Article:
Eftaxias, Giorgos, Weilenmann, Mirjam and Colbeck, Roger (2023) Advantages of Multicopy Nonlocality Distillation and Its Application to Minimizing Communication Complexity. Physical Review Letters. 100201. ISSN: 1079-7114

https://doi.org/10.1103/PhysRevLett.130.100201

Advantages of multi-copy nonlocality distillation and its application to minimizing communication complexity

Giorgos Eftaxias,[1,2,*] Mirjam Weilenmann,[3,†] and Roger Colbeck[2,‡]

[1] *Quantum Engineering Centre for Doctoral Training, University of Bristol, Bristol BS8 1FD, United Kingdom*
[2] *Department of Mathematics, University of York, York YO10 5DD, UK*
[3] *Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria*

(Dated: 6th March, 2023)

Nonlocal correlations are a central feature of quantum theory, and understanding why quantum theory has a limited amount of nonlocality is a fundamental problem. Since nonlocality also has technological applications, e.g., for device-independent cryptography, it is useful to understand it as a resource and, in particular, whether and how different types of nonlocality can be interconverted. Here we focus on nonlocality distillation which involves using several copies of a nonlocal resource to generate one with more nonlocality. We introduce several distillation schemes which distil an extended part of the set of nonlocal correlations including quantum correlations. Our schemes are based on a natural set of operational procedures known as wirings that can be applied regardless of the underlying theory. Some are sequential algorithms that repeatedly use a two-copy protocol, while others are genuine three-copy distillation protocols. In some regions we prove that genuine three-copy protocols are strictly better than two-copy protocols. By applying our new protocols we also increase the region in which nonlocal correlations are known to give rise to trivial communication complexity. This brings us closer to an understanding of the sets of nonlocal correlations that can be recovered from information-theoretic principles, which, in turn, enhances our understanding of what is special about quantum theory.

INTRODUCTION

A bound on the strength of correlations realisable between pairs of measurement inputs and outputs in any local theory was first shown by Bell [1, 2]. This bound is exceeded in quantum theory and there are even stronger correlations theoretically possible without enabling signalling [3, 4]. One way to better understand quantum theory is to consider it in light of possible alternative theories, which can be compared in terms of the correlations they can create, and the implications access to such correlations would have. For instance, it is known that theories that permit strong enough correlations have trivial communication complexity [5]. Furthermore, nonlocal correlations have found applications in cryptography, where they form a necessary resource for device-independent quantum key distribution [6–9] and randomness expansion [10–12], for example. Since non-local correlations serve as resources for information processing, it is natural to ask about their interconvertability. In this work we look at non-locality distillation [13], i.e., whether with access to several copies of some non-local resource we can generate stronger ones, which would have implications for the study of device-independent tasks in noisy regimes, for instance.

Non-locality distillation is often analysed in terms of *wirings* [13–19], which means interacting with systems by choosing inputs and receiving and processing outcomes from those systems. This has the advantage that, firstly, the distillation procedures apply to non-local quantum correlations no matter how complicated the system these have been obtained from and, secondly, these procedures are applicable beyond quantum theory. A general theory will prescribe various different ways to measure systems (in quantum theory, for instance, a measurement is described by a POVM). Wirings form an operationally natural sub-class that can be performed in any generalized probabilistic theory (GPT) [20] (including quantum theory).

Previous work on non-locality distillation has focused on specific protocols for the distillation of 2 copies of a non-local resource (see e.g., [13–15, 17, 19]). The case of more copies remains largely open, with only few specific results [16, 18]. In part, this is because analysing non-locality distillation is challenging: distillation protocols act non-linearly on the correlations and hence cannot be easily optimised. Furthermore, applying a successful 2-copy protocol twice often decreases the non-locality again (see e.g. [14] for an exception). Hence, understanding 2-copy protocols provides little insight into the n-copy case.

In this Letter we describe a sequential adaptive algorithm that uses wirings to distil non-locality. We use this algorithm to explore the distillable region within the set of non-local correlations, and the amount of distillation possible. We demonstrate new wirings that allow distillation of correlations that cannot be distilled with any 2-copy wiring protocol.

Our results have implications for communication complexity. In this problem, Alice with input x and Bob

* giorgos.eftaxias@bristol.ac.uk
† mirjam.weilenmann@oeaw.ac.at
‡ roger.colbeck@york.ac.uk

with input y want to enable Alice to compute $f(x,y) : \{0,1\}^k \times \{0,1\}^m \to \{0,1\}$. We ask how much communication from Bob to Alice is required to do so. Communication complexity is said to be trivial if any such function (no matter how large k and m) can be computed using only one bit of communication. Shared maximally non-local resources are known to make communication complexity trivial in this sense [21]. A probabilistic notion of trivial communication complexity was introduced in [5] in which for any f we require the existence of $p > 1/2$ such that Alice can obtain the correct value of $f(x,y)$ with probability at least p for all x and y. In this paper, when we talk about trivial communication complexity we mean it in this probabilistic sense. A larger set of shared states that render communication complexity trivial were found in Refs. [5, 14]. Our results further enlarge this set, demonstrating advantages of wirings beyond two copies.

NON-LOCALITY AND WIRINGS

Correlations of inputs x, y and outputs a, b are described by conditional probability distributions $P(ab|xy)$, and we refer to these as a *box* or a *behaviour*. In the context of non-locality, we usually imagine these correlations as generated by two parties, Alice and Bob, who each choose an input (x and y respectively) and obtain an output (a and b respectively). The correlations they can generate according to any theory that is consistent with special relativity have to be *non-signalling*, meaning

$$\sum_b P(ab|xy) = \sum_b P(ab|xy') \quad \forall\ a, x, y, y',$$

and the same holds with the roles of Alice and Bob (i.e., a, x and b, y) exchanged. A box is called *local* if it can be written

$$P(ab|xy) = \sum_\lambda P(a|x\lambda)P(b|y\lambda)P(\lambda) \quad \forall\ a, b, x, y\,.$$

In the language of Bell inequalities, there is a variable Λ that takes the value λ with probability $P(\lambda)$. Boxes that cannot be written in this form are *non-local*.

In the case of two binary inputs and outputs, i.e., $a, b, x, y \in \{0,1\}$, the set of all local boxes is the convex hull of 16 local deterministic boxes $P_i^{\mathrm{L}}(ab|xy) = \delta_{a,\mu x\oplus\nu}\,\delta_{b,\sigma y\oplus\tau}$ for $\mu,\nu,\sigma,\tau \in \{0,1\}$, $i = 1+\tau+2\sigma+4\nu+8\mu$, and the set of all non-signalling boxes is the convex hull of these local boxes and 8 extremal non-local boxes [4, 22] $P_i^{\mathrm{NL}}(ab|xy) = \frac{1}{2}\delta_{a\oplus b, xy\oplus\mu x\oplus\nu y\oplus\sigma}$ for $\mu,\nu,\sigma \in \{0,1\}$, $i = 1+\sigma+2\nu+4\mu$. Up to symmetry, the Clauser-Horne-Shimony-Holt (CHSH) inequality [23] is the only one that restricts the set of local boxes. Non-locality can hence be quantified in terms of the CHSH value $\mathrm{CHSH}(P(ab|xy)) = E_{00}+E_{01}+E_{10}-E_{11}$, with $E_{xy} = P(a=b|xy) - P(a\neq b|xy)$.

Because we work in a black-box picture, the most general operation we consider for each party is a wiring. We describe here the deterministic wirings; the most general wirings are convex combinations of these. Consider a party with access to n-boxes with inputs x_j and outputs a_j with $j = 1,\ldots,n$. They "wire" these together to form a new box with input x and output a. The most general deterministic wiring comprises choosing a box to make the first input to and then making a chosen input, then using the output of that box to choose the second box and the input to that second box and so on. We label the i^{th} box chosen $j_i(x, a_{j_1}, \ldots, a_{j_{i-1}})$ and its input $x_{j_i}(x, a_{j_1}, \ldots, a_{j_{i-1}})$. The final outcome is chosen depending on the overall input and all previous outcomes $a(x, a_{j_1}, \ldots, a_{j_n})$. Thus, if Alice and Bob each do wirings on shares of n boxes, they generate a new box $P(ab|xy)$.

Our main question is then: *given several copies of a non-local box, are there wirings for Alice and for Bob such that the resulting box is more non-local than the original?* In the case of two non-signalling boxes each with binary inputs and outputs, the possible wirings have been fully characterised [24]. Nevertheless, even in this case, deciding whether these can result in more non-locality for a specific box is computationally intensive: there are 82 deterministic wirings that each party can perform for each input [24], leading to a total of 82^4 possibilities (one of the 82 for each input of each party). To make the computation more tractable, we optimise the wirings of one party with a linear program, while iterating over 82^2 wirings for the other (see Appendix A for more details). We use this linear programming technique to illustrate the regions in which distillation is possible for various 2-dimensional cross-sections (CSs) of the no-signalling polytope in Figure I. In this work we consider three regions:

$$\begin{aligned}
\text{CS I}: &\ \omega P_1^{\mathrm{NL}} + \frac{\eta}{2}(P_1^{\mathrm{L}} + P_6^{\mathrm{L}}) + (1-\omega-\eta)P^{\mathrm{O}} \\
\text{CS II}: &\ \omega P_1^{\mathrm{NL}} + \eta P_1^{\mathrm{L}} + (1-\omega-\eta)P^{\mathrm{O}} \\
\text{CS III}: &\ \omega P_1^{\mathrm{NL}} + \frac{\eta}{2}(P_1^{\mathrm{L}} + P_9^{\mathrm{L}}) + (1-\omega-\eta)P^{\mathrm{O}}\,,
\end{aligned} \tag{1}$$

where $P^{\mathrm{O}} = 3/4P_1^{\mathrm{NL}} + 1/4P_2^{\mathrm{NL}}$ is local and $\eta, \omega \geq 0$ with $\eta + \omega \leq 1$.

We analysed the distillability within these cross sections. Among the optimal protocols we recovered several that were previously known [15, 26]. The protocols of [15] (called $\mathrm{ABL}^+1, 2$) are sufficient to characterize the two-copy distillability in CS II (see Figure I), and CS III is two-copy non-distillable. The observation that ABL^+2 achieves no distillation in CS I shows that optimal protocols depend on the cross-section.

The above analysis is generally not useful for analysing whether repeated distillation of a box can lead to a certain CHSH-value. Applying a wiring that works for two boxes to two copies of the generated box often does not give a further increase in non-locality, in which case a switch of wirings is needed to distil further. While there are boxes that cannot be distilled at all with wirings (e.g. isotropic boxes [27]), the maximum CHSH value that can be distilled using multiple copies of a specific resource box is unknown. This means that we do not know how

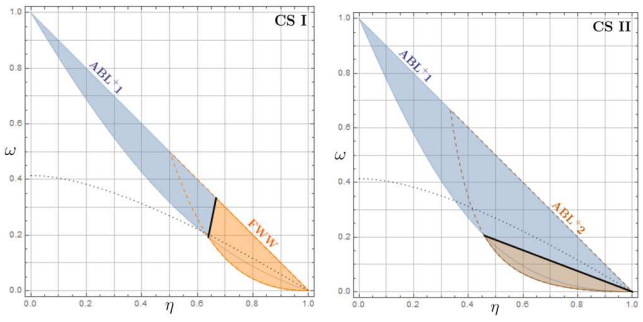

FIG. I. Protocols sufficient to characterise the two-copy distillability (both the distillable region and the strongest amplification) for two CSs (cf. Eq.(1)). The optimal two-copy protocols for CS II are the two protocols from [15] (ABL$^+$1,2), while for CS I the protocol of [13] (FWW) is optimal in some cases. The shading indicates where the corresponding protocol is optimal, with the boundary indicated by the black line (see Appendix A for details of the protocols). The dotted curve indicates the boundary of the set of correlations realisable in quantum theory (computed using the conditions in [3, 25]).

resourceful (multiple copies of) most non-local boxes are for information processing. For instance, shared boxes render communication complexity trivial if their initial CHSH value is greater than $\mathrm{CHSH}(P(ab|xy)) = 4\sqrt{\frac{2}{3}}$ [5]. The complete set of boxes that render commuication complexity trivial is unknown, although an additional region was found with the protocol of [14].

SEQUENTIAL ALGORITHMS FOR NON-LOCALITY DISTILLATION AND REDUCTION OF COMMUNICATION COMPLEXITY

While a repeated application of a successful 2-copy protocol often does not increase the non-locality further, there are various ways to combine different 2-copy wirings (see Appendix B). Here, we focus on the specific structure illustrated in Figure II. Our serial algorithm consists in optimising the wiring to be applied in every step, which is done in terms of a hybrid procedure of iterating over wirings and linear programming (see Appendix B for a detailed description of the algorithm). Applying our serial algorithm, we are able to extend the region of non-local boxes known to trivialise communication complexity, see Figure III.

Our algorithm furthermore provides us with a way to systematically derive new non-locality distillation protocols for multi-copy non-locality distillation. When performing two-steps of the serial algorithm, we find the three-copy protocol below to be successful.

In the first step, a box is created from two copies of a box P with inputs (outputs) labelled x_1, y_1 (a_1, b_1) and

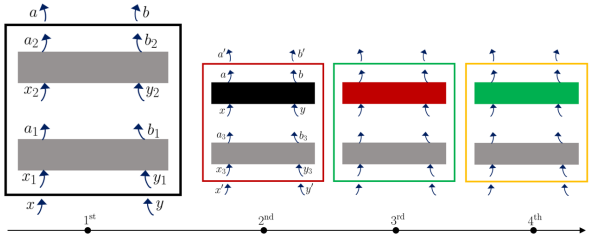

FIG. II. A serial architecture for combining nonlocal resources (gray) in a sequential manner. The first step on the left depicts the usual two-copy distillation scheme. Each subsequent iteration uses another copy of the original box and the previously generated one. Our sequential algorithm optimises the protocol at each round. See Appendix B for details.

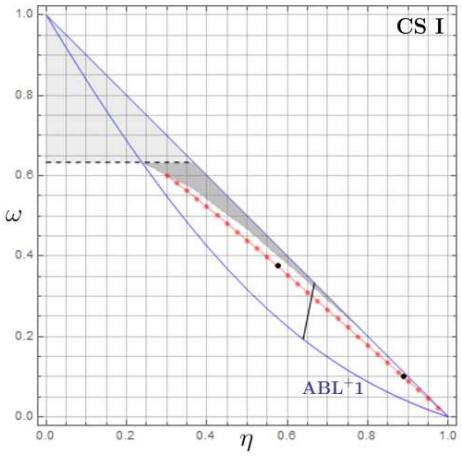

FIG. III. Region of trivial communication complexity in CS I. The light-gray part was identified in [5]. The dark-gray region includes boxes that trivialise communication complexity through (up to 4) iterations of ABL$^+$1. The red points (and everything on their right) collapse communication complexity using our serial algorithm. The black solid chord is that of Figure I (left) and indicates a change in protocol for the red points – see Appendix B for details, including analysis of the black points in the figure.

x_2, y_2 (a_2, b_2) respectively (first step in Figure II). Then this is wired to another copy of P, $P(a_3b_3|x_3y_3)$, using the functions

$$\begin{aligned} &x_1 = x = x',\ x_2 = x \oplus \bar{a}_1,\ a = a_1 \oplus a_2,\ x_3 = x\bar{a} \\ &y_1 = y = y',\ y_2 = yb_1,\ b = b_1 \oplus b_2,\ y_3 = y \oplus b, \\ &a' = a \oplus a_3,\ b' = b \oplus b_3, \end{aligned} \tag{2}$$

where $\oplus$ is the logical XOR and $\bar{z} = z \oplus 1$. This new protocol distils in CS II a strict superset of non-local boxes compared to the previously known 3-copy distillation protocol of [16] (in contrast to CS I where the protocol of [16] is superior). For completeness we introduce the protocol from [16] in Appendix C and we refer to it as HR. The region in which the new protocol distils in CS II is also shown in Appendix C.

GENUINE THREE-COPY DISTILLATION PROTOCOLS

When considering 3-copy distillation, the variety of possible protocols is vastly increased. In this case we can derive new protocols that outperform the previous ones in terms of the boxes for which they offer distillation. For this, we introduce a *genuine three-copy distillation protocol*, which is one that cannot be reduced to a concatenation of 2-copy protocols, i.e., is not of the form of Figure II. Consider the following wiring, where $\vee$ denotes the logical OR operation:

$$\begin{aligned} &x_1 = x_2 = \bar{x},\ x_3 = \bar{x}a_1 \vee \bar{x}a_2,\ a = a_1a_3 \vee a_2a_3 \vee \bar{a}_1\bar{a}_2\bar{a}_3,\\ &y_1 = y_2 = y,\ y_3 = yb_1 \vee yb_2 \vee \bar{y}\bar{b}_1\bar{b}_2,\\ &b = \bar{y}b_1b_3 \vee \bar{y}b_2b_3 \vee yb_1\bar{b}_3 \vee yb_2\bar{b}_3 \vee \bar{y}\bar{b}_1\bar{b}_2\bar{b}_3 \vee y\bar{b}_1\bar{b}_2b_3. \end{aligned} \tag{3}$$

We find larger regions of distillable boxes as compared to the two-copy case, see Figure IV. In CS III no 2-copy distillation is possible, while with 3 copies it is. Furthermore, the increase in the region of boxes that allow for distillation is considerably larger than that of HR (which is nearly indistinguishable from ABL$^+$1, see also Figure VIII in the Appendix).

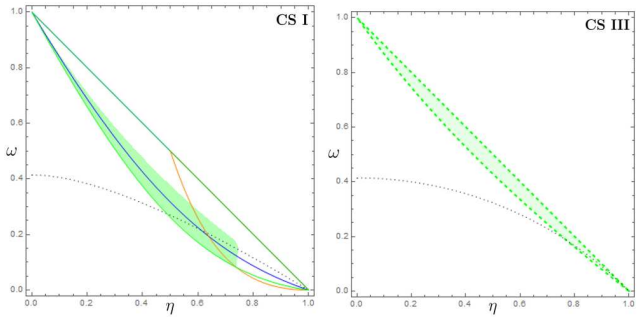

FIG. IV. Region of distillation by means of the 3-copy wiring of Eq.(3) bounded by the green lines. The blue and orange lines show the region of optimal 2-copy distillation in CS I, as in Figure I (left). The green shaded area in CS I depicts where our protocol leads to higher CHSH values than all previously known protocols (i.e., 2-copy and 3-copy FWW, ABL$^+$1, HR). In CS III no 2-copy non-locality distillation is possible and the ability to distil is unlocked only when given access to at least 3 copies of a non-local box where use of a genuine 3-copy protocol is imperative. The dotted curve indicates the boundary of the set of quantum-realizable correlations.

Additionally we find 3-copy protocols that increase the region where communication complexity is trivial. In particular

$$\begin{aligned} &x_1 = x_2 = x,\ x_3 = xa_2 \vee x\bar{a}_1 \vee \bar{x}\bar{a}_2a_1,\\ &a = a_3a_2 \vee a_3\bar{a}_1 \vee \bar{a}_3\bar{a}_2a_1,\ y_1 = y_2 = y,\ y_3 = yb_2 \vee y\bar{b}_1,\\ &b = b_3b_2 \vee b_3\bar{b}_1 \vee \bar{b}_3\bar{b}_2b_1. \end{aligned} \tag{4}$$

We illustrate the use of this protocol for trivialising communication complexity in Figure V. In addition, we find

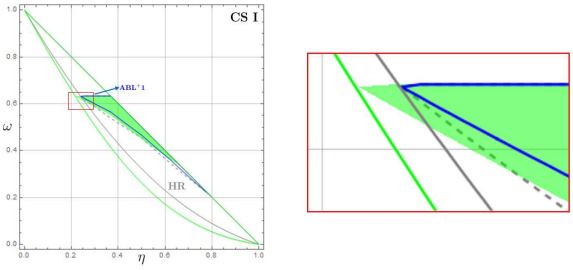

FIG. V. Regions of trivial communication complexity with various protocols. The green region is from repeated use of our genuine 3-copy protocol of Eq.(4), the blue bounded region is from repeated use of ABL$^+$1 and the dashed gray bounded region is from repeated use of HR. In the magnified view (right) we see a small region where our new 3-copy protocol outperforms HR and any possible 2-copy protocol.

that in CS I, starting from any point with $\omega > 0$ on the line $\omega = 1 - \eta$ we can distill arbitrarily close to a PR box by repeatedly iterating this protocol (see Appendix D). We observe, that as compared to using 2-copy protocols (even sequentially), 3-copy protocols provide further advantages.

Additionally, all the protocols introduced here, i.e., those of (2), (3) and (4) work in a full dimensional subset of the space of no-signalling correlations. This space is 8 dimensional for bipartite non-signalling boxes with binary inputs and outputs. The form of our distillation protocols (and many others in the literature) implies that the difference between the initial and final CHSH value is a polynomial in the parameters of the initial box $P(ab|xy)$ and hence continuous in these parameters. Thus, for any distillable point not on the boundary of the polytope, there exists an eight-dimensional ball around it that is also distillable.

CONCLUSIONS

We have found a genuine 3-copy protocol that distils nonlocality for boxes in which distillation with two copies is impossible and shown that there are 3-copy protocols that outperform *all* 2-copy protocols (and sequential applications thereof). For the latter we employed an optimization technique for 2-copy wiring protocols. Although this optimization furthers our understanding, it remains limited to cases with small numbers of inputs and outputs and there remains much more to discover about nonlocality distillation.

Whether the principle of non-trivial communication complexity [5] defines a closed set of correlations [28] that allows for a simple characterisation and lies well between quantum and non-signalling sets is an open question of interest for the foundations of quantum theory. Indeed, finding a sensible generalised probabilistic theory that leads to a set of correlations between the non-signalling

and quantum set with a simple geometric description has been a conundrum. The present work suggests that a better understanding of multi-copy non-locality distillation may give us insights into such a set, namely that of a GPT whose only restriction is imposed by the principle of non-trivial communication complexity. This would further advance the recent research program of experimentally ruling out generalised probabilistic theories due to the correlations they produce in networks [29, 30].

Some of our distillation protocols work within the set of quantum correlations (see Figure IV). [See also [31] for recent work aiming to distil quantum correlations.] Being wirings, they are much simpler to perform than entanglement distillation protocols [32]. It would be interesting to explore applications of these for information processing.

ACKNOWLEDGMENTS

GE is supported by the EPSRC grant EP/LO15730/1. MW is supported by the Lise Meitner Fellowship of the Austrian Academy of Sciences (project number M 3109-N). Some of the preliminary work for this project was performed using the Viking Cluster, a high performance computing facility at the University of York. We are grateful for computational support from the University of York High Performance Computing service, Viking, and the Research Computing team.

[1] J. S. Bell, "On the Einstein-Podolsky-Rosen paradox," Physics Physique Fizika **1**, 195–200 (1964).
[2] J. S. Bell, "The theory of local beables," in *Speakable and unspeakable in quantum mechanics* (Cambridge University Press, 1987) pp. 52–62.
[3] B. Cirel'son, "Quantum generalizations of Bell's inequality," Letters in Mathematical Physics **4**, 93–100 (1980).
[4] S. Popescu and D. Rohrlich, "Quantum nonlocality as an axiom," Foundations of Physics **24**, 379–385 (1994).
[5] G. Brassard, H. Buhrman, N. Linden, A. A. Méthot, A. Tapp, and F. Unger, "Limit on nonlocality in any world in which communication complexity is not trivial," Physical Review Letters **96**, 250401 (2006).
[6] A. K. Ekert, "Quantum cryptography based on Bell's theorem," Physical Review Letters **67**, 661–663 (1991).
[7] D. Mayers and A. Yao, "Quantum cryptography with imperfect apparatus," in *Proceedings of the 39th Annual Symposium on Foundations of Computer Science (FOCS-98)* (IEEE Computer Society, Los Alamitos, CA, USA, 1998) pp. 503–509.
[8] J. Barrett, L. Hardy, and A. Kent, "No signalling and quantum key distribution," Physical Review Letters **95**, 010503 (2005).
[9] S. Pironio, A. Acin, N. Brunner, N. Gisin, S. Massar, and V. Scarani, "Device-independent quantum key distribution secure against collective attacks," New Journal of Physics **11**, 045021 (2009).
[10] R. Colbeck, *Quantum and Relativistic Protocols For Secure Multi-Party Computation*, Ph.D. thesis, University of Cambridge (2007), also available as `arXiv:0911.3814`.
[11] S. Pironio, A. Acín, S. Massar, A. Boyer de la Giroday, D. N. Matsukevich, P. Maunz, S. Olmschenk, D. Hayes, L. Luo, T. A. Manning, and C. Monroe, "Random numbers certified by Bell's theorem," Nature **464**, 1021–4 (2010).
[12] R. Colbeck and A. Kent, "Private randomness expansion with untrusted devices," Journal of Physics A: Mathematical and Theoretical **44**, 095305 (2011).
[13] M. Forster, S. Winkler, and S. Wolf, "Distilling nonlocality," Physical Review Letters **102**, 120401 (2009).
[14] N. Brunner and P. Skrzypczyk, "Nonlocality distillation and postquantum theories with trivial communication complexity," Physical Review Letters **102**, 160403 (2009).
[15] J. Allcock, N. Brunner, N. Linden, S. Popescu, P. Skrzypczyk, and T. Vértesi, "Closed sets of nonlocal correlations," Physical Review A **80**, 062107 (2009).
[16] P. Høyer and J. Rashid, "Optimal protocols for nonlocality distillation," Physical Review A **82**, 042118 (2010).
[17] L.-Y. Hsu and K.-S. Wu, "Multipartite nonlocality distillation," Physical Review A **82**, 052102 (2010).
[18] X.-J. Ye, D.-L. Deng, and J.-L. Chen, "Nonlocal distillation based on multisetting bell inequality," Physical Review A **86**, 062103 (2012).
[19] G.-Z. Pan, C. Li, M. Yang, G. Zhang, and Z.-L. Cao, "Nonlocality distillation for high-dimensional correlated boxes," Quantum Information Processing **14**, 1321–1331 (2015).
[20] J. Barrett, "Information processing in generalized probabilistic theories," Physical Review A **75**, 032304 (2007).
[21] W. van Dam, "Implausible consequences of superstrong nonlocality," e-print quant-ph/0501159 (2005).
[22] B. S. Tsirelson, "Some results and problems on quantum Bell-type inequalities," Hadronic Journal Supplement **8**, 329–345 (1993).
[23] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed experiment to test local hidden-variable theories," Physical Review Letters **23**, 880–884 (1969).
[24] A. J. Short and J. Barrett, "Strong nonlocality: a trade-off between states and measurements," New Journal of Physics **12**, 033034 (2010).
[25] L. Masanes, "Necessary and sufficient condition for quantum-generated correlations," e-print arXiv:quant-ph/0309137 (2003).
[26] N. Brunner, D. Cavalcanti, A. Salles, and P. Skrzypczyk, "Bound nonlocality and activation," Physical Review Letters **106**, 020402 (2011).
[27] S. Beigi and A. Gohari, "Monotone measures for non-local correlations," IEEE Transactions on Information Theory **61**, 5185–5208 (2015).
[28] B. Lang, T. Vértesi, and M. Navascués, "Closed sets of correlations: answers from the zoo," Journal of Physics A: Mathematical and Theoretical **47**, 424029 (2014).
[29] M. Weilenmann and R. Colbeck, "Self-testing of physical theories, or, is quantum theory optimal with respect to some information-processing task?" Physical Review

Letters **125**, 060406 (2020).
[30] M. Weilenmann and R. Colbeck, "Toward correlation self-testing of quantum theory in the adaptive Clauser-Horne-Shimony-Holt game," Physical Review A **102**, 022203 (2020).
[31] S. G. Naik, G. L. Sidhardh, S. Sen, A. Roy, A. Rai, and M. Banik, "Distilling nonlocality in quantum correlations," e-print arXiv:2208.13976 (2022).
[32] C. H. Bennett, H. J. Bernstein, S. Popescu, and B. Schumacher, "Concentrating partial entanglement by local operations," Physical Review A **53**, 2046–2052 (1996).
[33] A. J. Short, S. Popescu, and N. Gisin, "Entanglement swapping for generalized nonlocal correlations," Physical Review A **73**, 012101 (2006).
[34] S. G. d. A. Brito, M. Moreno, A. Rai, and R. Chaves, "Nonlocality distillation and quantum voids," Physical Review A **100**, 012102 (2019).
[35] G. Eftaxias, *Theory-independent topics towards quantum mechanics: ψ-ontology and nonlocality distillation*, Ph.D. thesis, Quantum Engineering Centre for Doctoral Training, University of Bristol, UK, (2022).

Appendix A: Optimising over all two-copy non-locality distillation protocols

In order to establish whether a non-local box is amenable to 2-copy non-locality distillation, it is convenient (and due to the large number of possible protocols even necessary) to find ways to search and optimise over all such protocols. This can be achieved using Linear Programming. Specifically, while iterating over the extremal wirings of one party, we can optimise the operations of the other this way.

To see how this is possible, notice that the correlations obtained from wiring two boxes $Q_1(a_1b_1|x_1y_1)$, $Q_2(a_2b_2|x_2y_2)$ are

$$P(ab|xy) = \sum_{x_i,y_i,a_i,b_i} Q_1(a_1b_1|x_1y_1)Q_2(a_2b_2|x_2y_2)\chi_x(ax_1x_2|a_1a_2)\xi_y(by_1y_2|b_1b_2),$$

where $\chi_x(ax_1x_2|a_1a_2)$ and $\xi_y(by_1y_2|b_1b_2)$ describe Alice's and Bob's wirings upon receiving input x and y respectively. For a deterministic wiring, $\chi_x(ax_1x_2|a_1a_2) \in \{0,1\}$ for all a, a_1, a_2, x_1, x_2, and the wiring $x_1 = 0$, $x_2 = a_1$ and $a = a_1 \oplus a_2$ would correspond to $\chi(ax_1x_2|a_1a_2) = \delta_{x_1,0}\delta_{x_2,a_1}\delta_{a,a_1\oplus a_2}$, for example.

A wiring on Alice's side is made up of $|x|\cdot|a|$ vectors $\chi_x(a) = (\chi_x(ax_1x_2|a_1a_2))_{a_1a_2x_1x_2}$. In the case of 2-inputs and 2-outputs, these are straightforward to characterise since the wirings there are exactly the allowed measurements in a generalised probabilistic theory of non-local boxes [33]. Specifically, to have a valid wiring in this case, it is necessary and sufficient that the output distribution on any 2-input 2-output non-signalling box returns a valid probability distribution, i.e., for any $Q \in \{P_i^{\mathrm{L}}, P_j^{\mathrm{NL}}\}_{i,j}$

$$0 \leq \sum_{a_1,a_2,x_1,x_2} \chi_x(x_1x_2a|a_1a_2)Q(a_1a_2|x_1x_2) \leq 1 \qquad \forall x,a, \tag{A1}$$

$$\sum_{a,a_1,a_2,x_1,x_2} \chi_x(x_1x_2a|a_1a_2)Q(a_1a_2|x_1x_2) = 1 \qquad \forall x. \tag{A2}$$

These are linear constraints on the vectors $\chi_x(a)$.

Furthermore, CHSH$(P(ab|xy))$ is a linear function of the $P(ab|xy)$, which in turn is linear in $\chi_x(a)$. Thus, we can optimise the distilled non-locality over Alice's wirings with a linear program. Although this procedure works well when Alice and Bob each hold halves of two 2-input 2-output systems, going beyond this case presents several challenges:

1. With more than two systems the number of wirings on Bob's side significantly increases.

2. Sticking with two systems but increasing the number of inputs and outputs for each system significantly increases the number of wirings.

3. With more than two systems it is possible that the linear program optimizing over Alice's operations outputs a vector χ_x that is not a wiring.

The presence of such *non-wirings* for three systems was first noticed in [24]. In the main text we motivated the use of wirings based on maintaining validity of the results in any GPT. Allowing the non-wirings that come from such a linear program does not significantly alter the theory-independence in the sense that Eq.(A1) and Eq.(A2) are minimal requirements hence if no additional restrictions are placed on the theory any χ_x output by the linear program should be valid. Nevertheless it may be unnatural to allow non-wirings for Alice while restricting to wirings for Bob. Hence one would either like to add all the non-wirings valid in *any* theory to the set of Bob's possibilities, or remove non-wirings from the set of possible operations of Alice.

Wiring class	Number of wirings in class	Elements $\chi(a,a_1,a_2,x_1,x_2)=1$ if the following holds: (otherwise zero)	Label of wiring for each $\mu,\nu,\sigma,\delta,\epsilon\in\{0,1\}$
Constant	2	$x_1=x_2,\quad a=\mu$	$\mu+1$
One-sided	8	$x_1=x_2=\mu,\quad a=a_{\nu+1}\oplus\sigma$	$(4\mu+2\nu+\sigma+1)+2$
XOR-gated	8	$x_1=\mu,\quad x_2=\nu,\quad a=a_1\oplus a_2\oplus\sigma$	$(4\mu+2\nu+\sigma+1)+10$
AND-gated	32	$x_1=\mu,\quad x_2=\nu,$ $a=(a_1\oplus\sigma)(a_2\oplus\delta)\oplus\epsilon$	$(16\mu+8\nu+4\sigma+2\delta+\epsilon+1)+18$
Sequential	32	$x_{\mu+1}=\nu,\quad x_{(\mu\oplus1)+1}=a_{\mu+1}\oplus\sigma,$ $a=a_{(\mu\oplus1)+1}\oplus\delta a_{\mu+1}\oplus\epsilon$	$(16\mu+8\nu+4\sigma+2\delta+\epsilon+1)+50$

TABLE I. Labelling of 2-copy wirings. To iterate over all extremal wirings for Bob, we consider all combinations of $\xi_0(b)$, $\xi_1(b)$ from the above list, i.e., 82^2 wirings.

protocol name	wiring	analytic boundary of the region of distillation (ω as a function of η)	CHSH value of the distilled box
FWW [13]	$x_1=x_2=x$ $y_1=y_2=y$ $a=a_1\oplus a_2$ $b=b_1\oplus b_2$	$\omega=1-3\eta+2\sqrt{1-3\eta+3\eta^2}$ $\eta\in[1/2,1]$	$\frac{1}{2}\left[(1+\omega)^2-3\eta^2+6\eta(1+\omega)\right]$
ABL$^+$1 [15]	$x_1=x$ $y_1=y$ $x_2=x\oplus a_1\oplus1$ $y_2=yb_1$ $a=a_1\oplus a_2\oplus1$ $b=b_1\oplus b_2\oplus1$	$\omega=-\eta+\frac{1}{\sqrt{3}}\sqrt{3-4\eta+4\eta^2}$ $\eta\in[0,1]$	$\frac{1}{4}\left[3\omega^2+8\omega-\eta^2+\eta(4+6\omega)+5\right]$

TABLE II. Optimal 2-copy distillation protocols for CS I.

In the case of 2 copies of a box, in order to optimise the distilled non-locality over all wirings of Alice *and* Bob, we iterate over the 82^2 extremal wirings of Bob, as found in [33] and displayed in Table I, while optimizing Alice's wiring for each such choice with a linear program as described above.

Using this technique we can find whether there is a successful protocol for 2-copy non-locality distillation for any non-local box with two inputs and two outputs. In the following we illustrate this on CSs I and II (cf. Eq.(1)). In both cases, the full optimisation shows that two protocols are sufficient for characterising the region of 2-copy distillation in a CS. None of the points that are not distillable with either of these protocols can be distilled with any other 2-copy wiring there. In both CSs, we can choose non-locality distillation protocols from the literature to achieve this, i.e., known protocols are among the optimal ones when considering the region of distillation. Specifically, the region of distillation of CS I can be characterised in terms of the protocol from [13], which we call *FWW* here, as well as a protocol from [15], called *ABL$^+$1* here, which are both given in the Tables II and III. The parameters ω and η are chosen like in Figure I. Since the boundary of this region can be established as those boxes P for which 2-copy distillation leads to a box P' such that $\mathrm{CHSH}(P(ab|xy))=\mathrm{CHSH}(P'(ab|xy))$, this region can be characterised analytically.

The two CSs are displayed in Figure I. The black line where the two protocols work equally well is analytically characterised as

$$\omega=5\eta-3\,,\qquad \frac{1}{2}(1+\frac{1}{\sqrt{13}})\leq\eta\leq\frac{2}{3} \tag{A3}$$

in CS I and

$$\omega=-\frac{2\sqrt{6}-3}{5}(\eta-1)\,,\qquad \frac{1}{25}(9+\sqrt{6})\leq\eta\leq1 \tag{A4}$$

in CS II.

protocol name	wiring	analytic boundary of the region of distillation (ω as a function of η)	CHSH value of the distilled box
ABL$^+$2 [15]	$x_1 = x_2 = x$ $y_1 = y_2 = y$ $a = a_1 a_2$ $b = b_1 b_2$	$\omega = 3 - 11\eta + 2\sqrt{3 - 18\eta + 31\eta^2}$ $\eta \in [1/3, 1]$	$\frac{1}{8}\left[\omega^2 + 10\omega - 3\eta^2 + \eta(6 + 22\omega) + 13\right]$
ABL$^+$1 [15]	$x_1 = x$ $y_1 = y$ $x_2 = x \oplus a_1 \oplus 1$ $y_2 = y b_1$ $a = a_1 \oplus a_2 \oplus 1$ $b = b_1 \oplus b_2 \oplus 1$	$\omega = -\frac{4}{3}\eta + \frac{1}{3}\sqrt{9 - 18\eta + 25\eta^2}$ $\eta \in [0, 1]$	$\frac{1}{4}\left[3\omega^2 + 8\omega - 3\eta^2 + \eta(6 + 8\omega) + 5\right]$

TABLE III. Optimal 2-copy distillation protocols for CS II.

We remark here that previously, heuristics to simplify the optimisation over two-copy protocols have been proposed. For instance, the method in [34] suggests to reduce the search over 82^4 protocols to a manageable number of only 3152, by only considering protocols that preserve the PR-box, P_1^{NL}. Using linear programming, as proposed here, has the advantage that it takes all distillation protocols into account. In contrast, the heuristic from [34] discards various protocols, e.g., FWW and ABL$^+$2, that despite not preserving P_1^{NL}, are useful for non-locality distillation—they are even among the optimal 2-copy distillation protocols in CS I—so this shortcoming is pertinent.

Appendix B: Sequential non-locality distillation into the region of trivial communication complexity

In some situations we would like to distil non-locality up to a certain value that is useful for a specific task, e.g. because a particular CHSH score is needed in a device-independent scenario, or because we want to draw conclusions about the properties of those correlations, e.g. that they are unnatural since they imply that communication complexity is trivial. For this purpose, 2 copies of a non-local box are usually not sufficient. Since the repeated application of a fixed protocol is generally not successful in this respect either, it is natural to combine *different* protocols instead. There are various "architectures" that such combinations can take, two of which are displayed in Figure VI.

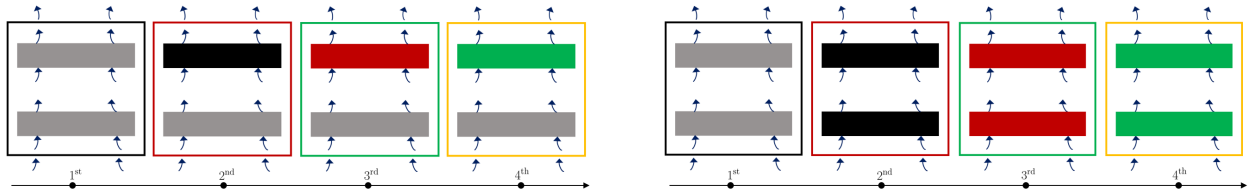

FIG. VI. Two architectures for combining an arbitrary number of resource boxes (gray) in a sequential manner. In each case, the purpose of our sequential algorithm is to find new optimal wirings in each round. Thus, the serial architecture on the left represents the *serial algorithm* introduced in Figure II. Similarly, the parallel architecture on the right will represent the *parallel algorithm*.

Analysing all of the wirings that are possible in such a multi-round procedure is computationally infeasible. We thus propose a sequential algorithm to (partially) optimise these procedures. This algorithm (in either version of Figure VI, serial or parallel or some alternatives, analysed more carefully in [35]) proceeds as follows:

(1) Optimise the wiring step by step using the procedure outlined in Appendix A. As figure of merit to be optimised we use the CHSH value here.

(2) Stop the procedure when either a certain round number is reached or when the CHSH value does not increase any further.

When applying the serial algorithm to the black points from Figure III, choosing the serial architecture turned out to be more effective than the parallel (in terms of distilled CHSH values). The tables below compare the findings of the serial algorithm with repeated iterations of other protocols. We can furthermore compare the different types of procedure. While we find that in CSs I and II, the serial procedure is more successful with respect to the increase

CS I, point $(\eta, \omega) = (0.888, 0.1)$, $\text{CHSH}_{init} = 2.2$						
CHSH_{final} , after # iterations				Serial Algorithm **STRATEGIES**		
iter #	two-copy $\text{ABL}^{+}1$, blindly repetitive	two-copy FWW, blindly repetitive	two-copy BS, blindly repetitive	Serial Algorithm	Alice's wiring $(\chi_{x=0}\ ,\ \chi_{x=1})$	Bob's wiring $(\chi_{y=0}\ ,\ \chi_{y=1})$
1	2.2815	2.3525	2.2812	2.3525	(12, 18)	(12, 18)
2	2.3837	2.5546	2.3823	2.4681	(12, 18)	(12, 18)
3	2.4964	2.7191	2.4918	2.5546	(12, 18)	(12, 18)
4	2.5885		2.5749	2.6186	(12, 18)	(12, 18)
5	2.5927			2.6729	(12, 78)	(74, 78)
6				2.7236	(70, 82)	(12, 82)
7				2.7706	(12, 78)	(74, 78)
8				2.8143	(70, 82)	(12, 82)
9				...	↻	↻
10				...	↻	↻
36				3.2683	(70, 82)	(12, 82)
41				3.2730	(12, 78)	(74, 78)

TABLE IV. Data about the lower black point of Figure III. The wirings are described using the labellings of the last column of Table I. The circular arrows denote the continued switching between the two strategies appearing on each side after the 4th iteration. The distilled CHSH values are recorded here as long as they increase. The yellow shaded entries compare final-CHSH values when each scheme has used 8 identical resource boxes. We observe that (37 copies of) the initial box trivializes communication complexity, a fact that only the serial algorithm reveals.

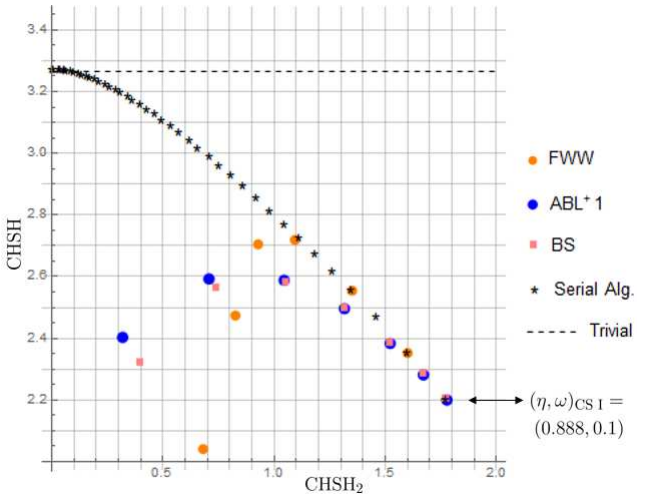

FIG. VII. Visualization of the data of Table IV (plus some further iterations that decrease the final-CHSH value). Here, the superiority of the serial algorithm – as opposed to the independent repetition of a fixed protocol– makes the initial box surpass the trivial communication complexity threshold (dashed line). The horizontal axis shows $\text{CHSH}_2 = E_{00} - E_{01} + E_{10} + E_{11}$.

in non-locality that is achieved, we have found other CSs where the parallel is favourable. For more details and the analysis of further types of procedures we refer to [35].

Notice also that, after a few iterations, we recover the same iteration of wiring strategies for each party in the two tables. This procedure corresponds to essentially exchanging the roles of the two players between iterations (and some bit-flips):

ODD iterations : $x_2 = x,\ x_1 = xa_2,\ a = a_1 \oplus a_2 \oplus 1,\ y_2 = y,\ y_1 = y \oplus b_2 \oplus 1,\ b = b_1 \oplus b_2 \oplus 1$
EVEN iterations : $x_2 = x,\ x_1 = x \oplus a_2,\ a = a_1 \oplus a_2 \oplus 1,\ y_2 = y,\ y_1 = y(b_2 \oplus 1),\ b = b_1 \oplus b_2 \oplus 1.$

CS I, point $(\eta, \omega) = (0.575, 0.375)$, $\text{CHSH}_{init} = 2.75$						
CHSH_{final}, after # iterations					Serial Algorithm **STRATEGIES**	
iter #	two-copy ABL$^+$1, blindly repetitive	two-copy FWW, blindly repetitive	two-copy BS, blindly repetitive	Serial Algorithm	Alice's wiring ($\chi_{x=0}$, $\chi_{x=1}$)	Bob's wiring ($\chi_{y=0}$, $\chi_{y=1}$)
1	2.9212	2.8212	2.9162	2.9212	(12, 78)	(74, 78)
2	3.0294		3.0096	3.0452	(70, 82)	(12, 82)
3				3.1327	↻	↻
4				3.1930	↻	↻
5				3.2324	↻	↻
6				3.2562	↻	↻
7				3.2683	(12, 78)	(74, 78)
8				3.2718	(70, 82)	(12, 82)

TABLE V. Data for the higher black point of Figure III. The wirings are described using the labellings of the last column of Table I. The circular arrows denote the continued switching between the two strategies appearing on each side after the 4th iteration. The distilled CHSH values are recorded here as long as they increase. The yellow shaded entries compare final CHSH values when each scheme has used 4 identical resource boxes. We observe that (8 copies of) the initial box trivializes communication complexity, and again, this is only revealed using the serial algorithm.

Cross Section	CHSH value of the distilled box
I	$\frac{1}{16}\left[\omega^3 - 5\eta^3 + 9\omega^2 + 31\omega + \eta^2(5+7\omega) + \eta(9+22\omega+5\omega^2) + 23\right]$
III	$\frac{1}{16}\left[\omega^3 + 7\eta^3 + 9\omega^2 + 31\omega + \eta^2(5+19\omega) + \eta(-3+18\omega+13\omega^2) + 23\right]$

TABLE VI. Final CHSH function after one iteration of the protocol of Equation (3), for the two cross sections of Figure IV.

Appendix C: 3-copy distillation in the literature

So far, the 3-copy non-locality distillation protocol that was so far known in the literature was introduced in [16]. This is specified by the following functions that make up the protocol HR:

Alice's side	Bob's side
$x_1 = x$	$y_1 = y$
$x_2 = x \oplus a_1$	$y_2 = y\bar{b}_1$
$x_3 = a_2\bar{a}_1 \oplus x(a_1 \oplus a_2 \oplus a_1 a_2)$	$y_3 = \bar{b}_1 \oplus b_2\bar{b}_1 \oplus y(\bar{b}_2 \oplus b_1 b_2)$
$a = a_1 \oplus a_2 \oplus a_3$	$b = b_1 \oplus b_2 \oplus b_3$

In some parts of CS I, this protocol outperforms the 2-copy distillation protocol ABL$^+$1 (around the point indicated with the star in Figure VIII). At the point indicated with the star, HR can distill non-locality while *no* 2-copy protocol can (thus, HR is also a *genuine* 3-copy protocol). However, the region around the starred point where this is possible is extremely small (see Figure VIII). This is different for our genuine 3-copy protocols (Equations (3) and (4)), for which this increase is considerable. Furthermore, we checked that HR, despite being a genuine 3-copy protocol, distills nothing in CS III, unlike our genuine 3-copy protocol that unlocks distillation there (Figure IV).

Appendix D: Further properties of the novel OR-gated protocols

In this section we present some extra features of the protocols introduced in Equations (3) and (4). The protocol of Equation (4) preserves the line (one dimensional convex combination)

$$\omega P_1^{\text{NL}} + (1-\omega)\frac{P_1^{\text{L}} + P_6^{\text{L}}}{2},$$

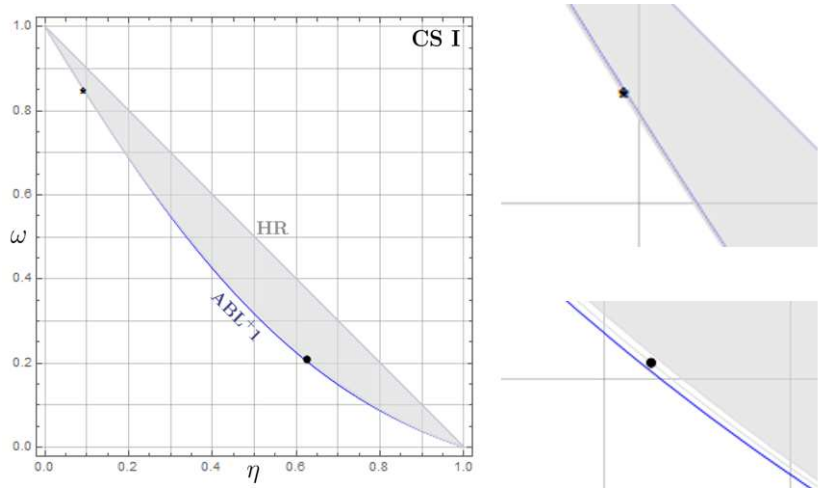

FIG. VIII. The blue (gray) boundary includes the boxes that are distillable by the ABL$^+$1 (HR). The gray region depicts the set where HR achieves higher distilled CHSH-values than ABL$^+$1. The bullet point corresponds to a box that is distillable by ABL$^+$1 but not by HR. Interestingly, the star (coordinates $(\eta, \omega) = (\frac{3}{32}, \frac{1}{32}(2\sqrt{227} - 3))$) corresponds to a box that is not distillable by ABL$^+$1 (so, not distillable by any two-copy protocol) but it can be distilled by HR.

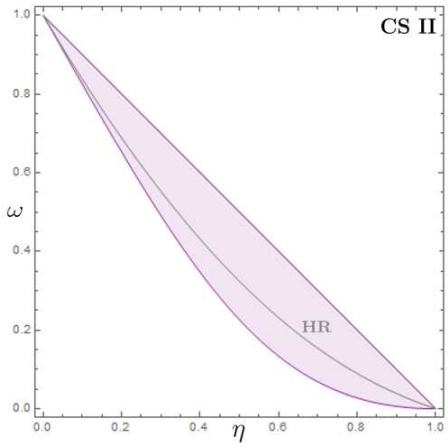

FIG. IX. Comparison of our new protocol (Equation (2)), displayed with the purple boundary, to the 3-copy protocol HR from [16], displayed in gray for CS II. The protocol of Equation (2) distils a strict superset of the boxes that HR distils and the non-locality increase at each point is also stronger than that of HR.

which is that subset of CS I corresponding to $\omega = 1 - \eta$. This means that an operation of the protocol maps any box belonging to that line, back to that line. Each iteration n, $n \geq 1$, of the protocol, updates the coordinate ω according to the recurrence relation

$$\omega_n = \frac{1}{4}\omega_{n-1}(7 - 4\omega_{n-1} + \omega_{n-1}^2) \quad , \quad \omega_0 = \omega. \tag{D1}$$

A plot showing the sequence of steps starting at $\omega_0 = 0.05$ is shown in Figure X. From the shape of the curves it is clear that for any initial $\omega \in (0, 1)$ repeated iterations allow us to generate a final box arbitrarily close to a PR box.

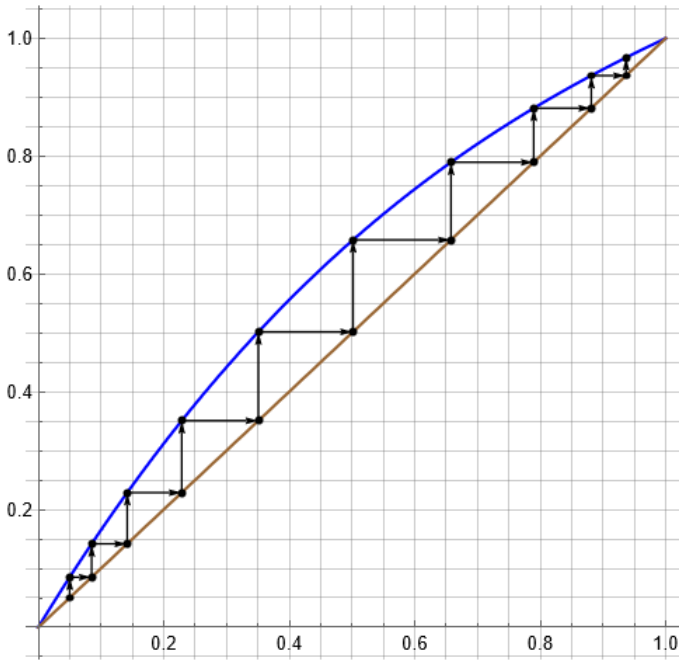

FIG. X. The blue curve depicts the function $f_1(\omega) = \frac{1}{4}\omega^2(7 - 4\omega + \omega^2)$ while the brown the $f_2(\omega) = \omega$, $\omega \in [0, 1]$. The black arrows lying in between represent all the steps from $n = 1$ to $n = 10$ of the recurrence relation (D1) for the case $\omega_0 = 0.05$.