Optimal bounds for correlations
via semidefinite programming

Peter Høyer and Jibran Rashid

Department of Computer Science, University of Calgary
2500 University Drive N.W., Calgary, AB, T2N 1N4 Canada. \{hoyer, jrashid\}@ucalgary.ca

Abstract

Nonlocal boxes (NLBs) are devices that are used to generate correlated bits between two spatially separated parties. If the generated correlation is weak, it can sometimes be distilled into a stronger correlation by repeated applications of the nonlocal box. We initiate here a study of the distillation of correlations for nonlocal boxes that output quantum states rather than classical bits. We give a semidefinite programming formulation of when distillation is possible under non-adaptive protocols. We propose a protocol for distillation and show that it asymptotically distills a class of correlated quantum nonlocal boxes to the value $\frac{1}{2}(3\sqrt{3} + 1) \approx 3.09876$, whereas in contrast, the optimal non-adaptive parity protocol for classical nonlocal boxes asymptotically distills only to the value 3.0. We then show that our protocol is an optimal non-adaptive protocol for 1, 2 and 3 copies by proving a matching upper bound. Our work demonstrates that some quantum nonlocal boxes exhibit stronger distillability than their classical counterpart and provides evidence that qNLBs are a stronger resource for nonlocality than NLBs.

1 Nonlocality distillation

Consider two parties, Alice and Bob, spatially separated and isolated, interested in jointly computing some boolean function $f(\cdot, \cdot)$. A third party, Charlie, provides Alice with an input $w_A$ (unbeknown to Bob) and Bob with an input $w_B$ (unbeknown to Alice) and challenges them to compute the bit $f(w_A, w_B)$. Charlie allows Alice and Bob to communicate, but charges for each and every bit communicated between Alice and Bob, whom therefore have pre-agreed upon a protocol that minimizing the amount of communication required for them to compute the bit $f(w_A, w_B)$. This is what we know as communication complexity.

It seems entirely impossible to jointly compute a non-trivial function if no information can be interchanged between Alice and Bob, and it is indeed one of the first results typically shown in any introduction to communication complexity. But as soon as one tweaks the models ever so slightly, surprising results are possible. This paper is about understanding one such tweaking and its implications for our understanding of computational resources and of the physical world.

A nonlocal box is a device shared between two parties that, when defined as we do here, in itself is incapable of transferring any information from Alice to Bob, or vice-versa, from Bob to Alice. Yet it renders all of communication complexity trivial. A nonlocal box takes as input two bits, a bit $x$ from Alice and a bit $y$ from Bob, and outputs two bits, a bit $a$ provided to Alice (and only Alice) and a bit $b$ provided to Bob (and only Bob). If the two input bits $x$ and $y$ from Alice and Bob equal $(0, 0)$, $(0, 1)$, or $(1, 0)$, the box
(by definition) provides Alice and Bob with identical bits. That is, they either both receive 0 or they both receive 1, each case happening with probability $\frac{1}{2}$. If the two parties both give the box a 1 as input, the box provides Alice and Bob with opposite bits $x$ and $y$, again each of the two cases 01 and 10 happening with equal probabilities $\frac{1}{2}$.

Such a nonlocal box is non-signalling. The marginal of the bit $a$ received by Alice is uniform irrespectively of whether Bob inputs 0 or 1 to the box, and she thus does not obtain any information about Bob’s input. Yet the parity of the two output bits $a$ and $b$ is perfectly correlated with the logical AND of the two input bits $x$ and $y$.

Nonlocal boxes were introduced by Popescu and Rohrlich in their seminal paper [24] and have since undergone extensive scrutiny. The box is a prospective computational resource when the two output bits exhibit a correlation that depends on the two input bits. The perfect nonlocal box (as defined above) is powerful enough to make communication complexity trivial, i.e., any boolean function may be computed by a single bit of communication between Alice and Bob [12]. Even if we modify the box so that, for each of the four possible inputs, it provides an output of the expected parity only with probability at least $\frac{3+\sqrt{6}}{6} \approx 0.908$, it would still be possible to compute any boolean function with bounded error using only a single bit of communication [6]!

The study of nonlocality is about understanding such correlations. Conditions that would imply such correlations are studied for instance through quantum strategies that violate Bell inequalities [11], the non-signalling principle [24], information causality [22], relativity [3] and macroscopic locality [29].

The study of nonlocal boxes is about understanding how useful such correlations are for computations and communication. Is there a universal correlation that can simulate all other correlations, how strong correlations can be generated in the physical world as we know it, and can correlations be made stronger through a distillation process? When studying these questions, we always require that the box is non-signalling, that is, it cannot in itself be used for communication.

Nonlocality distillation refers to the extent by which we can turn weak nonlocal boxes into more pure nonlocal boxes through a protocol. We have gained some understanding of when nonlocality can be distilled [13, 14, 16, 8, 2, 20, 15], when it cannot [25] and when it appears in bound form [7]. In general, the results suggest that distillation is only possible under special favorable circumstances and that large classes of nonlocal boxes are not distillable.

In this paper, we explore the scenario when we allow a nonlocal box to produce not only correlated bits as output, but correlated physical systems as output. If we produce correlated quantum states as output, can we then distill beyond what can be done with classical bits as output? A quantum nonlocal box, abbreviated qNLB, takes as input a joint quantum state and outputs a joint quantum state. A priori, such a model may not obey our non-signalling requirement since any unitary $U_{AB}$ not on the form $U_A \otimes U_B$ allows for signalling [5, 23]. It thus may appear that a quantum generalization of the NLB model would always allow for signalling, but this only holds true if we restrict the maps to be unitary. Quantum nonlocal boxes that satisfy the non-signalling requirement and allow for quantum states as output are possible when we drop the requirement of the box being unitary. Such boxes have previously been studied under the notion of causal maps, completely positive trace-preserving maps, and non-signalling operations [21, 23, 4, 18, 10]. Here we initiate a systematic study of such boxes in terms of nonlocality.

As our main result, we show that qNLBs exhibit strictly stronger nonlocality distillation than NLBs when restricting to non-adaptive distillation protocols.
Theorem 1 Quantum nonlocal boxes exhibit stronger nonlocality distillation for non-adaptive protocols than the optimal non-adaptive parity protocol for classical nonlocal boxes.

A distillation protocol using \( n \) nonlocal boxes is said to be non-adaptive if Alice is required to provide her input \( x \) to all \( n \) nonlocal boxes and Bob is required to provide his input \( y \) to all \( n \) boxes. We prove our main theorem by setting up a semi-definite programming framework [28] for analyzing non-adaptive protocols for qNLB distillation. We then use this framework to define and give a protocol for qNLB distillation and show that it outperforms the optimal non-adaptive protocol for classical nonlocal boxes [20]. We also show that our protocol is an optimal non-adaptive protocol for the class of correlated qNLBs, given 2 and 3 copies by proving a matching lower bound.

2 Distillation protocols

We define the value of a nonlocal box as the sum of the biases that the parity of the box agrees with the logical and of the input bits, over all four possible inputs,

\[
V = \sum_{x,y \in \{0,1\}} \Pr[a \oplus b = x \text{ and } y] - \sum_{x,y \in \{0,1\}} \Pr[a \oplus b \neq x \text{ and } y].
\]  

(1)

Brunner and Skrzypczyk considered and analyzed in [8] a class of NLBs that has only one-sided errors and labelled them correlated NLBs.

Definition 1 A correlated NLB maps the three inputs 00, 01 and 10 to the output 00 with probability \( \frac{1}{2} \), and to the output 11 with complementary probability \( \frac{1}{2} \). It maps the input 11 to either of the two outputs 01 and 10 with equal probabilities \( \frac{p}{2} \), and to either of the two outputs 00 and 11 with equal probabilities \( \frac{1-p}{2} \). Here \( p \in [0,1] \) denotes the probability that, on input 11, the output of the NLB is of odd parity.

The value of a correlated box is \( 3 + p - (1-p) = 2(1+p) \), and the value of a perfect NLB is 4.

Consider now that Alice and Bob share \( n \) instances of a correlated nonlocal box, all with the same parameter \( p \). Their goal is to simulate the behaviour of a correlated nonlocal box with a better parameter \( p' > p \) by using some pre-agreed upon protocol. They may use the \( n \) NLB instances as well as shared randomness, but are not allowed to communicate. If their protocol achieves a higher value \( p' \) than \( p \), we call the protocol a distillation protocol.

A correlated NLB can be asymptotically distilled to a perfect NLB by an adaptive protocol [15] as follows. Consider a single execution of a correlated NLB with input bits \( x \) and \( y \) and output bits \( a \) and \( b \). If the two inputs are both 0, a correlated NLB may output an incorrect correlation, whereas, if at least one of the two inputs is 0, the output is always correctly of even parity. Viewed from the perspective of the output bits \( a \) and \( b \), if the parity \( a \oplus b \) is odd, we can conclude that the two inputs bit were both 1, and that the output therefore is correct. Only if the output \( a \oplus b \) is even, can we not conclude with certainty that the output is correct. An adaptive protocol can use this one-sidedness of error to distill to the asymptotically optimal value of 4 by patiently waiting till the first time a usage of the correlated NLB yields an output of odd parity. This can be detected distributively (but not locally), and once detected, the protocol adaptively (and distributively) adjusts further usages of the NLBs so that all future outputs are of even parity. The eventual distributive detection of an output of odd parity reveals that the input bits were both 1, and the lack of an
output pair of odd parity indicates that at least one of the two input bits were 0. An odd parity output will eventually occur, allowing us to asymptotically distill to the optimal value 4.

A non-adaptive protocol can in contrast not distill to the value 4. By not allowing for adaptiveness, the distributive detection of the parity of the output can not be fed back into the system, and the protocol then fails in taking full advantage of the knowledge it possesses. A non-adaptive protocol must patiently wait till all outputs are produced, at which stage its best strategy is to take the parity of a certain number \(k\) of its outputs [20].

**Theorem 2** ([20]) The value attainable by any non-adaptive protocol using at most \(n\) correlated NLBs is upper bounded by

\[
V = \begin{cases} 
3 - (q - p)^n & \text{if } 0 \leq p < \frac{1}{2} \\
2(1 + p) & \text{if } \frac{1}{2} \leq p \leq 1,
\end{cases}
\]

and this value is attainable by the parity protocol.

In this work, we consider the case that the NLBs take quantum states as input and produce quantum states as output.

**Definition 2 (Quantum nonlocal box)** A quantum nonlocal box (qNLB) \(Q\) takes as input a product state \(|\psi_{xy}\rangle \in \{|00\}, \{01\}, \{10\}, \{11\}\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B\) and outputs a state \(\rho_{xy} \in \mathcal{H}_A \otimes \mathcal{H}_B\) such that for every map \(\Gamma_A : \mathcal{H}_A \rightarrow \mathcal{H}_A\) and \(\Gamma_B : \mathcal{H}_B \rightarrow \mathcal{H}_B\) the following two no-signalling conditions hold,

\[
\begin{align*}
\text{Tr}_A Q\big( (\Gamma_A \otimes 1) |\psi_{xy}\rangle \langle \psi_{xy}| \big) &= \text{Tr}_A Q\big( |\psi_{xy}\rangle \langle \psi_{xy}| \big) \\
\text{Tr}_B Q\big( (1 \otimes \Gamma_B) |\psi_{xy}\rangle \langle \psi_{xy}| \big) &= \text{Tr}_B Q\big( |\psi_{xy}\rangle \langle \psi_{xy}| \big),
\end{align*}
\]

In particular, we consider the class of correlated qNLB that generalizes the class of correlated NLBs.

**Definition 3** A correlated qNLB maps the three inputs \(|00\rangle, |01\rangle\ and \(|10\rangle\ to the pure state \(|\psi\rangle = \frac{1}{\sqrt{2}} (|00\rangle + |11\rangle)\, and maps the input \(|11\rangle\ to the mixed state \(\rho = p|\phi\rangle \langle \phi| + q|\psi\rangle \langle \psi|\, where \(|\phi\rangle = \frac{1}{\sqrt{2}} (|01\rangle + |10\rangle)\ is a superposition over the two odd-parity states, \(p \in [0, 1]\) a probability, and \(q = 1 - p\) the complementary probability.

Given that Alice and Bob share \(n\) copies of a correlated qNLB and measure observables \(A_x\) and \(B_y\) with eigenvalues \(\pm 1\) for input bits \(x\) and \(y\), respectively, the value attained for the CHSH inequality [11] is

\[
V = \langle \psi| \otimes n (A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0)|\psi\rangle \otimes n - \text{Tr}(A_1 \otimes B_1 \rho \otimes n). \tag{3}
\]

A qNLB is at least as powerful as an NLB: For any value of \(p\), Alice and Bob can use a correlated qNLB to simulate the correlation of a correlated NLB by simply measuring each of their outputs in the computational basis. In this paper, we formally prove that qNLBs are strictly more powerful in extracting nonlocality than are NLBs. We establish our main Theorem 1 by giving an explicit non-adaptive protocol that attains a higher distilled value for correlated qNLBs than the optimal parity protocol attains for correlated NLBs. The amount of distillability achievable by non-adaptive protocols for NLBs is characterized in [20], here specialized to correlated NLBs as Theorem 2 above.

We summarize the known results on non-adaptive distillation of classical and quantum correlated nonlocal boxes in Tables 1 and 2. For NLBs, we have complete knowledge: correlated NLBs are non-adaptively distillable if and only if \(0 < p < \frac{1}{2}\), and the parity protocol of Forster et al. [15] is an optimal non-adaptive protocol [20]. For qNLBs, we show here that correlated qNLBs are non-adaptively distillable.
NLB distill? | Value
--- | ---
\( p = 0 \) | no | 2
\( 0 < p < \frac{1}{2} \) | yes | \( 3 - (q - p)^n \)
\( \frac{1}{2} \leq p < 1 \) | no | \( 2(1 + p) \)

Table 1: Non-adaptive distillation of correlated NLBs is possible if and only if \( 0 < p < \frac{1}{2} \), for which they can be asymptotically distilled to the value 3.

| pNLB distill? | Value
--- | ---
\( p = 0 \) | no | 2
\( 0 < p < \frac{1}{2} \) | yes | \( (3 + (q - p)^n) \cos(\phi) + \frac{1}{2}(1 - (q - p)^n) \)
\( p = \frac{1}{2} \) | no for \( n \leq 3 \) | \( \frac{1}{2}(3\sqrt{3} + 1) \)
\( \frac{1}{2} \leq p < \frac{2}{3} \) | no for \( n \leq 3 \) | \( 3 \cos(\phi) - q \cos(3\phi) + p \)
\( \frac{2}{3} \leq p < 1 \) | no for \( n \leq 3 \) | \( 2(1 + p) \)
\( p = 1 \) | no | 4

Table 2: Non-adaptive distillation of correlated qNLBs is possible when \( 0 < p < \frac{1}{2} \), for which they can be asymptotically distilled to the value \( \frac{1}{2}(3\sqrt{3} + 1) \approx 3.09876 \). When \( \frac{1}{2} \leq p < 1 \), non-adaptive distillation is not possible using at most 3 qNLBs. Measurement angle \( \phi \) depends on \( p \) and is defined in Eq. 5.

when \( 0 < p < \frac{1}{2} \), and that correlated qNLBs can not be non-adaptively distilled when \( \frac{1}{2} \leq p < 1 \) if we Alice and Bob are allowed to use at most 3 qNLBs. When \( \frac{1}{2} \leq p < 1 \), we show that the single-usage qNLB protocol of Piani et al. [23] is optimal among all non-adaptive protocols using at most 3 qNLBs.

The values attainable are plotted in Figure 1. When \( 0 < p < \frac{2}{3} \), qNLBs achieves a strictly larger value than NLBs for any fixed value of \( n \). For \( 0 < p < \frac{1}{2} \), qNLBs can be asymptotically distilled to \( \frac{1}{2}(3\sqrt{3} + 1) \approx 3.09876 \), whereas NLBs can only be asymptotically distilled to the value 3.

### 3 Our distillation protocol

We propose the following protocol \( P \) for non-adaptively distilling correlated qNLBs.

**Protocol \( P \)** Let Alice and Bob share \( n \) identical copies of a correlated qNLB of parameter \( p \) and let them receive input bits \( x \) and \( y \), respectively. Their observables \( A_x \) and \( B_y \) are given by

\[
A_x = \cos\left(\frac{\phi}{2} + x\phi\right) Z + (-1)^x \sin\left(\frac{\phi}{2} + x\phi\right) X
\]

\[
B_y = \cos\left(\frac{\phi}{2} + y\phi\right) Z - (-1)^y \sin\left(\frac{\phi}{2} + y\phi\right) X.
\]

The operators \( Z \) and \( X \) for the observables \( A_x \) and \( B_y \) are chosen based on the value of \( p \).

\[
Z = \begin{cases} 
\sigma_z^n & \text{if } 0 < p < \frac{1}{2} \\
\sigma_z \otimes 1^{\otimes n-1} & \text{if } \frac{1}{2} \leq p \leq 1 
\end{cases}
\]

and

\[
X = \begin{cases} 
\sigma_x^n & \text{if } 0 < p < \frac{1}{2} \\
\sigma_x \otimes 1^{\otimes n-1} & \text{if } \frac{1}{2} \leq p \leq 1.
\end{cases}
\]
The measurement angle $\phi$ depends on $p$ and is chosen such that it maximizes the value attained for the CHSH inequality,

$$\cos^2 (\phi(p)) = \begin{cases} 
\frac{1}{4} \left( \frac{3 + (q-p)^n}{1 + (q-p)^n} \right) & \text{if } 0 < p < \frac{1}{2} \\
\frac{1 + q}{4q} & \text{if } \frac{1}{2} \leq p < \frac{2}{3} \\
1 & \text{if } \frac{2}{3} \leq p \leq 1.
\end{cases}$$

The observables chosen by Alice and Bob in Eq. 4 in Protocol $P$ depends on the probability $p$. If $0 < p < \frac{1}{2}$, Alice and Bob non-trivially use all $n$ available qNLBs. They view those $n$ qNLBs as a single qNLB and each applies an observable given by global measurement angle in a two dimensional space spanned by the two observables $\sigma_z \otimes^n$ and $\sigma_x \otimes^n$ (see Eq. 1). Thus viewed as a two-dimensional rotation, our chosen observables can be seen as a generalization of the measurements in the protocol of Piani et al. [23] for a single qNLB.

If $\frac{1}{2} \leq p \leq 1$, Alice and Bob effectively choose to use only a single qNLB by applying the identity observable $1 = \left( \begin{array}{cc} 1 & 0 \\ 0 & 1 \end{array} \right)$ on all but the first qNLB. The output bits $a$ and $b$ of Alice and Bob depend only on the output bits of the first qNLB. Therefore, an alternative protocol achieving the same values as ours, can be constructed in which Alice and Bob first choose to use a number $k$ of qNLBs, discard the remaining $n - k$ qNLBs, and then each apply an observable on the $k$ selected qNLBs as in Protocol $P$. When $0 < p < \frac{1}{2}$, they pick $k = n$, and when $p \geq \frac{1}{2}$, they pick $k = 1$.

Having specified the four observables $A_0, A_1, B_0,$ and $B_1$, we compute the value attained by our Protocol $P$ by plugging into Eq. 3.

**Lemma 3** Protocol $P$ attains the value

$$V = \begin{cases} (3 + (q-p)^n) \cos(\phi) + \frac{1}{2} (1 - (q-p)^n) & \text{if } 0 < p < \frac{1}{2} \\
2(1+q) \cos(\phi) + p & \text{if } \frac{1}{2} \leq p < \frac{2}{3} \\
2(1 + p) & \text{if } \frac{2}{3} \leq p \leq 1.
\end{cases}$$
A complete proof of Lemma 3 is given in Appendix A. In the next two sections, we prove that our protocol is an optimal non-adaptive protocol for any \(0 < p \leq 1\) and any \(n \leq 3\).

4 Protocol is optimal for a single copy

We now show that no other protocol can achieve a higher value \(V\) than ours when Alice and Bob are given a single qNLB. When \(n = 1\), the expression for the value \(V\) given in Eq. 3 simplifies to

\[
\langle \psi | A_0 \otimes B_0 | \psi \rangle + \langle \psi | A_0 \otimes B_1 | \psi \rangle + \langle \psi | A_1 \otimes B_0 | \psi \rangle - p \langle \phi | A_1 \otimes B_1 | \phi \rangle - q \langle \psi | A_1 \otimes B_1 | \psi \rangle.
\]

Using that the two Bell states \(|\psi\rangle\) and \(|\phi\rangle\) (given in Definition 3) can be locally mapped to each other, \(|\psi\rangle = (1 \otimes \sigma_x) |\phi\rangle\), we rewrite the optimization problem in terms of a single state \(|\psi\rangle\),

\[
V = \langle \psi | A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0 - p(A_1 \otimes \sigma_x B_1 \sigma_x) - q(A_1 \otimes B_1) | \psi \rangle,
\]

allowing us to apply Tsirelson’s conversion between observables and vectors [26, 27], as done in Wehner [28].

Lemma 4 (Tsirelson [26, 27]) Let \(A_0, \ldots, A_{m-1}\) and \(B_0, \ldots, B_{m-1}\) be observables with eigenvalues in the interval \([-1, 1]\). Then for any state \(|\psi\rangle\) shared between Alice and Bob, there exist real unit vectors \(x_0, \ldots, x_{m-1}\) and \(y_0, \ldots, y_{m-1}\) such that

\[
\langle \psi | A_i \otimes B_j | \psi \rangle = x_i \cdot y_j
\]

for all \(0 \leq i, j < m\). Conversely, for any set of real unit vectors, \(x_0, \ldots, x_{m-1}\) and \(y_0, \ldots, y_{m-1}\), and any maximally entangled state \(|\psi\rangle\), there exist observables \(A_i\) and \(B_j\) with eigenvalues \(\pm 1\) such that Eq. 8 holds for all \(0 \leq i, j < m\).

We define five vectors, one vector for each of Alice’s two observables \(A_0\) and \(A_1\), one for Bob’s observable \(B_0\), and two vectors for Bob’s observable \(B_1\),

\[
\begin{align*}
x_0 &= (A_0 \otimes 1) |\psi\rangle & y_0 &= (1 \otimes B_0) |\psi\rangle & z_0 &= (1 \otimes B_1) |\psi\rangle \\
x_1 &= (A_1 \otimes 1) |\psi\rangle & y_1 &= (1 \otimes B_1) |\psi\rangle & z_1 &= (1 \otimes (\sigma_x B_1 \sigma_x)) |\psi\rangle.
\end{align*}
\]

Let \(G = [g_{ij}]\) be the Gram Matrix of the five vectors \(\{x_0, x_1, y_0, z_0, z_1\}\),

\[
G = \begin{pmatrix}
x_0 \cdot x_0 & x_0 \cdot x_1 & x_0 \cdot y_0 & x_0 \cdot z_0 & x_0 \cdot z_1 \\
x_1 \cdot x_0 & x_1 \cdot x_1 & x_1 \cdot y_0 & x_1 \cdot z_0 & x_1 \cdot z_1 \\
y_0 \cdot x_0 & y_0 \cdot x_1 & y_0 \cdot y_0 & y_0 \cdot z_0 & y_0 \cdot z_1 \\
z_0 \cdot x_0 & z_0 \cdot x_1 & z_0 \cdot y_0 & z_0 \cdot z_0 & z_0 \cdot z_1 \\
z_1 \cdot x_0 & z_1 \cdot x_1 & z_1 \cdot y_0 & z_1 \cdot z_0 & z_1 \cdot z_1
\end{pmatrix},
\]

and set \(W\) to be the weight matrix

\[
W = \begin{pmatrix}
0 & 0 & 1 & 1 & 0 \\
0 & 0 & 1 & -q & -p \\
1 & 1 & 0 & 0 & 0 \\
1 & -q & 0 & 0 & 0 \\
0 & -p & 0 & 0 & 0
\end{pmatrix}.
\]
Finding an upper bound on the value $V$ in Eq. 7 then becomes equivalent to finding an upper bound on the primal value of the semidefinite program (SDP)

$$\max_G \frac{1}{2} \text{Tr}(GW)$$

subject to $G \succ 0$

$$g_{ii} = 1 \text{ for all } i \in \{1, \ldots, 5\}.$$  

The constraint $G \succ 0$ ensures that $G$ is a Gram matrix, and the constraints that the diagonal entries of $G$ are equal to 1, ensure that the five vectors are of unit norm. From any valid solution to the primal, we can extract a set of five observables via Tsirelson’s correspondence and construct a protocol that has the same value as the primal solution, and vice-versa, from any protocol, we can extract a set of five vectors, the Gram matrix of which is a primal solution having the same value as the value attained by the protocol.

We prove our upper bound on the primal value by giving a feasible solution to the dual of value equal to the value $V$ in Eq. 6. To conclude that our dual solution is feasible, we need to show that a particular matrix $M$ is positive semi-definite. Rather than attempting conveying a technical analysis of the roots of the matrix $M$’s characteristic polynomial, we shall instead break the matrix $M$ into smaller parts and repeatedly apply the following simple observation about the eigenvalues of a matrix of dimension $2 \times 2$.  

**Observation 5** A real-valued $2 \times 2$ matrix is positive semi-definite if and only if it has a non-negative diagonal entry and its determinant is non-negative.

To see this, notice that a symmetric real-valued matrix is positive semi-definite if and only if one of its two eigenvalues is non-negative and the product of its two eigenvalues is non-negative, which holds if and only if it has a non-negative diagonal entry and its determinant is non-negative.

**Lemma 6** The dual value of the SDP given in Eq. 9 is upper bounded by the value attained by Protocol $P$ for $n = 1$ given in Eq. 6.

**Proof** Let $b = (1, 1, 1, 1, 1)$ be a vector in $\mathbb{R}^5$. The dual of the primal SDP in Eq. 9 is

$$\min_{\lambda} \lambda \cdot b^T$$

subject to $K = 2 \text{diag} \lambda - W \succ 0$,  

where $\lambda$ is a vector in $\mathbb{R}^5$ and matrix $\text{diag} \lambda$ is of dimension $5 \times 5$ containing $\lambda_i$ in the $i^{th}$ diagonal entry and zeroes off-diagonal.

First consider the range $\frac{2}{3} \leq p \leq 1$. The dual solution

$$\lambda = (1, p, 1, \frac{p}{2}, \frac{p}{2})$$

has value $\lambda \cdot b^T = 2(1 + p)$, matching the value of the protocol given in Eq. 6. To show that the constraint $K \succ 0$ for the dual problem is satisfied, express matrix $K = 2 \text{diag} \lambda - W$ as the sum of two matrices,

$$K = K_1 + K_2 = \begin{pmatrix}
2 & 0 & -1 & -1 & 0 \\
0 & p & -1 & (1 - p) & 0 \\
-1 & -1 & 2 & 0 & 0 \\
-1 & (1 - p) & 0 & p & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix} + \begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & p & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & p & 0 & 0 & p
\end{pmatrix}.  $$
Matrix $K_2$ is a scaled projection with eigenvalues 0 and $2p$ and is therefore positive semi-definite. For matrix $K_1$, ignore its fifth row and column, which are zero, and conjugate the remaining $4 \times 4$ submatrix of $K_1$ by $\frac{1}{\sqrt{2}} \left( \begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right) \otimes 1$, yielding the submatrix

$$
\begin{pmatrix}
1 & -1 & 0 & 0 \\
-1 & 1 & 0 & 0 \\
0 & 0 & 3 & 1 \\
0 & 0 & 1 & 2p - 1
\end{pmatrix}.
$$

The upper-left $2 \times 2$ block is positive semi-definite, and, by Observation 5, the lower-right block is positive semi-definite when $3(2p - 1) \geq 1$, which holds when $p \geq \frac{2}{3}$. We have shown that matrix $K$ is the sum of two positive semi-definite matrices, and it is therefore positive semi-definite.

Next consider the range $0 < p < \frac{2}{3}$. The dual solution

$$
\lambda = \cos(\phi)(1, q, 1, q, 0) + (0, \frac{p}{2}, 0, 0, \frac{p}{2})
$$

has value $2(1 + q) \cos(\phi) + p$, matching the value of the protocol given in Eq. 6 for $n = 1$. (When $n = 1$, the expression in Eq. 4 for the range $0 < p < \frac{1}{2}$ simplifies to the expression for the range $\frac{1}{2} \leq p < \frac{2}{3}$.) It remains to show that the constraint $K \succeq 0$ is satisfied. Proceeding as in the case $\frac{2}{3} \leq p \leq 1$, we write

$$
K = K_1 + K_2 = 
\begin{pmatrix}
2 \cos(\phi) & 0 & -1 & -1 & 0 \\
0 & 2q \cos(\phi) & -1 & q & 0 \\
-1 & -1 & 2 \cos(\phi) & 0 & 0 \\
-1 & q & 0 & 2q \cos(\phi) & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix} + 
\begin{pmatrix}
0 & 0 & 0 & 0 & 0 \\
0 & p & 0 & 0 & p \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & p & 0 & 0 & p
\end{pmatrix},
$$

and conjugate the upper-left $4 \times 4$ submatrix of $K_1$ by $\frac{1}{\sqrt{2}} \left( \begin{array}{cc} 1 & -1 \\ -1 & 1 \end{array} \right) \otimes 1$, this time yielding the block matrix

$$
\begin{pmatrix}
2 \cos(\phi) - 1 & -1 & 0 & 0 \\
-1 & q(2 \cos(\phi) + 1) & 0 & 0 \\
0 & 0 & 2 \cos(\phi) + 1 & 1 \\
0 & 0 & 1 & q(2 \cos(\phi) - 1)
\end{pmatrix}.
$$

The two diagonal entries $2 \cos(\phi) + 1$ and $q(2 \cos(\phi) + 1)$ are non-negative since both $\cos(\phi)$ and $q$ are non-negative. Both blocks have the same determinant $4q \cos^2(\phi) - (1 + q)$ which equals zero. (When $n = 1$, the expression in Eq. 5 for the range $0 < p < \frac{1}{2}$ simplifies to the expression for the range $\frac{1}{2} \leq p < \frac{2}{3}$.) Applying Observation 5, we conclude that $K$ is positive semi-definite.

We have proved that the dual SDP is feasible and has a solution of value no larger than the value attained by the protocol. By Tsirelson’s correspondence, the protocol yields a feasible solution to the primal SDP of the same value as the protocol. These three values must therefore be equal. We conclude that our protocol is optimal for $n = 1$ and that the measurement angle specified by Eq. 5 is optimal.

5 Protocol $\mathcal{P}$ is optimal for 2 and 3 copies

In the preceding section, we show that no protocol can achieve a value higher than our Protocol $\mathcal{P}$ when given only a single copy of a qNLB. We show that the same statement holds true for 2 and 3 copies of a
qNLB: Among all non-adaptive protocols for distillation using at most 3 copies of a qNLB, none attains a value strictly higher than the value attained by our Protocol \( P \) using the same number of copies of a qNLB.

**Theorem 7** Protocol \( P \) is optimal among all non-adaptive protocols using at most 3 copies of a qNLB.

The proofs of the cases with multiple copies follow the outline we use in the simple single-copy case, except that now some of the steps become significantly more involved. We first give a general construction of a primal SDP and its dual SDP for any number \( n \) of copies of a qNLB, stated as Eqs. [14] and [15] in Appendix B below. The size of the dual SDP grows exponential in the number \( n \) of copies of qNLBs. In Appendix C below, we give a complete analytical proof of its value when \( n \) is at most 3, implying Theorem 7 stated above.

Our proof technique is general and should in principle be extendable to any fixed higher value of \( n \). More desirable, however, is to discover a method for analyzing our dual SDP for all values of \( n \) simultaneously. It seems plausible that such a generic proof technique should exist, but finding one has thus far eluded us. The dual SDP has an appealing representation in which we have been able to maintain many symmetries and letting it have an almost algorithmic structure. If a proof of optimality for any \( n \) could be found, it would imply that we could make a statement equally strong to the NLB case, thus proving that our Protocol \( P \) is optimal among all non-adaptive protocols.

### 6 Discussion

In our current work we have shown that if we restrict out attention to non-adaptive protocols, qNLBs offer improved distillation over NLBs. A generalization of our SDP approach that allows for adaptive protocols may reveal a similar improvement for adaptive protocols. This may imply distillability for correlations that are currently not known to be distillable and at the same time an increased understanding of correlations that violate principles such as information causality.

An objection against investigating the strength of correlations through the NLB model may be its hypothetical nature. Physical theories such as quantum mechanics do not generate the correlations that NLBs allow. A further generalization of the NLB model is also subject to these objections. Apart from the appeal of providing stronger distillation protocols, we provide an alternate perspective on the NLB model that serves to alleviate these concerns.

Consider the scenario depicted in Figure [2]. David who wants to compute a boolean function \( f(x, y) \), provides Alice and Bob with a description of \( f \) and the partitioned inputs \( x \) and \( y \). Alice and Bob may now use another trusted party Charlie who simulates the actions of a NLB/qNLB. This allows Alice and Bob to determine and transmit \( a \) and \( b \) to David such that \( f(x, y) = a \oplus b \). The three parties are able to help in computing the function \( f \) without any of them having access to complete information. Alice and Bob know the function, but not the complete input, nor its value, while Charlie knows the input without knowing the function being computed. The framework is analogous to the idealised secure scenario for two-party computation model considered by Yao [30]. The trust assumptions for Charlie may be unrealistic for cryptographic protocols but the model highlights the notion that it is possible to consider NLB correlations as a physical model rather than only as a hypothetical resource [6, 9]. An additional generalization may allow Alice and Bob to share an entangled state and transmit it to Charlie. Exploring such relationships between nonlocal boxes and cryptographic primitives and problems is one of the most exciting research directions.
Figure 2: A possible physical realization of the quantum nonlocal box. David who wants to compute a boolean function $f(x, y)$ transmits the description of $f$ as well as $x$ and $y$ to Alice and Bob respectively. Alice and Bob share the state $|xy\rangle$ with Charlie, who implements and distributes the results of the no-signalling map $Q(|xy\rangle)$. Finally, Alice and Bob send $a$ and $b$ to David, such that $f(x, y) = a \oplus b$.

Acknowledgements

We thank Cătălin Dohotaru for useful discussions. This work was supported by the Canadian Institute for Advanced Research (CIFAR), Canada’s Natural Sciences and Engineering Research Council (NSERC), the Canadian Network Centres of Excellence for Mathematics of Information Technology and Complex Systems (MITACS), and QuantumWorks.

References


A Proof of Lemma 3

In this appendix, we prove that our Protocol $\mathcal{P}$ given in Section 3 attains the value given by Eq. 3.

Lemma 8 For a mixed state $\rho = p|\phi\rangle\langle\phi| + q|\psi\rangle\langle\psi|$, where $p \in [0, 1]$ is a probability and $q = 1 - p$ the complementary probability, the following trace relations hold.

$$\begin{align*}
\text{Tr}(\sigma_z^\otimes n \otimes \sigma_z^\otimes n \rho^\otimes n) &= (q - p)^n \\
\text{Tr}(\sigma_x^\otimes n \otimes \sigma_x^\otimes n \rho^\otimes n) &= 1 \\
\text{Tr}(\sigma_z^\otimes n \otimes \sigma_x^\otimes n \rho^\otimes n) &= \text{Tr}(\sigma_x^\otimes n \otimes \sigma_z^\otimes n \rho^\otimes n) = 0.
\end{align*}$$

Proof First consider the case $n = 1$,

$$\begin{align*}
\text{Tr}(\sigma_z \otimes \sigma_z \rho) &= q\langle \psi | \sigma_z \otimes \sigma_z | \psi \rangle + p\langle \phi | \sigma_z \otimes \sigma_z | \phi \rangle = q - p \\
\text{Tr}(\sigma_x \otimes \sigma_x \rho) &= q\langle \psi | \sigma_x \otimes \sigma_x | \psi \rangle + p\langle \phi | \sigma_x \otimes \sigma_x | \phi \rangle = q + p = 1 \\
\text{Tr}(\sigma_x \otimes \sigma_z \rho) &= \text{Tr}(\sigma_z \otimes \sigma_x \rho) = q\langle \psi | \sigma_x \otimes \sigma_x | \psi \rangle + p\langle \phi | \sigma_x \otimes \sigma_x | \phi \rangle = 0.
\end{align*}$$

For $n > 1$, using that the operators are separable, rewrite $\text{Tr}(\sigma_1^\otimes n \otimes \sigma_2^\otimes n \rho^\otimes n) = (\text{Tr}(\sigma_1 \otimes \sigma_2 \rho))^n$ for all Pauli operators $\sigma_1$ and $\sigma_2$, and apply the case $n = 1$. 

We now prove Lemma 3 by substituting the appropriate expected values for each term in Equation 3.

Proof of Lemma 3 Let Alice and Bob share $n$ identical copies of a correlated qNLB and receive input bits $x$ and $y$ respectively. Application of Protocol $\mathcal{P}$ with observables $A_x$ and $B_y$ yields the following expectation values for inputs 00, 01, 10, 11,

$$\begin{align*}
\langle \psi | \otimes^n A_0 \otimes B_0 | \psi \rangle^\otimes n &= \cos^2 \left( \frac{\phi}{2} \right) - \sin^2 \left( \frac{\phi}{2} \right) = \cos \left( \frac{\phi}{2} \right), \\
\langle \psi | \otimes^n A_0 \otimes B_1 | \psi \rangle^\otimes n &= \cos \left( \frac{\phi}{2} \right) \cos \left( \frac{3\phi}{2} \right) + \sin \left( \frac{\phi}{2} \right) \sin \left( \frac{3\phi}{2} \right) = \cos \left( \frac{\phi}{2} \right)
\end{align*}$$

We reap the benefits of obtaining a lower value for the above inputs by obtaining a higher increase in the value for input 11 when $0 < p < \frac{2}{3}$. Applying Lemma 3, the expectation value for input 11 for $0 < p < \frac{1}{2}$
is given by,

\[
\text{Tr} \left( A_1 \otimes B_1 \rho^{\otimes n} \right) = \cos^2 \left( \frac{3\phi}{2} \right) \text{Tr} \left( \sigma_x^{\otimes n} \otimes \sigma_x^{\otimes n} \rho^{\otimes n} \right) - \sin^2 \left( \frac{3\phi}{2} \right) \text{Tr} \left( \sigma_x^{\otimes n} \otimes \sigma_z^{\otimes n} \rho^{\otimes n} \right)
\]

\[
= (q - p)^n \cos^2 \left( \frac{3\phi}{2} \right) - \sin^2 \left( \frac{3\phi}{2} \right)
\]

\[
= \frac{1}{2} ((1 + (q - p)^n) - (1 - (q - p)^n)) \cos^2 \left( \frac{3\phi}{2} \right)
\]

\[
- \frac{1}{2} ((1 + (q - p)^n) + (1 - (q - p)^n)) \sin^2 \left( \frac{3\phi}{2} \right)
\]

\[
= \frac{1}{2} ((1 + (q - p)^n) \cos(3\phi) - (1 - (q - p)^n)).
\]  \(\text{(12)}\)

Equation 12 yields the value for \( \frac{1}{2} \leq p \leq 1 \) if we fix \( n = 1 \). The expression is simplified by choosing \( \phi \) as specified in Protocol \( P \) and applying the following trigonometric equivalence,

\[
\cos(3\phi) = \begin{cases} 
-2(q-p)^n \cos(\phi) & \text{if } 0 < p < \frac{1}{2} \\
\frac{p-q}{\sqrt{1-q^2}} \cos(\phi) & \text{if } \frac{1}{2} \leq p < \frac{2}{3} \\
1 & \text{if } \frac{2}{3} \leq p \leq 1.
\end{cases}
\]  \(\text{(13)}\)

The value \( V \) attained by Protocol \( P \) is obtained by substituting the expectation values 11 and 12 in Equation 3, which gives

\[
V = \langle \psi | \otimes^n (A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0) | \psi \rangle^{\otimes n} - \text{Tr}(A_1 \otimes B_1 \rho^{\otimes n})
\]

\[
= 3 \cos(\phi) - \frac{1}{2} ((1 + (q - p)^n) \cos(3\phi) - (1 - (q - p)^n)).
\]

To complete the proof substitute Equation 13 in the expression for \( V \) and simplify to obtain

\[
V = \begin{cases} 
(3 + (q - p)^n) \cos(\phi) + \frac{1}{2}(1 - (q - p)^n) & \text{if } 0 < p < \frac{1}{2} \\
2(1 + q) \cos(\phi) + p & \text{if } \frac{1}{2} \leq p < \frac{2}{3} \\
2(1 + p) & \text{if } \frac{2}{3} \leq p \leq 1.
\end{cases}
\]

\[\Box\]

The value attained by Protocol \( P \) when Alice and Bob share \( n \) identical copies of a correlated qNLB is strictly greater than the value attained by the optimal classical protocol for \( 0 < p \leq \frac{1}{2} \). To verify the claim we need to show that the following inequality holds for \( 0 < p \leq \frac{1}{2} \),

\[
3 - l < 3 + l \sqrt{\frac{3 + l}{1 + l} + \frac{1 - l}{2}},
\]

where \( l = (q - p)^n \) with \( l \) ranging between \( 0 \leq l < 1 \). The inequality may be simplified to obtain,

\[
4(1 + l) < (3 + l)\sqrt{3 + l}.
\]

If we substitute \( k = 3 + l \), with \( k \) ranging between \( 3 \leq k < 4 \), we obtain the inequality,

\[
4k - k\sqrt{k} - 8 < 0.
\]
The inequality is verified by checking that the expression on the left hand side is negative for $0 < k < 4$ and has roots at $k$ equal to 4. The limit of $(q - p)^n$ as $n$ approaches infinity is 0 for $0 < p \leq \frac{1}{2}$. We conclude that Protocol $\mathcal{P}$ asymptotically distills correlated qNLBs to the value $\frac{1}{2}(3\sqrt{3} + 1) \approx 3.09876$ for $p$ less than a half.

B Constructing the $n$ copy SDP

Let Alice and Bob share $n$ identical copies of a correlated qNLB and receive input bits $x$ and $y$ respectively. Alice and Bob apply the observables $A_x$ and $B_y$ respectively, as specified in Protocol $\mathcal{P}$. Recall that the value attained for the CHSH inequality is

$$V = \langle \psi |^{\otimes n} (A_0 \otimes B_0 + A_0 \otimes B_1 + A_1 \otimes B_0) | \psi \rangle^{\otimes n} - \text{Tr}(A_1 \otimes B_1 \rho^{\otimes n}).$$

We define $N = 2^n + 3$ vectors, one vector for each of Alice’s two observables $A_x$, one for Bob’s observable $B_0$, and $2^n$ vectors for Bob’s observable $B_1$. Let the $2^n$ vectors $z_s$ be indexed by a length $n$ bit string $s$ in $\{0, 1\}^n$ and define $X_s = \sigma_x^{s_1} \otimes \sigma_x^{s_2} \otimes \cdots \otimes \sigma_x^{s_n}$ so that,

$$x_0 = (A_0 \otimes 1^{\otimes n}) | \psi \rangle^{\otimes n},$$
$$x_1 = (A_1 \otimes 1^{\otimes n}) | \psi \rangle^{\otimes n},$$
$$y_0 = (1^{\otimes n} \otimes B_0) | \psi \rangle^{\otimes n},$$
$$z_s = z_{s_1s_2...s_n} = (1^{\otimes n} \otimes (X_s B_1 X_s)) | \psi \rangle^{\otimes n}.$$

Let $G = [g_{ij}]$ be the Gram Matrix of the $N$ vectors $\{x_0, x_1, y_0, z_0^n, z_0^{n-1}, \ldots, z_{1^n}\}$. Set $W$ to be the symmetric weight matrix

$$w_{ij} = w_{ji} = \begin{cases} 1 & \text{if } (i, j) \in \{(1, 3), (1, 4), (2, 3)\} \\ -q^n - |s| |p| |s| & \text{if } i = 2 \text{ and } j = 3 + |s| \\ 0 & \text{otherwise}, \end{cases}$$

where $|s|$ is the Hamming weight of the bit string $s$. Let $s, s', t$ and $t'$ be length $n$ bit strings in $\{0, 1\}^n$ such that $s \neq s'$ and $t \neq t'$. Optimizing the value $V$ attained by Protocol $\mathcal{P}$ is then equivalent to finding an optimal primal solution to the following SDP.

$$\max_G \frac{1}{2} \text{Tr}(GW)$$
subject to $G \succ 0$

$$g_{ii} = 1 \text{ for all } i \in \{1, \ldots, N\}$$
$$g_{3+|s|,3+|s'|} = g_{3+|t|,3+|t'|} \iff s \oplus s' = t \oplus t'.$$

We already encountered the first two set of constraints in the primal for the single copy SDP in Section 4. These constraints ensure that the matrix $G$ is a Gram matrix and the $N$ vectors used in its construction have unit norm. The new set of constraints are derived from the $(2^n - 1)(2^n - 1)$ inner product restrictions of the form $z_s \cdot z_{s'} = z_t \cdot z_{t'}$ on the $z_s$ vectors.

To obtain the dual, we define vector $\lambda'$ in $\mathbb{R}^M$, with $M = 2^n + 3 + (2^n - 1)(2^n - 1)$, where the first $N = 2^n + 3$ components contribute to the solution value of the dual and the remaining entries correspond
to the additional constraints. To distinguish between these two different roles we partition \( \lambda' \) into two component vectors \( \mu \) and \( \tau \) such that,

\[
\mu_i = \begin{cases} 
\lambda'_i & \text{if } 1 \leq i \leq N \\
0 & \text{if } N < i \leq M
\end{cases}
\quad \text{and} \quad
\tau_i = \begin{cases} 
0 & \text{if } 1 \leq i \leq N \\
\lambda'_i & \text{if } N < i \leq M.
\end{cases}
\]

Define vector \( b \in \mathbb{R}^M \) and let \( b_i = 1 \) for \( i \leq N \) and 0 otherwise. Given four unique length \( n \) bit strings \( s, s', t \) and \( t' \), for each constraint of the form \( s \oplus s' = t \oplus t' \), define a matrix \( H_k \) for \( N < k \leq M \),

\[
h_{ij} = \begin{cases} 
1 & \text{if } i = 3 + |s|, j = 3 + |s'| \text{ such that } s \oplus s' = t \oplus t' \\
-1 & \text{if } i = 3 + |t|, j = 3 + |t'| \text{ such that } s \oplus s' = t \oplus t' \\
0 & \text{otherwise}.
\end{cases}
\]

The Lagrangian for the problem is given by

\[
\mathcal{L}(G, \lambda', Z) = \frac{1}{2} \text{Tr}(GW) + \text{Tr}(ZG) + \text{Tr}(\text{diag}(\mu) - \text{diag}(\mu)G) - \sum_{k=N+1}^{M} \tau_k \text{Tr}(H_k G)
\]

\[
= \lambda' \cdot b + \text{Tr} \left( \left( \frac{1}{2} W + Z - \text{diag}(\mu) - \sum_{k=N+1}^{M} \tau_k H_k \right) G \right),
\]

where \( \lambda' \) and \( Z \succ 0 \) are the dual variables. The dual function is then given by

\[
g(\lambda', Z) = \sup_G \mathcal{L}(G, \lambda', Z) = \begin{cases} 
\lambda' \cdot b & \text{if } \frac{1}{2} W + Z - \text{diag}(\mu) - \sum_{k=N+1}^{M} \tau_k H_k = 0 \\
+\infty & \text{otherwise}.
\end{cases}
\]

The dual problem may be stated as follows,

\[
\min_{\lambda'} \lambda' \cdot b
\]

subject to \( \frac{1}{2} W + Z - \text{diag}(\mu) - \sum_{k=N+1}^{M} \tau_k H_k = 0 \) and \( Z \succ 0 \).

We simplify the formulation by removing variable \( Z \) and defining \( \lambda = 2 \lambda' \) to obtain,

\[
\min_{\lambda'} \lambda' \cdot b
\]

subject to \( K = 2 \left( \text{diag}(\mu) - \sum_{k=N+1}^{M} \tau_k H_k \right) - W \succ 0 \). (15)

In the following section we utilize the above formulation of the dual to show that Protocol \( P \) is the optimal non-adaptive protocol for Alice and Bob when they have access to 2 or 3 copies of correlated qNLBs.

### C Optimal Dual Solutions for 2 and 3 Copies

The main idea we use to show optimality for the 2 and 3 copy cases, as in the single copy case is to break up the constraint matrix \( K \) into a sum of matrices \( K = W_{\text{head}} + W_{\text{tail}} \) and show that each matrix
is positive semi-definite. We decompose $K$ such that there is a fixed size $4 \times 4$ matrix $W_{\text{head}}$, while the matrix $W_{\text{tail}}$ has size $(2^n + 1) \times (2^n + 1)$. We begin by defining the a cut-off value $x$ that determines the decomposition of $K$ into $W_{\text{head}}$ and $W_{\text{tail}}$.

$$x = \begin{cases} \frac{1}{2}(1 + (q - p)^n) & \text{if } 0 < p \leq \frac{1}{2} \\ \frac{1}{2}(1 - p) & \text{if } \frac{1}{2} < p < 1. \end{cases} \tag{16}$$

Define the matrix,

$$W_{\text{head}} = \begin{pmatrix} \lambda_1 & 0 & -1 & -1 \\ 0 & l_1 & -1 & x \\ -1 & -1 & \lambda_3 & 0 \\ -1 & x & 0 & l_2 \end{pmatrix}. \tag{17}$$

The diagonal entries $\lambda_1$ and $\lambda_3$ are exactly the first and third components of the dual solution vector $\lambda$, while the entries $l_1$ and $l_2$ only have a partial contribution to the entries $\lambda_2$ and $\lambda_4$. Next we determine the diagonal values of $W_{\text{head}}$ and show that the matrix is positive semi-definite for these values.

**Lemma 9** The dual value for matrix $W_{\text{head}}$ is given by,

$$V' = \begin{cases} \sqrt{\frac{(1 + x)^3}{x}} & \text{if } 0 < p < \frac{2}{3} \\ \frac{3 - x}{x} & \text{if } \frac{2}{3} \leq p \leq 1. \end{cases}$$

**Proof** We fix $\lambda_1 = \lambda_3$ and $l_1 = l_2$ and choose the diagonal entries as follows,

$$\lambda_1 = \begin{cases} \sqrt{\frac{1 + x}{x}} & \text{if } 0 < p < \frac{2}{3} \\ \frac{2}{x} & \text{if } \frac{2}{3} \leq p \leq 1 \end{cases} \quad \text{and} \quad l_1 = \begin{cases} x\lambda_1 & \text{if } 0 < p < \frac{2}{3} \\ 1 - x & \text{if } \frac{2}{3} \leq p \leq 1. \end{cases}$$

The dual value attained by $W_{\text{head}}$ is verified by adding the diagonal entries. To prove that the matrix is positive semi-definite we conjugate the matrix $W_{\text{head}}$ by $H \otimes 1$,

$$(H \otimes 1)W_{\text{head}}(H \otimes 1) = \begin{pmatrix} \lambda_1 - 1 & -1 & 0 & 0 \\ -1 & l_1 + x & 0 & 0 \\ 0 & 0 & \lambda_1 + 1 & 1 \\ 0 & 0 & 1 & l_1 - x \end{pmatrix}. $$

The matrix is positive semi-definite if $(\lambda_1 - 1)(l_1 + x) \geq 1$ and $(\lambda_1 + 1)(l_1 - x) \geq 1$. Since both these inequalities hold for our choice of $\lambda_1$ and $l_1$, the matrix $W_{\text{head}}$ is positive semi-definite. \hfill \square

Unfortunately, we do not obtain a fixed size matrix $W_{\text{tail}}$ similar to $W_{\text{head}}$ that works for all $n$. To provide an overview of the dual constraints involved, we begin by giving a detailed construction of the $W_{\text{tail}}$ matrix for the 3 copy case. Let $|\Lambda\rangle = |\psi\rangle^{\otimes 3}$. We define the vectors $z_{\lambda}$ as follows,

$$
\begin{align*}
z_{000} &= 1^{\otimes 3} \otimes B_1|\Lambda\rangle \\
z_{001} &= 1^{\otimes 3} \otimes (1 \otimes 1 \otimes \sigma_x)B_1(1 \otimes 1 \otimes \sigma_x)|\Lambda\rangle \\
z_{010} &= 1^{\otimes 3} \otimes (1 \otimes \sigma_x \otimes 1)B_1(1 \otimes \sigma_x \otimes 1)|\Lambda\rangle \\
z_{001} &= 1^{\otimes 3} \otimes (1 \otimes \sigma_x \otimes \sigma_x)B_1(1 \otimes \sigma_x \otimes \sigma_x)|\Lambda\rangle \\
z_{100} &= 1^{\otimes 3} \otimes (\sigma_x \otimes 1 \otimes 1)B_1(\sigma_x \otimes 1 \otimes 1)|\Lambda\rangle \\
z_{101} &= 1^{\otimes 3} \otimes (\sigma_x \otimes 1 \otimes \sigma_x)B_1(\sigma_x \otimes 1 \otimes \sigma_x)|\Lambda\rangle \\
z_{110} &= 1^{\otimes 3} \otimes (\sigma_x \otimes \sigma_x \otimes 1)B_1(\sigma_x \otimes \sigma_x \otimes 1)|\Lambda\rangle \\
z_{111} &= 1^{\otimes 3} \otimes (\sigma_x \otimes \sigma_x \otimes \sigma_x)B_1(\sigma_x \otimes \sigma_x \otimes \sigma_x)|\Lambda\rangle.
\end{align*}
$$

17
The following inner product constraints apply on the vectors $z_s$ due to their definition. These are exactly the additional constraints required for the 3 copy dual solution.

$$
\begin{align*}
    z_{001} \cdot z_{000} &= z_{110} \cdot z_{111} = z_{011} \cdot z_{010} = z_{100} \cdot z_{101} \\
    z_{010} \cdot z_{000} &= z_{101} \cdot z_{111} = z_{100} \cdot z_{110} = z_{001} \cdot z_{011} \\
    z_{011} \cdot z_{000} &= z_{100} \cdot z_{111} = z_{001} \cdot z_{010} = z_{110} \cdot z_{101} \\
    z_{100} \cdot z_{000} &= z_{011} \cdot z_{111} = z_{100} \cdot z_{101} = z_{001} \cdot z_{011} \\
    z_{101} \cdot z_{000} &= z_{010} \cdot z_{111} = z_{001} \cdot z_{010} = z_{011} \cdot z_{110} \\
    z_{110} \cdot z_{000} &= z_{011} \cdot z_{110} = z_{100} \cdot z_{010} = z_{101} \cdot z_{011}
\end{align*}
$$

The dual constraint matrix $K$ for the 3 copy case is given by,

$$
K = \begin{pmatrix}
\lambda_1 & 0 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \lambda_2 & -1 & q^3 & q^2 & q^2 & q^2 & q^2 & q^2 & q^2 \\
-1 & -1 & \lambda_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-1 & q^3 & 0 & \lambda_4 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{12} + \lambda_{19} + \lambda_{26} & \lambda_5 & 0 & 0 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{13} + \lambda_{20} + \lambda_{27} & -\lambda_{21} & \lambda_6 & 0 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{14} + \lambda_{21} + \lambda_{28} & -\lambda_{27} & -\lambda_{19} & \lambda_7 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{15} + \lambda_{22} + \lambda_{29} & -\lambda_{23} & -\lambda_{24} & -\lambda_{32} & \lambda_8 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{16} + \lambda_{23} + \lambda_{30} & -\lambda_{29} & -\lambda_{25} & -\lambda_{31} & -\lambda_{26} & \lambda_9 & 0 \\
0 & q^2 & 0 & \lambda_{17} + \lambda_{24} + \lambda_{31} & -\lambda_{18} & -\lambda_{22} & -\lambda_{30} & -\lambda_{20} & -\lambda_{28} & \lambda_{10} \\
0 & p^3 & 0 & \lambda_{18} + \lambda_{25} + \lambda_{32} & -\lambda_{17} & -\lambda_{16} & -\lambda_{15} & -\lambda_{14} & -\lambda_{13} & -\lambda_{12} & \lambda_{11}
\end{pmatrix}
$$

where only the lower triangular matrix is shown for the constraints. We decompose $K$ into $W_{\text{head}}$ which contains contribution only from the upper left $4 \times 4$ block matrix with the remaining entries contained in $W_{\text{tail}}$.

$$
K = W_{\text{head}} + W_{\text{tail}} = \begin{pmatrix}
\lambda_1 & 0 & -1 & -1 \\
0 & l_1 & -1 & x \\
-1 & -1 & \lambda_3 & 0 \\
-1 & x & l_2 & 0
\end{pmatrix}
$$

$$
= \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & k_1 & q^3 - x & q^2 & q^2 & q^2 & q^2 & q^2 & q^2 & q^2 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{12} + \lambda_{19} + \lambda_{26} & \lambda_5 & 0 & 0 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{13} + \lambda_{20} + \lambda_{27} & -\lambda_{21} & \lambda_6 & 0 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{14} + \lambda_{21} + \lambda_{28} & -\lambda_{27} & -\lambda_{19} & \lambda_7 & 0 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{15} + \lambda_{22} + \lambda_{29} & -\lambda_{23} & -\lambda_{24} & -\lambda_{32} & \lambda_8 & 0 & 0 \\
0 & q^2 & 0 & \lambda_{16} + \lambda_{23} + \lambda_{30} & -\lambda_{29} & -\lambda_{25} & -\lambda_{31} & -\lambda_{26} & \lambda_9 & 0 \\
0 & q^2 & 0 & \lambda_{17} + \lambda_{24} + \lambda_{31} & -\lambda_{18} & -\lambda_{22} & -\lambda_{30} & -\lambda_{20} & -\lambda_{28} & \lambda_{10} \\
0 & p^3 & 0 & \lambda_{18} + \lambda_{25} + \lambda_{32} & -\lambda_{17} & -\lambda_{16} & -\lambda_{15} & -\lambda_{14} & -\lambda_{13} & -\lambda_{12} & \lambda_{11}
\end{pmatrix}
$$

18
where $\lambda_2 = l_1 + k_1$ and $\lambda_4 = l_2 + k_2$. In the following Lemma 11 we construct specific dual solution matrices that satisfy the constraint matrices of the above form for 2 and 3 copies of correlated qNLBs. The proof of Lemma 11 for the case $n = 3$ utilizes the following generalization of Observation 5.

**Theorem 10** (Corollary 7.2.4 in [19]) Let $A$ be a $n \times n$ Hermitian matrix, and let

$$p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_{n-m}t^{n-m}$$

(19)

be the characteristic polynomial of $A$. Suppose that $0 \leq m \leq n$ and $a_{n-m} \neq 0$. Then $A$ is positive semi-definite if and only if $a_k \neq 0$ for all $n - m \leq k \leq n$ and $a_k a_{k+1} < 0$ for $k = n - m, \ldots, n - 1$. We define $a_n \equiv 1$.

Even though the formulation of Lemma 11 applies to the general $n$ copy case, we prove it only for 2 and 3 copy case due to the complexity of the off-diagonal constraints.

**Lemma 11** The matrix $W_{\text{tail}}$ attains a dual solution value $1 - x$, for $n = 2$ and $n = 3$, where $x$ is defined in Equation 16.

**Proof** First consider the case $n = 2$, for the range $\frac{1}{2} \leq p < 1$. The matrix

$$W_{\text{tail}} = W_1 - W_2$$

$$= p \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} \\ 0 & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} \\ 0 & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} \\ p & 0 & 0 & 0 & 0 \end{pmatrix} - p \begin{pmatrix} 0 & q & -q & -q & -p \\ q & 0 & 0 & 0 & 0 \\ -q & 0 & 0 & 0 & 0 \\ -q & 0 & 0 & 0 & 0 \\ -p & 0 & 0 & 0 & 0 \end{pmatrix},$$

attains a dual solution value equal to $1 - x = p$. It remains to show that $W_{\text{tail}} \succeq 0$. The matrix $W_{\text{tail}}$ has rank 2 and its row space is spanned by its first two rows. The upper left $2 \times 2$ submatrix $\left( \begin{array}{cc} -\frac{q}{2} & -\frac{q}{2} \\ -\frac{q}{2} & -\frac{q}{2} \end{array} \right)$ is positive semi-definite by Observation 5 since $q \leq \frac{1}{2}$. It follows that $W_{\text{tail}}$ is positive semi-definite as well.

Next consider the range $0 < p < \frac{1}{2}$, for $n = 2$. We define the matrix $W_{\text{tail}}$ as

$$W_{\text{tail}} = W_1 - W_2$$

$$= p \begin{pmatrix} 2q & 0 & 0 & 0 & 0 \\ 0 & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} \\ 0 & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} \\ 0 & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} & -\frac{q}{2} \\ p & 0 & 0 & 0 & 0 \end{pmatrix} - p \begin{pmatrix} 0 & p & -q & -q & -p \\ p & 0 & 0 & 0 & 0 \\ -q & 0 & 0 & 0 & 0 \\ -q & 0 & 0 & 0 & 0 \\ -p & 0 & 0 & 0 & 0 \end{pmatrix},$$

which attains a dual solution value $2pq = 1 - x$. The matrix has rank 2 and its row space is spanned by its first two rows. The upper left $2 \times 2$ submatrix $\left( \begin{array}{cc} 2q & -p \\ -p & \frac{q}{2} \end{array} \right)$ is positive semi-definite by Observation 5 since $p < q$. It follows that $W_{\text{tail}}$ is positive semi-definite as well.

For the case $n = 3$, we begin by removing the two zero rows and columns from the $W_{\text{tail}}$ matrix as specified in Equation 18. We further restrict columns of the same Hamming weight to be equal. This corresponds to equating the three cases each for weight 1 and 2 identified by the entries $q^2 p$ and $q p^2$ respectively. The reduction in the number of constraints $\lambda$ allows us to consider a $5 \times 5$ matrix. A construction for the case
$p \leq \frac{1}{2}$ yields the matrix

$$
\begin{pmatrix}
    v_1 & v_2 & v_3 & v_4 & v_5 \\
    p(3q^2 + p^2) & -3qp^2 & q^2p & qp^2 & p^3 \\
    -3qp^2 & \frac{9}{2}q^2p^3 & -\frac{3}{2}q^2p^2 & \frac{3}{4}q^2p^3 & -\frac{3}{4}q^2p^2 \\
    q^2p & -\frac{3}{4}q^2p^3 & \frac{1}{2}q^2p^2 & \frac{1}{4}q^2p^3 & \frac{q^2p}{2}q^2p^2 - \frac{q^2p}{q^2 + p^2} \\
    qp^2 & \frac{3}{2}q^2p^3 & \frac{1}{2}q^2p^2 & \frac{1}{4}q^2p^3 & \frac{1}{2}q^2p^2 \\
    p^3 & -\frac{3}{4}q^2p^2 & \frac{3}{2}q^2p^2 - q^2p^2 & \frac{1}{4}q^2p^2 & \frac{3q^2 - 4q^2p^2 + 2p^4}{q^2 + p^2}
\end{pmatrix}.
$$

(20)

The rank of $W_{\text{tail}}$ may be further decreased by noting that $v_1 = 3v_3 + v_5$ and $v_2 = -3v_4$. We now choose to consider the rank 3 matrix spanned by $v_1, v_3$ and $v_4$, given by

$$
W_{\text{tail}} = \begin{pmatrix}
p(3q^2 + p^2) & q^2p & qp^2 \\
q^2p & \frac{1}{2}q^2p^2 & \frac{1}{4}q^2p^3 \\
qp^2 & \frac{1}{4}q^2p^3 & \frac{1}{2}q^2p^2
\end{pmatrix}.
$$

(21)

The dual value for Matrix $[21]$ is given by $p(3q^2 + p^2)$ which equals $1 - x$. A similar construction for the case $\frac{1}{2} < p < 1$ yields the matrix

$$
\begin{pmatrix}
v_1 & v_2 & v_3 & v_4 & v_5 \\
p & q^3 - q & q^3 - q & q^2p & q^3 - q \\
q^3 - q & q^3 - q & q^3 - q & q^3 - q & q^3 - q \\
q^2p & q^2p & q^2p & q^2p & q^2p \\
qp^2 & q^2p & q^2p & q^2p & q^2p \\
p^3 & p^3 & p^3 & p^3 & p^3
\end{pmatrix}.
$$

(22)

The rank of $W_{\text{tail}}$ may be further decreased by noting that $v_1 = 3v_3 + v_5$ and $v_2 = -2v_3 - v_4$. We now choose to consider the rank 3 matrix spanned by $v_1, v_3$ and $v_4$, given by

$$
W_{\text{tail}} = \begin{pmatrix}
p & q^2p & qp^2 \\
q^2p & \frac{1}{2}q^2p^2 & \frac{1}{4}q^2p^3 \\
qp^2 & \frac{1}{4}q^2p^3 & \frac{1}{2}q^2p^2
\end{pmatrix}.
$$

The dual value for Matrix $[22]$ is $p$, which equals $1 - x$. The fact that the Matrices $[21]$ and $[22]$ are positive semi-definite may be verified by application of Theorem $[10]$ to the characteristic polynomials of these matrices.

The final part of our analysis constitutes the proof of Theorem $[7]$ which is obtained by combining the dual solution values for both the matrices $W_{\text{head}}$ and $W_{\text{tail}}$.

**Proof of Theorem $[7]$**: We prove that the Protocol $[7]$ is optimal for 2 and 3 copies by combining the dual values from Lemmas $[9]$ and $[11]$ No distillation for the range $\frac{1}{2} < p \leq 1$ implies the dual solution values are the same for both $n = 2$ and $n = 3$. Also, the value attained matches the value attained by Protocol $[7]$ and is therefore tight. For $\frac{2}{3} < p \leq 1$, we have

$$
V = 3 - x + 1 - x \\
= 4 - 2x \\
= 2(1 + p).
$$

20
For $\frac{1}{2} < p \leq \frac{2}{3}$, we obtain
\[
V = \sqrt{\frac{(1+x)^3}{x}} + 1 - x \\
= \sqrt{\frac{(2-p)^3}{1-p}} + p \\
= (2-p)\sqrt{\frac{2-p}{1-p}} + p \\
= 2(2-p)\cos(\phi) + p \\
= 3\cos(\phi) - q\cos(3\phi) + p.
\]

For $0 < p \leq \frac{1}{2}$ and $n = 2$,
\[
V = \left(\frac{3+(q-p)^2}{2}\right) \sqrt{\frac{3+(q-p)^2}{1+(q-p)^2} + \frac{1-(q-p)^2}{2}} \\
= (3+(q-p)^2)\cos(\phi) + \frac{1}{2}(1-(q-p)^2) \\
= (3+(q-p)^2)\cos(\phi) + 2pq.
\]

Finally, for $0 < p \leq \frac{1}{2}$ and $n = 3$,
\[
V = \left(\frac{3+(q-p)^3}{2}\right) \sqrt{\frac{3+(q-p)^3}{1+(q-p)^3} + \frac{1-(q-p)^3}{2}} \\
= (3+(q-p)^3)\cos(\phi) + \frac{1}{2}(1-(q-p)^3).
\]

The formulation we use obtain for the matrices $W_{\text{head}}$ and $W_{\text{tail}}$ indicate that Protocol $\mathcal{P}$ is the optimal non-adaptive distillation protocol for correlated qNLBs. It seems reasonable to think that the solution value obtained in Lemma 11 holds for all values of $n$. One possible route in proving this, is to obtain a general form for the off-diagonal entries of the matrix $W_{\text{tail}}$. 

\[\square\]