 2222-scenarios for finite detection efficiency and visibility, respectively, with $Q_{ab} = (2 + (-1)^{a \oplus b} \sqrt{2})/8$. Following the procedure of Ref. [22], we compute the thresholds above which the DW rate is guaranteed to be positive and include them in Tab. 1. These are then higher but so are the CC-based critical values (see also App. D.1)—with the CC attack proving itself again to be very effective.

5.2.4 Noisy preprocessing

On the other hand, by performing noisy preprocessing (random bit-flip) of her key-setting outcome [17], Alice may improve the robustness of the DIQKD protocol [22, 23], with the tolerable noise-levels dropping then to $\eta_{\text{DW}}^{\uparrow} \approx 90.30\%$ and $V_{\text{DW}}^{\uparrow} \approx 83.83\%$ [22], see the the column ‘noisy’ in Tab. 1 for the corresponding 2333- and 2322-scenarios, respectively (and similarly for the 2233- and 2222-scenarios). However, after incorporating the noisy preprocessing step into the CC attack—see App. D.1 for analytic expressions, also for the setting in which Alice and Bob bin their ‘no-click’ outcomes randomly—we obtain strict lower bounds on the tolerable noise-thresholds for the DW rate⁹ as $\eta_{\text{crit}} \approx 88.52\%$ and $V_{\text{crit}} \approx 80.85\%$, which again leave only a couple of percent for potential improvement.

5.2.5 Arbitrary preprocessing

Crucially, any binning of ‘no-clicks’ or noisy preprocessing strategy constitutes just a special case of a stochastic map $A \rightarrow A'$ in Eq. (4), while the CC attack allows us, in fact, to determine an upper bound on the one-way key rate (4) that applies for *any preprocessing*. In particular, by resorting to heuristic methods (see App. D.1.3 for details) we maximise the upper bound (6) over all maps $A \rightarrow A'$ and $A \rightarrow M$, as in Eq. (4), so that it yields critical thresholds on detection efficiency and visibility, $\tilde{\eta}_{\text{crit}}$ and \tilde{V}_{crit} , that are universally valid for any one-way protocol given a particular correlation p_{AB}^{obs} .

As shown in the column ‘any’ of Tab. 1, we observe that for both 2322- and 2222-scenarios the uni-

DIQKD protocols involving maximally entangled states

<i>Preprocessing:</i>	two-way		one-way				
			<i>any</i>	<i>noisy</i>		<i>none</i>	
	Finite visibility V						
Scenario:	V_{crit}	$V_{\text{a.d.}}^\uparrow$	\check{V}_{crit}	V_{crit}	V_{DW}^\uparrow	V_{crit}	V_{DW}^\uparrow
2322	74.45	88.0	80.85	80.85	83.83	83.00	85.70
2222	78.36	84.6	88.52	88.52	92.38	90.61	93.76
Finite detection efficiency η							
Scenario:	η_{crit}	$\eta_{\text{a.d.}}^\uparrow$	$\check{\eta}_{\text{crit}}$	η_{crit}	$\eta_{\text{DW}}^\uparrow$	η_{crit}	$\eta_{\text{DW}}^\uparrow$
2333	85.36*	93.7	85.36	88.52*	90.30	89.16*	90.78
2233	87.87*	91.7	92.64	93.59*	95.84	94.80*	96.62

* with deterministic binning of ‘no-click’ events.

Table 1: **Critical visibilities** V_{crit} **and detection efficiencies** η_{crit} (in %) derived with help of the CC attack, below which no DIQKD protocol can be made secure when relying on measurements of maximally entangled states within each of the $m_A m_B n_A n_B$ -scenarios listed. The left-most critical values apply to all DIQKD protocols and are compared against the thresholds attained by two-way protocols involving advantage distillation (a.d.) [42]. The critical noise parameters are tightened for one-way protocols, for which a stricter upper bound on key rate (6) applies and varies between various strategies of data preprocessing: *any* (found by heuristic search), *noisy* [22–24] or *none*. The latter two cases are compared against the thresholds determined by lower-bounding the Devetak-Winter (DW) rate (5), which can be computed for all the scenarios considered [22]. All the stated values are valid in presence of coherent attacks, apart from the a.d.-based thresholds [42] (second column) that are derived in presence of collective attacks only.

versal values \check{V}_{crit} coincide with the CC-based critical visibilities obtained for noisy preprocessing, which can thus be considered optimal in terms of robustness against the CC attack. For finite detection efficiency, we are able to find analytically the minimal $\check{\eta}_{\text{crit}} = \frac{1}{4}(2 + \sqrt{2}) \approx 85.36\%$ for the 2333-scenario, as it is attained by a preprocessing strategy in which only the map $A \rightarrow M$ in Eq. (4) is required, with Alice publicly announcing in each round under the binary variable M whether or not she observes a ‘no-click’. From the perspective of our CC attack, as ‘no-clicks’ may happen only in the rounds in which Eve distributes a local correlation and has perfect knowledge of A , announcing M does not provide any ‘extra’ information to Eve, but helps Bob perform the EC—diminishing $H(A|B)$. In the 2233-scenario the situation is slightly different: although a similar strategy of signalling ‘no-clicks’ proves better than deterministic binning followed by noisy preprocessing, the optimal preprocessing that is most robust against the CC attack, yielding $\check{\eta}_{\text{crit}} \approx 92.64\%$, actually corresponds to randomly binning the ‘no-click’ events before performing noisy preprocessing. We elaborate on these findings in App. D.1.3.

5.3 Two-way CHSH protocols involving maximally entangled states

5.3.1 Finite detection efficiency

In order to establish critical noise thresholds that hold for *all DIQKD protocols*, i.e. also ones exploiting two-way communication, we resort to the CC-based up-

per bound (7) based on the intrinsic information (8). However, we write it in terms of the conditional mutual information as

$$r_{2\text{-way}}(A \leftrightarrow B) \leq I(A: B|F), \quad (32)$$

for any given mapping $p(F|E)$ that Eve applies on her random variable, $E \rightarrow F$, within the CC attack.

As in the one-way case, we consider firstly the noisy correlations (16) that incorporate finite detection efficiency $\eta < 1$, with Q_{ab}^{xy} violating again maximally the CHSH inequality. However, we are then unable to find (via an extensive numerical search) a map such that $I(A: B|F) = 0$ for any $\eta > \eta_{\text{loc}} \approx 82.8\%$, as defined below Eq. (28). In particular, for every choice of $p(F|E)$ we make, the upper bound (32) can be made vanishing only trivially, i.e. when the noisy correlation (16) becomes local.

We then consider a simpler version of the protocol in which, again, Alice and Bob bin deterministically their ‘no-click’ outcomes before performing any two-way processing of their bits, which is the case in current two-way DIQKD protocols involving an *advantage distillation* (a.d.) procedure [42]. Then, by using the map $p(F|E)$ proposed by us in [41], see also App. D.2.1, we arrive at an upper bound (32) of the form:

$$r_{2\text{-way,det}}(A' \leftrightarrow B') \leq \tilde{p} \sum_{a' \neq b'} h \left[\frac{q^{N\mathcal{L}} Q_{00} + Q_{a'b'}^\eta}{\tilde{p}} \right] - \tilde{p} H \left\{ \frac{Q_{01}^\eta}{\tilde{p}}, \frac{Q_{10}^\eta}{\tilde{p}}, \frac{q^{N\mathcal{L}} Q_{00}}{\tilde{p}}, \frac{q^{N\mathcal{L}} Q_{11}}{\tilde{p}} \right\} \quad (33)$$

where A' (B') is now the key-setting outcome of Alice (Bob) after deterministic binning of ‘no-clicks’, $Q_{ab}^\eta :=$

$\eta Q_{ab} + \eta \bar{\eta} Q_{11}$, $\tilde{p} := \eta + 2\eta \bar{\eta} Q_{11} + (q^{N_C} - \eta)(Q_{00} + Q_{11})$, and we define the entropy of any probability vector $(p_i)_i$, satisfying $\forall_i : p_i \geq 0$ and $\sum_i p_i = 1$, as $H\{(p_i)_i\} := -\sum_i p_i \log_2 p_i$. Substituting for the correlation Q_{ab}^{xy} that yields: maximal CHSH violation, $Q_{ab} = \delta_{ab}/2$, and the maximal local weight (28); we obtain the two-way equivalent of Eq. (29) as

$$r_{2\text{-way,det}}(A' \leftrightarrow B') \leq \eta \left(2 \left(1 + \sqrt{2} \right) \eta - 2\sqrt{2} - 1 \right) \times \left(1 - h \left[\frac{\bar{\eta}}{1 - 2 \left(1 + \sqrt{2} \right) \bar{\eta}} \right] \right), \quad (34)$$

which exhibits a zero at $\eta_{\text{crit}} = \frac{1}{4} (2 + \sqrt{2}) \approx 85.36\%$.

Hence, by convexity of the two-way upper bound (7) (see [41] and App. A), this implies that $r_{2\text{-way,det}} = 0$ for any $\eta_{\text{loc}} \approx 82.8\% \leq \eta \leq \eta_{\text{crit}}$, and no DIQKD protocol is possible⁹ within this range of η despite the shared correlation being non-local. We include the above CC-based η_{crit} in the first column of Tab. 1, where it consistently lower-bounds all the best-known tolerable detection efficiencies derived for one-way protocols⁹, $\eta_{\text{DW}}^\uparrow$ [22], as well as for two-way protocols involving the a.d. procedure, $\eta_{\text{a.d.}}^\uparrow$ [42]¹⁰. Moreover, it coincides exactly with $\tilde{\eta}_{\text{crit}}$, so that for the above 2333-scenario no two-way protocol⁹ may be more robust against our CC attack than the one-way protocol with optimal preprocessing. This is not the case for the 2233-scenario with $x^*, y^* \in \{0, 1\}$ and $Q_{ab} = (2 + (-1)^{a \oplus b} \sqrt{2})/8$, which upon being substituted into Eq. (33) leads to $\eta_{\text{crit}} = 3(1 - 1/\sqrt{2}) \approx 87.87\%$ that, however, is only 4% away from best-known threshold $\eta_{\text{a.d.}}^\uparrow \approx 91.7\%$ [42]. The derivation can be found in App. D.2.1, where we also deal with the special case of $x^* = y^* = 1$, for which a different $p(F|E)$ must be chosen for the upper bound (32) to provide a non-trivial $\eta_{\text{crit}} \approx 87.47\% (> \eta_{\text{loc}})$.

5.3.2 Finite visibility

As in the case of one-way protocols, we repeat the above construction when finite visibility ($V < 1$) is considered instead, and no binning procedure is necessary. However, this corresponds to the special case of correlations considered by us already in [41] (see Eq. (14) therein with $\theta = \pi/4$), which leads to

$$r_{2\text{-way}}(A \leftrightarrow B) \leq \frac{(2 + \sqrt{2})V - 3\sqrt{2} + 2}{2(2 - \sqrt{2})} \times \left(1 - h \left[\frac{4V - 2\sqrt{2}}{(2 + \sqrt{2})V - 3\sqrt{2} + 2} \right] \right) \quad (35)$$

that ceases to be positive below $V_{\text{crit}} \approx 74.45\%$ (see the top-left entry of Tab. 1). As a result, within the

¹⁰Note that the two-way advantage distillation (a.d.) procedure [42] is superior over the one-way protocol with noisy preprocessing [63] only for the 2222 and 2233 scenarios considered in Tab. 1, for which indeed $V_{\text{a.d.}}^\uparrow < V_{\text{DW}}^\uparrow$ and $\eta_{\text{a.d.}}^\uparrow < \eta_{\text{DW}}^\uparrow$.

range $1/\sqrt{2} \leq V \leq V_{\text{crit}}$ there is strictly *no* possibility for any standard DIQKD protocol to yield positive keys, while the correlations remain non-local [41]. We also construct the equivalent of the upper bound (35) for the 2222-scenario, in which case $V_{\text{crit}} \approx 78.36\%$ —see App. D.2.2 but also Tab. 1 where the threshold value for the a.d.-protocol is also listed, $V_{\text{a.d.}}^\uparrow \approx 84.6\%$ [42], which according to the CC attack may thus be improved only by at most $\approx 6\%$.

5.4 One-way CHSH protocols involving partially entangled states

Since the seminal work of Eberhard [48] it is well known that in order to get the highest robustness to finite detection efficiency in observing Bell-violation, one should consider correlations obtained by measuring partially entangled states, i.e. as in Eq. (17) with $\theta \neq \pi/2$ and, in particular, in the limit $\theta \rightarrow 0$. On the contrary, this is not the case when finite visibility $V < 1$ is considered instead, as then setting $\theta = \pi/2$ always yields the highest CHSH violation. That is why, we repeat the above one-way key analysis for $\eta < 1$ where (as already stated in Sec. 5.1.1) we choose the measurements of Alice and Bob such that the CHSH functional (18) is maximised for a given value of η and θ . Each maximal value of the functional, on the other hand, allows us then to directly compute valid⁹ lower-bounds on the attainable DW rate (5), also when accounting for noisy preprocessing [22].

In Fig. 4, we present the corresponding thresholds, $\eta_{\text{DW}}^\uparrow$ [22], above which the DW rate is guaranteed to be positive (*dot-dashed lines*) as a function of the angle θ defining the the partially entangled state (17). Crucially, we compare these with the critical detection efficiencies, η_{crit} , obtained with help of the CC attack (*solid lines*). We compute the latter by resorting to the upper bound (26) and substituting for the correlation $Q_{ab}^{xy}(\theta, \eta)$ maximising Eq. (18), which in turn specifies the maximal local weight $q^{\mathcal{L}}(\theta, \eta)$ obtained via a linear program. However, in the limit of the partially entangled state approaching its separable form (see App. B.2.2), we evaluate analytically $q^{\mathcal{L}}(\theta \rightarrow 0, \eta) = 1 - \eta(3\eta - 2)$. As this is the limit in which the highest robustness to imperfect detection is exhibited, this allows us to determine analytically (see App. D.3) the minimal η_{crit} allowed by the CC attack as $3/4 = 75\%$ and $(\sqrt{21} - 3)/2 \approx 79.13\%$ for deterministic binning without and with inclusion of noisy preprocessing, respectively—see *red* and *blue solid curves* in Fig. 4 and their values at $\theta \rightarrow 0$, while the values at $\theta = \pi/2$ consistently coincide with the ones stated in Tab. 1.

We observe that the noisy preprocessing that introduces bit-flip errors onto the bit-string of Alice in an almost uniform manner ($\mathbf{p} \rightarrow 1/2$), which is known to significantly lower the threshold $\eta_{\text{DW}}^\uparrow$ [22–24], actually

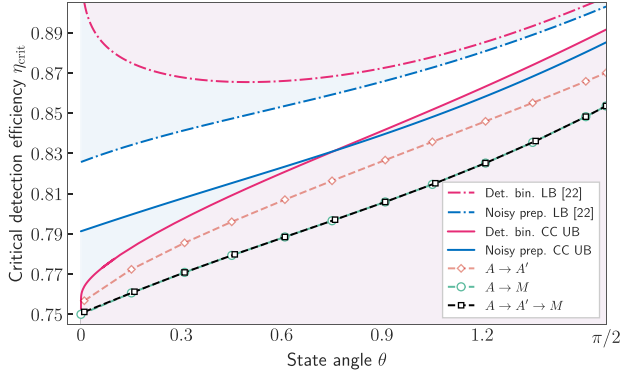


Figure 4: **Critical detection efficiencies as a function of the θ -angle parametrising the partially entangled state** involved in a one-way DIQKD protocol. *Dot-dashed curves* describe thresholds, η_{DW}^\dagger , above which the DW rate (5) is guaranteed to be positive [22], while *solid curves* denote critical values, η_{crit} , below which the CC attack excludes the possibility of key distillation; in both cases the inconclusive outcomes are binned deterministically without (red) or with (blue) inclusion of noisy (with $p \rightarrow 1/2$) preprocessing. The critical efficiencies may be further diminished by optimising heuristically over the preprocessing strategies, i.e. stochastic maps in Eq. (4) that generally encompass the operations of Alice: manipulating somehow her ternary output ($A \rightarrow A'$), publicly announcing some form of her preprocessed variable ($A \rightarrow M$), or both ($A \rightarrow A' \rightarrow M$).

improves the effectiveness of the CC attack for small θ -angles—note the *blue line* crossing the *red line* in Fig. 4. As a result, for deterministic binning and noisy preprocessing the CC attack provides a very stringent restriction, $79.13\% \leq \eta_{\text{DW}} \leq 82.57\%$, on any potential improvement of the minimal tolerable detection efficiency—see the narrow $\approx 3\%$ gap at $\theta \rightarrow 0$ between *blue solid and dot-dashed curves* in Fig. 4.

As the attainable threshold $\eta_{\text{DW}}^\dagger = 82.57\%$ of [22] has recently been improved by Brown *et al.* [35] and Masini *et al.* [47], we present the corresponding two best-known thresholds in Tab. 2. We compare them explicitly against the critical efficiencies allowed by the CC attack, which we are able to evaluate having access to the exact correlations used, and the particular bit-flip strength p employed at the noisy-preprocessing stage in [35, 47], thanks to the courtesy of the authors. Strikingly, the CC attack leaves only a $\approx 1\%$ gap for potential improvement, while the CC-based upper bound remains tight for the whole region of detection efficiencies with positive key rates—see Fig. 5 in which we compare it explicitly with the lower bound on the DW rate (5) established in Ref. [35].

However, the state-of-the-art proofs of the DW rate (5), e.g. [35, 47], require one to somehow bin the ‘no-click’ events and perform noisy preprocessing on the binary raw-data. Hence, one may still ask the question by how much could the thresholds in Tab. 2 be still improved, if novel derivations of one-way key rates were possible that allow Alice to perform any preprocessing map on her ternary variable A . That is

DIQKD protocols involving partially entangled states

η_{crit}	η_{DW}^\dagger	Reference
79.04%	80.00%	Brown <i>et al.</i> [35]
79.15%	80.26%	Masini <i>et al.</i> [47]

Table 2: **Critical detection efficiencies η_{crit} (in %)** determined by the CC attack for the shared correlations (and rates of bit-flip errors applied within noisy preprocessing) that lead to the best-known thresholds, η_{DW}^\dagger , above which the DW rate (5) is guaranteed to be positive [35, 47]. The CC attack proves that there is hardly any room for improvement of these state-of-the-art threshold values, given the data preprocessing (binning ‘no-clicks’ + bit-flip errors) employed.

why, for the correlations considered in Fig. 4, we also compute the critical thresholds determined by the CC attack that are, however, minimised over all meaningful preprocessing strategies, i.e. $A \rightarrow A' \rightarrow M$ appearing in Eq. (4). We observe that, as in the case of maximally entangled states, from the point of view of the CC attack it is always optimal for Alice to announce the inconclusive rounds (via the map $A \rightarrow M$)—see *black-squared* and *green-circled curves* coinciding in Fig. 4—while in the limit $\theta \rightarrow 0$ it is sufficient to solely bin the ‘no-clicks’. As a result, bearing in mind that Fig. 4 considers particular θ -parametrised family of correlations, we conclude that the analytic value $\eta_{\text{crit}} = 3/4$ constitutes then a fundamental bound on the detection efficiency, below which no one-way DIQKD protocol may be possible. Note that it is strictly larger than $\eta_{\text{loc}} = 2/3$ below which the corre-

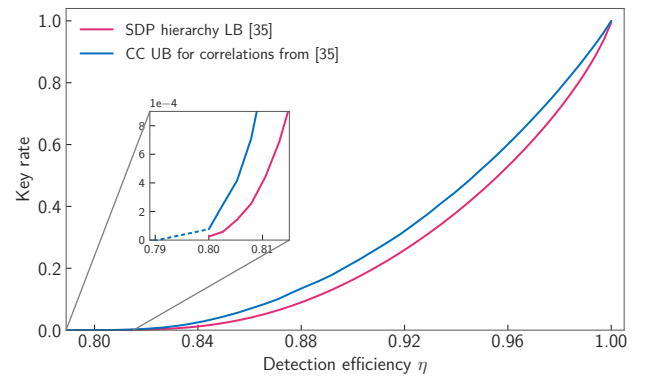


Figure 5: **CC-based upper bound derived for the protocol of Ref. [35] with noisy preprocessing, as compared with the lower bound on the DW rate (5) established therein.** For each optimal correlation and bit-flip probability p that maximises the lower bound (LB) at a given η (*red line*) [35], we compute the upper bound (UB) on the rate above which the CC attack invalidates the security (*blue line*). Within the inset we magnify the region of critical detection efficiencies that appear in Tab. 2. Note that due to the correlations of Ref. [35] being provided only for the region of positive LBs, $\eta \geq \eta_{\text{DW}}^\dagger = 80.00\%$, the CC-based UB for the region $\eta < \eta_{\text{DW}}^\dagger$ (and, hence, the critical value $\eta_{\text{crit}} = 79.04\%$ in Tab. 2) is computed (*dashed line* within the inset) using the same correlation and bit-flip probability p as for η_{DW}^\dagger .

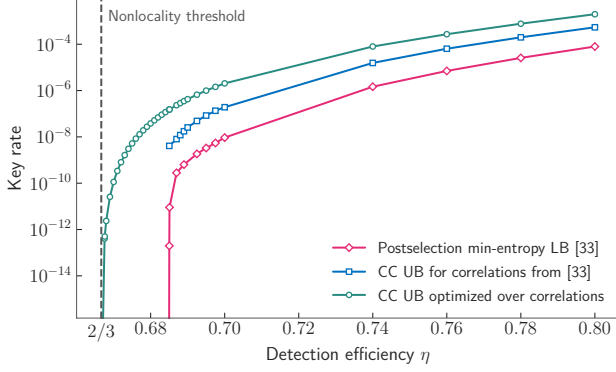


Figure 6: **Lower bound on the DW rate (5) compared with the CC-based upper bound for the protocol of Ref. [33] involving postselection.** For each optimal correlation and the acceptance probability of postselection determined by Xu *et al.* [33] to maximise the lower bound (LB) at a given η (*red diamonds*), we compute the upper bound (UB) on the rate above which the CC attack invalidates the security (*blue squares*). Moreover, by performing brute-force optimisation of the shared correlation and the acceptance probability to maximise the CC-based UB instead (*green circles*), we observe that the CC attack in principle allows for positive rates within the whole non-local range $\eta \geq \eta_{\text{loc}} = 2/3$.

lations cease to be non-local [48].

5.4.1 Robustness improvement by postselection

Nonetheless, it has been recently demonstrated that the thresholds stated in Tab. 2 can be decreased to $\approx 68.5\%$, if Alice and Bob perform postselection of their raw data for the key settings [33]. However, it is unclear if the derived bounds using this postselection are valid for general attacks where Eve exploits correlations among realisations of the protocol. In fact, there are situations in which some entropy is left in the postselected data when Eve applies any i.i.d. attack, but there is an attack using correlations between two realisations of the experiment for which Eve can perfectly predict the postselected outputs [34].

In Fig. 6, we present for the protocol of Xu *et al.* [33] lower bounds on the rates accompanied by upper bounds based on the CC attack, as a function of the detection efficiency. The protocol corresponds to the 2333-scenario in which the ‘no-click’ events are binned again onto a predetermined outcome, say ‘1’, while Alice and Bob then separately decide whether to accept or discard each of ‘1’s contained within their bit-string of key-generation rounds with probability q or $1 - q$, respectively. This does not open the detection loophole, as no postselection is performed within the rounds used to assure the nonlocality of the correlation. Moreover, Alice and Bob reveal publicly only the information whether each bit is accepted, irrespectively of its actual value (accepting simply all ‘0’s).

The *red curve* in Fig. 6 corresponds to the lower bound on the DW rate obtained by the authors of Ref. [33] by approximating the von Neumann entropy

with min-entropy in Eq. (5) and optimising the acceptance probability q . For each point (*dots/squares* in Fig. 6), see App. D.4, we evaluate the upper bound that follows from the CC attack for the corresponding correlation and the optimal value of q utilised by Xu *et al.* [33]—see the *blue curve* in Fig. 6. Moreover, we also maximise by brute-force heuristic methods the so-determined CC-based upper bound over all correlations (and the acceptance probability q for each correlation), in order to determine the *green curve* in Fig. 6. We observe that the CC attack disallows any significant improvement of the already very low rate, $\lesssim 10^{-4}$ for any $\eta \leq 80\%$, however, it suggests that positive key rates could be potentially attained with postselection for the whole non-local range $\eta \geq 2/3$ (disallowing any non-i.i.d. attacks though [34]).

5.5 One-way protocols with more than two settings and outcomes

In this last section, we would like to emphasise that the CC attack can be applied efficiently to any protocol by just following its consecutive stages, also when it involves correlations with larger number of settings and outcomes on both sides. In order to do so, we consider the two DIQKD schemes recently analysed by Gonzales-Ureta *et al.* [49] with correlations obtained by measuring a maximally entangled two-quart state ($\frac{1}{2} \sum_{i=1}^4 |ii\rangle$) within the 4522- and 3444-scenarios. In particular, the correlations employed in Ref. [49] correspond then to ones that exhibit robustness to noise (again, finite detection efficiency and visibility) when violating I_{4422}^4 [65] and I_{234} [66] Bell inequalities, respectively. In the former case the correlation Q_{4422} introduced in Ref. [64] is used, while in the latter the correlation Q_{234} that leads to the maximal quantum violation of $I_{234} = 9$ [49]. In Tab. 3, we compare the resulting thresholds, V_{DW}^\uparrow and $\eta_{\text{DW}}^\uparrow$, above which the DW rate (5) has been proven to be positive by Gonzales-Ureta *et al.* [49], against the corresponding critical values on visibility and detection efficiency that the CC attack allows for, V_{crit} and η_{crit} .

**DIQKD protocols
with more settings and outcomes**

Scenario	Correlation	V_{crit}	V_{DW}^\uparrow	η_{crit}	$\eta_{\text{DW}}^\uparrow$
4522	Q_{4422} [64]	85.35	93.96	86.78	94.74
3444	Q_{234} [49]	85.29	91.04	88.28	92.18

Table 3: **Critical visibilities V_{crit} and detection efficiencies η_{crit} (in %) below which no DIQKD is possible due to the CC attack, for protocols considered by Gonzales-Ureta *et al.* [49] that employ correlations being obtained by measuring two maximally entangled ququarts and violating I_{4422}^4 [65] and I_{234} [66] Bell inequalities within 4522- and 3444-scenarios, respectively. The CC-based critical values are compared against the thresholds, V_{DW}^\uparrow and $\eta_{\text{DW}}^\uparrow$, above which the DW rate (5) has been proven to be positive [49].**

The CC-based values suggest that the thresholds obtained by Gonzales-Ureta *et al.* [49] may be potentially improved below 90%, but not beyond 85%.

6 Conclusions

We have introduced the convex-combination (CC) attack as an easy-to-use tool to compute upper bounds on asymptotic rates in one-way and two-way DIQKD protocols. This in turn allows one to quickly establish critical noise parameters (here, detection efficiency and visibility) below which these upper bounds vanish and, hence, the CC attack disallows any DIQKD to be possible.

By applying the CC attack to one-way and two-way protocols involving either maximally or partially entangled states, as well as ones including a postselection stage or relying on correlations with more than two measurement settings and outcomes, we have demonstrated that despite its simple construction—decomposition of a given quantum correlation into a ‘local’ and a ‘more non-local’ part—the CC attack turns out to be very efficient in proving that the current thresholds on noise tolerance established with help of state-of-the-art security proofs are already very close to critical noise values, below which the CC attack invalidates the security. It is worth stressing that computing the upper bounds on the key rates, or equivalently, on Eve’s entropy, in the CC attack is very simple, in particular, much simpler than computing lower bounds. In light of the heuristic results derived in this work, the CC attack appears to be also a versatile tool to benchmark lower bounds to entropies obtained with existing techniques, such as the hierarchies of [35, 36].

We have successfully applied the CC attack in its simplest form, in which the ‘non-local’ (but quantum) correlation within the convex decomposition is fixed, while the ‘local’ contribution can then be chosen to maximise its weight within the decomposition by means of a linear program. On one hand, it may not be generally true that the strategy of maximising the probability of distributing the local correlation, for which the eavesdropper perfectly knows the outcomes, is actually optimal from the point of view of providing the tightest upper bound on the key rate. On the other, we have pessimistically assumed the eavesdropper not to possess any information about the outcomes in case the ‘more non-local’ correlation is distributed. Although the linear program can be straightforwardly adapted to utilise multiple non-local point within the decomposition, it would be desirable to generalise the construction, so that it actually includes optimisation over the non-local points e.g. by approximating the quantum set of correlations with sufficient accuracy from outside by means of a convergent hierarchy of relaxations [31, 32]. We leave the above interesting developments of our CC attack

for future work.

Note Added. Upon completion of this manuscript, we have learned that the CC attack has been applied to a two-way protocol by Yu-Zhe Zhang *et al.* [67], while incorporating the optimisation of the non-local point(s) in the CC decomposition with help of the NPA-hierarchy [31], as suggested above.

Acknowledgements

We thank Erik Woodhead, Stefano Pironio, Feihu Xu, Peter Brown, Jędrzej Kaniewski, Karol Horodecki and Filip Rozpędek for fruitful comments. We are also very grateful to the authors of Refs. [35, 47] and [33] for providing us with an explicit numerical form of the correlations employed in the respective protocols, especially to Yu-Zhe Zhang and Yi-Zheng Zhen for an extensive discussion about the application of the CC attack to the protocol involving post-selection [33].

We acknowledge support from the Government of Spain (TRANQI, FUNQIP, NextGeneration EU/PRTR and Severo Ochoa CEX2019-000910-S), Fundació Cellex, Fundació Mir-Puig, Generalitat de Catalunya (CERCA Program), the ERC AdG CERQUTE, the AXA Chair in Quantum Information Science, the EU Quantum Flagship project QRANGE and QSNP, and the Foundation for Polish Science within the “Quantum Optical Technologies” project carried out within the International Research Agendas programme cofinanced by the European Union under the European Regional Development Fund. MBJ acknowledges funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 847517. This project has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 754510.

References

- [1] Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, Stefano Pironio, and Valerio Scarani. “Device-independent security of quantum cryptography against collective attacks”. *Phys. Rev. Lett.* **98**, 230501 (2007).
- [2] Stefano Pironio, Antonio Acín, Nicolas Brunner, Nicolas Gisin, Serge Massar, and Valerio Scarani. “Device-independent quantum key distribution secure against collective attacks”. *New J. Phys.* **11**, 045021 (2009).
- [3] Claude E. Shannon. “Communication theory of secrecy systems”. *The Bell System Technical Journal* **28**, 656–715 (1949).
- [4] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner.

- “Bell nonlocality”. *Rev. Mod. Phys.* **86**, 419–478 (2014).
- [5] Jonathan Barrett, Lucien Hardy, and Adrian Kent. “No signaling and quantum key distribution”. *Phys. Rev. Lett.* **95**, 010503 (2005).
- [6] Antonio Acín, Nicolas Gisin, and Lluís Masanes. “From Bell’s theorem to secure quantum key distribution”. *Phys. Rev. Lett.* **97**, 120405 (2006).
- [7] Antonio Acín, Serge Massar, and Stefano Pironio. “Efficient quantum key distribution secure against no-signalling eavesdroppers”. *New J. Phys.* **8**, 126–126 (2006).
- [8] Yi Zhao, Chi-Hang Fred Fung, Bing Qi, Christine Chen, and Hoi-Kwong Lo. “Quantum hacking: Experimental demonstration of time-shift attack against practical quantum-key-distribution systems”. *Phys. Rev. A* **78**, 042333 (2008).
- [9] Feihu Xu, Bing Qi, and Hoi-Kwong Lo. “Experimental demonstration of phase-remapping attack in a practical quantum key distribution system”. *New J. Phys.* **12**, 113026 (2010).
- [10] Lars Lydersen, Carlos Wiechers, Christoffer Wittmann, Dominique Elser, Johannes Skaar, and Vadim Makarov. “Hacking commercial quantum cryptography systems by tailored bright illumination”. *Nat. Photonics* **4**, 686–689 (2010).
- [11] Ilja Gerhardt, Qin Liu, Antía Lamas-Linares, Johannes Skaar, Christian Kurtsiefer, and Vadim Makarov. “Full-field implementation of a perfect eavesdropper on a quantum cryptography system”. *Nat. Commun.* **2**, 349 (2011).
- [12] Valerio Scarani, Helle Bechmann-Pasquinucci, Nicolas J. Cerf, Miloslav Dušek, Norbert Lütkenhaus, and Momtchil Peev. “The security of practical quantum key distribution”. *Rev. Mod. Phys.* **81**, 1301–1350 (2009).
- [13] Rotem Arnon-Friedman, Frédéric Dupuis, Omar Fawzi, Renato Renner, and Thomas Vidick. “Practical device-independent quantum cryptography via entropy accumulation”. *Nat. Commun.* **9**, 459 (2018).
- [14] Gláucia Murta, Suzanne B. van Dam, Jérémy Ribeiro, Ronald Hanson, and Stephanie Wehner. “Towards a realization of device-independent quantum key distribution”. *Quantum Sci. Technol.* **4**, 035011 (2019).
- [15] René Schwonnek, Koon Tong Goh, Ignatius W. Primaatmaja, Ernest Y.-Z. Tan, Ramona Wolf, Valerio Scarani, and Charles C.-W. Lim. “Device-independent quantum key distribution with random key basis”. *Nat Commun* **12**, 2880 (2021).
- [16] Igor Devetak and Andreas Winter. “Distillation of secret key and entanglement from quantum states”. *Proc. R. Soc. Lond. A* **461**, 207–235 (2005).
- [17] Renato Renner, Nicolas Gisin, and Barbara Kraus. “Information-theoretic security proof for quantum-key-distribution protocols”. *Phys. Rev. A* **72**, 012332 (2005).
- [18] Rotem Arnon-Friedman. “Device-independent quantum information processing”. *Springer Theses* (2020).
- [19] Yanbao Zhang, Honghao Fu, and Emanuel Knill. “Efficient randomness certification by quantum probability estimation”. *Phys. Rev. Research* **2**, 013016 (2020).
- [20] John F. Clauser, Michael A. Horne, Abner Shimony, and Richard A. Holt. “Proposed experiment to test local hidden-variable theories”. *Phys. Rev. Lett.* **23**, 880–884 (1969).
- [21] Antonio Acín, Serge Massar, and Stefano Pironio. “Randomness versus nonlocality and entanglement”. *Phys. Rev. Lett.* **108**, 100402 (2012).
- [22] Erik Woodhead, Antonio Acín, and Stefano Pironio. “Device-independent quantum key distribution with asymmetric CHSH inequalities”. *Quantum* **5**, 443 (2021).
- [23] Melvyn Ho, Pavel Sekatski, Ernest Y.-Z. Tan, Renato Renner, Jean-Daniel Bancal, and Nicolas Sangouard. “Noisy preprocessing facilitates a photonic realization of device-independent quantum key distribution”. *Phys. Rev. Lett.* **124**, 230502 (2020).
- [24] Pavel Sekatski, Jean-Daniel Bancal, Xavier Valcarce, Ernest Y.-Z. Tan, Renato Renner, and Nicolas Sangouard. “Device-independent quantum key distribution from generalized CHSH inequalities”. *Quantum* **5**, 444 (2021).
- [25] Robert König, Renato Renner, and Christian Schaffner. “The operational meaning of min- and max-entropy”. *IEEE Trans. Inf. Theory* **55**, 4337–4347 (2009).
- [26] Lluís Masanes, Stefano Pironio, and Antonio Acín. “Secure device-independent quantum key distribution with causally independent measurement devices”. *Nat Commun* **2**, 238 (2011).
- [27] Olmo Nieto-Silleras, Stefano Pironio, and Jonathan Silman. “Using complete measurement statistics for optimal device-independent randomness evaluation”. *New J. Phys.* **16**, 013035 (2014).
- [28] Jean-Daniel Bancal, Lana Sheridan, and Valerio Scarani. “More randomness from the same data”. *New J. Phys.* **16**, 033011 (2014).
- [29] Alejandro Máttar, Paul Skrzypczyk, Jonatan Bohr Brask, Daniel Cavalcanti, and Antonio Acín. “Optimal randomness generation from optical Bell experiments”. *New J. Phys.* **17**, 022003 (2015).
- [30] Jan Kołodyński, Alejandro Máttar, Paul Skrzypczyk, Erik Woodhead, Daniel Cavalcanti, Konrad Banaszek, and Antonio Acín. “Device-independent quantum key distribution with single-photon sources”. *Quantum* **4**, 260 (2020).

- [31] Miguel Navascués, Stefano Pironio, and Antonio Acín. “Bounding the set of quantum correlations”. *Phys. Rev. Lett.* **98**, 010401 (2007).
- [32] Miguel Navascués, Stefano Pironio, and Antonio Acín. “A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations”. *New Journal of Physics* **10**, 073013 (2008).
- [33] Feihu Xu, Yu-Zhe Zhang, Qiang Zhang, and Jian-Wei Pan. “Device-independent quantum key distribution with random postselection”. *Phys. Rev. Lett.* **128**, 110506 (2022).
- [34] Le Phuc Thinh, Gonzalo de la Torre, Jean-Daniel Bancal, Stefano Pironio, and Valerio Scarani. “Randomness in post-selected events”. *New Journal of Physics* **18**, 035007 (2016).
- [35] Peter Brown, Hamza Fawzi, and Omar Fawzi. “Device-independent lower bounds on the conditional von Neumann entropy” (2021). [arXiv:2106.13692](https://arxiv.org/abs/2106.13692).
- [36] Peter Brown, Hamza Fawzi, and Omar Fawzi. “Computing conditional entropies for quantum correlations”. *Nat Commun* **12**, 575 (2021).
- [37] Ernest Y.-Z. Tan, René Schwonnek, Koon Tong Goh, Ignatius William Primaatmaja, and Charles C.-W. Lim. “Computing secure key rates for quantum cryptography with untrusted devices”. *npj Quantum Inf* **7**, 1–6 (2021).
- [38] Eneet Kaur, Mark M Wilde, and Andreas Winter. “Fundamental limits on key rates in device-independent quantum key distribution”. *New J. Phys.* **22**, 023039 (2020).
- [39] Matthias Christandl, Roberto Ferrara, and Karol Horodecki. “Upper bounds on device-independent quantum key distribution”. *Phys. Rev. Lett.* **126**, 160501 (2021).
- [40] Rotem Arnon-Friedman and Felix Leditzky. “Upper bounds on device-independent quantum key distribution rates and a revised Peres conjecture”. *IEEE Trans. Inf. Theory* **67**, 6606–6618 (2021).
- [41] Máté Farkas, Maria Balanzó-Juandó, Karol Łukanowski, Jan Kołodyński, and Antonio Acín. “Bell nonlocality is not sufficient for the security of standard device-independent quantum key distribution protocols”. *Phys. Rev. Lett.* **127**, 050503 (2021).
- [42] Ernest Y.-Z. Tan, Charles C.-W. Lim, and Renato Renner. “Advantage distillation for device-independent quantum key distribution”. *Phys. Rev. Lett.* **124**, 020502 (2020).
- [43] Imre Csiszár and János Körner. “Broadcast channels with confidential messages”. *IEEE Trans. Inf. Theory* **24**, 339–348 (1978).
- [44] Ueli Maurer. “Secret key agreement by public discussion from common information”. *IEEE Trans. Inf. Theory* **39**, 733–742 (1993).
- [45] Rudolf Ahlswede and Imre Csiszár. “Common randomness in information theory and cryptography. I. Secret sharing”. *IEEE Trans. Inf. Theory* **39**, 1121–1132 (1993).
- [46] Eneet Kaur, Karol Horodecki, and Siddhartha Das. “Upper bounds on device-independent quantum key distribution rates in static and dynamic scenarios”. *Phys. Rev. Appl.* **18**, 054033 (2022).
- [47] Michele Masini, Stefano Pironio, and Erik Woodhead. “Simple and practical DIQKD security analysis via BB84-type uncertainty relations and Pauli correlation constraints”. *Quantum* **6**, 843 (2022).
- [48] Philippe H. Eberhard. “Background level and counter efficiencies required for a loophole-free Einstein-Podolsky-Rosen experiment”. *Phys. Rev. A* **47**, R747–R750 (1993).
- [49] Junior R. Gonzales-Ureta, Ana Predojević, and Adán Cabello. “Device-independent quantum key distribution based on Bell inequalities with more than two inputs and two outputs”. *Phys. Rev. A* **103**, 052436 (2021).
- [50] Daniel Collins and Nicolas Gisin. “A relevant two qubit Bell inequality inequivalent to the CHSH inequality”. *J. Phys. A: Math. Gen.* **37**, 1775–1787 (2004).
- [51] Stefano Pironio, Lluís Masanes, Anthony Leverrier, and Antonio Acín. “Security of device-independent quantum key distribution in the bounded-quantum-storage model”. *Phys. Rev. X* **3**, 031007 (2013).
- [52] Xiongfeng Ma and Norbert Lutkenhaus. “Improved data post-processing in quantum key distribution and application to loss thresholds in device independent QKD”. *Quantum Information and Computation* **12**, 203–214 (2012).
- [53] Ignatius W. Primaatmaja, Koon Tong Goh, Ernest Y.-Z. Tan, John T.-F. Khoo, Shouvik Ghorai, and Charles C.-W. Lim. “Security of device-independent quantum key distribution protocols: a review”. *Quantum* **7**, 932 (2023).
- [54] Ernest Y.-Z. Tan, Pavel Sekatski, Jean-Daniel Bancal, René Schwonnek, Renato Renner, Nicolas Sangouard, and Charles C.-W. Lim. “Improved DIQKD protocols with finite-size analysis”. *Quantum* **6**, 880 (2022).
- [55] Ueli Maurer and Stefan Wolf. “The intrinsic conditional mutual information and perfect secrecy”. In *Proceedings of IEEE International Symposium on Information Theory*. IEEE (1997).
- [56] Matthias Christandl, Artur Ekert, Michał Horodecki, Paweł Horodecki, Jonathan Oppenheim, and Renato Renner. “Unifying classical and quantum key distillation”. In Vadhan, S.P. (eds) *Theory of Cryptography*. TCC 2007. Volume 4392 of *Lecture Notes in Computer Science*

- ence, pages 456–478. Berlin, Heidelberg (2007). Springer.
- [57] Marek Winczewski, Tamoghna Das, and Karol Horodecki. “Limitations on a device-independent key secure against a nonsignaling adversary via squashed nonlocality”. *Phys. Rev. A* **106**, 052612 (2022).
- [58] David Avis, Hiroshi Imai, Tsuyoshi Ito, and Yuya Sasaki. “Two-party Bell inequalities derived from combinatorics via triangular elimination”. *J. Phys. A* **38**, 10971–10987 (2005).
- [59] Boris S. Cirel’son. “Quantum generalizations of Bell’s inequality”. *Letters in Mathematical Physics* **4**, 93–100 (1980).
- [60] Stephen Boyd and Lieven Vandenbergh. “Convex optimization”. Cambridge University Press. (2004).
- [61] Víctor Zapatero and Marcos Curty. “Long-distance device-independent quantum key distribution”. *Sci Rep* **9**, 1–18 (2019).
- [62] N. David Mermin. “The EPR Experiment—Thoughts about the “Loophole””. *Ann. N.Y. Acad. Sci.* **480**, 422–427 (1986).
- [63] Erik Woodhead, Jędrzej Kaniewski, Boris Bourdoncle, Alexia Salavrakos, Joseph Bowles, Antonio Acín, and Remigiusz Augusiak. “Maximal randomness from partially entangled states”. *Phys. Rev. Research* **2**, 042028 (2020).
- [64] Tamás Vértesi, Stefano Pironio, and Nicolas Brunner. “Closing the detection loophole in Bell experiments using qudits”. *Phys. Rev. Lett.* **104**, 060401 (2010).
- [65] Nicolas Brunner and Nicolas Gisin. “Partial list of bipartite Bell inequalities with four binary settings”. *Phys. Lett. A* **372**, 3162–3167 (2008).
- [66] Adán Cabello. ““All versus nothing” inseparability for two observers”. *Phys. Rev. Lett.* **87**, 010403 (2001).
- [67] Yu-Zhe Zhang, Yi-Zheng Zhen, and Feihu Xu. “Upper bound on device-independent quantum key distribution with two way classical postprocessing under individual attack”. *New Journal of Physics* **24**, 113045 (2022).
- [68] Daniel Collins, Nicolas Gisin, Noah Linden, Serge Massar, and Sandu Popescu. “Bell inequalities for arbitrarily high-dimensional systems”. *Phys. Rev. Lett.* **88**, 040404 (2002).

Appendices

Within the appendices, we firstly provide in App. A an explicit construction of the tripartite state and the measurements allowing the eavesdropper Eve to implement any individual attack. In App. B, we then show how to obtain analytic expressions for the maximal local weight $q^{\mathcal{L}}$ utilised within the CC attack for the CHSH-based protocols subject to finite visibility ($V < 1$) and detection efficiency ($\eta < 1$) that involve maximally entangled states, but also partially entangled states (with $\theta \rightarrow 0$ in Eq. (17)) when $\eta < 1$. Throughout our work—see the beginning of Sec. 5 for a discussion of this choice—we consider the version of the CC attack in which Eve uses only one nonlocal correlation, $p_{AB}^{\mathcal{NL}}(a, b|x, y) \equiv Q_{AB}(a, b|x, y)$, which corresponds to the probability distribution of Alice and Bob registering measurement outcomes a and b in the noiseless scenario of $\eta = V = 1$. In App. C, we demonstrate how to construct the upper bounds on one-way key rates based on this choice of the CC attack for the two noise models considered and various preprocessing strategies: *random* and *deterministic* binning of non-detection events, with and without *noisy preprocessing*. This allows us to explicitly derive in App. D the thresholds on the tolerable noise parameters, in particular, in App. D.1 for the protocols involving maximally entangled states with the resulting (numerical) values presented in Tab. 1 of the main text. We achieve this analytically for all particular preprocessing strategies considered, and semi-analytically when including the optimisation over all potential preprocessing maps. We then generalise the above analysis to two-way protocols in App. D.2 (also ones that involve multiple key settings [15] for $V < 1$), as well as one-way protocols involving partially entangled states in App. D.3. The latter case we study in more detail in App. D.4, where we allow further for postselection of some events, as proposed in Ref. [33].

A Explicit form of the state and measurements in individual attacks

In this section, we provide an explicit construction for a shared tripartite state, and measurements for Alice, Bob and Eve for achieving the individual attack in Eq. (3). We start from the observed correlation, and write it as a convex combination of quantum correlations:

$$p_{AB}^{\text{obs}}(a, b|x, y) = \sum_{\lambda} q(\lambda) p_{AB}(a, b|x, y, \lambda). \quad (36)$$

That is, there exist a state ρ_{AB} on a Hilbert space $\mathcal{H}_A \otimes \mathcal{H}_B$ and measurements $\{M_a^x\}$ and $\{N_b^y\}$ on \mathcal{H}_A and \mathcal{H}_B , respectively, such that

$$p_{AB}^{\text{obs}}(a, b|x, y) = \text{Tr}\{\rho_{AB}(M_a^x \otimes N_b^y)\}. \quad (37)$$

Moreover, $q(\lambda)$ is a probability distribution and there exist states ρ_{AB}^λ on a Hilbert space $\mathcal{H}_A^\lambda \otimes \mathcal{H}_B^\lambda$ and measurements $\{(M^\lambda)_a^x\}$ and $\{(N^\lambda)_b^y\}$ on \mathcal{H}_A^λ and \mathcal{H}_B^λ , respectively, such that

$$p_{AB}(a, b|x, y, \lambda) = \text{Tr}\{\rho_{AB}^\lambda[(M^\lambda)_a^x \otimes (N^\lambda)_b^y]\}. \quad (38)$$

It is clear from the convex structure of the quantum set of correlations that for every valid convex decomposition of the form Eq. (36), if $q(\lambda)$, the states ρ_{AB}^λ , and the measurements $\{(M^\lambda)_a^x\}$ and $\{(N^\lambda)_b^y\}$ are known, then one can build the state ρ_{AB} and the measurements $\{M_a^x\}$ and $\{N_b^y\}$. In particular, one can pick the Hilbert spaces $\mathcal{H}_A = \bigoplus_\lambda \mathcal{H}_A^\lambda$ and $\mathcal{H}_B = \bigoplus_\lambda \mathcal{H}_B^\lambda$, and define the state $\rho_{AB} = \bigoplus_\lambda q(\lambda)\rho_{AB}^\lambda$ on $\mathcal{H}_A \otimes \mathcal{H}_B$. This state together with the measurements $M_a^x = \bigoplus_\lambda (M^\lambda)_a^x$ on \mathcal{H}_A and $N_b^y = \bigoplus_\lambda (N^\lambda)_b^y$ on \mathcal{H}_B gives rise to the correlation $p_{AB}^{\text{obs}}(a, b|x, y)$.

We now construct a tripartite state ρ_{ABE} on $\mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_E$ and measurement operators E_e on \mathcal{H}_E such that the resulting tripartite correlation in the individual attack reads

$$p_{ABE}(a, b, e|x, y) = \text{Tr}\{\rho_{ABE}(M_a^x \otimes N_b^y \otimes E_e)\} = \sum_\lambda q(\lambda)p(e|\lambda)p_{AB}(a, b|x, y, \lambda), \quad (39)$$

for some arbitrary distributions $p(e|\lambda)$. We simply choose the state

$$\rho_{ABE} = \bigoplus_\lambda q(\lambda)\rho_{AB}^\lambda \otimes |\lambda\rangle\langle\lambda|_E, \quad (40)$$

where $\{|\lambda\rangle\}$ is an orthonormal basis on \mathcal{H}_E , and the measurement operators

$$E_e = \sum_\lambda p(e|\lambda)|\lambda\rangle\langle\lambda| \quad (41)$$

on \mathcal{H}_E . It follows that

$$\text{Tr}\{\rho_{ABE}(M_a^x \otimes N_b^y \otimes E_e)\} = p_{ABE}(a, b, e|x, y) = \sum_\lambda q(\lambda)p(e|\lambda)p_{AB}(a, b|x, y, \lambda) \quad (42)$$

as required, and clearly, $\sum_e p_{ABE}(a, b, e|x, y) = p_{AB}^{\text{obs}}(a, b|x, y)$ for all a, b, x, y . Last, we note that $p(e|\lambda)$ can be chosen in a way that e preserves the information about λ (given that the alphabet size of e is large enough).

B Analytical evaluation of the maximal local weight $q^\mathcal{L}$ in the CC attack

B.1 Correlations yielding maximal CHSH-violation subject to finite visibility

In this section, we provide the analytical form of the local weight (30) in the CC attack for the 2322 CHSH-based protocol subject to finite visibility. The correlation in the noise-free case is the one obtained by Alice and Bob sharing the state $|\Phi^+\rangle$, i.e. setting $\theta = \pi/2$ in Eq. (17), on which they perform CHSH-optimal dichotomic measurements (27) with outcomes $a, b \in \{0, 1\}$, and reads

$$Q_{AB}(a, b|x, y) = \begin{cases} \frac{1}{4} \left[1 + \frac{(-1)^{a+b+xy}}{\sqrt{2}} \right] & \text{if } x, y \in \{0, 1\} \\ \frac{1}{2} \delta_{a,b} & \text{if } (x, y) = (0, 2) \\ \frac{1}{4} & \text{if } (x, y) = (1, 2). \end{cases} \quad (43)$$

We consider noisy versions of this correlation with finite visibility $V \in [0, 1]$, i.e. the uniform noise as specified in Eq. (15) with $n_A = n_B = 2$, i.e.,

$$p_{AB}^{\text{obs}}(a, b|x, y) = V Q_{AB}(a, b|x, y) + \frac{1-V}{4}. \quad (44)$$

The CC attack for this protocol consists of the convex combination of a local correlation $p_{AB}^\mathcal{L}(a, b|x, y)$ and the noise-free correlation $Q_{AB}(a, b|x, y)$, such that the observed correlation is of the form:

$$p_{AB}^{\text{obs}}(a, b|x, y) = q^\mathcal{L} p_{AB}^\mathcal{L}(a, b|x, y) + (1 - q^\mathcal{L}) Q_{AB}(a, b|x, y). \quad (45)$$

From equating the above expression for $p_{AB}^{\text{obs}}(a, b|x, y)$ it follows that

$$p_{AB}^\mathcal{L}(a, b|x, y) = \tilde{V} Q_{AB}(a, b|x, y) + \frac{1 - \tilde{V}}{4}, \quad (46)$$

where $\tilde{V} \in [0, 1]$ and $q^\mathcal{L} = (1 - V)/(1 - \tilde{V})$. Therefore, maximising $q^\mathcal{L}$ simply corresponds to maximising \tilde{V} in Eq. (46), such that $p_{AB}^\mathcal{L}(a, b|x, y)$ is local. The result of this maximisation is the *local visibility* $V^\mathcal{L}$, which in turn fully characterises the CC attack for this protocol with $q^\mathcal{L} = (1 - V)/(1 - V^\mathcal{L})$ when $V \geq V^\mathcal{L}$ and otherwise $q^\mathcal{L} = 1$. Even though the maximisation is a linear program, it is possible to solve it explicitly as all the facets of the polytope in the 2322-scenario are known analytically [50]. In particular, all the facets that correspond to non-trivial constraints (i.e. do not correspond to the positivity and normalisation of the probabilities) are of the CHSH-type:

$$-2 \leq \langle A_{x_0} B_{y_0} \rangle + \langle A_{x_0} B_{y_1} \rangle + \langle A_{x_1} B_{y_0} \rangle - \langle A_{x_1} B_{y_1} \rangle \leq 2 \quad x_0, x_1 \in \{0, 1\}, y_0, y_1 \in \{0, 1, 2\}, \quad (47)$$

where $\langle A_x B_y \rangle = p(0, 0|x, y) + p(1, 1|x, y) - p(0, 1|x, y) - p(1, 0|x, y)$ are the correlators of a given $p(a, b|x, y)$. That is, a correlation in the 2322-scenario is local if and only if it satisfies all the inequalities in Eq. (47).

Let us denote the correlators of $Q_{AB}(a, b|x, y)$ by $\langle A_x B_y \rangle^{\mathcal{NL}}$, and those of $p_{AB}^\mathcal{L}(a, b|x, y)$ by $\langle A_x B_y \rangle^\mathcal{L}$. It is easy to see that

$$\langle A_x B_y \rangle^\mathcal{L} = \tilde{V} \langle A_x B_y \rangle^{\mathcal{NL}}. \quad (48)$$

Therefore, finding the maximal \tilde{V} such that $p_{AB}^\mathcal{L}(a, b|x, y)$ is local corresponds to finding the maximal \tilde{V} such that

$$-\frac{2}{\tilde{V}} \leq \mathcal{S}_{x_0, x_1, y_0, y_1}^{\mathcal{NL}} \leq \frac{2}{\tilde{V}} \quad \forall x_0, x_1 \in \{0, 1\}, \forall y_0, y_1 \in \{0, 1, 2\}, \quad (49)$$

where we have defined

$$\mathcal{S}_{x_0, x_1, y_0, y_1}^{\mathcal{NL}} = \langle A_{x_0} B_{y_0} \rangle^{\mathcal{NL}} + \langle A_{x_0} B_{y_1} \rangle^{\mathcal{NL}} + \langle A_{x_1} B_{y_0} \rangle^{\mathcal{NL}} - \langle A_{x_1} B_{y_1} \rangle^{\mathcal{NL}}. \quad (50)$$

A straightforward computation yields

$$\begin{aligned} \mathcal{S}_{0,1,0,1}^{\mathcal{NL}} &= 2\sqrt{2} & \mathcal{S}_{0,1,0,2}^{\mathcal{NL}} &= 1 + \sqrt{2} & \mathcal{S}_{0,1,1,0}^{\mathcal{NL}} &= -\sqrt{2} & \mathcal{S}_{0,1,1,2}^{\mathcal{NL}} &= 1 \\ \mathcal{S}_{0,1,2,0}^{\mathcal{NL}} &= 1 & \mathcal{S}_{0,1,2,1}^{\mathcal{NL}} &= 1 + \sqrt{2} & \mathcal{S}_{1,0,0,1}^{\mathcal{NL}} &= 0 & \mathcal{S}_{1,0,0,2}^{\mathcal{NL}} &= \sqrt{2} - 1 \\ \mathcal{S}_{1,0,1,0}^{\mathcal{NL}} &= 0 & \mathcal{S}_{1,0,1,2}^{\mathcal{NL}} &= -1 & \mathcal{S}_{1,0,2,0}^{\mathcal{NL}} &= 1 & \mathcal{S}_{1,0,2,1}^{\mathcal{NL}} &= 1 - \sqrt{2}. \end{aligned} \quad (51)$$

It is clear from the first equation that the maximal \tilde{V} (denoted by $V^\mathcal{L}$) is bounded by $V^\mathcal{L} \leq 1/\sqrt{2}$. Furthermore, substituting $\tilde{V} = 1/\sqrt{2}$ into Eq. (49) also implies that $p_{AB}^\mathcal{L}(a, b|x, y)$ with visibility $\tilde{V} = 1/\sqrt{2}$ is local, and therefore $V^\mathcal{L} \geq 1/\sqrt{2}$. Hence, we get that $V^\mathcal{L} = 1/\sqrt{2}$, fully characterising the CC attack for this protocol and determining the local weight as

$$q^\mathcal{L} = \min \left\{ 1, \frac{1 - V}{1 - \frac{1}{\sqrt{2}}} \right\}. \quad (52)$$

B.2 Correlations yielding maximal CHSH-violation subject to finite detection efficiency

In this section, we determine explicitly the maximal local weights in the CC attack for the CHSH-based 2333-protocol subject to finite detection efficiency. As in Eq. (16) of the main text, the lossy observed correlation is given by

$$p_{AB}^{\text{obs}}(a, b|x, y) = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 Q_{00}^{xy} & \eta^2 Q_{01}^{xy} & \eta \bar{\eta} Q_0^x \\ 1 & \eta^2 Q_{10}^{xy} & \eta^2 Q_{11}^{xy} & \eta \bar{\eta} Q_1^x \\ \hline \emptyset & \bar{\eta} \eta Q_0^y & \bar{\eta} \eta Q_1^y & \bar{\eta}^2 \end{array}, \quad (53)$$

where $Q_{ab}^{xy} := Q(a, b|x, y)$ denotes the ideal, noiseless correlation with marginals $Q_a^x = \sum_b Q_{ab}^{xy}$ and $Q_b^y = \sum_a Q_{ab}^{xy}$ for Alice and Bob, respectively. Specifically, we calculate the maximal local weight $q^\mathcal{L}(\theta = \pi/2, \eta)$ for protocols involving maximally entangled states, discussed in Secs. 5.2&5.3, for which $Q(a, b|x, y)$ is given by Eq. (43) above, as well as $q^\mathcal{L}(\theta \rightarrow 0, \eta)$ for protocols involving partially entangled states, discussed in Sec. 5.4, where the Q -probabilities take a more complicated form discussed in App. B.2.2 below.

Consider first the 2233-scenario. The complete characterisation of the local polytope in terms of facet (Bell) inequalities becomes more complicated than in the 2222-scenario as the number of such inequalities is 1116 [50]. However, they may still be checked for violation with the help of some symbolic computation software. In general, the corresponding facet inequalities can be cast into three categories [50]: 36 ‘‘trivial’’ inequalities ensuring non-negativity of probabilities, 648 CHSH-like inequalities (resulting from the original CHSH inequality by some relabelling of measurements, outcomes and parties), and 432 CGLMP-like inequalities [68] (also all equivalent

under some choice of relabelling). All of them impose constraints on conditional probabilities, assuring the resulting correlation to admit a local hidden-variable model.

On the other hand, we note that for a given correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ observed by Alice and Bob, and a particular lossless correlation $p_{AB}^{\mathcal{L}}(a, b|x, y)$ distributed by Eve in the ‘‘nonlocal’’ rounds of the CC attack, the local correlation in Eq. (12) must satisfy the following equality:

$$p_{AB}^{\mathcal{L}}(a, b|x, y) = \frac{p_{AB}^{\text{obs}}(a, b|x, y) - (1 - q^{\mathcal{L}}) Q_{AB}(a, b|x, y)}{q^{\mathcal{L}}}, \quad (54)$$

where $q^{\mathcal{L}}$ is to be maximised and we set $p_{AB}^{\mathcal{L}} = Q_{AB}$. Although the maximal value of $q^{\mathcal{L}}$ can always be determined numerically by the linear program, one may equivalently treat $q^{\mathcal{L}}$ as a free parameter and verify what is its maximal value such that none of the aforementioned inequalities is violated by the correlation in Eq. (54). Importantly, in this way not only the maximal value of $q^{\mathcal{L}}$ is determined, but also the particular (facet) inequality may be identified, i.e. the facet of the local polytope on which the correlation (54) then resides in the correlation space when $q^{\mathcal{L}}$ is maximal. As all the Bell inequalities are linear in probabilities [4], i.e.

$$\sum_{a,b,x,y} b_{a,b,x,y} p_{AB}^{\mathcal{L}}(a, b|x, y) \leq 0, \quad (55)$$

with $b_{a,b,x,y} \in \mathbb{R}$, it can be rearranged into an inequality for $q^{\mathcal{L}}$ using Eq. (54), and expressing $p_{AB}^{\text{obs}}(a, b|x, y)$ and $Q_{AB}(a, b|x, y)$ as functions of η and θ .

B.2.1 Maximally entangled states

Focusing first on the 2233-scenario with $\theta = \pi/2$ in (17) and standard CHSH-optimal measurements $x, y \in \{0, 1\}$ (27), we observe that for all values of η , the relevant inequality imposing locality of the correlation $p_{AB}^{\mathcal{L}}$ in Eq. (54) is of the CHSH type, does not involve non-detection events, and after simplifying reads:

$$-p_A^{\mathcal{L}}(1|0) - p_B^{\mathcal{L}}(1|0) + p_{AB}^{\mathcal{L}}(1, 1|0, 0) + p_{AB}^{\mathcal{L}}(1, 1|0, 1) + p_{AB}^{\mathcal{L}}(1, 1|1, 0) - p_{AB}^{\mathcal{L}}(1, 1|1, 1) \leq 0 \quad (56)$$

with $p_A^{\mathcal{L}}$ and $p_B^{\mathcal{L}}$ being the marginal distributions. We can evaluate the relevant conditional probabilities by calculating the corresponding terms for the correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ observed by Alice and Bob from (53) as

$$\begin{aligned} p_A^{\text{obs}}(1|0) &= p_B^{\text{obs}}(1|0) = \frac{\eta}{2}, \\ p_{AB}^{\text{obs}}(1, 1|0, 0) &= p_{AB}^{\text{obs}}(1, 1|0, 1) = p_{AB}^{\text{obs}}(1, 1|1, 0) = \eta^2 \frac{2 + \sqrt{2}}{8}, \\ p_{AB}^{\text{obs}}(1, 1|1, 1) &= \eta^2 \frac{2 - \sqrt{2}}{8}. \end{aligned} \quad (57)$$

Hence, substituting for the local distribution $p_{AB}^{\mathcal{L}}$ and its marginals into Eq. (56) according to Eq. (54), with observed and ideal ($\eta = 1$) correlations specified as above, we obtain the desired upper bound on the local weight within the CC attack, i.e. $q^{\mathcal{L}} \leq (1 - \eta) (1 + (3 + 2\sqrt{2}) \eta)$, so that we can write explicitly the maximal local weight in the lossy 2233-scenario utilising maximally entangled states as

$$q^{\mathcal{L}} = \min \left\{ 1, (1 - \eta) \left(1 + (3 + 2\sqrt{2}) \eta \right) \right\}. \quad (58)$$

In the 2333-scenario Bob uses an additional measurement $B_2 = \sigma_z$ identical to A_0 , so that these are correlated and, hence, most efficient in generating the key. Although the dimensionality of the correlation space is then formally increased, such an added setting does not impose any further locality constraints on the resulting shared correlation. In particular, as the inequality (56) remains then the only relevant, the above analysis similarly applies. For completeness, however, we verify this numerically by running explicitly the linear program that consistently outputs maximal values of $q^{\mathcal{L}}$ according to Eq. (58) also in the 2333-scenario considered.

B.2.2 Partially entangled states

Moreover, focusing further on the 2233-scenario but examining correlations determined by the partially entangled states (17) and measurements chosen to maximise the CHSH functional (18), in the limiting case $\theta \rightarrow 0$, we

observe the relevant inequality imposing locality to be the same one as for the maximally entangled states (56). In this case, we have

$$\begin{aligned}
p_A^{\text{obs}}(1|0) &= \frac{\eta}{2}(1 - \cos \theta), \\
p_B^{\text{obs}}(1|0) &= \frac{\eta}{2}(1 - z_+ \cos \theta), \\
p_{AB}^{\text{obs}}(1, 1|0, 0) &= \frac{\eta^2}{4}(1 + z_+)(1 - \cos \theta), \\
p_{AB}^{\text{obs}}(1, 1|0, 1) &= \frac{\eta^2}{4}(1 + z_-)(1 - \cos \theta), \\
p_{AB}^{\text{obs}}(1, 1|1, 0) &= \frac{\eta^2}{4} \left[1 + z_+ \cos \phi_A - (z_+ + \cos \phi_A) \cos \theta + \sqrt{1 - z_+^2} \sin \phi_A \sin \theta \right], \\
p_{AB}^{\text{obs}}(1, 1|1, 1) &= \frac{\eta^2}{4} \left[1 + z_- \cos \phi_A - (z_- + \cos \phi_A) \cos \theta - \sqrt{1 - z_-^2} \sin \phi_A \sin \theta \right],
\end{aligned} \tag{59}$$

with

$$\begin{aligned}
P &:= \alpha \eta + \alpha \bar{\eta} \cos \theta, & Q &:= \eta \cos \phi_A + \bar{\eta} \cos \theta, \\
R &:= \eta \sin \phi_A \sin \theta, & z_{\pm} &:= \frac{P \pm Q}{\sqrt{(P \pm Q)^2 + R^2}},
\end{aligned} \tag{60}$$

and ϕ_A characterising the optimal measurement A_1 (cf. [22] where this notation is introduced)

$$A_1 = \cos(\phi_A) \sigma_z + \sin(\phi_A) \sigma_x. \tag{61}$$

The corresponding conditional probabilities Q_{AB} can be obtained by setting $\eta = 1$ in (59), and depend on θ and on ϕ_A due to the optimisation of measurements. Using these expressions we arrive at:

$$\begin{aligned}
& -\eta_{\mathcal{L}} \left(1 - \frac{1 + z_+}{2} \cos \theta \right) + \frac{\chi_{\mathcal{L}}^2}{4} \left[(2 + z_+ + z_-)(1 - \cos \theta) \right. \\
& \quad \left. + (z_+ - z_-)(\cos \phi_A - \cos \theta) + \left(\sqrt{1 - z_+^2} + \sqrt{1 - z_-^2} \sin \theta \sin \phi_A \right) \right] \leq 0,
\end{aligned} \tag{62}$$

where

$$\eta_{\mathcal{L}} := \frac{\eta + q^{\mathcal{L}} - 1}{q^{\mathcal{L}}} \quad \text{and} \quad \chi_{\mathcal{L}}^2 := \frac{\eta^2 + q^{\mathcal{L}} - 1}{q^{\mathcal{L}}}. \tag{63}$$

Expanding the inequality (62) in the lowest order of θ and ϕ_A we have

$$\left[\frac{q^{\mathcal{L}} - 1 - \eta(2 - 3\eta)}{4q^{\mathcal{L}}} + \mathcal{O}(\phi_A^2) \right] \theta^2 + \mathcal{O}(\theta^3) \leq 0. \tag{64}$$

Hence, in the limit $\theta \rightarrow 0$, in which also $\phi_A \rightarrow 0$ [22] and the lowest-order term $\sim \theta^2$ dominates, we obtain a general upper bound on local weight: $q^{\mathcal{L}} \leq 1 - \eta(3\eta - 2)$. Thus, we conclude that the maximal local weight for the CC attack in the lossy 2233 protocol utilising partially entangled states with $\theta \rightarrow 0$ and measurements chosen to maximise the CHSH-violation is given by

$$q^{\mathcal{L}} = \min \{1, 1 - \eta(3\eta - 2)\}, \tag{65}$$

with $q^{\mathcal{L}} = 1$ certifying the observed correlation to be local for $\eta \leq \eta_{\text{loc}} = 2/3$ —in consistency with Ref. [48].

Similarly to the $\theta = \pi/2$ -case above, we confirm for completeness that adding an extra key setting of Bob, $B_2 = \sigma_z$, in the 2333-scenario does not affect the above analysis. In particular, we compute numerically the maximal local weights with the linear program (13), which match then exactly the expression (65), as expected.

C Constructing the upper bounds on one-way key rates with help of the CC attack

In this section, we calculate the *error correction* (EC) and *privacy amplification* (PA) terms appearing in the upper bound on the one-way key rate (6) for the finite visibility ($V < 1$) and detection efficiency ($\eta < 1$) noise models, while considering particular preprocessing strategies that Alice may apply to her raw data. Specifically,

we consider here three types of them referenced in Tab. 1: one trivial case, i.e. in which Alice does not transform her outcome at all; and two cases in which she converts the ternary variable A into a binary variable A' , so that the preprocessing map $p_{A'|A}$ corresponds then to a 2×3 stochastic matrix \mathcal{S} , i.e. *deterministic binning* of the non-detection event \emptyset with and without *noisy preprocessing* (performing also a bit-flip with some probability on the resulting binary variable). Moreover, we discuss two additional cases not shown in Tab. 1 that involve *random binning*, i.e. the non-detection event \emptyset is randomly binned to one of the two measurement outcomes, with and without noisy preprocessing. While these preprocessing strategies appear most commonly in literature, the methodology described here may naturally be adapted to other protocols and preprocessing schemes.

The preprocessing strategies considered here make no use of the publicly announced random variable M appearing in the upper bound (6) and, hence, we drop for our purposes the M -conditioning and rewrite the r.h.s. of Eq. (6) as

$$r_{1\text{-way},\bullet}^{\uparrow} := H(A'|E)_{\bullet} - H(A'|B)_{\bullet}, \quad (66)$$

where for simplicity we also omit the notation $(A \rightarrow B|A')$ and instead introduce another subscript \bullet , within which we will denote the particular preprocessing map $A \rightarrow A'$ being employed—e.g. “det”/“rand” or “n.p.” for deterministic/random binning or noisy preprocessing, respectively. In the following, we refer to $H(A'|B)$ as the EC-term and to $H(A'|E)$ as the PA-term.

Having determined the local weight of the correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ —which includes all the possible inputs x and y —the calculation of the EC- and PA-terms depends only on the tripartite distribution conditioned the key settings x^* and y^* , i.e. $p_{ABE}(a, b, e|x^*, y^*)$. Therefore, for simplicity, we adopt the following notation, dropping the $\{x^*, y^*\}$ labels:

$$\begin{aligned} P_{ab} &= VQ(a, b|x^*, y^*) + \frac{1-V}{4}, \\ P_a^A &= VQ(a|x^*) + \frac{1-V}{2}, \quad P_b^B = VQ(b|y^*) + \frac{1-V}{2}, \end{aligned} \quad (67)$$

where the corresponding marginals satisfy $P_a^A = \sum_b P_{ab}$ and $P_b^B = \sum_a P_{ab}$, and $Q(a, b|x^*, y^*)$, $Q(a|x^*)$, $Q(b|y^*)$ denote the ideal probabilities of obtaining measurement outcomes a and b in case of perfect detection efficiency and visibility, $\eta = V = 1$, after Alice and Bob have chosen the key settings x^* and y^* .

The introduction of the P-probabilities allows us to consider finite detection efficiency and finite visibility at the same time, as we can write the shared correlation (53) for the key settings, $p_{AB}^{\text{obs}}(a, b|x^*, y^*)$ as

$$p_{AB}(a, b) = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 P_{00} & \eta^2 P_{01} & \eta\bar{\eta}P_0^A \\ 1 & \eta^2 P_{10} & \eta^2 P_{11} & \eta\bar{\eta}P_1^A \\ \hline \emptyset & \bar{\eta}\eta P_0^B & \bar{\eta}\eta P_1^B & \bar{\eta}^2 \end{array}, \quad (68)$$

where $\bar{\eta} := 1 - \eta$ and we have dropped the ‘obs’ superscript for simplicity. To recover the purely noisy correlation it suffices to set $\eta = 1$ in (68), in which case the outcomes \emptyset don’t occur, whereas to obtain the purely lossy correlation (53) it suffices to replace the P-probabilities with Q-probabilities as they become equal if one sets $V = 1$ in (67). Lastly, note that this always yields the marginal distribution of Bob as

$$p_B(b) = \sum_a p_{AB}(a, b) = \begin{cases} \eta P_0^B, & \text{if } b = 0 \\ \eta P_1^B, & \text{if } b = 1 \\ \bar{\eta}, & \text{if } b = \emptyset \end{cases}, \quad (69)$$

being trivially independent of the preprocessing map $p_{A'|A}$ applied by Alice.

C.1 Calculation of the EC-term $H(A'|B)$

C.1.1 No preprocessing

If Alice performs no preprocessing, then simply $A \equiv A'$ and

$$\begin{aligned} H(A'|B)_{\text{no-prep}} &= H(A|B) = \sum_b p_B(b) H(A|B=b) \\ &= \eta P_0^B H(A|B=0) + \eta P_1^B H(A|B=1) + \bar{\eta} H(A|B=\emptyset), \end{aligned} \quad (70)$$

where $H(A|B = b)$ is the entropy of Alice's outcome conditioned on Bob measuring b . Each $H(A|B = b)$ can be evaluated with the help of the conditional probability

$$p_{A|B}(a|b) = \frac{p_{AB}(a,b)}{p_B(b)} = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta \frac{P_{00}}{P_0^B} & \eta \frac{P_{01}}{P_1^B} & \eta P_0^A \\ \hline 1 & \eta \frac{P_{10}}{P_0^B} & \eta \frac{P_{11}}{P_1^B} & \eta P_1^A \\ \hline \emptyset & \bar{\eta} & \bar{\eta} & \bar{\eta} \end{array}, \quad (71)$$

which is obtained by dividing each column of Tab. (68) by the corresponding marginal probability of Bob in Eq. (69). The columns of Tab. (71) determine then the conditional entropy $H(A|B)$, i.e. the EC-term, as after defining the entropy of a probability vector as $H\{(p_i)_i\} := -\sum_i p_i \log_2 p_i$ for any $\sum_i p_i = 1$, it can be just written as a sum of entropies for each of the columns, i.e.:

$$H(A'|B)_{\text{no-prep}} = \eta P_0^B H\left\{\eta \frac{P_{00}}{P_0^B}, \eta \frac{P_{10}}{P_0^B}, \bar{\eta}\right\} + \eta P_1^B H\left\{\eta \frac{P_{01}}{P_1^B}, \eta \frac{P_{11}}{P_1^B}, \bar{\eta}\right\} + \bar{\eta} H\{\eta P_0^A, \eta P_1^A, \bar{\eta}\}. \quad (72)$$

C.1.2 Deterministic binning

We now consider the case when Alice deterministically bins every no-click event \emptyset . Without loss of generality, we may assume she always interprets it as the 0-outcome. This formally corresponds to her applying a stochastic map, see also Eq. (20) of the main text, of the form

$$\mathcal{S}_{\text{det}} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix} \quad (73)$$

to Tab. (68), so that the resulting shared correlation then reads

$$p_{A'B}(a',b)_{\text{det}} = \begin{array}{c|ccc} a' \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 P_{00} + \eta \bar{\eta} P_0^B & \eta^2 P_{01} + \eta \bar{\eta} P_1^B & \bar{\eta} \eta P_0^A + \bar{\eta}^2 \\ \hline 1 & \eta^2 P_{10} & \eta^2 P_{11} & \eta \bar{\eta} P_1^A \end{array}, \quad (74)$$

whose first row is obtained by summing the first and the third row of Tab. (68).

Again, in order to determine the EC-term we compute Alice's conditional probability distribution by dividing the columns of Tab. (22) by the corresponding marginal probabilities (69), i.e.:

$$p_{A'|B}(a'|b)_{\text{det}} = \frac{p_{A'B}(a',b)_{\text{det}}}{p_B(b)} = \begin{array}{c|ccc} a' \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta \frac{P_{00}}{P_0^B} + \bar{\eta} & \eta \frac{P_{01}}{P_1^B} + \bar{\eta} & \eta P_0^A + \bar{\eta} \\ \hline 1 & \eta \frac{P_{10}}{P_0^B} & \eta \frac{P_{11}}{P_1^B} & \eta P_1^A \end{array}, \quad (75)$$

which allows to directly compute the relevant conditional entropy in case Alice bins deterministically:

$$H(A'|B)_{\text{det}} = \eta P_0^B h\left[\frac{\eta P_{10}}{P_0^B}\right] + \eta P_1^B h\left[\frac{\eta P_{11}}{P_1^B}\right] + \bar{\eta} h[\eta P_1^A], \quad (76)$$

where $h[x] := -x \log_2 x - (1-x) \log_2 (1-x)$ is the binary entropy function.

C.1.3 Deterministic binning with noisy preprocessing (bit-flip)

In case Alice applies further noisy preprocessing [22–24] to her bit-string, she simply flips the value of each bit with probability \mathfrak{p} after having binned them deterministically. This corresponds to her applying instead a stochastic matrix:

$$\mathcal{S}_{\text{det+n.p.}} = \begin{pmatrix} 1-\mathfrak{p} & \mathfrak{p} & 1-\mathfrak{p} \\ \mathfrak{p} & 1-\mathfrak{p} & \mathfrak{p} \end{pmatrix}, \quad (77)$$

which consistently reproduces the one of deterministic binning in Eq. (73) (and Eq. (20)) when letting $\mathfrak{p} \rightarrow 0$. On the other hand, it follows that the conditional distribution of Alice can thus be obtained by “mixing” the two rows of Tab. (75) with weights $1-\mathfrak{p}$ and \mathfrak{p} , respectively, accounting for the bit-flip errors, i.e.:

$$p_{A'|B}(a'|b)_{\text{det+n.p.}} = \begin{array}{c|ccc} a' \setminus b & 0 & 1 & \emptyset \\ \hline 0 & (1-\mathfrak{p}) \left(\eta \frac{P_{00}}{P_0^B} + \bar{\eta} \right) + \mathfrak{p} \left(\eta \frac{P_{10}}{P_0^B} \right) & (1-\mathfrak{p}) \left(\eta \frac{P_{01}}{P_1^B} + \bar{\eta} \right) + \mathfrak{p} \left(\eta \frac{P_{11}}{P_1^B} \right) & (1-\mathfrak{p}) \left(\eta P_0^A + \bar{\eta} \right) + \mathfrak{p} \left(\eta P_1^A \right) \\ \hline 1 & \mathfrak{p} \left(\eta \frac{P_{00}}{P_0^B} + \bar{\eta} \right) + (1-\mathfrak{p}) \left(\eta \frac{P_{10}}{P_0^B} \right) & \mathfrak{p} \left(\eta \frac{P_{01}}{P_1^B} + \bar{\eta} \right) + (1-\mathfrak{p}) \left(\eta \frac{P_{11}}{P_1^B} \right) & \mathfrak{p} \left(\eta P_0^A + \bar{\eta} \right) + (1-\mathfrak{p}) \left(\eta P_1^A \right) \end{array}. \quad (78)$$

Evaluating now the conditional entropy based on the above conditional distribution, we obtain the EC-term as

$$H(A'|B)_{\text{det+n.p.}} = \eta P_0^B h \left[(1-p) \left(\eta \frac{P_{00}}{P_0^B} + \bar{\eta} \right) + p \left(\frac{\eta P_{10}}{P_0^B} \right) \right] + \eta P_1^B h \left[(1-p) \left(\eta \frac{P_{01}}{P_1^B} + \bar{\eta} \right) + p \left(\frac{\eta P_{11}}{P_1^B} \right) \right] + \bar{\eta} h \left[(1-p) (\eta P_0^A + \bar{\eta}) + p (\eta P_1^A) \right], \quad (79)$$

which, as expected, reproduces $H(A'|B)_{\text{det}}$ in Eq. (76) after letting $p \rightarrow 0$.

C.1.4 Random binning

We also consider the case when Alice rather randomly bins her ternary variable A , in particular, she assigns each outcome \emptyset with equal probability to either outcome '0' or '1'. The corresponding stochastic matrix applied by Alice to Tab. (68) then reads

$$\mathcal{S}_{\text{rand}} = \begin{pmatrix} 1 & 0 & \frac{1}{2} \\ 0 & 1 & \frac{1}{2} \end{pmatrix}, \quad (80)$$

so that the third row of Tab. (68) gets redistributed equally (with a factor of 1/2) over the first two rows, i.e.:

$$p_{A'|B}(a', b)_{\text{rand}} = \begin{array}{c|ccc} a' \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 P_{00} + \frac{1}{2} \eta \bar{\eta} P_0^B & \eta^2 P_{01} + \frac{1}{2} \eta \bar{\eta} P_1^B & \bar{\eta} \eta P_0^A + \frac{1}{2} \bar{\eta}^2 \\ \hline 1 & \eta^2 P_{10} + \frac{1}{2} \eta \bar{\eta} P_0^B & \eta^2 P_{11} + \frac{1}{2} \eta \bar{\eta} P_1^B & \bar{\eta} \eta P_1^A + \frac{1}{2} \bar{\eta}^2 \end{array}. \quad (81)$$

As before, we determine then the probability distribution of Alice's outcomes conditioned on Bob's as

$$p_{A'|B}(a'|b)_{\text{rand}} = \begin{array}{c|ccc} a' \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta \frac{P_{00}}{P_0^B} + \frac{1}{2} \bar{\eta} & \eta \frac{P_{01}}{P_1^B} + \frac{1}{2} \bar{\eta} & \eta P_0^A + \frac{1}{2} \bar{\eta} \\ \hline 1 & \eta \frac{P_{10}}{P_0^B} + \frac{1}{2} \bar{\eta} & \eta \frac{P_{11}}{P_1^B} + \frac{1}{2} \bar{\eta} & \eta P_1^A + \frac{1}{2} \bar{\eta} \end{array}, \quad (82)$$

with the help of which we calculate the EC-term applicable to the case of random binning:

$$H(A'|B)_{\text{rand}} = \eta P_0^B h \left[\frac{\eta P_{00}}{P_0^B} + \frac{\bar{\eta}}{2} \right] + \eta P_1^B h \left[\frac{\eta P_{01}}{P_1^B} + \frac{\bar{\eta}}{2} \right] + \bar{\eta} h \left[\eta P_0^A + \frac{\bar{\eta}}{2} \right]. \quad (83)$$

C.1.5 Random binning with noisy preprocessing (bit-flip).

Finally, as before for deterministic binning, we consider the case in which Alice, apart from randomly binning the \emptyset -outcome, applies also noisy preprocessing [22, 23] to the resulting bit, i.e. flips its value with probability p . This then corresponds to her applying to Tab. (68) the stochastic matrix

$$\mathcal{S}_{\text{rand+n.p.}} = \begin{pmatrix} 1-p & p & \frac{1}{2} \\ p & 1-p & \frac{1}{2} \end{pmatrix}, \quad (84)$$

which consistently reproduces the one of random binning (80) when letting $p \rightarrow 0$. As before, the conditional probability distribution of Alice can then be obtained by just "mixing" the two rows of Tab. (81) (describing the case of random binning) with probabilities p and $1-p$, i.e.:

$$p_{A'|B}(a'|b)_{\text{rand+n.p.}} = \begin{array}{c|ccc} a' \setminus b & 0 & 1 & \emptyset \\ \hline 0 & (1-p) \eta \frac{P_{00}}{P_0^B} + p \eta \frac{P_{10}}{P_0^B} + \frac{1}{2} \bar{\eta} & (1-p) \eta \frac{P_{01}}{P_1^B} + p \eta \frac{P_{11}}{P_1^B} + \frac{1}{2} \bar{\eta} & (1-p) \eta P_0^A + p \eta P_1^A + \frac{1}{2} \bar{\eta} \\ \hline 1 & (1-p) \eta \frac{P_{10}}{P_0^B} + p \eta \frac{P_{00}}{P_0^B} + \frac{1}{2} \bar{\eta} & (1-p) \eta \frac{P_{11}}{P_1^B} + p \eta \frac{P_{01}}{P_1^B} + \frac{1}{2} \bar{\eta} & (1-p) \eta P_1^A + p \eta P_0^A + \frac{1}{2} \bar{\eta} \end{array}, \quad (85)$$

so that the relevant conditional entropy constituting the EC-term reads:

$$H(A'|B)_{\text{rand+n.p.}} = \eta P_0^B h \left[(1-p) \eta \frac{P_{00}}{P_0^B} + p \eta \frac{P_{10}}{P_0^B} + \frac{1}{2} \bar{\eta} \right] + \eta P_1^B h \left[(1-p) \eta \frac{P_{01}}{P_1^B} + p \eta \frac{P_{11}}{P_1^B} + \frac{1}{2} \bar{\eta} \right] + \bar{\eta} h \left[(1-p) \eta P_0^A + p \eta P_1^A + \frac{1}{2} \bar{\eta} \right]. \quad (86)$$

C.2 Calculation of the PA-term $H(A'|E)$

As within the CC attack Eve knows whether a local or a nonlocal correlation is being distributed to Alice and Bob, the entropy of the variable A (which describes the output of the measurement used by Alice for key distribution) conditioned on Eve's knowledge is given by the convex mixture of local and non-local contributions. Moreover, Eve not only knows when a local distribution is shared by the parties, but also knows then perfectly the outcomes A and B of Alice and Bob, respectively. Hence, the contribution of the local distribution to the conditional entropy of A is zero, unless Alice performs a non-deterministic preprocessing $p_{A'|A} \equiv \mathcal{S}$ of the outcome A that introduces some randomness, so that the knowledge of Eve about the resulting variable A' is no longer perfect.

In order to determine the conditional entropy $H(A'|E)$, we must only track Eve's knowledge of Alice's outputs. Hence, without loss of generality, we can assume Eve to hold a random variable E taking four values $e \in \{\{\tilde{e}\}, \{?\}\}$, where $\{?\}$ means that she distributed a nonlocal correlation and has no knowledge of Alice's output, while values $\tilde{e} \in \{0, 1, \emptyset\}$ correspond to the perfect knowledge of Alice's output A , which Eve possesses after distributing a local correlation (so that always $\tilde{e} = a$).

As a consequence, we can generally write the PA-term for the CC attack as

$$H(A'|E) = \sum_e p(E = e) H(A'|E = e) = (1 - q^\mathcal{L}) H(A'|E = ?) + q^\mathcal{L} \sum_{\tilde{e}} p(E = \tilde{e}|\mathcal{L}) H(A'|E = \tilde{e}), \quad (87)$$

where $q^\mathcal{L}$ is the local weight, so that $p(E = ?) = 1 - q^\mathcal{L}$ and $p(E = \tilde{e}) = q^\mathcal{L} p(E = \tilde{e}|\mathcal{L})$, with $p(E = \tilde{e}|\mathcal{L})$ denoting the probability of Eve recording \tilde{e} given she has distributed a local correlation.

As Eve has perfect knowledge of Alice's outcome, the conditional entropy within the "local rounds" is

$$\begin{aligned} H(A'|E = \tilde{e}) &= - \sum_{a'} p_{A'}(a'|E = \tilde{e}) \log_2 [p_{A'}(a'|E = \tilde{e})] = - \sum_{a', a} \mathcal{S}_{a'a} p_A(a|E = \tilde{e}) \log_2 \left[\sum_a \mathcal{S}_{a'a} p_A(a|E = \tilde{e}) \right] \\ &= - \sum_{a'} \mathcal{S}_{a'\tilde{e}} \log_2 \mathcal{S}_{a'\tilde{e}} =: H(\mathcal{S}_{\tilde{e}}), \end{aligned} \quad (88)$$

where we have used the fact that $p_A(a|E = \tilde{e}) = \delta_{a,\tilde{e}}$, and defined above $H(\mathcal{S}_{\tilde{e}}) := H[A']_{p(A'|A=\tilde{e})}$ as the entropy of the distribution described by the \tilde{e} -column of the stochastic matrix \mathcal{S} , which equivalently represents the randomness (entropy) of the preprocessed variable A' when $A = \tilde{e}$.

We can further simplify the expression (87) by expanding the probability $p(E = \tilde{e}|\mathcal{L})$, after realising that $p(E = \tilde{e}|\mathcal{L}) = p_A^\mathcal{L}(a = \tilde{e})$, where $p_A^\mathcal{L}(a) = \sum_b p_{AB}^\mathcal{L}(a, b|x^*, y^*)$ is the Alice's marginal of the local correlation. As the convex decomposition of the observed correlation (12) naturally carries over onto the marginal, i.e.:

$$p_A(a) = q^\mathcal{L} p_A^\mathcal{L}(a) + (1 - q^\mathcal{L}) p_A^{\mathcal{NL}}(a), \quad (89)$$

we can then explicitly compute

$$p_A^\mathcal{L}(a) = \begin{cases} \eta_\mathcal{L} P_a^A - \frac{1-q^\mathcal{L}}{q^\mathcal{L}} (1 - V) (Q_a^A - \frac{1}{2}) & \text{if } a \in \{0, 1\} \\ 1 - \eta_\mathcal{L} =: \bar{\eta}_\mathcal{L} & \text{if } a = \emptyset \end{cases}, \quad (90)$$

after substituting for the observed Alice's marginal, $p_A(a) = \sum_b p_{AB}(a, b)$, according to the lossy correlation (68), while the nonlocal contribution in Eq. (89) corresponds to the noiseless $p_A^{\mathcal{NL}}(a) = Q_a^A$. We also define $\eta_\mathcal{L}$ as above in Eq. (63), which should be understood as the effective "local" detection efficiency.

Finally, we arrive at the expression for the PA-term as

$$H(A'|E) = (1 - q^\mathcal{L}) H(A'|E = ?) + q^\mathcal{L} [p_A^\mathcal{L}(0) H(\mathcal{S}_{.0}) + p_A^\mathcal{L}(1) H(\mathcal{S}_{.1}) + p_A^\mathcal{L}(\emptyset) H(\mathcal{S}_{.\emptyset})], \quad (91)$$

where we should recall that $H(A'|E = ?)$ is the conditional entropy of Alice's outputs applicable whenever Eve distributes the nonlocal correlation within the CC attack, i.e. $p_{AB}^{\mathcal{NL}}(a) = Q_a^A$, and Alice preprocesses the outcome A onto A' according to the map \mathcal{S} .

In what follows, we calculate in detail the PA-term (91) for the preprocessing strategies of Alice listed in Tab. 1, as considered above in the evaluation of the EC-term, as well as the other two cases of random binning with and without noisy preprocessing.

C.2.1 No preprocessing or any deterministic binning

Whenever the matrix \mathcal{S} describes a stochastic map that is deterministic, i.e. contains only 0s or 1s as its entries, the whole second term in Eq. (91) identically vanishes. On the other hand, as within nonlocal rounds

inconclusive outcomes never occur, and so any operations on \emptyset -outcomes are never performed, any binning strategy does not affect the first term in Eq. (91). Thus, we can write the PA-term in absence of preprocessing or for any deterministic binning as

$$H(A'|E)_{\text{no-prep}} = H(A'|E)_{\text{det}} = (1 - q^{\mathcal{L}}) H(A'|E = ?) = (1 - q^{\mathcal{L}}) h[\mathbf{Q}_0^{\text{A}}]. \quad (92)$$

C.2.2 Deterministic binning with noisy preprocessing (bit-flip)

In case Alice decides to further “noisy preprocess” her outcomes after having binned \emptyset to 0, then the overall preprocessing map she applies corresponds to the \mathcal{S} -matrix introduced in Eq. (77). As a result, the first term in Eq. (91) can be obtained from Eq. (92) after including a bit-flip occurring with probability \mathfrak{p} , while the second term in Eq. (91) is then no longer zero, as the entropy for each column of \mathcal{S} equals now $h[\mathfrak{p}]$. Thus, the full PA-term (91) then reads

$$H(A'|E)_{\text{det+n.p.}} = (1 - q^{\mathcal{L}}) h[(1 - \mathfrak{p}) \mathbf{Q}_0^{\text{A}} + \mathfrak{p} \mathbf{Q}_1^{\text{A}}] + q^{\mathcal{L}} h[\mathfrak{p}]. \quad (93)$$

C.2.3 Random binning

As before, the first term in Eq. (91) is unaffected by any binning of the \emptyset -outcomes and, hence, also when Alice bins these randomly. However, the second term in Eq. (91) must now be evaluated based on the stochastic matrix \mathcal{S} given in Eq. (80), which is no longer deterministic—its last column yields a non-trivial contribution. Hence, for random binning of inconclusive outcomes we obtain

$$H(A'|E)_{\text{rand}} = (1 - q^{\mathcal{L}}) h[\mathbf{Q}_0^{\text{A}}] + q^{\mathcal{L}} \bar{\eta}_{\mathcal{L}} h\left[\frac{1}{2}\right] = (1 - q^{\mathcal{L}}) h[\mathbf{Q}_0^{\text{A}}] + 1 - \eta. \quad (94)$$

C.2.4 Random binning with noisy preprocessing (bit-flip)

In case Alice decides to further “noisy preprocess” her outcomes after having binned \emptyset randomly to 0 and 1, she, in fact, implements the stochastic matrix \mathcal{S} given in Eq. (84). Within the first term of Eq. (91) one has to account for the bit-flip occurring with probability \mathfrak{p} and arrives at the same expression as in Eq. (93). Whereas for the second term, we note that the entropy of the first two columns of \mathcal{S} in Eq. (84) is then $h[\mathfrak{p}]$, while the entropy of the last column is 1. Therefore, we have

$$H(A'|E)_{\text{rand+n.p.}} = (1 - q^{\mathcal{L}}) h[(1 - \mathfrak{p}) \mathbf{Q}_0^{\text{A}} + \mathfrak{p} \mathbf{Q}_1^{\text{A}}] + (\eta - 1 + q^{\mathcal{L}}) h[\mathfrak{p}] + 1 - \eta. \quad (95)$$

D CC-based upper bounds on key rates and the resulting noise thresholds

D.1 One-way protocols involving maximally entangled states

In this section we utilise the formulae derived in App. C for the EC- and PA-terms under particular preprocessing strategies of Alice, in order to determine the corresponding upper bounds (66) on the one-way key rates for the standard CHSH-based 2333- and 2233-protocols (in the finite detection efficiency model), as well as 2322- and 2222-protocols (in the finite visibility model). The goal is to determine analytically the tolerable noise thresholds below which no key rate can be distilled, some of which are listed in Table 1 for specific preprocessing strategies.

In this section we assume that the parties ideally measure the pure, maximally entangled state with $\theta = \pi/2$ in (17) via projective measurements (27) with the measurement settings $x^*, y^* = \{0, 2\}$ being used for key generation in the 2333- and 2322-protocols, and any settings $x^*, y^* \in \{0, 1\}$ in the 2233- and 2222-protocols. While the obtained results for the EC- and PA-terms hold generally for any η and V , we consider here specifically two cases of purely lossy correlations (with $V = 1$) and of purely noisy correlations (with $\eta = 1$).

D.1.1 Finite detection efficiency

Given perfect visibility ($V = 1$) but imperfect detection efficiency ($\eta < 1$), the correlation (68) used for the key generation simplifies to the purely lossy one (53) with all $\mathbf{P}_{ab} = \mathbf{Q}_{ab}$. Moreover, in case of the 2333-protocol we have from Eq. (43) that $\mathbf{Q}_{ab} = \frac{1}{2} \delta_{a,b}$ with marginals $\mathbf{Q}_a^{\text{A}} = \mathbf{Q}_b^{\text{B}} = \frac{1}{2}$, whereas for the 2233-protocol $\mathbf{Q}_{00} = \mathbf{Q}_{11} = (2 - \sqrt{2})/8$ and $\mathbf{Q}_{01} = \mathbf{Q}_{10} = (2 + \sqrt{2})/8$ if $x^* = y^* = 1$, and $\mathbf{Q}_{00} = \mathbf{Q}_{11} = (2 + \sqrt{2})/8$, $\mathbf{Q}_{01} = \mathbf{Q}_{10} = (2 - \sqrt{2})/8$ otherwise (with marginal probabilities also always equal to $1/2$).

No preprocessing. In absence of any preprocessing map, we use Eqs. (72) and (92) to calculate $r_{1\text{-way,no-prep}}^\uparrow = H(A'|E)_{\text{no-prep}} - H(A'|B)_{\text{no-prep}}$, which after substituting also for the Q-probabilities of the 2333-protocol and the optimal local weight (58) reads

$$r_{1\text{-way,no-prep}}^\uparrow(\eta) = \eta \left[4\eta + \log_2 \eta - 2\sqrt{2}(1-\eta) \right] + 2\log_2(1-\eta) - 3\eta, \quad (96)$$

and leads to the critical detection efficiency

$$\eta_{\text{crit}}^{\text{no-prep}} \approx 91.85\%. \quad (97)$$

Following the same steps for the 2233-protocol one arrives at the formula for $r_{1\text{-way,no-prep}}^\uparrow$ with a zero at $\eta_{\text{crit}}^{\text{no-prep}} \approx 96.90\%$ irrespectively of the particular choice of key settings $x^*, y^* \in \{0, 1\}$.

Deterministic binning. In case Alice applies deterministic binning as her preprocessing strategy, we use Eqs. (76) and (92) to calculate $r_{1\text{-way,det}}^\uparrow = H(A'|E)_{\text{det}} - H(A'|B)_{\text{det}}$ instead, which after substituting for the Q-probabilities of the 2333-protocol and the optimal local weight (58) reads

$$r_{1\text{-way,det}}^\uparrow(\eta) = \eta \left(3\eta + 2\sqrt{2}\eta + \frac{\log_2 \eta}{2} - 2\sqrt{2} \right) - \frac{1-\eta}{2} \left[-(2-\eta)\log_2(2-\eta) - \eta\log_2(1-\eta) \right] - 1 - \eta, \quad (98)$$

and leads to the critical detection efficiency

$$\eta_{\text{crit}}^{\text{det}} \approx 89.16\%. \quad (99)$$

The fact that $\eta_{\text{crit}}^{\text{det}} < \eta_{\text{crit}}^{\text{no-prep}}$ suggests that the binning procedure of the inconclusive outcomes is indeed beneficial for the parties to be able to tolerate lower detection efficiencies. Following the same steps for the 2233-protocol one arrives at the formula for $r_{1\text{-way,det}}^\uparrow$ with a zero at $\eta_{\text{crit}}^{\text{det}} \approx 94.80\%$ irrespectively of the choice of key settings.

Deterministic binning with noisy preprocessing. If Alice decides to apply noisy preprocessing apart from binning deterministically her inconclusive outcomes \emptyset , we have $r_{1\text{-way,det+n.p.}}^\uparrow = H(A'|E)_{\text{det+n.p.}} - H(A'|B)_{\text{det+n.p.}}$ that can be calculated using Eqs. (79) and (93), so that after substituting for the Q-probabilities of the 2333-protocol it reads

$$r_{1\text{-way,det+n.p.}}^\uparrow(\eta) = 1 - q^\mathcal{L} + q^\mathcal{L} h[\mathfrak{p}] - \frac{\eta}{2} h[\mathfrak{p}] - \frac{\eta}{2} h[(1-\mathfrak{p})(1-\eta) + \mathfrak{p}\eta] - (1-\eta) h \left[(1-\mathfrak{p}) \left(1 - \frac{\eta}{2} \right) + \mathfrak{p} \left(\frac{\eta}{2} \right) \right]. \quad (100)$$

Substituting then for the optimal local weight, $q^\mathcal{L}$ in Eq. (58), one can verify that the critical detection efficiency gets smaller with the bit-flip probability approaching $\mathfrak{p} \rightarrow \frac{1}{2} \pm$. Although in such a regime the upper bound and, hence, any attainable rate is severely suppressed, in order to determine its lowest possible positive value we expand $r_{1\text{-way,det+n.p.}}^\uparrow$ in δ after substituting for $\mathfrak{p} = \frac{1}{2} \pm \delta$, i.e.:

$$r_{1\text{-way,det+n.p.}}^\uparrow(\eta) = \frac{2\eta(\eta^2 + 2\sqrt{2}\eta + 4\eta - 2\sqrt{2} - 4)}{\ln 2} \delta^2 + O(\delta^3), \quad (101)$$

which allows us to locate the zero at

$$\eta_{\text{crit}}^{\text{det+n.p.}} = \sqrt{10 + 6\sqrt{2}} - 2 - \sqrt{2} \approx 88.52\%. \quad (102)$$

Following the same steps for the 2233-protocol one arrives at the formula for $r_{1\text{-way,det+n.p.}}^\uparrow$ with a zero at $\eta_{\text{crit}}^{\text{det+n.p.}} = 2 \left(\sqrt{8 + 5\sqrt{2}} - 2 - \sqrt{2} \right) \approx 93.59\%$ irrespectively of the choice of key settings.

Random binning. In case Alice applies random binning as her preprocessing strategy, we use Eqs. (83) and (94) to calculate $r_{1\text{-way,rand}}^\uparrow = H(A'|E)_{\text{rand}} - H(A'|B)_{\text{rand}}$, which after substituting for the Q-probabilities of the 2333-protocol and the optimal local weight (58) reads

$$r_{1\text{-way,rand}}^\uparrow(\eta) = \frac{\eta}{2} \left(\eta(\log_2(1+\eta) - \log_2(1-\eta)) + \log_2(1-\eta) + \log_2(1+\eta) - 2(1-\eta) \left(3 + 2\sqrt{2} \right) \right), \quad (103)$$

and leads to the critical detection efficiency

$$\eta_{\text{crit}}^{\text{rand}} \approx 88.34\%. \quad (104)$$

Following the same steps for the 2233-protocol one arrives at the formula for $r_{1\text{-way,rand}}^\uparrow$ with a zero at $\eta_{\text{crit}}^{\text{rand}} \approx 94.03\%$ irrespectively of the choice of key settings.

Random binning with noisy preprocessing. If Alice decides to apply noisy preprocessing apart from randomly binning her inconclusive outcomes \emptyset , we have $r_{1\text{-way,rand+n.p.}}^\uparrow = H(A'|E)_{\text{rand+n.p.}} - H(A'|B)_{\text{rand+n.p.}}$ that can be calculated using Eqs. (86) and (95), so that after substituting for the Q-probabilities of the 2333-protocol it reads

$$r_{1\text{-way,rand+n.p.}}^\uparrow(\eta) = 1 - q^\mathcal{L} + (\eta - 1 + q^\mathcal{L}) h[\mathbf{p}] - \eta h\left[\mathbf{p}\eta + \frac{1}{2}(1 - \eta)\right]. \quad (105)$$

Substituting then for the optimal local weight, $q^\mathcal{L}$ in Eq. (58), one can verify that the critical detection efficiency gets smaller with the bit-flip probability approaching $\mathbf{p} \rightarrow \frac{1}{2}_\pm$. Hence, similarly to the “det+n.p.” case, we can determine its lowest possible positive value by expanding $r_{\text{rand+n.p.}}^\uparrow$ in δ after substituting also for $\mathbf{p} = \frac{1}{2} \pm \delta$, i.e.:

$$r_{1\text{-way,rand+n.p.}}^\uparrow(\eta) = \frac{2\eta(\eta(\eta + 2\sqrt{2} + 3) - 2\sqrt{2} - 3)\delta^2}{\log(2)} + O(\delta^3), \quad (106)$$

which implies

$$\eta_{\text{crit}}^{\text{rand+n.p.}} = \frac{1}{2} \left(\sqrt{29 + 20\sqrt{2}} - 3 - 2\sqrt{2} \right) \approx 87.01\%. \quad (107)$$

Following the same steps for the 2233-protocol one arrives at the formula for $r_{\text{rand+n.p.}}^\uparrow$ with a zero at $\eta_{\text{crit}}^{\text{rand}} = \sqrt{23 + 16\sqrt{2}} - 3 - 2\sqrt{2} \approx 92.64\%$ irrespectively of the choice of key settings.

D.1.2 Finite visibility

Given perfect detection efficiency ($\eta = 1$) but imperfect visibility ($V < 1$), one should consider for key generation the correlation (68) after setting $\eta = 1$ instead, where now in the case of the 2333-protocol: $P_{ab} = V\frac{1}{2}\delta_{a,b} + \frac{1-V}{4}$ with marginals $P_a^A = P_b^B = \frac{1}{2}$; whereas for the 2233-protocol: $P_{00} = P_{11} = V(2 - \sqrt{2})/8 + \frac{1-V}{4}$ and $P_{01} = P_{10} = V(2 + \sqrt{2})/8 + \frac{1-V}{4}$ if $x^* = y^* = 1$, and $P_{00} = P_{11} = V(2 + \sqrt{2})/8 + \frac{1-V}{4}$, $P_{01} = P_{10} = V(2 - \sqrt{2})/8 + \frac{1-V}{4}$ otherwise (with marginal probabilities also always equal to $\frac{1}{2}$). Consistently, all the noiseless Q-probabilities specified previously when dealing with purely lossy correlations can be recovered by setting $V = 1$ in all the P-probabilities listed above.

No preprocessing. In absence of any preprocessing map the upper bound on the one-way rate can be calculated again using Eqs. (72) and (92) as $r_{1\text{-way,no-prep}}^\uparrow = H(A'|E)_{\text{no-prep}} - H(A'|B)_{\text{no-prep}}$, which after substituting for the P- and Q-probabilities of the 2333-protocol, $\eta = 1$, and the optimal local weight (52) reads

$$r_{1\text{-way,no-prep}}^\uparrow(V) = V(2 + \sqrt{2}) - h\left[\frac{1+V}{2}\right] - 1 - \sqrt{2}, \quad (108)$$

and vanishes at the critical visibility:

$$V_{\text{crit}}^{\text{no-prep}} \approx 83.00\%. \quad (109)$$

Following the same steps for the 2222-protocol one arrives at the formula for $r_{\text{no-prep}}^\uparrow$ with a zero at $V_{\text{crit}}^{\text{no-prep}} \approx 90.61\%$, irrespectively of the particular choice of key settings $x^*, y^* \in \{0, 1\}$.

Noisy preprocessing. If Alice decides to apply noisy preprocessing to her binary variable, the upper bound (66) on the one-way rate can be evaluated by setting $\eta = 1$ in either Eqs. (79)&(93) or Eqs. (86)&(95) as $r_{1\text{-way,n.p.}}^\uparrow = H(A'|E)_{\bullet+\text{n.p.}} - H(A'|B)_{\bullet+\text{n.p.}}$. Because for the perfect detection efficiency the inconclusive outcomes never occur, formulae derived assuming any binning of the \emptyset -outcomes are valid upon setting $\eta = 1$. Hence, substituting also for the P- and Q-probabilities of the 2322-protocol, we arrive at

$$r_{1\text{-way,n.p.}}^\uparrow(V) = 1 - q^\mathcal{L} + q^\mathcal{L} h[\mathbf{p}] - h\left[\left(\mathbf{p} - \frac{1}{2}\right)V + \frac{1}{2}\right]. \quad (110)$$

Substituting then for the optimal local weight, $q^\mathcal{L}$ in Eq. (52), one can again verify that the critical visibility gets smaller with the bit-flip probability approaching $\mathbf{p} \rightarrow \frac{1}{2}_\pm$. Although in such a regime the upper bound, and hence any attainable rate, is severely suppressed, in order to determine its lowest possible positive value we expand $r_{\text{n.p.}}^\uparrow$ in δ after substituting also for $\mathbf{p} = \frac{1}{2} \pm \delta$, i.e.:

$$r_{1\text{-way,n.p.}}^\uparrow(V) = \frac{2(V^2 - (1 - V)(2 + \sqrt{2}))}{\ln 2} \delta^2 + O(\delta^3), \quad (111)$$

which allows us to locate the zero at

$$V_{\text{crit}}^{\text{n.p.}} = \sqrt{\frac{7}{2} + 2\sqrt{2}} - 1 - \frac{1}{\sqrt{2}} \approx 80.85\%. \quad (112)$$

Following the same steps for the 2222-protocol, one arrives at the formula for $r_{\text{n.p.}}^\uparrow$ with a zero at $V_{\text{crit}}^{\text{n.p.}} = \sqrt{10 + 6\sqrt{2}} - 2 - \sqrt{2} \approx 88.52\%$, irrespectively of the choice of key settings.

D.1.3 Optimisation over all preprocessing maps

In this section, we discuss the results obtained for the computation of the *general* upper bounds applicable to one-way key rates (4)—the asymptotic one-way key rates optimised over all preprocessing strategies including not just the stochastic mapping applied by Alice, $p_{A'|A}$, on her variable A , but also the extra message M she prepares by applying $p_{M'|A'}$ on A' and sends publicly to Bob, i.e.:

$$r_{1\text{-way}}(A \rightarrow B) \leq r_{1\text{-way}}^\uparrow := \max_{p_{A'|A}, p_{M'|A'}} I(A':B|M) - I(A':E|M), \quad (113)$$

where the so-defined $r_{1\text{-way}}^\uparrow$ corresponds to the upper bound (6) being now crucially maximised over all the $p_{A'|A}$ and $p_{M'|A'}$ preprocessing maps, including ones with $|A'| > 2$ or $|M| > 2$. In particular, we perform the maximisation in Eq. (113) numerically by means of heuristic methods, despite dealing with a non-convex optimisation problem. This allows us to, at least numerically, determine $r_{1\text{-way}}^\uparrow$ while firstly accounting for finite detection efficiency ($\eta < 1$ with purely lossy correlation (53) being shared) within the CHSH-based 2333- and 2233-protocols. In a similar manner, we then consider the visibility to be finite instead ($V < 1$), and determine $r_{1\text{-way}}^\uparrow$ for the corresponding CHSH-based 2322- and 2222-protocols. These upper bounds can then be used to determine *universal* noise thresholds $\check{\eta}_{\text{crit}}$ and \check{V}_{crit} , below which *no* key can be distilled with one-way communication—these appear in Table 1 (column ‘any’) and in Fig. 4 of the main text. However, in order to perform the numerical optimisation, we must decide on the number of outcomes for the discrete random variables A' and M , which, in principle, can be as large as possible. We proceed phenomenologically, i.e. in each case we raise the outcome-number by one, until the moment we can conclude that no further increase is necessary.

Finally, let us emphasise that we perform the optimisation here over preprocessing strategies for the same CHSH-optimal correlations, for which the thresholds in App. D.1 were derived, utilising maximally entangled states $|\Phi^+\rangle$ and standard CHSH measurements (27).

Finite detection efficiency. In our optimisation we directly seek the minimal detection efficiency, η , such that $r_{1\text{-way}}^\uparrow = 0$, which then constitutes the desired universal threshold, $\check{\eta}_{\text{crit}}$, below which no key can be extracted with one-way communication. We first consider the 2333-protocol with measurement settings $x^*, y^* = \{0, 2\}$, and then briefly discuss the 2233-protocol case with $x^*, y^* \in \{0, 1\}$.

We start by optimising the bound (113) only over the $p_{A'|A}$ maps, while disregarding the $p_{M'|A'}$ maps. We perform the maximisation over all corresponding stochastic matrices $\mathcal{S}_{A \rightarrow A'}$ of size $|A'| \times |A|$, where we vary the outcome number $2 \leq |A'| \leq 7$ ($|A| = 3$ is fixed with $A \in \{0, 1, \emptyset\}$ by the lossy correlation $p_{AB}(a, b)$ in Eq. (53) considered). Independently of the outcome number $|A'|$, we always arrive at the critical detection efficiency:

$$\check{\eta}_{\text{crit}}^{A \rightarrow A'} \approx 87.0105\%, \quad (114)$$

which coincides up to our best-achieved numerical precision with the critical efficiency attainable with random binning followed by noisy preprocessing, i.e. $\eta_{\text{crit}}^{\text{rand+n.p.}}$ in Eq. (107). Hence, as the critical efficiency (114) applies to all preprocessing strategies of Alice $p_{A'|A}$, we conjecture that the strategy of random binning combined with noisy preprocessing is the optimal form of defense against the CC attack, when Alice is not utilising the message M sent to Bob. However, we observe that the numerical optimisation converges to different stochastic matrices leading to the critical efficiency (114), suggesting that the optimal preprocessing strategy is then not unique.

Secondly, we incorporate into the maximisation of (113) the optimisation over both $p_{A'|A}$ and $p_{M'|A'}$, which correspond to some choice of $\mathcal{S}_{A \rightarrow A'}$ and $\mathcal{S}_{A' \rightarrow M}$ stochastic matrices of dimensions $|A| \times |A'|$ and $|A'| \times |M|$, respectively. Allowing the outcome numbers to range in $3 \leq |A'| \leq 5$ and $2 \leq |M| \leq 5$, we surprisingly observe that the upper bound $r_{1\text{-way}}^\uparrow$ can be increased thanks to inclusion of the $p_{M'|A'}$ mapping. This, in turn, allows to lower the required critical efficiency to

$$\check{\eta}_{\text{crit}}^{A \rightarrow A' \rightarrow M} = \check{\eta}_{\text{crit}}^{A \rightarrow M} \approx 85.3553\%. \quad (115)$$

Furthermore, we observe this to be possible when considering already $|A'| = 3$, $|M| = 2$ (but not for $|A'| = 2$, $|M| = 2$ for which the value (114) is recovered).

However, as noted in Eq. (115), we find that the inclusion of the map $p_{A'|A}$ is then unnecessary—it is sufficient for Alice to use the “raw” A -variable for the key and send some A -dependent message M to Bob as the best preprocessing strategy. In particular, we establish the same critical efficiency (115) by considering now only $\mathcal{S}_{A \rightarrow M}$ with $2 \leq |M| \leq 5$, implying that the outcome number $|M| = 2$ is already sufficient, as the optimal maps always possess the crucial feature of effectively “singling out” the \emptyset -outcome of Alice within the message send to Bob. In particular, it is always optimal for Alice to simply announce to Bob whether she has a conclusive outcome or not, which is the bit of information to be encoded in M . Indeed, from the perspective of the CC attack this does not provide any extra information to Eve, as she knows whether a local or non-local correlation was distributed to the parties, while $A = \emptyset$ never occurs in the latter case. Hence, as Eve perfectly knows the outcomes of Alice whenever a local correlation is shared, she also always knows whenever Alice records any inconclusive outcome \emptyset . Note also that Alice announcing the \emptyset outcomes does not constitute postselection—these rounds are not discarded by the parties and therefore the announcements do not lead to a violation of the detection loophole.

As we now show, the critical value (115) can be, in fact, analytically proven for the preprocessing strategy described above—corresponding to Alice applying on her key variable $A \in \{0, 1, \emptyset\}$ the stochastic matrix

$$\mathcal{S}_{A \rightarrow M} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad (116)$$

such that the binary $M \in \{\checkmark, \emptyset\}$ takes the value $M = \emptyset$ if $A = \emptyset$, and $M = \checkmark$ if Alice records a conclusive outcome.

To calculate the resulting bound $r_{p_{M|A}}^\dagger = I(A:B|M) - I(A:E|M)$ for $p_{M|A} \equiv \mathcal{S}_{A \rightarrow M}$ in Eq. (116), let us first consider the mutual information between Alice and Bob conditioned on M . We can construct the tripartite probability distribution $p_{MAB}(m, a, b)$ by first augmenting the lossy correlation (53) shared by Alice and Bob in key generation rounds with an extra “dummy” random variable \tilde{A} perfectly correlated to A ,

$$p_{\tilde{A}AB}(\tilde{a}, a, b) = \begin{array}{c|ccccccccc} \tilde{a} \setminus a, b & 0, 0 & 0, 1 & 0, \emptyset & 1, 0 & 1, 1 & 1, \emptyset & \emptyset, 0 & \emptyset, 1 & \emptyset, \emptyset \\ \hline 0 & \eta^2 Q_{00} & \eta^2 Q_{01} & \eta\bar{\eta}Q_0^A & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & \eta^2 Q_{10} & \eta^2 Q_{11} & \eta\bar{\eta}Q_1^A & 0 & 0 & 0 \\ \hline \emptyset & 0 & 0 & 0 & 0 & 0 & 0 & \bar{\eta}\eta Q_0^B & \bar{\eta}\eta Q_1^B & \bar{\eta}^2 \end{array}, \quad (117)$$

and transform it by applying $\mathcal{S}_{\tilde{A} \rightarrow M}$ of Eq. (116) (adding together the first two rows) to obtain the desired

$$p_{MAB}(m, a, b) = \begin{array}{c|ccccccccc} m \setminus a, b & 0, 0 & 0, 1 & 0, \emptyset & 1, 0 & 1, 1 & 1, \emptyset & \emptyset, 0 & \emptyset, 1 & \emptyset, \emptyset \\ \hline \checkmark & \eta^2 Q_{00} & \eta^2 Q_{01} & \eta\bar{\eta}Q_0^A & \eta^2 Q_{10} & \eta^2 Q_{11} & \eta\bar{\eta}Q_1^A & 0 & 0 & 0 \\ \hline \emptyset & 0 & 0 & 0 & 0 & 0 & 0 & \bar{\eta}\eta Q_0^B & \bar{\eta}\eta Q_1^B & \bar{\eta}^2 \end{array}. \quad (118)$$

Consistently, the marginal distribution of the message variable M reads (summing the rows in Tab. (118)),

$$p_M(m) = \begin{cases} \eta, & \text{if } m = \checkmark \\ \bar{\eta}, & \text{if } m = \emptyset \end{cases}, \quad (119)$$

as it effectively denotes whether a detection event occurred or not. With help of Tab. (118) we then calculate

$$\begin{aligned} I(A:B|M) &= \eta I(A:B|M = \checkmark) + \bar{\eta} I(A:B|M = \emptyset) \\ &= \eta \left(h[\eta Q_0^A] + H\{\eta Q_0^B, \eta Q_1^B, \bar{\eta}\} - H\{\eta Q_{00}, \eta Q_{01}, \eta\bar{\eta}Q_0^A, \eta Q_{10}, \eta Q_{11}, \eta\bar{\eta}Q_1^A\} \right), \end{aligned} \quad (120)$$

where $I(A:B|M = \emptyset) = 0$, as $H(A|M = \emptyset) = 0$ and $H(B|M = \emptyset) = H(A, B|M = \emptyset)$. Substituting further the Q-distribution for the 2333-protocol, we finally obtain

$$I(A:B|M) = \eta \left(h\left[\frac{\eta}{2}\right] + H\left\{\frac{\eta}{2}, \frac{\eta}{2}, \bar{\eta}\right\} - H\left\{\frac{\eta}{2}, \frac{\bar{\eta}}{2}, \frac{\eta}{2}, \frac{\bar{\eta}}{2}\right\} \right) = \eta h\left[\frac{\eta}{2}\right] - \eta\bar{\eta}. \quad (121)$$

We now turn to the mutual information between Alice and Eve conditioned on M , which we calculate in a similar manner after identifying the tripartite distribution $p_{MAE}(m, a, e)$. We first, however, note that Eq. (90) implies that

$$p_{AE}(a, e = \tilde{e}) = q^{\mathcal{L}} p_A^{\mathcal{L}}(a), \quad p_{AE}(a, e = ?) = (1 - q^{\mathcal{L}}) Q_a^A \quad (122)$$

where $\tilde{e} \in \{0, 1, \emptyset\}$ is the outcome of Eve when she distributes a local correlation. Hence, by introducing again an auxiliary variable \tilde{A} that is perfectly correlated to A , we can write the overall distribution as

$$p_{\tilde{A}AE}(\tilde{a}, a, e) = \begin{array}{|c|c|c|c|c|c|} \hline \tilde{a} \setminus a, e & 0, 0 & 0, ? & 1, 1 & 1, ? & \emptyset, \emptyset \\ \hline 0 & q^{\mathcal{L}}\eta_{\mathcal{L}}\mathbb{Q}_0^A & (1 - q^{\mathcal{L}})\mathbb{Q}_0^A & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & q^{\mathcal{L}}\eta_{\mathcal{L}}\mathbb{Q}_1^A & (1 - q^{\mathcal{L}})\mathbb{Q}_1^A & 0 \\ \hline \emptyset & 0 & 0 & 0 & 0 & q^{\mathcal{L}}(\bar{\eta}_{\mathcal{L}}) \\ \hline \end{array}, \quad (123)$$

and apply the stochastic map $\mathcal{S}_{\tilde{A} \rightarrow M}$ of Eq. (116) onto the auxiliary variable to obtain the desired distribution:

$$p_{MAE}(m, a, e) = \begin{array}{|c|c|c|c|c|c|} \hline m \setminus a, e & 0, 0 & 0, ? & 1, 1 & 1, ? & \emptyset, \emptyset \\ \hline \checkmark & q^{\mathcal{L}}\eta_{\mathcal{L}}\mathbb{Q}_0^A & (1 - q^{\mathcal{L}})\mathbb{Q}_0^A & q^{\mathcal{L}}\eta_{\mathcal{L}}\mathbb{Q}_1^A & (1 - q^{\mathcal{L}})\mathbb{Q}_1^A & 0 \\ \hline \emptyset & 0 & 0 & 0 & 0 & q^{\mathcal{L}}(\bar{\eta}_{\mathcal{L}}) \\ \hline \end{array}. \quad (124)$$

As the distribution (124) consistently yields the same marginal distribution (119) for the message variable M , the conditional mutual information can be similarly split into

$$\begin{aligned} I(A:E|M) &= \eta I(A:E|M = \checkmark) + \bar{\eta} I(A:E|M = \emptyset) \\ &= \eta \left(h[\eta\mathbb{Q}_0^A] + H \left\{ \frac{q^{\mathcal{L}}\eta_{\mathcal{L}}}{\eta}\mathbb{Q}_0^A, \frac{q^{\mathcal{L}}\eta_{\mathcal{L}}}{\eta}\mathbb{Q}_1^A, \frac{1 - q^{\mathcal{L}}}{\eta} \right\} \right. \\ &\quad \left. - H \left\{ \frac{q^{\mathcal{L}}\eta_{\mathcal{L}}}{\eta}\mathbb{Q}_0^A, \frac{(1 - q^{\mathcal{L}})}{\eta}\mathbb{Q}_0^A, \frac{q^{\mathcal{L}}\eta_{\mathcal{L}}}{\eta}\mathbb{Q}_1^A, \frac{(1 - q^{\mathcal{L}})}{\eta}\mathbb{Q}_1^A \right\} \right), \end{aligned} \quad (125)$$

where $I(A:E|M = \emptyset) = 0$, as it is always the single outcome \emptyset being transmitted between Alice and Eve when $M = \emptyset$, carrying zero information on its own. Substituting further the trivial marginals $\mathbb{Q}_0^A = \mathbb{Q}_1^A = \frac{1}{2}$ that apply to the 2333-protocol, we have

$$I(A:E|M) = \eta h\left[\frac{\eta}{2}\right] + q^{\mathcal{L}} - 1 = \eta h\left[\frac{\eta}{2}\right] + \bar{\eta}(1 + (3 + 2\sqrt{2})\eta) - 1, \quad (126)$$

where we substituted already for the optimal local weight $q^{\mathcal{L}}$ according to Eq. (58).

Finally, we arrive at the desired upper bound for the one-way key rate based on the CC attack applicable when Alice does not preprocess her outcome A , but rather reveals the rounds in which she obtained \emptyset by transmitting a message M prepared by applying $p_{M|A} \equiv \mathcal{S}_{A \rightarrow M}$ of Eq. (116) to A , i.e.:

$$r_{p_{M|A}}^{\uparrow} = I(A:B|M) - I(A:E|M) = 1 - \bar{\eta}(1 + (3 + 2\sqrt{2})\eta) - \eta\bar{\eta}, \quad (127)$$

which vanishes at the value:

$$\eta_{\text{crit}}^{A \rightarrow M} = \frac{1}{4} (2 + \sqrt{2}) \approx 85.3553\%, \quad (128)$$

which coincides, indeed, with the numerically obtained critical efficiency in Eq. (115), optimised over all preprocessing strategies of Alice. Interestingly, it further coincides with the critical efficiency (143), which we derive below by applying the CC attack to the (intrinsic-information-based) upper bound that accounts for *two-way* communication, but assumes symmetric deterministic binning of inconclusive outcomes by both Alice and Bob.

For completeness, let us just summarize the results obtained for the 2233-protocol. Considering preprocessing strategies in which Alice applies only the $p_{A'|A}$ map in Eq. (113) (with $2 \leq |A'| \leq 5$) we obtain with help of numerical heuristic methods the following critical efficiency

$$\check{\eta}_{\text{crit}}^{A \rightarrow A' \rightarrow M} = \check{\eta}_{\text{crit}}^{A \rightarrow A'} \approx 92.6380\% \quad (129)$$

irrespective of the key settings used by Alice and Bob ($x^*, y^* \in \{0, 1\}$), which coincides with the expression obtained previously in Sec. D.1.1 when Alice resorts to random binning of her outcome, followed by noisy preprocessing. Nevertheless, we find numerically multiple stochastic matrices allowing to attain the value (129), which suggests that the optimal preprocessing strategy is not unique from the perspective of the CC attack.

As already noted in Eq. (129), in contrast to the case of the 2333-protocol, we do not observe any improvement of the critical efficiency (129) (lowering its value) by allowing Alice also to perform arbitrary maps $p_{M|A'}$ in Eq. (113) and letting $2 \leq |A'|, |M| \leq 5$. On the other hand, if one disregards the mapping $A \rightarrow A'$ and allows only for the map $A \rightarrow M$ with M being the public message, then we observe that, similarly to the 2333-protocol, it is (numerically) optimal for Alice to just signal the occurrences of the inconclusive outcomes \emptyset . Such a strategy when considering the CC attack leads to the following critical efficiency:

$$\eta_{\text{crit}}^{A \rightarrow M} = \frac{4(3 + 2\sqrt{2})}{2(5 + 4\sqrt{2}) + \sqrt{2} \log_2[3 + 2\sqrt{2}]} \approx 93.5910\%, \quad (130)$$

which we obtain analytically following the same procedure as for the 2333-protocol above, irrespectively of the key settings used by Alice and Bob. Note that $\check{\eta}_{\text{crit}}^{A \rightarrow A' \rightarrow M} < \eta_{\text{crit}}^{A \rightarrow M}$, so the CC attack suggests for the 2233-protocol that the best preprocessing strategy for Alice is to apply $p_{A'|A}$ that implements random binning of inconclusive outcomes, followed by noisy preprocessing of all the resulting outcomes (i.e. the stochastic matrix $\mathcal{S}_{\text{rand+n.p.}}$ in Eq. (84) with $\mathbf{p} \rightarrow \frac{1}{2}_{\pm}$).

Finite visibility. In analogy to the previous section, we perform an optimisation in which we directly seek the minimal visibility, V , such that $r_{1\text{-way}}^{\uparrow} = 0$, which then constitutes the desired universal threshold, \check{V}_{crit} , below which again no key can be extracted with one-way communication. We first consider the 2322-protocol with measurement settings $x^*, y^* = \{0, 2\}$, and then briefly discuss the 2222-protocol case with $x^*, y^* \in \{0, 1\}$.

We first optimise the bound (113) only over the $p_{A'|A}$ maps, while disregarding the $p_{M|A'}$ maps. We perform the maximisation over all corresponding stochastic matrices $\mathcal{S}_{A \rightarrow A'}$ of size $|A'| \times |A|$, where we vary the outcome number $2 \leq |A'| \leq 6$ ($|A| = 2$ is fixed with $A \in \{0, 1\}$ by the noisy correlation $p_{AB}(a, b)$ considered). Independently of the outcome number $|A'|$, we always arrive at the critical visibility:

$$\check{V}_{\text{crit}}^{A \rightarrow A' \rightarrow M} = \check{V}_{\text{crit}}^{A \rightarrow A'} \approx 80.8530\%, \quad (131)$$

which coincides up to our best-achieved numerical precision with the critical visibility attainable with noisy preprocessing, see $V_{\text{crit}}^{\text{n.p.}}$ in Eq. (112). Hence, as the critical efficiency (131) applies to all preprocessing strategies of Alice $p_{A'|A}$, we conjecture that the strategy of noisy preprocessing is the optimal form of defense against the CC attack, when the parties observe the purely noisy correlation and Alice is not utilising the message M sent to Bob. Still, note that the numerical optimisation converges to different stochastic matrices leading to the critical visibility (131), suggesting that the optimal preprocessing strategy is then not unique.

As already noted in Eq. (131), in contrast to the case of the purely lossy 2333-protocol discussed in the previous section, we do not observe any improvement of the critical visibility (131) (lowering its value) by allowing Alice also to perform arbitrary maps $p_{M|A'}$ in Eq. (113) and letting $2 \leq |A'|, |M| \leq 4$. Furthermore, allowing only for the map $p_{A|M}$ to be performed by Alice on her ‘raw’ variable A , we arrive at the critical visibility

$$\check{V}_{\text{crit}}^{A \rightarrow M} \approx 82.9995\%, \quad (132)$$

which coincides up to our best-achieved numerical precision with the critical visibility obtained by performing no preprocessing by Alice, see $V_{\text{crit}}^{\text{no-prep}}$ in Eq. (109). We are thus led to the conclusion that for the lossless case of finite visibility, the inclusion of the publicly announced variable M serves no purpose against the CC attack, as it cannot lower the attainable critical visibility.

For completeness, let us also cite the results obtained for the 2222-protocol. Considering preprocessing strategies in which Alice applies either only $p_{A'|A}$ map in Eq. (113) (with $2 \leq |A'| \leq 6$) or in which Alice applies also the $p_{M|A'}$ map in Eq. (113) (with $2 \leq |A'|, |M| \leq 4$), we find that the critical visibility coincides with the one obtained with noisy preprocessing— $V_{\text{crit}}^{\text{n.p.}}$ specified below Eq. (112)—i.e.:

$$\check{V}_{\text{crit}}^{A \rightarrow A' \rightarrow M} = \check{V}_{\text{crit}}^{A \rightarrow A'} \approx 88.5238\%, \quad (133)$$

again with the optimisation arriving at different optimal stochastic matrices, suggesting that they are not unique. Considering preprocessing strategies with only the $p_{A|M}$ map (letting $2 \leq |M| \leq 5$), we find

$$\check{V}_{\text{crit}}^{A \rightarrow M} \approx 90.6075\%, \quad (134)$$

which, similarly as for the 2322-protocol, coincides to our best numerical precision with the critical visibility obtained by performing no preprocessing by Alice, see $V_{\text{crit}}^{\text{no-prep}}$ below Eq. (109). The above thresholds (133-134) apply irrespectively of the key settings used by Alice and Bob within the 2222-protocol ($x^*, y^* \in \{0, 1\}$).

D.2 Two-way protocols involving maximally entangled states

D.2.1 Finite detection efficiency and deterministic binning of ‘no-clicks’

As noted in the main text, for the purely lossy correlation (53) we are unable to find a non-trivial upper bound on the two-way key rate, $r_{2\text{-way}}(A \leftrightarrow B)$ in Eq. (32) of the main text, by resorting to the conditional mutual information $I(A : B|F)$ and heuristically searching over all possible stochastic maps $E \rightarrow F$ applied on Eve’s variable. However, we provide a non-trivial upper bound under an additional assumption that both Alice and Bob *bin* their non-detection events \emptyset in a deterministic fashion, that is, whenever a detection failure \emptyset occurs they simply interpret it always as the 0-outcome (or equivalently always as 1).

Such an assumption is motivated by the fact that the binning procedure of the inconclusive outcomes naturally arises in DIQKD protocols whose security is based on two-outcome Bell inequalities, both in one-way protocols [22] and two-way protocols based on advantage distillation [42]. As a consequence, the upper bound determined by us applies to all such protocols or, generally speaking, to any protocol in which the parties decide to deterministically bin their data before performing any other operations, also ones requiring two-way communication. Importantly, the binning is performed not only to test the Bell-violation, but also in the key-generation rounds.

Still, note that the binning procedure does not change the nature of the protocol, which remains 2333 or 2233 depending whether or not, respectively, Bob uses an extra setting for the key distillation. This is because any *preprocessing* of the data, an example of which is binning, is performed by the parties *after* they record their strings of outcomes, which include also the \emptyset -events and are all correlated with the string in hands of Eve. Therefore, the decomposition (12) of the CC attack must be valid before binning (or any preprocessing) and, hence, the maximal local weight allowed within the attack, $q^{\mathcal{L}}$, remains to be given by Eq. (58).

After determining the local weight, subsequent calculations of the upper bound on the key rate depend only on the tripartite correlation $p_{ABE}(a, b, e|x^*, y^*)$ conditioned on Alice and Bob choosing the key settings x^* and y^* that includes also the eavesdropper Eve performing the CC attack. In order to write down this correlation, we first notice that whenever Alice and/or Bob record \emptyset , this may only happen within the protocol rounds in which Eve distributes a local correlation (and knows perfectly every outcome), as whenever she distributes a nonlocal correlation Alice and Bob observe Q_{AB} , which has perfect detection efficiency. Consequently, the entries in the rows (2-5) of Tab. (138) stated below contain only non-zero diagonal elements.

In particular, the probability that both Alice and Bob register a *conclusive* outcome, which we label by the variables $\tilde{a}, \tilde{b} \in \{0, 1\}$, and Eve has no knowledge about the result (having distributed a nonlocal correlation) reads:

$$p_{ABE}(\tilde{a}, \tilde{b}, ?|x^*, y^*) = p_E(?) p_{AB|E}(\tilde{a}, \tilde{b}|x^*, y^*, ?) = (1 - q^{\mathcal{L}}) Q_{\tilde{a}\tilde{b}}, \quad (135)$$

where $p_E(?) = q^{\mathcal{NL}} = 1 - q^{\mathcal{L}}$ is just the probability of Eve distributing a non-local correlation, while

$$p_{AB|E}(\tilde{a}, \tilde{b}|x^*, y^*, ?) = p_{AB}^{\mathcal{NL}}(\tilde{a}, \tilde{b}|x^*, y^*) = Q_{\tilde{a}\tilde{b}} \quad (136)$$

is the lossless correlation producing conclusive outcomes, and we used the simplified notation $Q_{ab} = Q_{ab}^{x^*y^*}$.

Moreover, as for conclusive outcomes $p_{AB}(\tilde{a}, \tilde{b}|x^*, y^*) = p_{ABE}(\tilde{a}, \tilde{b}, ?|x^*, y^*) + p_{ABE}(\tilde{a}, \tilde{b}, (\tilde{a}, \tilde{b})|x^*, y^*)$, we obtain the missing expression for the correlations applicable when Eve perfectly knows \tilde{a} and \tilde{b} as

$$p_{ABE}(\tilde{a}, \tilde{b}, (\tilde{a}, \tilde{b})|x^*, y^*) = p_{AB}(\tilde{a}, \tilde{b}|x^*, y^*) - (1 - q^{\mathcal{L}}) Q_{\tilde{a}\tilde{b}} = (\eta^2 - 1 + q^{\mathcal{L}}) Q_{\tilde{a}\tilde{b}}, \quad (137)$$

which allows us then to fully write out the desired tripartite correlation:

$$p_{ABE}(a, b, e|x^*, y^*) = \begin{array}{c|cccc} e \setminus a, b & \tilde{a}, \tilde{b} & \tilde{a}, \emptyset & \emptyset, \tilde{b} & \emptyset, \emptyset \\ \hline (\tilde{a}, \tilde{b}) & (\eta^2 - 1 + q^{\mathcal{L}}) Q_{\tilde{a}\tilde{b}} & 0 & 0 & 0 \\ (\tilde{a}, \emptyset) & 0 & \eta\bar{\eta} Q_{\tilde{a}}^A & 0 & 0 \\ (\emptyset, \tilde{b}) & 0 & 0 & \eta\bar{\eta} Q_{\tilde{b}}^B & 0 \\ (\emptyset, \emptyset) & 0 & 0 & 0 & \bar{\eta}^2 \\ ? & (1 - q^{\mathcal{L}}) Q_{\tilde{a}\tilde{b}} & 0 & 0 & 0 \end{array}. \quad (138)$$

The correlations shared by Alice and Bob after they perform the binning ($\emptyset \rightarrow 0$) procedure can be simply obtained from Tab. (138) by adding every *column* involving one (or more) \emptyset -outcomes to the corresponding one in which \emptyset is (are) replaced by 0.

From now on we turn our attention to Eve and propose a preprocessing strategy $E \rightarrow F$ that leads to a non-trivial upper bound on the key rate. Although what follows is the best strategy that we have found, note that any $E \rightarrow F$ map gives a valid upper bound $r_{2\text{-way}}(A \leftrightarrow B) \leq I(A:B|F)$ and we do not exclude the possibility that there exists a map leading to a tighter bound. First, we have Eve bin her variable deterministically, just like the honest parties. On her part, this corresponds to defining a new variable $\tilde{e} \in \{(0, 0), (0, 1), (1, 0), (1, 1), ?\}$, whose probabilities are determined by Tab. (138) in the similar manner, i.e. by adding *rows* involving any \emptyset to ones in which the \emptyset -outcome is replaced by 0. The resulting correlation obtained after all the three parties

perform binning reads:

$$p_{AB\tilde{E}}(a, b, \tilde{e}|x^*, y^*) = \begin{array}{c|cccc} \tilde{e} \setminus a, b & 0, 0 & 0, 1 & 1, 0 & 1, 1 \\ \hline (0, 0) & \frac{(\eta^2 - 1 + q^\mathcal{L}) Q_{00} + \eta\bar{\eta} (Q_0^A + Q_0^B) + \bar{\eta}^2}{\eta^2} & 0 & 0 & 0 \\ \hline (0, 1) & 0 & \frac{(\eta^2 - 1 + q^\mathcal{L}) Q_{01} + \eta\bar{\eta} Q_1^B}{\eta^2} & 0 & 0 \\ \hline (1, 0) & 0 & 0 & \frac{(\eta^2 - 1 + q^\mathcal{L}) Q_{10} + \eta\bar{\eta} Q_1^A}{\eta^2} & 0 \\ \hline (1, 1) & 0 & 0 & 0 & (\eta^2 - 1 + q^\mathcal{L}) Q_{11} \\ \hline ? & (1 - q^\mathcal{L}) Q_{00} & (1 - q^\mathcal{L}) Q_{01} & (1 - q^\mathcal{L}) Q_{10} & (1 - q^\mathcal{L}) Q_{11} \end{array}. \quad (139)$$

We now transform the variable of Eve, $\tilde{E} \rightarrow F$, by applying the post-processing map proposed by us in Ref. [41], which takes the form of a stochastic matrix $P_{F|\tilde{E}}$ given by

$$P_{F|\tilde{E}} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad (140)$$

in order to determine the resulting tripartite correlation

$$p_{ABF}(a, b, f|x^*, y^*) = \sum_{\tilde{e}} p_{F|\tilde{E}}(f|\tilde{e}) p_{AB\tilde{E}}(a, b, \tilde{e}|x^*, y^*), \quad (141)$$

where $p_{F|\tilde{E}}(f|\tilde{e}) = [P_{F|\tilde{E}}]_{f\tilde{e}}$ are the entries in Eq. (140). By applying the map $P_{F|\tilde{E}}$, Eve keeps her outcomes $\tilde{e} \in \{(0, 0), (1, 1)\}$ intact while uniformly mixing the ‘‘other’’ outcomes $\{(0, 1), (1, 0), ?\}$. The distribution p_{ABF} is thus constructed by adding together the three relevant rows of Tab. (139) corresponding to the ‘‘other’’ outcomes, which gives us

$$p_{ABF}(a, b, f|x^*, y^*) = \begin{array}{c|cccc} f \setminus a, b & 0, 0 & 0, 1 & 1, 0 & 1, 1 \\ \hline (0, 0) & \frac{(\eta^2 - 1 + q^\mathcal{L}) Q_{00} + \eta\bar{\eta} (Q_0^A + Q_0^B) + \bar{\eta}^2}{\eta^2} & 0 & 0 & 0 \\ \hline (1, 1) & 0 & 0 & 0 & (\eta^2 - 1 + q^\mathcal{L}) Q_{11} \\ \hline \text{other} & (1 - q^\mathcal{L}) Q_{00} & \frac{\eta^2 Q_{01} + \eta\bar{\eta} Q_1^B}{\eta^2} & \frac{\eta^2 Q_{10} + \eta\bar{\eta} Q_1^A}{\eta^2} & (1 - q^\mathcal{L}) Q_{11} \end{array}. \quad (142)$$

The above choice allows us to calculate a non-trivial upper bound on the two-way key rate introduced in Eq. (32) of the main text, i.e. $r_{2\text{-way}}(A \leftrightarrow B) \leq I(A:B|F)$, by evaluating the conditional mutual information for the distribution p_{ABF} in Eq. (142). In what follows, we do this for the 2333- and 2233-scenarios of interest, in which the Q-probabilities in Tab. (142) are determined by the CHSH-optimal measurements (27) performed on a shared maximally entangled state $|\Phi^+\rangle$ —and are given by Eq. (43) depending on the key settings x^*, y^* .

2333-protocol. Substituting the Q-probabilities of the 2333-protocol with $x^*, y^* = \{0, 2\}$, as well as the form of the optimal local weight $q^\mathcal{L}$ in Eq. (58), into the tripartite correlation (142), we find $I(A:B|F) = 0$ at

$$\eta_{\text{crit}} = \frac{1}{4} (2 + \sqrt{2}) \approx 85.36\%, \quad (143)$$

which we state in Tab. 1 of the main text, see the column labelled ‘two-way’.

As a consequence, η_{crit} constitutes a lower bound on the detection efficiency required by any (even) two-way DIQKD protocol based on the 2333-scenario (with $x^*, y^* = \{0, 2\}$ key-settings), under the assumption that both parties perform deterministic binning of their non-detection events prior to any preprocessing of their data. In comparison, employing a concrete two-way protocol of advantage distillation has been shown, under the same assumption and additionally restricting to collective attacks, to require detection efficiency $\eta_{\text{a.d.}} = 93.7\%$ [42].

2233-protocol. Performing an analogous calculation for the 2233-protocol, we find that the upper bound depends on the specific choice of key settings. If the parties choose $x^*, y^* = \{0, 0\}, \{0, 1\}$ or $\{1, 0\}$, we find $I(A:B|F) = 0$ for p_{ABF} in Eq. (142) at

$$\eta_{\text{crit}} = 3(1 - 1/\sqrt{2}) \approx 87.87\%, \quad (144)$$

which we state in Tab. 1 in the ‘two-way’ column. In comparison, the corresponding upper bound on tolerable detection efficiency in two-way DIQKD protocols (when restricting to collective attacks) obtained for the same lossy correlations supplemented by deterministic binning and advantage distillation is $\eta_{\text{a.d.}} = 91.7\%$ [42].

However, if the parties choose $x^*, y^* = \{1, 1\}$ as the key settings, the mapping $p(F|\tilde{E})$ introduced in Eq. (140) is no longer sufficient to make $I(A:B|F)$ vanish for $\eta < 1$ in the nonlocal regime. This can be fixed noting that the map $p(F|\tilde{E})$ should now rather equally mix the outcomes of Eve: $\tilde{e} = \{(0, 0), (1, 1), ?\}$; and not $\tilde{e} = \{(0, 1), (1, 0), ?\}$, as before. This formally corresponds to permuting the columns of the map $p(F|\tilde{E})$ in Eq. (140) or equivalently adding now the three rows denoting $\tilde{e} = \{(0, 0), (1, 1), ?\}$ in Tab. (139) when computing p_{ABF} in Eq. (142). In this way, one obtains the desired equivalent of critical detection efficiency (144) reading

$$\eta_{\text{crit}} = \frac{1}{24} \left(24 - 3\sqrt{2} + \sqrt{6(32\sqrt{2} - 45)} \right) \approx 87.47\%, \quad (145)$$

which is slightly lower than for other choices of key settings. This follows from the fact that the choice of $x^*, y^* = \{1, 1\}$ leads to a higher probability of Alice and Bob having different outcomes, which doesn’t fall in line with their symmetric, deterministic binning. It turns out, however, that if Bob and Alice bin \emptyset deterministically, but one of them to 0 and the other to 1, the threshold can again be shown to read $\eta_{\text{crit}} = 3(1 - 1/\sqrt{2})$, as in Eq. (144).

D.2.2 Finite visibility

In this section, we consider the case of the visibility being finite ($V < 1$) instead, for which we now study not only the 2222- and 2322-scenarios, but also for completeness the 2422-scenario introduced in Ref. [15]. Similarly to the above, we compute non-trivial upper bounds on two-way key rates in the form of Eq. (32) of the main text, $r_{2\text{-way}}(A \leftrightarrow B) \leq I(A:B|F)$, by identifying sufficient forms of the conditional mutual information $I(A:B|F)$. Since in the scenario that we are considering Alice and Bob announce their inputs, the tripartite correlations from which Alice and Bob attempt to extract a secure key can be written as in Eq. (11) of the main text, from which one can then compute the necessary conditional mutual information after choosing a suitable post-processing map, $E \rightarrow F$, for Eve. Nonetheless, let us note for completeness that we are primarily reproducing here calculations from our Ref. [41].

2322-scenario. Consider here the CHSH-based DIQKD protocol with an added key setting for Bob. Let us recall that the ideal correlation shared between Alice and Bob is given by Eq. (43), i.e.:

$$Q_{AB}(a, b|x, y) = \begin{cases} \frac{1}{4} \left[1 + \frac{(-1)^{a+b+xy}}{\sqrt{2}} \right], & \text{if } x, y \in \{0, 1\} \\ \frac{1}{2} \delta_{a,b}, & \text{if } (x, y) = (0, 2), \\ \frac{1}{4}, & \text{if } (x, y) = (1, 2) \end{cases}, \quad (146)$$

while the observed (noisy) correlation reads $p_{AB}^{\text{obs}}(a, b|x, y) = VQ_{AB}(a, b|x, y) + \frac{1-V}{4}$, see Eqs. (15) or (67). Similarly to Eq. (140), we use the CC attack with the post-processing for Eve given by $p_{F|E}(f|e) = (P_{F|E})_{f,e}$, where

$$P_{F|E} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \end{pmatrix}, \quad (147)$$

and the variables $e \in \{(0, 0), (0, 1), (1, 0), (1, 1), ?\}$ are transformed onto $f \in \{(0, 0), (1, 1), \text{“other”}\}$.

We can now compute $I(A:B|F)$ to obtain an upper bound (32) on the two-way key rate, i.e:

$$r_{2\text{-way}} \leq \frac{1 + \sqrt{2}}{2} \left\{ \log_2 \left[\frac{j(V)^{j(V)} k(V)^{k(V)}}{[j(V) + k(V)]^{j(V)+k(V)}} \right] + j(V) + k(V) \right\}, \quad (148)$$

where $j(V) := (1 - V)(\sqrt{2} - 1)$ and $k(V) := 2(\sqrt{2}V - 1)$. One can verify that the upper bound is zero at $V_{\text{crit}} = \frac{1}{17}(7 + 4\sqrt{2}) \approx 74.45\%$ ($> \frac{1}{\sqrt{2}} = V_{\text{loc}}$), as stated in Tab. 1 in the ‘two-way’ column.

2222-scenario. Consider now the case where Alice and Bob make two measurements each, that is, $x, y \in \{0, 1\}$ and there is no extra key setting for Bob. The ideal correlation shared between them is then given by the appropriate case in Eq. (146) (or Eq. (43)) and reads

$$Q_{AB}(a, b|x, y) = \frac{1}{4} \left[1 + \frac{(-1)^{a+b+xy}}{\sqrt{2}} \right]. \quad (149)$$

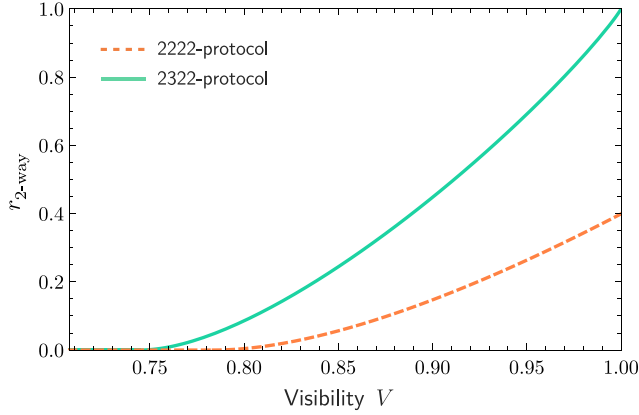


Figure 7: Upper bounds on the two-way key rates for the 2222 and the 2322-protocols based on the CHSH inequality.

As before, we consider a noisy version of this correlation with visibility $V < 1$, and we apply the CC attack with the same post-processing of Eve given by (147). In this case, the two-way key rate is upper-bounded by

$$r_{2\text{-way}} \leq \frac{1 + \sqrt{2}}{2} \left\{ \log_2 \left[\frac{\tilde{j}(V)^{\tilde{j}(V)} \tilde{k}(V)^{\tilde{k}(V)}}{[\tilde{j}(V) + \tilde{k}(V)]^{\tilde{j}(V) + \tilde{k}(V)}} \right] + \tilde{j}(V) + \tilde{k}(V) \right\}, \quad (150)$$

where $\tilde{j}(V) := (1 - V/\sqrt{2})(\sqrt{2} - 1)$ and $\tilde{k}(V) := (1 + 1/\sqrt{2})(\sqrt{2}V - 1)$. One can verify that the bound is zero at $V_{\text{crit}} = \frac{3}{7}(2\sqrt{2} - 1) \approx 78.36\%$ ($> \frac{1}{\sqrt{2}} = V_{\text{loc}}$), as stated in Tab. 1 in the ‘two-way’ column. The difference between the 2222-protocol and the 2322-protocol bounds are shown in Fig. 7, and it is clear that from the perspective of the honest users, the 2322-protocol performs better at all visibilities.

2422-scenario. Let us consider as well the CHSH-based DIQKD protocol with two added key settings for Bob, $y \in \{0, 1, 2, 3\}$, as proposed in Ref. [15], i.e.: with one setting $y = 2$ chosen again to be correlated with the setting $x = 0$ of Alice, but also another extra setting $y = 3$ correlated with the setting $x = 1$ of Alice instead. The ideal correlation shared between Alice and Bob is then given by

$$Q_{AB}(a, b|x, y) = \begin{cases} \frac{1}{4} \left[1 + \frac{(-1)^{a+b+xy}}{\sqrt{2}} \right], & \text{if } x, y \in \{0, 1\} \\ \frac{1}{2} \delta_{a,b}, & \text{if } (x, y) = (0, 2) \text{ or } (1, 3) \\ \frac{1}{4}, & \text{if } (x, y) = (0, 3) \text{ or } (1, 2) \end{cases} \quad (151)$$

The key setting pairs (x^*, y^*) are therefore either $(0, 2)$ or $(1, 3)$, and it is clear that both of these choices give rise to the same upper bound as the 2322-protocol, that is, Eq. (148).

D.3 One-way protocols involving partially entangled states

D.3.1 Finite detection efficiency

In this section, we utilise the formulae derived in App. C for the EC- and PA-terms in the scenario of finite detection efficiency ($\eta < 1$) and the purely lossy correlation (53) being shared, in order to determine the CC-based upper bounds on one-way key rates when Alice and Bob now ideally measure the partially entangled state (17) parametrised by θ , while performing projective measurements that maximise the CHSH violation (18)—the setting we introduced in Sec. 5.1.1 of the main text.

Given particular preprocessing strategies or optimising over these, we compute then numerically the corresponding thresholds on the tolerable detection efficiency, $\eta_{\text{crit}}(\theta)$, that are depicted in Fig. 4 of the main text as a function of the state-parameter θ . However, we are also importantly able to determine analytically the lowest possible thresholds applicable in the regime of $\theta \rightarrow 0$ that is known to exhibit highest robustness to imperfect detection [48]—these correspond to the smallest critical values presented in the left-most part of Fig. 4, i.e the values at which all the corresponding curves start from at $\theta = 0$.

Specifically, we treat here two preprocessing strategies of Alice applied on the purely lossy correlation (53), namely, deterministic binning with and without noisy preprocessing. Crucially, for these two choices it is beneficial for Alice and Bob to tune θ , in order to lower the critical detection efficiencies set by the CC attack. In contrast, for random binning of inconclusive outcomes (or their absence in the finite visibility model) it is the maximally-entangled case of $\theta = \pi/2$ discussed in the preceding section that remains optimal.

Moreover, we focus here on the 2333-protocol, as it generally exhibits lower noise thresholds than the 2233-protocol. In such a case, the parties generate the key from the lossy correlation (53) using measurement settings $x^*, y^* = \{0, 2\}$, in which the Q-probabilities then simply read

$$Q_{00} = \cos^2 \frac{\theta}{2}, \quad Q_{11} = \sin^2 \frac{\theta}{2}, \quad Q_{01} = Q_{10} = 0, \quad Q_0^A = Q_0^B = \cos^2 \frac{\theta}{2}, \quad Q_1^A = Q_1^B = \sin^2 \frac{\theta}{2}. \quad (152)$$

However, let us emphasise that other measurement settings, which are chosen to maximise the CHSH functional (18), determine the maximal local weight $q^{\mathcal{L}}$ employed within the CC attack that now also depends on the θ -parameter of the partially entangled state. Although for the important case of $\theta \rightarrow 0$ we possess an analytic expression for $q^{\mathcal{L}}$, see Eq. (65), for any other θ we may still evaluate efficiently the maximal local weight via the linear program (13), in particular, when computing all the curves that represent thresholds on tolerable detection efficiency in Fig. 4 as a function of θ .

Deterministic binning. In case Alice applies deterministic binning as her preprocessing strategy, we use again Eqs. (76) and (92) to compute the upper bound on the one-way rate, i.e. $r_{1\text{-way,det}}^\uparrow = H(A'|E)_{\text{det}} - H(A'|B)_{\text{det}}$. However, in contrast to Eq. (98), we now substitute for the Q-probabilities (152) that apply to the 2333-protocol involving partially entangled states, and for the optimal local weight applicable when $\theta \rightarrow 0$, i.e. $q^{\mathcal{L}}$ in Eq. (65), in order to obtain

$$r_{1\text{-way,det}}^\uparrow(\eta, \theta) \underset{\theta \rightarrow 0}{=} \eta(3\eta - 2) h \left[\cos^2 \frac{\theta}{2} \right] - \eta \sin^2 \frac{\theta}{2} h[\eta] - (1 - \eta) h \left[\eta \sin^2 \frac{\theta}{2} \right], \quad (153)$$

which we may expand further in the limit of $\theta \rightarrow 0$, as follows:

$$r_{1\text{-way,det}}^\uparrow(\eta, \theta) \underset{\theta \rightarrow 0}{=} \frac{\eta\theta^2 (-3 + 4\eta(1 + \ln 4) - \eta \ln(1 - \eta) + \ln(\eta(1 - \eta)/64))}{\ln 16} + \frac{\eta\theta^2}{\ln 16} (6 - 8\eta) \ln \theta + \mathcal{O}(\theta^4). \quad (154)$$

Now, it becomes evident that as $\theta \rightarrow 0$ it is the second term above that dominates over the first term, as the ratio of the former to the latter is proportional to $\ln \theta$ and diverges in that limit. Therefore, it must be made zero by choosing adequately η if $r_{\text{det},\theta \rightarrow 0}^\uparrow$ is to exhibit a root as $\theta \rightarrow 0$. Hence, this proves that $\lim_{\theta \rightarrow 0} \{r_{1\text{-way,det}}^\uparrow(\eta, \theta)\}$ may be vanishing only at the critical detection efficiency:

$$\eta_{\text{crit}}^{\text{det}}(\theta \rightarrow 0) = 75\%, \quad (155)$$

which is indeed clearly observed in Fig. 4—see the *solid red curve* at $\theta = 0$.

Deterministic binning with noisy preprocessing. If Alice decides to apply noisy preprocessing apart from deterministically binning her inconclusive outcomes \emptyset , we similarly to Eq. (100) compute the upper bound on the one-way rate with help of Eqs. (79) and (93) as $r_{1\text{-way,det+n.p.}}^\uparrow = H(A'|E)_{\text{det+n.p.}} - H(A'|B)_{\text{det+n.p.}}$. However, this time after substituting for the Q-probabilities (152) that apply for the 2333-protocol with partially entangled states, and the optimal local weight valid in the $\theta \rightarrow 0$ limit, i.e. $q^{\mathcal{L}}$ in Eq. (65), we firstly verify that the critical detection efficiency gets smaller with the bit-flip probability approaching $\mathfrak{p} \rightarrow \frac{1}{2}_\pm$. Hence, substituting further for $\mathfrak{p} = \frac{1}{2} \pm \delta$ and expanding in small δ , we get

$$r_{1\text{-way,det+n.p.}}^\uparrow(\eta, \theta) \underset{\theta \rightarrow 0}{=} \frac{4\eta \sin^2 \frac{\theta}{2}}{\ln 2} (-4 + \eta(4 + \eta) + (-2 + \eta(2 + \eta)) \cos \theta) \delta^2 + \mathcal{O}(\delta^4), \quad (156)$$

which we expand further in the $\theta \rightarrow 0$ limit to obtain

$$r_{1\text{-way,det+n.p.}}^\uparrow(\eta, \theta) \underset{\theta \rightarrow 0}{=} \left(\frac{2\eta(-3 + 3\eta + \eta^2)\theta^2}{\ln 2} + \mathcal{O}(\theta^4) \right) \delta^2 + \mathcal{O}(\delta^4). \quad (157)$$

As a result, we may now explicitly identify that $r_{1\text{-way,det+n.p.}}^\uparrow$ evaluated for $\theta \rightarrow 0$ vanishes when $-3 + 3\eta + \eta^2$ is zero, which exhibits a positive root at

$$\eta_{\text{crit}}^{\text{det+n.p.}}(\theta \rightarrow 0) = \frac{\sqrt{21} - 3}{2} \approx 79.13\%. \quad (158)$$

The above threshold value corresponds importantly to the starting point of the *solid blue curve* at $\theta = 0$ in Fig. 4, which, as claimed in the main text, we can now state analytically.

Optimising over all preprocessing maps. Finally, in a similar manner to App. D.1.3, we compute *general* upper bounds (4) that apply to one-way key rates independently of the preprocessing strategy employed when the parties share partially entangled states and observe lossy correlations (53). In particular, we further perform by heuristic numerical methods the optimisation in Eq. (4) over all stochastic maps $p_{A'|A}$ applied by Alice on her variable A , as well as stochastic maps $p_{M|A'}$ resulting in an extra message M sent publicly to Bob.

The upper bounds can be then translated onto universal thresholds on tolerable detection efficiency, $\check{\eta}_{\text{crit}}(\theta)$, below which no key can be distilled with one-way communication. These appear in Fig. 4 of main text as *dashed lines* with *diamonds*, *circles*, and *squares* corresponding to the optimization being performed, respectively: over the mappings $A \rightarrow A'$ and $A \rightarrow M$, and both of them simultaneously. Let us emphasise that we perform the optimisation over preprocessing strategies for the same CHSH-optimal correlations for which the thresholds with deterministic binning of inconclusive outcomes (with and without noisy preprocessing) in Fig. 4 were derived.

Strikingly, it follows from Fig. 4 that it is not only the special case of $\theta = \pi/2$, previously discussed in App. D.1.3, but actually independently of the θ -angle parametrising the partially entangled state, when it is sufficient for Alice to utilize only the $A \rightarrow M$ mapping and omit the $A \rightarrow A'$ preprocessing in order to achieve highest robustness against the CC attack. Moreover, we find that it is again always sufficient to consider $\mathcal{S}_{A \rightarrow M}$ in Eq. (116) as the stochastic map $A \rightarrow M$, i.e. the strategy in which Alice effectively signals the occurrence of inconclusive events to Bob. Furthermore, the preprocessing-optimized critical thresholds $\check{\eta}_{\text{crit}}$ obtained in the $\theta \rightarrow \pi/2$ limit (right-most values of *dashed lines* in Fig. 4) consistently coincide with the values (114) and (115) determined in App. D.1.3 for protocols utilising maximally entangled states.

D.4 One-way protocols involving partially entangled states with postselection

In this last section, we derive an upper bound on the DW rate, Eq. (5) of the main text, that now incorporates a postselection (PS) step, being used to certify the security of the protocol considered in Ref. [33], i.e.:

$$r_{\text{DW}}^{\text{PS}} := p_{\mathcal{V}_p} [H(A|E, \mathcal{V}_p) - H(A|B, \mathcal{V}_p)] \geq p_{\mathcal{V}_p} [H_{\min}(A|E, \mathcal{V}_p) - H(A|B, \mathcal{V}_p)], \quad (159)$$

where \mathcal{V}_p indicates successful postselection occurring with probability $p_{\mathcal{V}_p}$; $H(A|E, \dots)$ denotes the von Neumann entropy conditioned on the information possessed by the most general quantum eavesdropper Eve (hence the roman letter E instead of an italic E that would correspond to a classical random variable), which can in turn be lower-bounded by the min-entropy $H_{\min}(A|E, \dots)$. The CC attack, in which Eve holds a classical variable E , allows us to directly compute the upper bound on the PA-term, i.e.:

$$H_{\min}(A|E, \mathcal{V}_p) \leq H(A|E, \mathcal{V}_p) \leq H(A|E, \mathcal{V}_p), \quad (160)$$

however, we drop for simplicity the conditioning on the postselected subset \mathcal{V}_p in what follows.

In particular, we write the upper bound determined by the CC attack as just $\frac{r_{\text{DW}}^{\text{PS}}}{p_{\mathcal{V}_p}} \leq H(A|E) - H(A|B)$, and determine the analytic form of the corresponding EC- and PA-terms, $H(A|B)$ and $H(A|E)$, respectively. These may then be evaluated explicitly given a particular form of the observed correlation $p_{AB}^{\text{obs}}(a, b|x, y)$ —here, see Sec. 5.4.1 of the main text, we consider the purely lossy correlation (16,53) with $\eta < 1$ and $V = 1$ in Eq. (68).

D.4.1 The EC-term $H(A|B)$

We adopt the notation of Ref. [33] and consider the 2333-protocol with the honest users, Alice and Bob, ideally sharing a partially entangled two-qubit state:

$$|\psi(\theta)\rangle_{\text{AB}} = \cos \theta |00\rangle + \sin \theta |11\rangle. \quad (161)$$

Alice uses two measurement settings $x \in \{1, 2\}$, while Bob three settings $y \in \{1, 2, 3\}$, each corresponding to a projective measurement $\{\Pi(\phi), \mathbf{1} - \Pi(\phi)\}$ parametrised by the angle ϕ , where

$$\Pi(\phi) = \begin{pmatrix} \cos(\phi/2)^2 & \cos(\phi/2) \sin(\phi/2) \\ \cos(\phi/2) \sin(\phi/2) & \sin(\phi/2)^2 \end{pmatrix}. \quad (162)$$

This is equivalent to considering projective measurements $\{\frac{1 \pm \Pi(\phi)}{2}\}$ and $\Pi(\phi) = \cos \phi \sigma_z + \sin \phi \sigma_x$. For future reference, the key settings are set to $x^* = 1$, $y^* = 3$.

The lossy correlation shared by Alice and Bob is defined in Eqs. (16) and (53) for finite detection efficiency η (and $\bar{\eta} := 1 - \eta$) as—dropping the (x, y) -dependence and the ‘obs’ superscript for convenience, as in Eq. (68):

$$p_{AB}(a, b) = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 Q_{00} & \eta^2 Q_{01} & \eta \bar{\eta} Q_0^A \\ 1 & \eta^2 Q_{10} & \eta^2 Q_{11} & \eta \bar{\eta} Q_1^A \\ \hline \emptyset & \bar{\eta} \eta Q_0^B & \bar{\eta} \eta Q_1^B & \bar{\eta}^2 \end{array}, \quad (163)$$

where \emptyset denotes the no-click event, while Q_{ab} are the ideal probabilities observed in the absence of losses, i.e.:

$$Q_{ab} = \text{Tr}\{A_{a|x^*} \otimes B_{b|y^*} |\psi(\theta)\rangle_{AB}\langle\psi(\theta)|\} \quad (164)$$

with

$$A_{a|x^*} = \delta_{a,0}\Pi(\phi_{x^*}) + \delta_{a,1}(1 - \Pi(\phi_{x^*})), \quad B_{b|y^*} = \delta_{b,0}\Pi(\phi_{y^*}) + \delta_{b,1}(1 - \Pi(\phi_{y^*})). \quad (165)$$

The marginals are defined as $Q_0^A = Q_{00} + Q_{01}$, $Q_1^A = Q_{10} + Q_{11}$, and similarly for Bob.

After Alice bins deterministically $\emptyset \rightarrow 1$ (but not Bob [52]), the shared correlation reads

$$p_{AB}^{\text{det}}(a, b) = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & \eta^2 Q_{00} & \eta^2 Q_{01} & \eta\bar{\eta}Q_0^A \\ \hline 1 & \eta^2 Q_{10} + \eta\bar{\eta}Q_0^B & \eta^2 Q_{11} + \eta\bar{\eta}Q_1^B & \eta\bar{\eta}Q_1^A + \bar{\eta}^2 \end{array}. \quad (166)$$

She then performs the postselection step, i.e. she keeps the events when she measured ‘1’ with probability q and discards them with probability $1 - q$. Bob does the same but, importantly, he postselects both ‘1’ and \emptyset results, because of \emptyset being equivalent to ‘1’ with respect to postselection (which should be done as if Bob also binned his data [52]). This corresponds to both of them defining a new variable ‘D’ signifying discarded bits. We will ignore these discarded outcomes, but let us for completeness write down the full correlation including the ‘D’ outcome. In order to do so, let us express (166) for short as

$$p_{AB}^{\text{det}}(a, b) := \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & q_{00} & q_{01} & q_{0\emptyset} \\ \hline 1 & q_{10} & q_{11} & q_{1\emptyset} \end{array}, \quad (167)$$

so that the ‘full’ postselected (fPS) correlation then reads

$$p_{AB}^{\text{fPS}}(a, b) = \begin{array}{c|cccc} a \setminus b & 0 & 1 & \emptyset & \text{D} \\ \hline 0 & q_{00} & q q_{01} & q q_{0\emptyset} & (1 - q)(q_{01} + q_{0\emptyset}) \\ \hline 1 & q q_{10} & q^2 q_{11} & q^2 q_{1\emptyset} & q(1 - q)(q_{11} + q_{1\emptyset}) \\ \hline \text{D} & (1 - q) q_{10} & q(1 - q) q_{11} & q(1 - q) q_{1\emptyset} & (1 - q)^2(q_{11} + q_{1\emptyset}) \end{array}. \quad (168)$$

Throwing away the discarded bits, i.e. columns and rows of Tab. (168) labelled by ‘D’, we arrive at the desired postselected (PS) correlation:

$$p_{AB}^{\text{PS}}(a, b) = \frac{1}{P(\mathcal{V}_p)} \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & q_{00} & q q_{01} & q q_{0\emptyset} \\ \hline 1 & q q_{10} & q^2 q_{11} & q^2 q_{1\emptyset} \end{array}, \quad (169)$$

where the probability of successful postselection is then given by

$$\begin{aligned} P(\mathcal{V}_p) &:= q_{00} + q(q_{01} + q_{0\emptyset} + q_{10}) + q^2(q_{11} + q_{1\emptyset}) \\ &= \eta^2 Q_{00} + q[\eta^2(Q_{01} + Q_{10}) + \eta\bar{\eta}(Q_0^A + Q_0^B)] + q^2[\eta^2 Q_{11} + \eta\bar{\eta}(Q_1^A + Q_1^B) + \bar{\eta}^2], \end{aligned} \quad (170)$$

with the full expression obtained after substituting explicitly for $p_{AB}^{\text{det}}(a, b)$ according to Eq. (166).

In general, the postselected marginals of Bob read

$$p_B^{\text{PS}}(b) = \frac{1}{P(\mathcal{V}_p)} \begin{cases} q_{00} + q q_{10} & \text{if } b = 0 \\ q q_{01} + q^2 q_{11} & \text{if } b = 1 \\ q q_{0\emptyset} + q^2 q_{1\emptyset} & \text{if } b = \emptyset \end{cases}, \quad (171)$$

and the relevant conditional probability distribution is obtained by dividing the columns of (169) by the corresponding terms in (171), i.e.:

$$p_{AB}^{\text{PS}}(a|b) = \begin{array}{c|ccc} a \setminus b & 0 & 1 & \emptyset \\ \hline 0 & q_{00}/(q_{00} + q q_{10}) & q q_{01}/(q q_{01} + q^2 q_{11}) & q q_{0\emptyset}/(q q_{0\emptyset} + q^2 q_{1\emptyset}) \\ \hline 1 & q q_{10}/(q_{00} + q q_{10}) & q^2 q_{11}/(q q_{01} + q^2 q_{11}) & q^2 q_{1\emptyset}/(q q_{0\emptyset} + q^2 q_{1\emptyset}) \end{array}. \quad (172)$$

Hence, we can now evaluate explicitly the conditional entropy $H(A|B)$, i.e. the EC-term, as:

$$H(A|B) = \sum_b p_B^{\text{PS}}(b) H(A|B = b) = p_B^{\text{PS}}(0) h\left[\frac{q_{00}}{q_{00} + q q_{10}}\right] + p_B^{\text{PS}}(1) h\left[\frac{q q_{01}}{q q_{01} + q^2 q_{11}}\right] + p_B^{\text{PS}}(\emptyset) h\left[\frac{q q_{0\emptyset}}{q q_{0\emptyset} + q^2 q_{1\emptyset}}\right], \quad (173)$$

where $h[x]$ is again the binary entropy function.

D.4.2 The PA-term $H(A|E)$

Let us recall again that within the CC attack here considered Eve distributes a noiseless correlation in each non-local round that occur with probability $q^{\mathcal{NL}} := 1 - q^{\mathcal{L}}$. In contrast, whenever she distributes a local correlation with probability $q^{\mathcal{L}}$ (the local weight), she perfectly knows the outcome of both Alice and Bob. Hence, denoting generally by $P_{ab}^{\mathcal{L}}$ the resulting correlation shared by Alice and Bob within the local rounds after they bin their \emptyset -outcomes onto ‘1’, we may write the overall tripartite correlation specifying the CC attack as

$$p_{ABE}(a, b, e) = q^{\mathcal{NL}} \begin{bmatrix} Q_{00} & Q_{01} \\ Q_{10} & Q_{11} \end{bmatrix} \delta_{e,?} + q^{\mathcal{L}} \begin{bmatrix} P_{00}^{\mathcal{L}} \delta_{e,00} & P_{01}^{\mathcal{L}} \delta_{e,01} \\ P_{10}^{\mathcal{L}} \delta_{e,10} & P_{11}^{\mathcal{L}} \delta_{e,11} \end{bmatrix}, \quad (174)$$

where the random variable of Eve, E , consists of two bits (one for Alice and one for Bob) and an extra outcome ‘?’ representing her lack of knowledge. Note that, as it will be clear from the calculation below, without loss of generality we could have also ignored the second bit of Eve, because in the following we are interested only in the correlations between her and Alice—assumed to be the party performing PA within the one-way paradigm of key distribution. Moreover, recall that, after tracing out Eve, Alice and Bob must recover their correct shared correlation that corresponds the case of finite detection efficiency η and binning \emptyset -outcomes onto ‘1’—obtained by adding the last column onto the second one and similarly then for the rows in Eq. (163). Hence, we also have a constraint that:

$$p_{AB}(a, b) = q^{\mathcal{NL}} \begin{bmatrix} Q_{00} & Q_{01} \\ Q_{10} & Q_{11} \end{bmatrix} + q^{\mathcal{L}} \begin{bmatrix} P_{00}^{\mathcal{L}} & P_{01}^{\mathcal{L}} \\ P_{10}^{\mathcal{L}} & P_{11}^{\mathcal{L}} \end{bmatrix} \equiv \begin{bmatrix} \eta^2 Q_{00} & \eta^2 Q_{01} + \eta \bar{\eta} Q_0^A \\ \eta^2 Q_{10} + \bar{\eta} \eta Q_0^B & \eta^2 Q_{11} + \bar{\eta} \eta (Q_1^B + Q_1^A) + \bar{\eta}^2 \end{bmatrix}, \quad (175)$$

which specifies unambiguously the correlation $P_{ab}^{\mathcal{L}}$ distributed by Eve in the local rounds.

Now, we consider the tripartite correlation (174) after both Alice and Bob postselect their ‘1’ outcomes with probability q , as in Eq. (169), i.e.:

$$p_{ABE}^{\text{PS}}(a, b, e) = \frac{1}{P(\mathcal{V}_p)} \left(q^{\mathcal{NL}} \begin{bmatrix} Q_{00} & q Q_{01} \\ q Q_{10} & q^2 Q_{11} \end{bmatrix} \delta_{e,?} + q^{\mathcal{L}} \begin{bmatrix} P_{00}^{\mathcal{L}} \delta_{e,00} & q P_{01}^{\mathcal{L}} \delta_{e,01} \\ q P_{10}^{\mathcal{L}} \delta_{e,10} & q^2 P_{11}^{\mathcal{L}} \delta_{e,11} \end{bmatrix} \right), \quad (176)$$

where the probability of successful postselection, $P(\mathcal{V}_p)$, can be naturally decomposed as

$$P(\mathcal{V}_p) = P(\mathcal{V}_p, \mathcal{NL}) + P(\mathcal{V}_p, \mathcal{L}) = P(\mathcal{NL}) P(\mathcal{V}_p | \mathcal{NL}) + P(\mathcal{L}) P(\mathcal{V}_p | \mathcal{L}), \quad (177)$$

with $P(\mathcal{L}) = q^{\mathcal{L}}$, $P(\mathcal{NL}) = q^{\mathcal{NL}}$ by the definition of local weight, and

$$P(\mathcal{V}_p | \mathcal{NL}) = Q_{00} + q(Q_{01} + Q_{10}) + q^2 Q_{11}, \quad (178)$$

$$P(\mathcal{V}_p | \mathcal{L}) = P_{00}^{\mathcal{L}} + q(P_{01}^{\mathcal{L}} + P_{10}^{\mathcal{L}}) + q^2 P_{11}^{\mathcal{L}}, \quad (179)$$

being the probabilities of successful postselection given that, respectively, the non-local or the local box is distributed. In particular, we have that

$$\begin{aligned} P(\mathcal{V}_p) &= q^{\mathcal{NL}} (Q_{00} + q(Q_{01} + Q_{10}) + q^2 Q_{11}) + q^{\mathcal{L}} (P_{00}^{\mathcal{L}} + q(P_{01}^{\mathcal{L}} + P_{10}^{\mathcal{L}}) + q^2 P_{11}^{\mathcal{L}}) \\ &= \eta^2 Q_{00} + q [\eta^2 (Q_{01} + Q_{10}) + \bar{\eta} \eta (Q_0^A + Q_0^B)] + q^2 [\eta^2 Q_{11} + \bar{\eta} \eta (Q_1^A + Q_1^B) + \bar{\eta}^2], \end{aligned} \quad (180)$$

where we consistently reproduce the expression (170) after imposing the constraint (175) on $P_{ab}^{\mathcal{L}}$.

Tracing out Bob from (176), we obtain

$$p_{AE}^{\text{PS}}(a, e) = \frac{1}{P(\mathcal{V}_p)} \begin{array}{c|ccccc} a \setminus e & 00 & 01 & 10 & 11 & ? \\ \hline 0 & q^{\mathcal{L}} P_{00}^{\mathcal{L}} & q^{\mathcal{L}} q P_{01}^{\mathcal{L}} & 0 & 0 & q^{\mathcal{NL}} (Q_{00} + q Q_{01}) \\ \hline 1 & 0 & 0 & q^{\mathcal{L}} q P_{10}^{\mathcal{L}} & q^{\mathcal{L}} q^2 P_{11}^{\mathcal{L}} & q^{\mathcal{NL}} (q Q_{10} + q^2 Q_{11}) \end{array}, \quad (181)$$

where, as expected, Eve is perfectly correlated with Alice in the local rounds, so that $\forall e \neq ? : H(A|E = e) = 0$.

As a result, the desired entropy of Alice’s outcomes conditioned on Eve is fully determined by the case when Eve distributes a non-local correlation, which occurs with probability

$$p_E(?) = P(\mathcal{NL} | \mathcal{V}_p) = \frac{P(\mathcal{NL}, \mathcal{V}_p)}{P(\mathcal{V}_p)} = q^{\mathcal{NL}} \frac{P(\mathcal{V}_p | \mathcal{NL})}{P(\mathcal{V}_p)}, \quad (182)$$

so that the PA-term reads

$$H(A|E) = p_E(?) H(A|E = ?) = q^{\mathcal{NL}} \frac{P(\mathcal{V}_p | \mathcal{NL})}{P(\mathcal{V}_p)} h \left[\frac{Q_{00} + q Q_{01}}{P(\mathcal{V}_p | \mathcal{NL})} \right], \quad (183)$$

being defined completely by the last column of the correlation (181).