Witnessing environment dimension through temporal correlations

Lucas B. Vieira[1,2], Simon Milz[3,2,1], Giuseppe Vitagliano[4], and Costantino Budroni[5,2,1]

[1]Institute for Quantum Optics and Quantum Information (IQOQI), Austrian Academy of Sciences, Boltzmanngasse 3, 1090 Vienna, Austria

[2]Faculty of Physics, University of Vienna, Boltzmanngasse 5, 1090 Vienna, Austria

[3]School of Physics, Trinity College Dublin, Dublin 2, Ireland

[4]Vienna Center for Quantum Science and Technology, Atominstitut, TU Wien, 1020 Vienna, Austria

[5]Department of Physics "E. Fermi" University of Pisa, Largo B. Pontecorvo 3, 56127 Pisa, Italy

We introduce a framework to compute upper bounds for temporal correlations achievable in open quantum system dynamics, obtained by repeated measurements on the system. As these correlations arise by virtue of the environment acting as a memory resource, such bounds are witnesses for the minimal dimension of an effective environment compatible with the observed statistics. These witnesses are derived from a hierarchy of semidefinite programs with guaranteed asymptotic convergence. We compute non-trivial bounds for various sequences involving a qubit system and a qubit environment, and compare the results to the best known quantum strategies producing the same outcome sequences. Our results provide a numerically tractable method to determine bounds on multi-time probability distributions in open quantum system dynamics and allow for the witnessing of effective environment dimensions through probing of the system alone.

1 Introduction

Physical systems are never truly isolated, and inevitably interact with their surrounding environment [7, 55]. As a consequence, information within the system leaks away into its surroundings, leading to entanglement between the system and the environment. In many instances, this leaked information may be partially recovered at a later time, leading to non-Markovian dynamics, i.e., non-negligible memory effects and complex correlations in time [8, 48, 56].

Like their spatial counterpart, temporal correlations in quantum mechanics fundamentally differ from those that can be observed in the classical case. Such differences between classical and quantum tempo-

Lucas B. Vieira: lucas.vieira@oeaw.ac.at
Simon Milz: milzs@tcd.ie
Giuseppe Vitagliano: giuseppe.vitagliano@tuwien.ac.at
Costantino Budroni: costantino.budroni@unipi.it

ral correlations have been noted since the works of Leggett and Garg [26, 40, 41, 68]. In the presence of memory, a clear distinction between an underlying process – carrying temporal correlations – and the measurement process – probing said correlations – can be obtained by employing correlation kernels [1, 42] or higher order quantum maps [54] for their description.

Broadly speaking, the dimension of a physical system, i.e., the number of perfectly distinguishable states, imposes fundamental constraints over the temporal correlations it is able to produce [12]. In this sense, the physical dimension acts as a memory resource, restricting the amount of information stored about the past that is capable of affecting the future. This memory constraint leads to different behaviors (i.e., different achievable temporal correlations) in classical, quantum, or more general physical theories [12, 27, 35, 46, 61, 67]. In particular quantum memories are known to allow for a larger set of correlations than classical ones of the same size [12, 13, 67]. However, this advantage only holds for restricted memory size, i.e., if the memory size is *unrestricted*, all correlations compatible with a time-ordered causal structure may be achieved with either classical or quantum memories [27, 35]. Therefore, understanding the nature of dimensional constraints on observable correlations sheds light on fundamental differences between our descriptions of physical systems, as well as their connection with causality.

These limitations imposed by the dimensionality of physical systems have been exploited for the construction of "dimension witnesses", inequalities which, when violated, certify the minimum dimension of the system compatible with made observations [9, 11, 29, 31, 58, 60, 61]. In a similar spirit, our approach allows for the computation of upper bounds on the temporal correlations achievable in an open system dynamics with an environment of bounded dimension, so that any violations of these bounds certify the minimum dimension of the effective environment. Employing techniques from entanglement detection [25, 49], these bounds are obtained by exploiting the inherent symmetries of the problem. In particular, we relax the
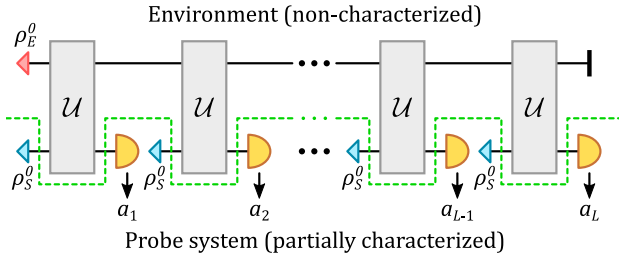
Figure 1: Diagram of the sequential measurement protocol, involving a partially characterized probe system and a non-characterized environment, here separated by the dashed line. The sequence of measurement outcomes $\boldsymbol{a} = a_1 \ldots a_L$ is obtained by repeated preparations of a probe state ρ_S^0, fixed and the same at every time step, which is left to interact with the environment, then followed by measurements (semicircles). The environment acts as a *memory* resource, capable of establishing long-term correlations between measurements.

a priori non-linear problem of computing maximum joint probabilities in sequential measurements to a numerically tractable hierarchy of semidefinite programs (SDPs) [5]. Both the computational accessibility of the final formulation of the problem, as well as the non-triviality of the resulting bounds, are then demonstrated for paradigmatic examples, showing that, indeed, joint probability distributions obtained from probing an open system alone provide viable means to deduce dimensional properties of the a priori experimentally inaccessible environment it is coupled to.

The paper is organized as follows. In Sec. 2, we introduce the temporal sequence protocol, the notion of temporal correlations, and their mathematical description. Sec. 3 discusses the application of temporal correlations to the task of characterizing open systems and their environment. Sec. 4 describes the semidefinite program we constructed to bound temporal correlations, with Sec. 5 presenting the numerical results we obtained in their optimization. In Sec. 6 we comment on the various challenges involved in realizing these numerical optimizations, and how we have overcome them. In Sec. 7 we comment on several additional technical details involving our work, with Sec. 8 covering our conclusions and future outlook.

2 The measurement protocol

We assume throughout that an experimenter is able to prepare the system in a known state and perform measurements in a certain basis. Its dynamics, on the other hand, are not under experimental control, and are governed by its inevitable interaction with the environment. Typically, this environment is a much larger system, generally inaccessible, and featuring complex dynamics. Together, system and environment undergo closed, i.e., unitary evolution. Their interaction leads to an imprinting of information on

the probe system, which can be used to learn something about the environment by means of the probe alone. In fact, this is a common way of making indirect measurement on (typically large many-body) systems that are not fully controlled [6, 14]. For example, a small probe can be used for estimating an unknown parameter of a larger (many-body) environment [22], in particular temperature [3, 21, 47, 57].

We now introduce the details of our scenario and the notation used. By $\mathcal{H}_{\mathrm{S}}$ and $\mathcal{H}_{\mathrm{E}}$ we denote the finite-dimensional Hilbert spaces of system and environment, with $d_{\mathrm{S}} = \dim \mathcal{H}_{\mathrm{S}}$ and $d_{\mathrm{E}} = \dim \mathcal{H}_{\mathrm{E}}$, and their joint space as $\mathcal{H}_{\mathrm{ES}} = \mathcal{H}_{\mathrm{E}} \otimes \mathcal{H}_{\mathrm{S}}$, with $d_{\mathrm{ES}} = d_{\mathrm{E}} d_{\mathrm{S}}$. The experiment involves L identical measurements on the probe system, at discrete time steps, each outcome $a_\ell \in \mathcal{A}$ collected into a sequence $\boldsymbol{a} = a_1 \ldots a_L$. For simplicity, we consider $\mathcal{A} = \{0, 1\}$ in the following; the generalization to arbitrary outcomes is straightforward. Measurements on the system are described by a Positive Operator-Valued Measure (POVM) $(E_S^a)_a$, i.e., $E_S^a \geq 0$ and $\sum_{a \in \mathcal{A}} E_S^a = \mathbb{1}_S$.

The system is initially in the state ρ_S^0, interacts with the environment via the global unitary $\mathcal{U}(\cdot) = U \cdot U^\dagger$, then it is measured and reprepared in the state ρ_S^0. This procedure is repeated a total of L times; see Fig. 1. The unitary operation is assumed to be the same at each iteration (corresponding to, for example the situation of a time-independent system-environment Hamiltonian and temporally equidistant measurements); see Sec. 7.1 for more details.

In the following, we consider a measure-and-prepare operation of the form

$$\mathcal{M}_a(\rho_{ES}) = \mathrm{tr}_S[\rho_{ES} \cdot (\mathbb{1}_E \otimes E_S^a)] \otimes \rho_{\mathrm{S}}^0, \quad (1)$$

with the generalization to arbitrary operations being straightforward. The probability of a sequence of outcomes can then be written as

$$\begin{aligned} p(\boldsymbol{a}|d_{\mathrm{E}}) &= p(a_1, \ldots, a_L|d_{\mathrm{E}}) \\ &= \mathrm{tr}\left[\mathcal{M}_{a_L} \circ \mathcal{U} \circ \cdots \circ \mathcal{M}_{a_1} \circ \mathcal{U}(\rho_{\mathrm{E}}^0 \otimes \rho_{\mathrm{S}}^0)\right], \end{aligned} \quad (2)$$

where $p(\boldsymbol{a}|d_{\mathrm{E}})$ denotes that it is obtained with an environment of dimension d_{E}. Our task is to establish upper bounds on such probabilities given a finite amount/dimension of "memory" d_{E}. The assumption of identical measurements is essential for capturing this notion of memory, as arbitrary time-dependent operations would impose no non-trivial restriction on the probabilities.

For simplicity, we consider the maximization of the probability of a given sequence $\boldsymbol{a}$, namely, to find $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$ such that $p(\boldsymbol{a}|d_{\mathrm{E}}) \leq \omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$, for all possible correlations generated by an environment of dimension d_{E}. The generalization to arbitrary linear functions of the distributions $(p(\boldsymbol{a}|d_{\mathrm{E}}))_{\boldsymbol{a}}$ is straightforward.

We formulate the problem via the Choi-Jamiołkowski (CJ) isomorphism [19, 37], which involves taking multiple copies of the original Hilbert
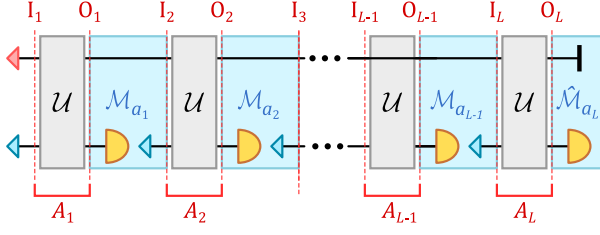
Figure 2: The protocol with each time step having its own set of input and output Hilbert spaces, denoted by the vertical dotted lines. Note that the input and output spaces of the $\mathcal{U}$ and $\mathcal{M}_a$ are interleaved: $\mathcal{U}$ has inputs I_ℓ and outputs O_ℓ, but for $\mathcal{M}_{a_\ell}$, inputs are O_ℓ and outputs $\mathsf{I}_{\ell+1}$.

space $\mathcal{H}_{\mathrm{ES}}$ to describe its time evolution; see Fig. 2. The ℓ-th unitary acts as $\mathcal{U} : \mathsf{I}_\ell \mapsto \mathsf{O}_\ell$, while the ℓ-th measurement, placed between unitaries, acts instead as $\mathcal{M}_a : \mathsf{O}_\ell \mapsto \mathsf{I}_{\ell+1}$. All spaces are isomorphic, i.e., $\mathsf{O} \cong \mathsf{I} \cong \mathcal{H}_{\mathrm{ES}}$. For convenience, the ℓ-th unitary's input and output spaces are jointly referred to as $\mathsf{A}_\ell = \mathsf{I}_\ell \otimes \mathsf{O}_\ell$. It is important to emphasize that the local input and output spaces of the $\mathcal{U}$ and $\mathcal{M}_a$ are interleaved, in the sense that the outputs of one are the inputs of the other (Fig. 2). For clarity, we shall use the cursive $\mathcal{I}$ and $\mathcal{O}$ to generically refer to each map's local input and output spaces: for unitaries, $\mathcal{I} = \mathsf{I}_\ell$ and $\mathcal{O} = \mathsf{O}_\ell$, and for measurements, $\mathcal{I} = \mathsf{O}_\ell$ and $\mathcal{O} = \mathsf{I}_{\ell+1}$.

Given a linear map $\Lambda : \mathcal{I} \mapsto \mathcal{O}$ between input and output Hilbert spaces $\mathcal{I}$ and $\mathcal{O}$, with $d_\mathcal{I} = \dim \mathcal{I}$ and $d_\mathcal{O} = \dim \mathcal{O}$, its Choi-Jamiołkowski representation [19, 37] is given by the matrix

$$\mathcal{C}(\Lambda) := \sum_{i,j=1}^{d_\mathcal{I}} |i\rangle\langle j| \otimes \Lambda(|i\rangle\langle j|) = \mathrm{id}_\mathcal{I} \otimes \Lambda(|\Omega\rangle\langle\Omega|), \quad (3)$$

where $|\Omega\rangle = \sum_{i=0}^{d_\mathcal{I}-1} |ii\rangle$ is a non-normally maximally entangled state. The matrix $\mathcal{C}(\Lambda)$ is typically referred to as a Choi matrix for the map Λ and, if normalized, also as its Choi state. Note that $\mathcal{C}(\Lambda)$ is of size $d_\mathcal{I} d_\mathcal{O} \times d_\mathcal{I} d_\mathcal{O}$. Since in our case input and output spaces are isomorphic, we have $d_\mathcal{I} = d_\mathcal{O} = d_{\mathrm{ES}}$. We note that the decomposition of sequential maps in Eq. (2) is equivalent to the standard formulation of temporal scenarios in the quantum comb [17], process matrix [18, 52], and process tensor [54] formalism.

3 Witnesses for open system dynamics

Given a probe system S, the first question we may ask is whether or not it is open, i.e., if it is interacting with an environment at all, or equivalently, whether $d_{\mathrm{E}} > 1$. To answer this question, we compute $\omega_1^{\boldsymbol{a}} := \max_{\rho_S^0, (E_S^a)_a} p(\boldsymbol{a}|d_{\mathrm{E}} = 1)$, giving the inequality

$$p(\boldsymbol{a}|d_{\mathrm{E}} = 1) \leq \omega_1^{\boldsymbol{a}}. \quad (4)$$

As an example, the maximum for the sequence $\boldsymbol{a} = 00101$ is $\omega_1^{\boldsymbol{a}} = (3/5)^3(2/5)^2 = 0.03456$, which we explain how to obtain shortly. If we perform an experiment and observe $p(\boldsymbol{a} = 00101) = 0.5$, then we must conclude the system is open. In other words, the inequality Eq. (4) acts as a witness for open systems. However, if no violation is observed, the experiment is inconclusive.

To compute this maximum, we first note that, for our choice of $\mathcal{M}_a$, since both ρ_S^0 and E_S^a are the same at each measurement, all outcomes are independent and identically distributed. Writing $q_a = \mathrm{tr}\left[\rho_S^0 E_S^a\right]$ as the probability of outcome a, with $\sum_a q_a = 1$, and using n_a as the number of occurrences of a symbol a in $\boldsymbol{a}$, we have

$$p(\boldsymbol{a}|d_{\mathrm{E}} = 1) = \prod_{\ell=1}^{L} q_{a_\ell} = \prod_{a \in \mathcal{A}} q_a^{n_a}, \qquad \sum_{a \in \mathcal{A}} n_a = L. \quad (5)$$

The global maximum $\omega_1^{\boldsymbol{a}}$ can be found analytically with standard techniques for optimizing $\{q_a\}_a$ (see App. C), and is given by

$$p(\boldsymbol{a}|d_{\mathrm{E}} = 1) \leq \omega_1^{\boldsymbol{a}} = \prod_{a \in \mathcal{A}} \left(\frac{n_a}{L}\right)^{n_a}, \quad (6)$$

with the convention $q_a^{n_a} = 1$ for $q_a = n_a = 0$. The maximum is independent of d_S and is obtained with $E_S^a = \frac{n_a}{L}\mathbb{1}_S$ and any ρ_S^0.

Similar principles apply when developing witnesses for the case of $d_{\mathrm{E}} > 1$, although the calculation of the maximum becomes nontrivial, as the probability no longer factorizes as in Eq. (5). Concretely, we want to obtain the tightest bounds of the form

$$p(\boldsymbol{a}|d_{\mathrm{E}}) \leq \omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}, \quad (7)$$

which hold for all possible realizations of the sequence $\boldsymbol{a}$ in an experiment. If a violation of Eq. (7) is observed, we can certify that the dimension is greater than d_{E}.

In the simple case of $d_{\mathrm{E}} = 1$, we were able to optimize over ρ_S^0 and $(E_S^a)_a$. In the general case, however, optimizing over all possible protocols (i.e. all possible ρ_S^0, $(E_S^a)_a$ and U) is difficult. Nevertheless, some general considerations can be made about how the bounds depend on these parameters; see Sec. 7.3. In the following, we assume fixed and well characterized preparations and measurements, in agreement with our initial assumptions on the probe system. The optimization we need to perform is, thus,

$$\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}} := \max_U p(\boldsymbol{a}|d_{\mathrm{E}}) \quad (8)$$

where assumptions about the initial state of the environment can be removed by convexity and symmetry arguments; see Sec. 7.2. In the following section, we explain how to compute the non-trivial maxima $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$ via convex optimization methods.

4 A hierarchy of semidefinite programs bounding temporal correlations

The maxima in Eq. (7) require that the probe state ρ_S^0 and measurements $(E_S^a)_a$ are characterized and fixed throughout the experiment, as discussed in the previous section. To obtain upper bounds for the maximum $\omega_{d_E}^a$, we formulate the problem as a hierarchy of semidefinite programs (SDPs) [5]. This class of optimization problems are ubiquitous in quantum information theory, offering strict convergence of solutions for many classes of problems, as well as having efficient algorithms widely available for solving them [5, 70]. For a recent compendium on SDPs in quantum information science, see [59, 64]. In the following, we explain the general steps taken in formulating our problem as an SDP, which can be solved numerically and, importantly, efficiently. A more detailed step-by-step formulation is included in App. A.

Under the CJ representation, the repeated applications of the unitary and measurements can each be written as single operators existing in a larger space, encompassing multiple time steps, with their input and output spaces interleaved. This is illustrated in Fig. 2. Via Eq. (3), we define

$$C_U := \frac{1}{d_{ES}} \mathcal{C}(\mathcal{U}),$$
$$M_a := \mathcal{C}(\mathcal{M}_a), \quad \text{and} \quad \hat{M}_a := \text{tr}_{\mathcal{O}}[M_a]. \tag{9}$$

With this, we can write the L repeated applications of $\mathcal{U}$ as $C_U^{\otimes L}$, and the measure-and-prepare operations as a single operator

$$X := (d_{ES})^L (\rho_E^0 \otimes \rho_S^0) \otimes M_{a_1} \otimes M_{a_2} \otimes \cdots \otimes \hat{M}_{a_L}, \tag{10}$$

where $\hat{M}_a$ is simply the final measurement without a re-preparation, obtained by partially tracing M_a over its output space. As the goal is to obtain a maximum, the initial environment state ρ_E^0 can be chosen to be pure by convexity arguments, fixed to be $|0\rangle\langle 0|_E$ by unitary invariance, and considered as part of the experimental setup X without loss of generality; see Sec. 7.2. We can then express the probability of $\boldsymbol{a}$ as

$$p(\boldsymbol{a}|d_E) = \text{tr}\left[X^{\text{T}} C_U^{\otimes L}\right]. \tag{11}$$

Note that by Eq. (3) and Eq. (9) C_U is positive, rank-1, and normalized, which makes $C_U^{\otimes L}$ a pure symmetric separable quantum state. In addition, each C_U satisfies $\text{tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{ES}$, which arises from the trace-preserving property of each individual unitary. With all of these observations, we reformulate Eq. (8)

as the equivalent optimization problem:

Given: d_E, d_S, ρ_E^0, ρ_S^0, $\{M_a\}_a$, $\boldsymbol{a}$,

with $X := (d_{ES})^L (\rho_E^0 \otimes \rho_S^0) \otimes M_{a_1} \otimes \cdots$
$$\cdots \otimes M_{a_{L-1}} \otimes \hat{M}_{a_L},$$

Find: $\omega_{d_E}^{\boldsymbol{a}} := \max_{C_U} \text{tr}\left[X^{\text{T}} C_U^{\otimes L}\right]$

Subject to: $C_U \geq 0, \quad \text{rank } C_U = 1,$
$$\text{tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{ES}. \tag{12}$$

As C_U enters in the objective function as a tensor power, and we further require it to be rank-1, the problem is both non-linear and non-convex with respect to C_U, and thus the above problem cannot be directly solved as an SDP. To reach an SDP *relaxation* of Eq. (12), we first transform it into a chain of equivalent problems.

The problem can be made convex by replacing $C_U^{\otimes L}$ with a separable state on the symmetric subspace

$$\widetilde{\Phi}_{A_1^L} := \sum_i p_i |\phi_i\rangle\langle\phi_i|^{\otimes L}, \tag{13}$$

with $p_i \geq 0$, $\sum_i p_i = 1$. By convexity, the maximum will be achieved for a rank-1 $\widetilde{\Phi}_{A_1^L}$, therefore this relaxation leaves the optimal value unchanged. Thus, we replace $C_U^{\otimes L}$ in Eq. (12) with $\widetilde{\Phi}_{A_1^L}$ satisfying

$$\widetilde{\Phi}_{A_1^L} \in \text{SEP}_L, \qquad P_L^+ \widetilde{\Phi}_{A_1^L} = \widetilde{\Phi}_{A_1^L}, \tag{14}$$

and the partial trace constraint of Eq. (12), where SEP_L is the set of fully separable L-partite states, and P_L^+ is the projector onto the symmetric subspace of L systems (see the definition in Eq. (36) in App. B.1). In fact, the constraints in Eq. (14) are exact, as it can be shown that all separable states in the symmetric subspace are of the form in Eq. (13); see App. A.1.1 for a proof.

So far, we transformed the original problem into an equivalent one, but due to the condition $\widetilde{\Phi}_{A_1^L} \in \text{SEP}_L$ it is not yet an SDP. This requirement can be obtained by relaxing the original problem via the quantum de Finetti theorem [15] (see also [20]), which tells us that $\widetilde{\Phi}_{A_1^L}$ can be approximated as the reduced state of a larger symmetric (and potentially entangled) state $\Phi_{A_1^N}$, such that $\widetilde{\Phi}_{A_1^L} \approx \text{tr}_{A_{L+1}^N}[\Phi_{A_1^N}]$, where $\text{tr}_{A_{L+1}^N}$ is the trace over systems $L+1, \ldots, N$.

While this relaxation *a priori* only provides an upper bound for the original problem, it establishes a hierarchy of approximate solutions, which are known to converge exactly to the separable set as N increases [25]. The de Finetti theorem, and analogous results for permutationally invariant (or exchangeable), rather than symmetric, operators have found broad applications in quantum information theory, from entanglement detection [24, 25, 49] to more general optimization problems, such as evaluating convex roofs of entanglement measures [66], constrained

bilinear optimization [4], dimension-bounded quantum games [38], as well as rank-constrained optimization [73]. All the above mentioned results exploit the basic idea of using either the symmetric subspace or permutation invariance to relax nonlinear constraints. In particular, Ref. [4] analyzed the optimization over pairs of quantum channels, basing their construction on permutation invariant operators, and showing the partial trace constraint necessary to translate the usual procedure (e.g., in the entanglement detection approach) from states to channels. Moreover, [73] introduced the idea of a rank-constrained optimization based on the symmetric subspace, which is central to impose the rank constraint in Eq. (12). Our construction is inspired by these works and could be derived starting from some of these results. However, it is more straightforward to provide a direct construction in terms of the de Finetti theorem; see App. A.2.

Ultimately, we arrive at an SDP relaxation of the original problem:

$$
\begin{aligned}
\textbf{Given:} \quad & d_{\mathrm{E}},\ d_{\mathrm{S}},\ \rho_{\mathrm{E}}^0,\ \rho_{\mathrm{S}}^0,\ \{M_a\}_a,\ \boldsymbol{a},\ N, \\
\text{with} \quad & X := (d_{\mathrm{ES}})^L (\rho_{\mathrm{E}}^0 \otimes \rho_{\mathrm{S}}^0) \otimes M_{a_1} \otimes \cdots \\
& \cdots \otimes M_{a_{L-1}} \otimes \hat{M}_{a_L}, \\
\text{and} \quad & \hat{M}_{a_L} := \mathrm{tr}_{\mathcal{O}}[M_{a_L}] \\
\textbf{Find:} \quad & \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N} := \max_{\Phi_{\mathrm{A}_1^N}} \mathrm{tr}\left[(X^{\mathrm{T}} \otimes \mathbb{1}_{\mathrm{A}_{L+1}^N}) \Phi_{\mathrm{A}_1^N} \right] \\
\textbf{Subject to:} \quad & \Phi_{\mathrm{A}_1^N} \geq 0, \\
& \mathrm{P}_N^+ \Phi_{\mathrm{A}_1^N} = \Phi_{\mathrm{A}_1^N}, \quad \mathrm{tr}\left[\Phi_{\mathrm{A}_1^N}\right] = 1, \\
& \mathrm{tr}_{\mathrm{O}_1}\left[\Phi_{\mathrm{A}_1^N}\right] = \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{tr}_{\mathrm{A}_1}\left[\Phi_{\mathrm{A}_1^N}\right].
\end{aligned}
\tag{15}
$$

Here, $N \geq L$ is the size of the symmetric state $\Phi_{\mathrm{A}_1^N}$ used in the approximation of the separable L-partite state $\Phi_{\mathrm{A}_1^L}$. The last constraint in Eq. (15) is required to ensure that $\Phi_{\mathrm{A}_1^N}$ represents a sequence of trace-preserving maps, and it can be enforced on a single step due to the symmetry constraint on $\Phi_{\mathrm{A}_1^N}$; see App. A.1.2. Optionally, one can also add further linear constraints to the SDP to improve its approximation to the separable set, e.g., entanglement witnesses [33, 65] or the positive partial transpose (PPT) criterion [53] $\Phi_{\mathrm{A}_1^N}^{\mathrm{T}_\alpha} \geq 0, \forall \alpha$, where $\Phi_{\mathrm{A}_1^N}^{\mathrm{T}_\alpha}$ is the partial transpose with respect to a bipartition α. These conditions are not necessary for convergence, but they may provide better results [49] at an extra computational cost.

Once approximate solutions $\tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N}$ are obtained from Eq. (15), they can be used to establish a convergent hierarchy of upper bounds on the temporal correlations for each sequence, i.e.,

$$
\begin{aligned}
& \omega_{d_{\mathrm{E}}}^{\boldsymbol{a}} \leq \cdots \leq \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N+1} \leq \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N} \leq \cdots \leq \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},L}, \\
& \text{with} \quad \lim_{N\to\infty} \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N} = \omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}.
\end{aligned}
\tag{16}
$$

Consequently, the above formulates a sequence of SDPs capable of approximating the maxima $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$ to arbitrary precision. Numerical results obtained from *any* of these SDPs (i.e., for any $N \geq L$) can be used to construct dimensional witnesses, which, when violated, certify the minimum dimension of the effective environment interacting with the system. A software implementation of the SDP in Eq. (15), however, is not straightforward even for small values of $\{d_{\mathrm{S}}, d_{\mathrm{E}}, L, N\}$, as the memory requirements quickly render the problem computationally intractable. Obtaining the numerical results presented in the next section, thus, required a significant amount of optimization; see Sec. 6 for details. A schematic outline of all steps undertaken to formulate the SDP and obtain the numerical results is presented in Fig. 3.

The asymptotic convergence of the hierarchy of bounds in Eq. (16) for the SDP in Eq. (15), without PPT constraints, is given by [16]

$$
|\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}} - \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N}| \leq \frac{L(L + (d_{\mathrm{ES}})^2 + 1)}{N + d_{\mathrm{ES}}}.
\tag{17}
$$

If PPT constraints are included, we only have partial analytical results for the asymptotic error. Numerical evidence and previous results involving PPT constraints [49] lead us to conjecture an asymptotic scaling of the form

$$
|\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}} - \tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N}| \leq f(L, d_{\mathrm{ES}})\, O\left(\frac{1}{N^2}\right),
\tag{18}
$$

where $f(L, d_{\mathrm{ES}})$ is a function of L and d_{E}. We leave the analysis of these asymptotic error bounds to a future investigation.

5 Numerical results

In this section, we discuss the numerical results we obtained for $\tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N}$. The SDPs were run with CVXPY [2, 23] using the solver SCS [50, 51], on a compute server with an Intel Xeon Gold 5218 24-core processor at 2.294 GHz, and with 128 GB of RAM. For $d_{\mathrm{E}} = 1$, optimization time was in the order of seconds, whereas $d_{\mathrm{E}} = 2$ scenarios required from 4 up to 14 hours to complete.

For the explicit computation, we chose the following measurement protocol: $\mathcal{A} = \{0, 1\}$, $\rho_{\mathrm{E}}^0 = |0\rangle\langle 0|_E$, $d_{\mathrm{S}} = 2$, $\rho_{\mathrm{S}}^0 = |0\rangle\langle 0|_S$ and $E_S^a = |a\rangle\langle a|_S$, so that

$$
\mathcal{M}_a(\rho_{\mathrm{ES}}) = \mathrm{tr}\left[\rho_{\mathrm{ES}} \cdot \mathbb{1}_E \otimes |a\rangle\langle a|_S\right] \otimes |0\rangle\langle 0|_S.
\tag{19}
$$

For this particular choice, the bounds $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$ are the same as for the isolated system case studied in [67], where explicit time evolutions were found through gradient descent techniques.

The analytical maxima for $d_{\mathrm{E}} = 1$ described in Sec. 3 served as a useful test for the soundness of our approach. We ran the SDP for all binary sequences

		Formulation of the problem	Mathematical representation	
Exact	Not SDP	Original problem	$\max_U\ p(\boldsymbol{a}|d_{\mathrm{E}})$	Intractable
		Choi-Jamiołkowski representation	$\max_{C_U}\ \mathrm{tr}\left[X^{\mathrm{T}}(C_U)^{\otimes L}\right]$	
		Convex relaxation	$\max_{C_U^L}\ \mathrm{tr}\left[X^{\mathrm{T}}C_U^L\right]\qquad C_U^L=\sum_i p_i(C_U^{(i)})^{\otimes L}$	
		Symmetric separable states relaxation	$\max_{\Phi_{A_1^L}}\ \mathrm{tr}\left[X^{\mathrm{T}}\Phi_{A_1^L}\right]\qquad \Phi_{A_1^L}\in\mathcal{P}_L^+\cap\mathrm{SEP}_L$	
Outer Approximation	SDP	Separable approximation (Quantum de Finetti)	$\max_{\Phi_{A_1^N}}\ \mathrm{tr}\left[(X^{\mathrm{T}}\otimes\mathbb{1}_{A_{L+1}^N})\Phi_{A_1^N}\right]\qquad \Phi_{A_1^N}\in\mathcal{P}_N^+$	Intractable
		Software implementation		
		Symmetric representation	$\max_\phi\ \mathrm{tr}\left[x\phi\right]\qquad\qquad\qquad \phi\in\mathcal{S}_N$	Tractable
		Sparse implementation	$\max_\phi\ \mathrm{tr}\left[x\phi\right]\qquad \phi_{\boldsymbol{t},\boldsymbol{t}'}=0\ \forall\ (\boldsymbol{t},\boldsymbol{t}')\notin\theta^*$	
		Numerical results		

Figure 3: Schematic of all steps undertaken for computing an upper bound for $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$ by formulating and solving the SDP problem. Detailed descriptions of each step are covered in Apps. A and B. The tractable/intractable labels, on the right, refer to the case $d_{\mathrm{E}}=2$.

L	N	$\boldsymbol{a}$	Without PPT	With PPT	$\omega_1^{\boldsymbol{a}}$
2	2	00	0.999996	0.999999	1
2	3	00	1.000000	1.000000	1
2	2	01	0.500000	0.250002	1/4
2	3	01	0.250007	0.250001	1/4
3	3	000	1.000010	1.000000	1
3	4	000	1.000000	0.999991	1
3	3	001	0.250005	0.148149	4/27
3	4	001	0.148148	0.148148	4/27
3	3	010	0.250005	0.148149	4/27
3	4	010	0.148148	0.148148	4/27
3	3	011	0.250005	0.148149	4/27
3	4	011	0.148148	0.148148	4/27

Table 1: Results of the SDP for $d_{\mathrm{E}}=1$, compared with the analytical maxima $\omega_1^{\boldsymbol{a}}$. Note that $4/27=0.\overline{148}$. It can be seen that either PPT constraints or an extension of one extra system (i.e. $N=L+1$) was sufficient to achieve the analytical maximum in this case. Similar results were obtained for $L=4$, omitted here for conciseness.

L	N	$\boldsymbol{a}$	Upper bound (SDP)	Achievable value (GD)
3	3	001	0.683477	0.437341
3	4	001	0.521219	0.437341
4	4	0010	0.512220	0.437142
4	4	0011	0.487058	0.362047
4	4	0100	0.494499	0.333147
4	4	0110	0.488837	0.361968
4	4	0001	0.492088	0.300545

Table 2: Comparison between SDP upper bounds and best achievable values known for the maximum $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$, for $d_{\mathrm{E}}=2$. Best known values were obtained through gradient descent (GD) [67] and act as lower bounds on the maximum, but are not suitable as witnesses. All results are in agreement, and for $\boldsymbol{a}=001$, we observe convergence of the bounds as N increases.

(up to $0\leftrightarrow1$ relabeling symmetry), for $L=2,3,4$, $N=L,L+1$. Every case was tested with and without the additional PPT constraints for comparison. Results for $L=2,3$ are shown in Table 1, in which it can be observed that, for $d_{\mathrm{E}}=1$, either the PPT constraints or a symmetric extension of a single extra system appear to be sufficient for achieving the exact analytical maximum.

We have also solved the SDP for the case $d_{\mathrm{S}}=d_{\mathrm{E}}=2$, for various sequences without a trivial maximum probability (i.e., $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}<1$), but only without the additional PPT constraint, as the requirements needed for its addition would have exceeded the available memory. Additionally, since the involved SDPs are only numerically tractable for $N\approx L$, and no convergence was observed for the values we managed to compute, we can only claim to have obtained upper bounds for the maximum $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$, as in Eq. (16).

As our choice of probe state and measurements (Eq. (19)) for $d_{\mathrm{S}}=d_{\mathrm{E}}=2$ reproduces the quan-

tum scenarios previously investigated in [67], as we show in Sec. 7.3, we can compare our upper bounds against the largest known achievable values, obtained through explicit time-evolutions found via gradient descent techniques. This allows us to estimate the range of values containing the maxima $\omega_2^{\boldsymbol{a}}$. These comparisons are shown in Table 2, where, for instance, we see $0.437341\leq\omega_2^{001}\leq0.521219$. As a concrete example of a violation of this bound, for $d_{\mathrm{S}}=2$ and $d_{\mathrm{E}}=3$, we may construct a unitary in $\mathcal{H}_{\mathrm{ES}}$ as follows:

$$U=|1\rangle\langle0|_E\otimes|0\rangle\langle0|_S+|2\rangle\langle1|_E\otimes|0\rangle\langle0|_S+|2\rangle\langle0|_E\otimes|1\rangle\langle1|_S$$
$$+|1\rangle\langle1|_E\otimes|1\rangle\langle1|_S+|0\rangle\langle2|_E\otimes|0\rangle\langle1|_S+|0\rangle\langle2|_E\otimes|1\rangle\langle0|_S.\tag{20}$$

For the ρ_{E}^0, ρ_{S}^0 and $E_S^{\boldsymbol{a}}$ we have chosen, the above unitary gives $p(001|d_{\mathrm{E}}=3)=1>\omega_2^{001}$, implying that the bound found for $d_{\mathrm{E}}=2$ is not only non-trivial, but actually witnesses environment dimension since it can indeed be violated by means of a larger environment ($d_{\mathrm{E}}=3$ in this case). In fact, this example certifies that $\omega_3^{001}=1$; see also Sec. 7.3.

6 Implementation

This section outlines the concrete steps we have taken to relax the original problem to a hierarchy of SDPs and to make these SDPs numerically tractable; see App. B for more details. As noted previously, a straightforward implementation is not computationally tractable, even for short sequences, due to the large number of variables and constraints involved.

We have addressed these additional numerical challenges by exploiting several properties of the problem. Firstly, the symmetric constraint imposed on $\Phi_{A_1^N}$ can be satisfied automatically by expressing $\Phi_{A_1^N}$ in terms of a basis for the symmetric subspace in the numerical implementation of the SDP. While this provides a significant reduction in the total number of involved variables and constraints, it is by itself insufficient to make the problem tractable.

However, thanks to our specific choice of initial states and measurements, the sparsity of the objective function could be exploited to further reduce the number of variables and constraints, by eliminating all of those that do not affect the objective, either directly or indirectly. This relaxation generally results in a sparse outer approximation of the original problem, i.e., to a value $\Omega_{d_E}^{a,N} \geq \tilde{\omega}_{d_E}^{a,N}$. In our particular case, as we show in App. B.2, this elimination procedure is exact, thus yielding an optimal value $\Omega_{d_E}^{a,N} = \tilde{\omega}_{d_E}^{a,N}$.

6.1 Choice of parameters

Before a concrete implementation, one must first choose the initial environment state ρ_E^0, as well as fixing the probe state and measurements. As the objective function is convex on ρ_E^0, and the optimization is over all unitaries, we may – without loss of generality – fix a pure initial state $\rho_E^0 = |0\rangle\langle 0|_E$ and eliminate any assumptions on the environment state; see Sec. 7.2. On the other hand, as assumptions on d_S, ρ_S^0 and $(E_S^a)_a$ are experiment-dependent, one must search for "proper" probe states and measurements on a case-by-case basis, but optimal choices, leading to a larger bound, can also be addressed in general terms; see Sec. 7.3. For the choice of sequence, in particular, if it is too simple relative to d_E, the maximum attainable probability may be trivial (i.e., equal to one), such that no optimization is required, and no witness can be constructed. Therefore, it is important to choose sequences which have a non-trivial maximum for a given d_E. For closed systems, this problem has been solved via the notion of "deterministic complexity" [67], which defines the minimum requirements for a sequence to be able to occur deterministically. In the open system case, the conditions for determinism involve not only the available memory, as was the case in the isolated system, but also the dimension of the system. We elaborate on these conditions in Sec. 7.3.

6.2 Symmetric representation

Effectively solving any of the SDPs in the hierarchy requires a significant amount of optimization, as a naive implementation quickly becomes computationally intractable. As a rough example, without any simplifications, $\Phi_{A_1^N}$ is a square matrix of size $(d_{ES}^2)^N$, which in the simplest non-trivial scenario, $d_E = d_S = 2$ and $N = L = 3$, already results in a 4096×4096 matrix, with over 16 million complex scalar variables. The partial trace constraint alone involves over 1 million linear equations between these variables, making the SDP numerically intractable in this naïve formulation.

In practice, several simplifications can be made; see App. B for full details. $\Phi_{A_1^N}$ and $X \otimes \mathbb{1}_{A_{L+1}^N}$ may be written directly in terms of an operator basis for the symmetric subspace, thus automatically satisfying the symmetry constraint, which significantly reduces the number of variables. Defining an isometry between the symmetric subspace of A_1^N, denoted by $\mathcal{S}_N$ (with dimension $\dim \mathcal{S}_N$), and a canonical basis for $\mathcal{S}_N$,

$$V = \sum_{\boldsymbol{t}} |\boldsymbol{t}\rangle_{\mathcal{S}_N} \langle \mathrm{sym}(\boldsymbol{t})|_{A_1^N}, \qquad (21)$$

we may cast all involved objects directly as $\dim \mathcal{S}_N \times \dim \mathcal{S}_N$ matrices:

$$\begin{aligned} \Phi_{A_1^N} &= \sum_{\boldsymbol{t},\boldsymbol{t}'} \phi_{\boldsymbol{t},\boldsymbol{t}'} |\mathrm{sym}(\boldsymbol{t})\rangle\langle\mathrm{sym}(\boldsymbol{t}')|_{A_1^N}, \\ \phi &= V \Phi_{A_1^N} V^\dagger, \text{ and} \\ x &= V(X^T \otimes \mathbb{1}_{A_{L+1}^N}) V^\dagger, \end{aligned} \qquad (22)$$

where $\boldsymbol{t}$ denotes a "type" for the canonical representation of symmetric states $|\mathrm{sym}(\boldsymbol{t})\rangle_{A_1^N}$; see [34]. Under this representation, the objective function becomes $\mathrm{tr}[x\phi]$, and the constraints may be formulated in terms of $\phi_{\boldsymbol{t},\boldsymbol{t}'}$ directly, without resorting to the the full space A_1^N. Note that $VV^\dagger = \mathbb{1}_{\mathcal{S}_N}$ but $V^\dagger V \neq \mathbb{1}_{A_1^N}$ (rather, $V^\dagger V$ is a projector onto $\mathcal{S}_N$). As the dimension of $\mathcal{S}_N$ is [34]

$$\dim \mathcal{S}_N = \binom{N + d_{ES}^2 - 1}{N}, \qquad (23)$$

for $d_{ES} = 4$ and $N = 3$ the matrix ϕ is of size 816×816, with around 600 thousand variables: a reduction to 4% of the original 16 million. With further work, the partial trace constraints can also be efficiently written directly in this representation, leading to $(d_{ES} \cdot \binom{(N-1)+d_{ES}^2-1}{N-1})^2$ equality constraints, resulting in approximately 300 thousand constraints in the $d_{ES} = 4$ and $N = 3$ scenario: a reduction to 28% of the original. This symmetric representation is explained in more detail in App. B.1.

6.3 Sparse implementation

While these are significant improvements, they still prove to be insufficient, as the dense representation

N		Symmetric	Sparse	Reduction to
3	**Variables**	665 856	3 566	0.54%
	Constraints	295 937	2 809	0.95%
4	**Variables**	15 023 376	35 688	0.24%
	Constraints	10 653 697	40 441	0.38%

Table 3: Comparison between the number of variables and linear constraints in the symmetric problem vs. its sparse implementation, for the sequence $\boldsymbol{a} = 001$. Our algorithm achieves a vast reduction in the number of variables and constraints, while still allowing exact solutions for our problem. Note that these numbers precede any further optimization, which could remove any remaining redundant constraints or variables.

of the symmetric problem involves too many variables and constraints to be solvable. As a concrete example, for the simplest case of $\boldsymbol{a} = 001$, $d_{\mathrm{E}} = 2$ and $L = N = 3$, the SDP solver SCS attempts to allocate a dense array of over 3 TB in size.

To address this, we have developed an algorithm that exploits any sparsity of the objective function, and – whenever possible – constructs a sparse relaxation of the original problem. This is achieved by iteratively selecting which variables of ϕ and which of its constraints are strictly necessary to solve the problem, due to their direct or indirect influence in the objective function. While generally a relaxation, due to the explicit structure of our problem this sparse implementation is exact, yielding $\tilde{\omega}_{d_{\mathrm{E}}}^{\boldsymbol{a},N}$. A detailed explanation of our algorithm is available in App. B.2.

For the state and measurements considered, our technique was capable of reducing the number of variables and constraints immensely, to less than 1% of the symmetric case (see Table 3). This allowed us to successfully compute upper bounds for various sequences up to $N = 4$ and $d_{\mathrm{E}} = 2$ (Table 2).

As can be observed in Table 3, the sparse problems can clearly be simplified further, by eliminating redundant variables and constraints. We opted for leaving such task to the numerical pre-solver, as this only took a few minutes of computing time. Solving the SDP for $N = 3$ was possible within minutes, but for $N = 4$, from 4 up to 14 hours were needed, depending on the sequence.

7 Discussion

7.1 Repeated unitaries

In the following, we discuss in which cases the condition of repeated unitaries is physically justified. To see this, imagine that the evolution of the system is governed by a time-independent Hamiltonian H_{SE}. The corresponding unitaries will be of the form $U(t) = e^{-iHt}$. This assumption of time-independence is always possible, as the environment may include any source of time dependence. Ideally, then, we would

have the same unitary if we perform the measure-and-prepare operations equally spaced in time, say, by an interval Δt. One should take into account, however, some uncertainty in the time measurement. Let us then assume that our choice of time for performing a measurement is distributed according to a distribution $q(t)$, centered around Δt. This means that instead of transforming our system according to the unitary $U(\Delta t)$, we transform it, on average, according to the mapping $\Lambda := \int \mathcal{U}(t) q(t) dt$, seemingly breaking our assumption of repeated unitaries. Since Λ is a valid completely positive trace preserving (CPTP) map, it can be dilated by means of a larger environment into a unitary $\mathcal{U}'$, namely $\Lambda(\rho_{\mathrm{SE}}) = \mathrm{tr}_{\mathrm{E}_1}[\mathcal{U}'(\rho_{\mathrm{SE}} \otimes |0\rangle\langle 0|_{\mathrm{E}_1})]$. To complete the argument that this situation can still be described by repeated unitaries, it is enough to show that the same can be done for multiple copies of it, namely, that the repeated operation Λ^L can be dilated to some unitary $\widetilde{\mathcal{U}}^L$. To do so, it is enough to provide at each time-step i a new environment E_i prepared in the correct initial state $|0\rangle\langle 0|_{\mathrm{E}_i}$. Let us compute explicitly the case for two time steps, the general case is straightforward. We define $W_{\mathrm{E}_1\mathrm{E}_2}$ the swap operator between systems E_1 and E_2, and we set $\widetilde{\mathcal{U}} := W_{\mathrm{E}_1\mathrm{E}_2} \circ \mathcal{U}'_{\mathrm{SEE}_1}$. Let us calculate, for simplicity

$$
\begin{aligned}
\mathrm{tr}_{\mathrm{E}_1\mathrm{E}_2}&[\mathcal{U}'_{\mathrm{SEE}_1} \circ W_{\mathrm{E}_1\mathrm{E}_2} \circ \mathcal{U}'_{\mathrm{SEE}_1}(\rho_{\mathrm{SE}} \otimes |0\rangle\langle 0|_{\mathrm{E}_1} \otimes |0\rangle\langle 0|_{\mathrm{E}_2})] \\
&= \mathrm{tr}_{\mathrm{E}_1\mathrm{E}_2}[\mathcal{U}'_{\mathrm{SEE}_1} \circ W_{\mathrm{E}_1\mathrm{E}_2}(\sigma_{\mathrm{SEE}_1} \otimes |0\rangle\langle 0|_{\mathrm{E}_2})] \\
&= \mathrm{tr}_{\mathrm{E}_1}[\mathcal{U}'_{\mathrm{SEE}_1}(\mathrm{tr}_{\mathrm{E}_2}[\sigma_{\mathrm{SEE}_2}] \otimes |0\rangle\langle 0|_{\mathrm{E}_1})] \\
&= \mathrm{tr}_{\mathrm{E}_1}[\mathcal{U}'_{\mathrm{SEE}_1}(\Lambda(\rho_{\mathrm{SE}}) \otimes |0\rangle\langle 0|_{\mathrm{E}_1})] = \Lambda^2(\rho_{\mathrm{SE}}),
\end{aligned}
$$
(24)

where $\sigma_{\mathrm{SEE}_1} := \mathcal{U}'_{\mathrm{SEE}_1}(\rho_{\mathrm{SE}} \otimes |0\rangle\langle 0|_{\mathrm{E}_1})$. The effect of the final $W_{\mathrm{E}_1\mathrm{E}_2}$ operation is to swap the space E_1 and E_2 that are irrelevant after the operations on SE have been performed. Evidently, this argument can straightforwardly be extended to more time steps. In summary, this means that the condition of "the same unitary" is satisfied as long as the time choice is always drawn from the same distribution. In other words, it is sufficient to always "probabilistically repeat" the same operation.

Finally, we remark that even if this procedure may use a very large environment, we are only interested in showing that we can always assume there *is* a unitary dynamics, with the notion of an effective environment, then, taking care of estimating the environment consistent with the observed statistics. Moreover, we also remark that time-independent operations are necessary, as unrestricted time-dependent operations can achieve arbitrarily long temporal correlations even with bounded memory size [46].

7.2 Effective environment and initial state

The environment of a quantum system, intended as all the physical systems surrounding and pos-

sibly interacting with it, is typically a very high-dimensional system, if not directly assumed to be infinite-dimensional. From this perspective, we want to make sense of the notion of *effective* environment. Consider a global transformation of the system and environment. As a first approximation, we can say that the global unitary is of the form $U_{\mathrm{SE}} \otimes U_{\mathrm{E}'}$, where U_{SE} is an entangling unitary between the system and the effective environment and $U_{\mathrm{E}'}$ is acting on the rest of the environment. At the same time, an evolution of the form $U_{\mathrm{SE}} \otimes U_{\mathrm{E}'}$ is just an approximation of the full evolution of the environment, as we expect its state to thermalize after some time interval. Nevertheless, from a physical perspective this approximation is still valid if the time required for a single run of the experiment, i.e., the measurement of the temporal sequence, is much shorter than the time needed for the effective environment to thermalize. This is to be expected, as the environment is composed of many particles that may interact with each other with different strengths. Underlying this expectation is the assumption – known as Markovian embedding [10, 71, 72] and frequently employed in the description and modeling of open quantum system dynamics [43–45, 63] – that the environment can always be split into two parts: a far environment, leading to irretrievable, memoryless information loss, and an effective environment, that can transport memory. Due to the irretrievable information loss, the dynamics of the system and the effective environment is then described by a general (non-unitary) CPTP map, a situation which, as discussed in the previous section, can again be dilated to repeated unitaries.

Reversing this perspective, namely, looking at the problem of characterizing the effective environment from temporal correlation experiments, we may say that this characterization is inherently dependent on the typical time-scales of a single experimental run. This difference in time scales and the eventual thermalization of the environment, however, are essential for repeating the experiment for collecting statistics, as it is required that the initial state of the environment is always (probabilistically) described by the same state ρ_{E}^0 in each experimental run. As previously mentioned, this is typically a thermal state, but it does not need to be characterized in our approach. In fact, since the SDP maximizes the temporal correlations, which are linear in the initial state of the environment, we know that the maximum is always achieved with a pure state. Up to local unitaries, we can thus always assume it to be $\rho_{\mathrm{E}}^0 = |0\rangle\!\langle 0|_E$. The bound calculated for this state is, then, valid for any possible initial state of the environment.

Finally, somewhat independent of the explicit experimental situation, we may see our setup as a question of simulation resources: What is the smallest dimension an environment coupling to a known system must have in order to reproduce observed statistics
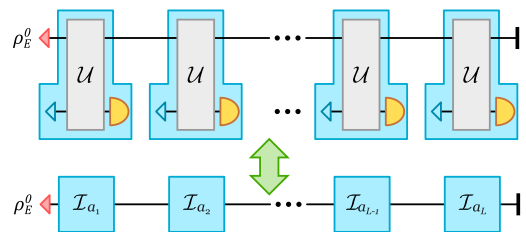

Figure 4: The open system and environment scenario can be equivalently interpreted as the dilation of the scenario involving sequential measurements $\mathcal{I}_a$ on an isolated environment system.

in a unitary way? Seen in this way, the results we present attribute a "simulation hardness" (in the sense of required environment dimension) to each sequence, that is agnostic with respect to concrete time scales or experimental limitations, but rather inherent to the respective sequence.

7.3 Conditions for deterministic realizations

The maximum probability for a given sequence increases as more memory is available, i.e.,

$$0 \leq \omega_1^{\boldsymbol{a}} \leq \omega_2^{\boldsymbol{a}} \leq \cdots \leq \omega_{d_{\mathrm{E}}}^{\boldsymbol{a}} \leq 1, \qquad (25)$$

as larger environments can always simulate the dynamics of smaller ones. Therefore, if the sequence $\boldsymbol{a}$ is sufficiently simple, its maximum may be trivial for a given d_{E}, i.e., $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}} = 1$.

The maxima $\omega_{d_{\mathrm{E}}}^{\boldsymbol{a}}$ depend on the sequence, as not all sequences are equally "difficult" to produce with a given amount of memory d_{E}. For example, it is easy to see that the sequence '000' can *always* be produced with unit probability, while the sequence '001' cannot be produced with unit probability if the environment is not of sufficient size [67]. This observation suggests a relevant notion of "complexity" of sequences, which offers a more fundamental relationship between $\boldsymbol{a}$ and d_{E}. Since our goal is to establish non-trivial maxima on the probabilities of sequences, the natural question to ask is how large should d_{E} be such that $\boldsymbol{a}$ can occur deterministically?

Understanding the conditions where this occurs allows us to pick only scenarios featuring non-trivial maxima. It is useful to recall the following

Definition. (Deterministic Complexity [67]) Let $\boldsymbol{a}$ be a sequence of measurement outcomes, produced by repeated identical measurements on an isolated d-dimensional system. The *deterministic complexity* of the sequence $\boldsymbol{a}$, or $\mathrm{DC}(\boldsymbol{a})$, is the minimum d such that $p(\boldsymbol{a}|d) = 1$ can occur.

$\mathrm{DC}(\boldsymbol{a})$ is a property of the sequences $\boldsymbol{a}$ and it is independent of the model (quantum or classical). First, we establish a correspondence between the single system and open system scenarios. Let $d_{\mathrm{S}} = |\mathcal{A}|$, so that

we may choose $E_S^a = |a\rangle\langle a|_S$ and $\rho_S^0 = |0\rangle\langle 0|_S$. Then, the changes in the environment state due to the unitary and measurements can be written succinctly in terms of the completely positive maps

$$\mathcal{I}_a(\rho_E) = \text{tr}_S\left[U(\rho_E \otimes |0\rangle\langle 0|_S)U^\dagger \cdot \mathbb{1}_E \otimes |a\rangle\langle a|_S\right]. \quad (26)$$

Therefore, under the action of $(\mathcal{I}_a)_{a\in\mathcal{A}}$, we may interpret the *environment* as a single isolated "memory" system in which the maps $\mathcal{I}_a$ act sequentially, as in [67]. The converse mapping, from the sequential measurements on an isolated memory system to the open system scenario, is the dilation map with the probe system acting as the ancilla and the environment as the memory. This construction is well known, but it is worth expanding it in detail in order to understand what are the conditions on the ancilla (i.e., the probe system) needed to implement it.

Let $n(\boldsymbol{a})$ denote the number of unique symbols appearing in $\boldsymbol{a}$, with $n(\boldsymbol{a}) \leq |\mathcal{A}|$, and let us assume $d_S \geq n(\boldsymbol{a})$. Now, consider an instrument given by the maps $\tilde{\mathcal{I}}_a$ of the form $\tilde{\mathcal{I}}_a(\rho_E) = K_a \rho_E K_a^\dagger$, for $(K_a)_{a\in\mathcal{A}}$ Kraus operators, i.e., $\sum_{a\in\mathcal{A}} K_a^\dagger K_a = \mathbb{1}_E$. The corresponding unitary arises from the $d_E d_S \times d_E$ isometry matrix

$$Q = \sum_{a\in\mathcal{A}} K_a \otimes |a\rangle_S, \quad (27)$$

which can always be completed into a unitary matrix U. Thus $\tilde{\mathcal{I}}_a$ can be written in the form of Eq. (26). As we have chosen $d_S = |\mathcal{A}|$, $\rho_S^0 = |0\rangle\langle 0|_S$ and $E_S^a = |a\rangle\langle a|_S$ in our implementation of the SDP (Sec. 6), there is a direct correspondence between the two scenarios; see Fig. 4. Therefore, the upper bounds obtained by the SDP in Eq. (15) can be compared with the known achievable values from [67], as shown in Table 2 and discussed in Sec. 5.

This construction applies to deterministic models as they have deterministic transitions between states, i.e., each is described by a single Kraus operator per outcome. We thus have

Proposition 1. *If both $d_E \geq \text{DC}(\boldsymbol{a})$ and $d_S \geq n(\boldsymbol{a})$, then there is a choice of probe state and measurements such that $\omega_{d_E}^{\boldsymbol{a}} = 1$.*

Note that the condition $d_S \geq n(\boldsymbol{a})$ is strictly required for deterministic production of $\boldsymbol{a}$. In fact, if $p(\boldsymbol{a}|d_E) = 1$, every symbol must occur deterministically. Then, at each step the system must be in a state ρ_S^a such that $\text{tr}\left[\rho_S^a E_S^a\right] = 1$. Thus, the measurements E_S^a must be able to perfectly discriminate between the states $\{\rho_S^a\}_a$, which is possible only if $d_S \geq n(\boldsymbol{a})$. We have established:

Proposition 2. *If $d_S < n(\boldsymbol{a})$, then $\omega_{d_E}^{\boldsymbol{a}} < 1$ for any choice of probe state or measurements, and any environment dimension d_E.*

In conclusion, these results tells us that nontrivial bounds appear for $d_E < \text{DC}(\boldsymbol{a})$, that we can compare

these bound with the single-system scenario for $d_S \geq n(\boldsymbol{a})$, and that $|\mathcal{A}| = 2$ and $L = 3$ is the smallest scenario displaying non-trivial memory effects.

7.4 Choice of sequence

As briefly noted in Section 7.3, in order to construct a witness for $d_E > d$, it is vital to choose a sequence $\boldsymbol{a}$ with sufficient deterministic complexity. If the sequence chosen is too short or too simple (i.e., if $\text{DC}(\boldsymbol{a}) \leq d$), then the bound $\omega_d^{\boldsymbol{a}}$ is trivial and the dimension d_E cannot be tested with such sequence.

However, due to the difficulty in computing these bounds, and the fact longer sequences generally become less probable—and therefore less likely to violate the witness inequality—generally one should choose sequences to be as short as possible while having deterministic complexity of the same size as the minimum environment dimension one wishes to witness.

As a concrete example, if we wish to witness $d_E > 3$, we should choose a sequence $\boldsymbol{a}$ with deterministic complexity $\text{DC}(\boldsymbol{a}) = 4$, e.g., $\boldsymbol{a} = 0001$, which is the shortest sequence satisfying this requirement. Then, a violation of the bound ω_3^{0001} informs us that, in fact, $d_E \geq 4$.

8 Conclusions and outlook

We presented a method to lower bound the dimension of the environment interacting with a probe system, based only on the statistics of measurements performed on the (partially characterized) system, and without any assumption on the environment or the dynamics. This is achieved via a hierarchy of semidefinite programs that upper bound temporal correlations achievable in various experimental scenarios, under the assumption of finite memory. Such bounds can be applied to the detection of the effective environment size in the dynamics of open systems, as well as a certification of the minimum size of an environment's dimension compatible with observations.

To keep the discussion simpler, we applied the optimization for a single sequence. It is straightforward to adapt the objective function to arbitrary linear functions of the full probability distribution $(p(\boldsymbol{a}|d_E))_{\boldsymbol{a}}$, as those appearing in the temporal inequalities derived in, e.g., [12, 35, 46, 61]. This may, in principle, lead to better witnesses. We leave the numerical explorations of this problem to a future investigation.

We assumed a joint unitary evolution between system and environment, which leads to a CJ representation of these maps given in terms of symmetric states. As explained in Sec. 7.1, this assumption can usually be justified on physical grounds. Nevertheless, a natural question to ask is how do our results change if we consider arbitrary CPTP maps? In that case, the SDP should be modified by replacing the symmetry constraint with permutation invariance, i.e.,

$V_\sigma \Phi_{A_1^N} V_\sigma^\dagger = \Phi_{A_1^N}$ for all permutations $\sigma \in \mathfrak{S}_N$; see Eq. (35). We expect that bounds for CPTP maps may be larger than those for unitary channels of the same dimension, especially for objective functions involving more than a single sequence, but numerical optimization is significantly more costly in this case, as the permutation invariant operator basis is of size [69, Ch. 7] $\binom{N+(d_{ES})^4-1}{N}$, in contrast to $\binom{N+(d_{ES})^2-1}{N}^2$ in the symmetric case. While it is also possible to write such CPTP maps in terms a dilation of the environment, this approach would likely result in more extraneous variables in the SDP, i.e., the terms in the subspace orthogonal to the dilation ancilla's initial state, possibly rendering the optimization intractable.

Additionally, the SDP in Eq. (15) and its related implementation techniques are, in fact, quite general, and can be applied to a wide range of scenarios beyond what we have considered here. As the operator X from Eq. (10) is fixed, any choice of intermediate operations between each unitary could be chosen. E.g., the initial system and environment states could be correlated, and the intermediate maps $\mathcal{M}_a$ could each be replaced by arbitrary joint operations, even time-dependent ones. Such approaches could then, for example, be used to bound observables for specific types of processes, e.g., quantum processes with only classical memory [30], by assuming additional entanglement breaking channels on the environment. Therefore, provided the problem is numerically tractable, our techniques are independent of what explicit measurements are chosen.

While the high dimensionality of the current formulation of the SDP quickly renders general numerical implementations intractable, our approach still offers new avenues for the subject of bounding temporal correlations, and their relationship to open-system dynamics. Ultimately, the techniques developed herein should be taken as a proof-of-concept for future developments and improvements. It remains to be seen whether more efficient numerical techniques, or even alternative outer approximations, are better suited for addressing such problems. Further investigation of these avenues is subject to future work.

Nevertheless, the success of our approach highlights the wealth of information contained in temporal correlations and the potential of new techniques for characterizing large complex systems by means of a small probe alone, by exploiting non-trivial properties of the temporal correlations achieved by systems of bounded size.

9 Acknowledgments

A Constructing the SDP relaxation

This appendix explains in detail the various steps to analytically formulate the maximization problem $\omega_{d_E}^{\boldsymbol{a}} := \max_U p(\boldsymbol{a}|d_E)$ as a semidefinite program, with the next appendix (App. B) focusing on its software implementation. Figure 5 provides a schematic outline of our approach.

		Formulation of the problem	Mathematical representation	
Exact	Not SDP	Original problem	$\max_U\ p(\boldsymbol{a}\|d_E)$	
		Choi-Jamiołkowski representation	$\max_{C_U} \operatorname{tr}\left[X^{\mathrm{T}}(C_U)^{\otimes L}\right]$	
		Convex relaxation	$\max_{C_U^L} \operatorname{tr}\left[X^{\mathrm{T}}C_U^L\right] \qquad C_U^L = \sum_i\ p_i(C_U^{(i)})^{\otimes L}$	
		Symmetric separable states relaxation	$\max_{\Phi_{A_1^L}} \operatorname{tr}\left[X^{\mathrm{T}}\Phi_{A_1^L}\right] \qquad \Phi_{A_1^L} \in \mathcal{P}_L^+ \cap \mathrm{SEP}_L$	
Outer Approximation	SDP	Separable approximation (Quantum de Finetti)	$\max_{\Phi_{A_1^N}} \operatorname{tr}\left[(X^{\mathrm{T}} \otimes \mathbb{1}_{A_{L+1}^N})\Phi_{A_1^N}\right] \qquad \Phi_{A_1^N} \in \mathcal{P}_N^+$	Intractable
		Software implementation		
		Symmetric representation	$\max_\phi \operatorname{tr}[x\phi] \qquad \phi \in \mathcal{S}_N$	
		Sparse implementation	$\max_\phi \operatorname{tr}[x\phi] \qquad \phi_{\boldsymbol{t},\boldsymbol{t}'} = 0 \ \forall\ (\boldsymbol{t},\boldsymbol{t}') \notin \theta^*$	Tractable
		Numerical results		

Figure 5: Schematic of all steps undertaken for computing an upper bound for $\omega_{d_E}^{\boldsymbol{a}}$ by formulating and solving the SDP problem. The tractable/intractable labels, on the right, refer to the case $d_E = 2$.

Let $\ell = 1, \dots, L$ enumerate the time steps, and let I_ℓ and O_ℓ be the input and output spaces of the ℓ-th unitary evolution $\mathcal{U}$, respectively. For convenience, we write $A_\ell := \mathrm{I}_\ell \otimes \mathrm{O}_\ell$ for the joint space at step ℓ, and $A_a^b := A_a \otimes A_{a+1} \otimes \cdots \otimes A_{b-1} \otimes A_b$ for the sequential spaces from a to b. We assume that all spaces are isomorphic, i.e., $\mathrm{I}_\ell \cong \mathrm{O}_\ell \cong \mathcal{H}_{\mathrm{ES}}$ for all ℓ, but we preserve labels for clarity. We can thus write the respective maps in the Choi-Jamiołkowski representation (see Eq. (3)) as:

$$C_U := \frac{1}{d_{ES}}\mathcal{C}(\mathcal{U}), \quad M_a := \mathcal{C}(\mathcal{M}_a), \quad \hat{M}_a := \mathrm{tr}_{\mathcal{O}}[M_a], \tag{28}$$

where $\mathcal{C}(\Lambda)$ is the Choi matrix of the map Λ and C_U is normalized such that it corresponds to a quantum state, which is useful during implementation of the symmetric representation of the problem (App. B.1). As before, we highlight that the local input ($\mathcal{I}$) and output ($\mathcal{O}$) spaces of these maps are interleaved: For the ℓ-th $\mathcal{U}$, $\mathcal{I} = \mathrm{I}_\ell$ and $\mathcal{O} = \mathrm{O}_\ell$, but for $\mathcal{M}_{a_\ell}$ we have $\mathcal{I} = \mathrm{O}_\ell$ and $\mathcal{O} = \mathrm{I}_{\ell+1}$. Since the final output state is discarded (i.e., we only are only concerned with the *probability* of outcome sequences), the final $\mathcal{O}$ is traced out, yielding $\hat{M}_a$ in the equation above. This structure of the spaces is illustrated in Fig. 6.
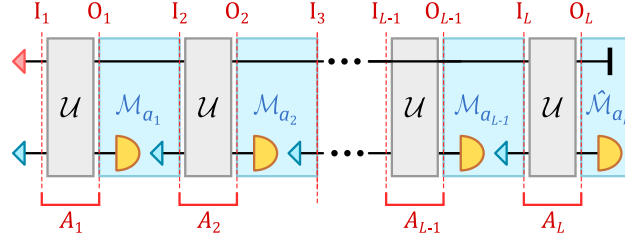

Figure 6: Diagram of the specific protocol discussed in this work, with each time step written in terms of distinct input and output Hilbert spaces, denoted by the dotted lines. Note that the input and output spaces of the $\mathcal{U}$ and $\mathcal{M}_a$ are interleaved: $\mathcal{U}$ has inputs I_ℓ and outputs I_ℓ, but for $\mathcal{M}_{a_\ell}$, inputs are O_ℓ and outputs $\mathrm{I}_{\ell+1}$.

We may now specify the probability of the sequence $\boldsymbol{a}$ as

$$p(\boldsymbol{a}|d_{\mathrm{E}}) = \mathrm{tr}\left[X^{\mathrm{T}}\left(\bigotimes_{\ell=1}^{L} C_U^{\mathrm{I}_\ell \mathrm{O}_\ell}\right)\right] = \mathrm{tr}\left[X^{\mathrm{T}}(C_U)^{\otimes L}\right] \tag{29}$$

$$X := (d_{\mathrm{ES}})^L \cdot (\rho_{\mathrm{E}}^0 \otimes \rho_{\mathrm{S}}^0)^{\mathrm{I}_1} \otimes M_{a_1}^{O_1 I_2} \otimes \cdots \otimes M_{a_{L-1}}^{O_{L-1} I_L} \otimes \hat{M}_{a_L}^{O_L}. \tag{30}$$

Where the correcting normalization factor $(d_{\mathrm{ES}})^L$ was incorporated into X (to make up for the normalization of C_U), and X^{T} denotes the transpose with respect to the basis chosen for the isomorphism. In order to obtain an upper-bound for $p(\boldsymbol{a}|d_{\mathrm{E}})$, our goal is to optimize Eq. (29) over all possible unitaries, in terms of C_U. As per the above definitions, it follows that C_U must satisfy the constraints:

- $C_U \geq 0$ and $\mathrm{rank}\, C_U = 1$, as it represents a unitary channel,

- $\mathrm{tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}$, as it is a trace preserving map

We may thus define our optimization problem as

Optimization Problem 1. *The initial formulation of the problem.*

$$\begin{aligned} \textbf{\textit{Given:}} \quad & \rho_E^0, \ \rho_S^0, \ \{M_a\}_a, \ \boldsymbol{a} \\ \textbf{\textit{Find:}} \quad & \max_{C_U} \ \mathrm{tr}\left[X^T(C_U)^{\otimes L}\right] \\ \textbf{\textit{Subject to:}} \quad & C_U \geq 0, \quad \mathrm{rank}\, C_U = 1, \quad \mathrm{tr}_{\mathcal{O}}[C_U] = \mathbb{1}_{\mathcal{I}}/d_{ES}. \end{aligned} \tag{31}$$

This may be immediately relaxed to a convex form, without affecting the maximum of the objective function, as

Optimization Problem 2. *The convex relaxation of the original problem.*

$$\begin{aligned} \textbf{\textit{Given:}} \quad & \rho_E^0, \ \rho_S^0, \ \{M_a\}_a, \ \boldsymbol{a} \\ \textbf{\textit{Find:}} \quad & \max_{\{|\phi_i\rangle\}_i} \ \mathrm{tr}\left[X^T\widetilde{\Phi}_{A_1^L}\right], \qquad \widetilde{\Phi}_{A_1^L} = \sum_i p_i \, |\phi_i\rangle\!\langle\phi_i|^{\otimes L}, \ p_i \geq 0, \ \sum_i p_i = 1 \\ \textbf{\textit{Subject to:}} \quad & \mathrm{tr}_{\mathcal{O}}[|\phi_i\rangle\!\langle\phi_i|] = \mathbb{1}_{\mathcal{I}}/d_{ES}, \quad |\phi_i\rangle \in \mathcal{H}_{ES} \otimes \mathcal{H}_{ES}. \end{aligned} \tag{32}$$

However, this optimization problem is non-linear, not only due to the tensor product in $|\phi_i\rangle\langle\phi_i|^{\otimes L}$, but also on the rank-1 nature of $|\phi_i\rangle\langle\phi_i|$. Let us now define our *target set* $\mathcal{T}$ as

$$\widetilde{\Phi}_{A_1^L} \in \mathcal{T} := \left\{ \sum_i p_i \, |\phi_i\rangle\langle\phi_i|^{\otimes L} \;\middle|\; |\phi_i\rangle \in \mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}}, \; p_i \geq 0, \; \sum_i p_i = 1, \; \mathrm{tr}_{\mathcal{O}}[|\phi_i\rangle\langle\phi_i|] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}} \right\}. \tag{33}$$

Elements of this set will not generally be rank-1, but the optimal solutions of Problems 1 and 2 belong to this set. Our goal now is to approach $\mathcal{T}$ by means of further relaxations, which is achieved by exploiting the symmetry of $\widetilde{\Phi}_{A_1^L}$.

A.1 The unitary channel constraints

The symmetric structure of $\widetilde{\Phi}_{A_1^L}$ requires all subspaces to be in the same local state. We can relax this by considering instead the separable set on L parties,

$$\Phi_{A_1^L} \in \mathrm{SEP}_L := \mathrm{conv}\left\{ \bigotimes_{\ell=1}^{L} Q_\ell \;\middle|\; Q_\ell \in \mathcal{B}(\mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}}), \; Q_\ell \geq 0, \; \mathrm{tr}\,[Q_\ell] = 1 \right\}, \tag{34}$$

such that all local spaces become independent. Here, conv denotes the convex hull of the set, and $\mathcal{B}$ the set of bounded operators. We must now find ways to restore rank-1 and permutation invariance constraints for the optima over this set. For now, we highlight that $\mathcal{T} \subset \mathrm{SEP}_L$, but simply switching from $\widetilde{\Phi}_{A_1^L} \in \mathcal{T}$ to $\Phi_{A_1^L} \in \mathrm{SEP}_L$ *does* change our problem, so that we must impose further constraints to the set SEP_L to restore the original problem in $\mathcal{T}$. The first step is restoring symmetry and optimality of rank-1 to $\Phi_{A_1^L}$. After this, we may restore the partial trace constraint.

Let $\sigma \in \mathfrak{S}_n$ be a permutation from the set of all permutations on n symbols $\mathfrak{S}_n$. We define

$$V_\sigma := \sum_{i_1,\cdots,i_n=1}^{d} |i_{\sigma^{-1}(1)},\cdots,i_{\sigma^{-1}(n)}\rangle\langle i_1,\cdots,i_n|, \tag{35}$$

$$P_n^+ := \frac{1}{n!} \sum_{\sigma \in \mathfrak{S}_n} V_\sigma \tag{36}$$

as, respectively, the permutation operator and projector onto the symmetric subspace. The symmetric subspace on n copies of $\mathcal{H}$ is defined as $\mathrm{Sym}_n(\mathcal{H}) := \{ |\psi\rangle \in \mathcal{H}^{\otimes n} \mid P_n^+ |\psi\rangle = |\psi\rangle \}$, and the space of symmetric operators is given by the set $\mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) := \{A \in \mathcal{B}(\mathcal{H}^{\otimes n}) \mid P_n^+ A = A\}$; see [34]. Note that, as ρ is Hermitian, the symmetry condition can be equivalently written either as $P_n^+ \rho = \rho$ or $P_n^+ \rho P_n^+ = \rho$. In fact, if $P_n^+ \rho = \rho$, then $\rho^\dagger = (P_n^+ \rho)^\dagger = \rho^\dagger P^+ = \rho P^+ = \rho \Rightarrow P_n^+ \rho P_n^+ = \rho$. Conversely, $\rho = P_n^+ \rho P_n^+ = (P_n^+)^2 \rho P_n^+ = P_n^+ \rho$. More details about the symmetric subspace can be found in App. B.1.

A.1.1 Ensuring symmetry and rank-1 at the optimum

By utilizing $\Phi_{A_1^L}$ as our variable as in Eq. (34), we have lost both symmetry and rank-1 guarantees at the optimum. However, both can be restored by ensuring $\Phi_{A_1^L}$ is in the symmetric subset of SEP_L; see, e.g., [39]. For completeness, we provide in the following an elementary proof that

$$\mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n = \left\{ \sum_i p_i \, |\phi_i\rangle\langle\phi_i|^{\otimes n} \;\middle|\; |\phi_i\rangle \in \mathcal{H}_{\mathrm{ES}} \otimes \mathcal{H}_{\mathrm{ES}}, \quad \langle\phi_i|\phi_i\rangle = 1, \quad p_i \geq 0, \quad \sum_i p_i = 1 \right\}, \tag{37}$$

which implies $\mathcal{T} \subset \mathrm{Sym}_L(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_L$. In fact, consider $\rho \in \mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n$. Since it is separable, we can write it as as a convex mixture of pure product states, i.e.,

$$\rho = \sum_i p_i \, |\Phi_i\rangle\langle\Phi_i|, \tag{38}$$

with $|\Phi_i\rangle$ a n-party pure product state, i.e., $|\Phi_i\rangle = \bigotimes_{j=1}^{n} |\phi_{i,j}\rangle$, with the $\{|\phi_{i,j}\rangle\}_j$ not necessarily equal for a given i. However, $\rho \in \mathrm{Sym}_n(\mathcal{B}(\mathcal{H}))$ implies $\mathrm{ran}(\rho) \subseteq \mathrm{Sym}_n(\mathcal{H})$, and if we prove that $\mathrm{ran}(|\Phi_i\rangle\langle\Phi_i|) \subseteq \mathrm{Sym}_n(\mathcal{H})$ for all i, then we find that $\{|\phi_{i,j}\rangle\}_j$ must be identical for each i, otherwise $|\Phi_i\rangle$ fails to be symmetric, namely, $|\Phi_i\rangle\langle\Phi_i| = |\phi_i\rangle\langle\phi_i|^{\otimes n}$ and, thus, $\rho = \sum_i p_i \, |\phi_i\rangle\langle\phi_i|^{\otimes n}$. It remains to prove that $\mathrm{ran}(|\Phi_i\rangle\langle\Phi_i|) \subseteq \mathrm{Sym}_n(\mathcal{H})$. To show this, we note that by Eq. (38) and positivity of ρ and $|\Phi_i\rangle\langle\Phi_i|$, we have $\mathrm{Ker}(\rho) = \cap_i \mathrm{Ker}(|\Phi_i\rangle\langle\Phi_i|)$. By Hermiticity, it follows that $\mathrm{ran}(\rho) = (\mathrm{Ker}(\rho))^\perp = \mathrm{span}(\cup_i \mathrm{ran}(|\Phi_i\rangle\langle\Phi_i|))$, which concludes the proof.

Importantly, Eq. (37), restores the symmetry and rank-1 properties for optimum solutions of Prob. 2 when $\Phi_{A_1^L}$ is restricted to $\mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n$. In light of this, in the following we use $\mathrm{SymSEP}_n := \mathrm{Sym}_n(\mathcal{B}(\mathcal{H})) \cap \mathrm{SEP}_n$.

A.1.2 The trace-preserving constraint

We now restore the trace-preserving constraint $\operatorname{tr}_{\mathcal{O}}[|\phi_i\rangle\langle\phi_i|] = \mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}$ for states expressed as in Eq. (37). This is established by proving that it is sufficient to satisfy this condition for a single party in the convex mixture.

Proposition 3. *For any* $\Phi_{A_1^L} \in \mathrm{SymSEP}_L$:

$$\operatorname{tr}_{\mathcal{O}_1}\Big[\Phi_{A_1^L}\Big] = \frac{\mathbb{1}_{\mathcal{I}_1}}{d_{ES}} \otimes \operatorname{tr}_{A_1}\Big[\Phi_{A_1^L}\Big] \quad\Longleftrightarrow\quad \operatorname{tr}_{\mathcal{O}_\ell}[|\phi_i\rangle\langle\phi_i|_{A_\ell}] = \frac{\mathbb{1}_{\mathcal{I}_\ell}}{d_{ES}} \otimes |\phi_i\rangle\langle\phi_i|^{\otimes L-1} \quad \forall \ell. \tag{39}$$

Proof. We provide a generalization of the arguments in [73, App. A]. From Eq. (37), if $\mathrm{P}_L^+\Phi_{A_1^L} = \Phi_{A_1^L}$, we know it admits a decomposition of the form

$$\Phi_{A_1^L} = \sum_i p_i \, |\phi_i\rangle\langle\phi_i|^{\otimes L}, \qquad p_i \geq 0, \quad \sum_i p_i = 1. \tag{40}$$

We introduce the one-party auxiliary map $\mathcal{E}(\cdot) = \operatorname{tr}_{\mathcal{O}}[\cdot] - \operatorname{tr}[\cdot]\,\mathbb{1}_{\mathcal{I}}/d_{\mathrm{ES}}$ implementing our partial trace constraint in $\mathcal{T}$, such that $(\mathcal{E}_{A_1} \otimes \mathrm{id}_{A_2} \otimes \mathrm{id}_{A_3^L})(\Phi_{A_1^L}) = 0$ for any $\Phi_{A_1^L} \in \mathrm{SymSEP}_L \cap \mathcal{T}$, and, due to symmetry, the same is true if the map had been applied to the second party instead. Defining $\tilde{\mathcal{E}}(\cdot) = [\mathcal{E}(\cdot)]^\dagger$, we may then write $(\mathcal{E}_{A_1} \otimes \tilde{\mathcal{E}}_{A_2} \otimes \mathrm{id}_{A_3^L})(\Phi_{A_1^L}) = 0$. As $\mathcal{E}$ acts on Hermitian operators and is Hermiticity-preserving, we have $\mathcal{E} = \tilde{\mathcal{E}}$ and the explicit distinction is only made for added clarity in the following proof. Applying the map to the convex mixture in Eq. (40), we obtain

$$\sum_i p_i \mathcal{E}_{A_1}(|\phi_i\rangle\langle\phi_i|_{A_1}) \otimes \tilde{\mathcal{E}}_{A_1}(|\phi_i\rangle\langle\phi_i|_{A_2}) \otimes |\phi_i\rangle\langle\phi_i|^{\otimes L-2} = \sum_i p_i E_i \otimes E_i^\dagger \otimes |\phi_i\rangle\langle\phi_i|^{\otimes L-2} = 0, \tag{41}$$

where $E_i = \mathcal{E}(|\phi_i\rangle\langle\phi_i|)$. We need this to be true for all E_i if we want to ensure $\Phi_{A_1^L}$ obeys our partial trace constraint globally in the convex mixture. To prove this is the case, let $G \in \mathbb{C}^{d_{\mathrm{ES}}^2 \times d_{\mathrm{ES}}^2}$, so that

$$\operatorname{tr}\left[(G \otimes G^\dagger \otimes \mathbb{1}_{A_3^L})\left(\sum_i p_i E_i \otimes E_i^\dagger \otimes |\phi_i\rangle\langle\phi_i|^{\otimes L-2}\right)\right] = \sum_i p_i |\operatorname{tr}[GE_i]|^2 \operatorname{tr}\left[|\phi_i\rangle\langle\phi_i|^{\otimes L-2}\right] = 0, \tag{42}$$

which must hold for any G. Since $p_i \geq 0$, $|\operatorname{tr}[GE_i]| \geq 0$, and $\operatorname{tr}\left[|\phi_i\rangle\langle\phi_i|^{\otimes L-2}\right] = 1$, this can only be true for all G if all $E_i = 0$. Furthermore, by the symmetry of each term in the mixture, it is then sufficient to ensure the constraint is satisfied in a single party. From Eq. (37), the converse statement follows trivially. $\qquad\square$

Thus, any state $\Phi_{A_1^L} \in \mathrm{SymSEP}_L$ satisfying $\operatorname{tr}_{O_1}\Big[\Phi_{A_1^L}\Big] = (\mathbb{1}_{I_1}/d_{\mathrm{ES}}) \otimes \operatorname{tr}_{A_1}\Big[\Phi_{A_1^L}\Big]$ is part of the target set $\mathcal{T}$ and vice versa.

A.1.3 The final rank-constrained problem over SEP

Given all of the previous results, we have now established a relation between the sets $\mathcal{T}$ and SEP_L,

$$\mathcal{T} = \left\{ \Phi_{A_1^L} \in \mathrm{SEP}_L \ \Big| \ \mathrm{P}_L^+\Phi_{A_1^L} = \Phi_{A_1^L}, \quad \operatorname{tr}_{O_1}\Big[\Phi_{A_1^L}\Big] = (\mathbb{1}_{I_1}/d_{\mathrm{ES}}) \otimes \operatorname{tr}_{A_1}\Big[\Phi_{A_1^L}\Big] \right\}, \tag{43}$$

which allows us to remove the non-linear dependence on the tensor product $(C_U)^{\otimes L}$, by replacing the search space with SymSEP_L, while obeying the linear constraints on the partial trace of one party. We thus obtain:

Optimization Problem 3. *Rank-1 and symmetric optimum (exact)*

$$
\begin{aligned}
\textbf{Given:} \quad & \rho_E^0, \ \rho_S^0, \ \{M_a\}_a, \ \boldsymbol{a} \\
\textbf{Find:} \quad & \max_{\Phi_{A_1^L} \in SEP_L} \operatorname{tr}\Big[X^T \Phi_{A_1^L}\Big] \\
\textbf{Subject to:} \quad & \Phi_{A_1^L} \geq 0, \quad \operatorname{tr}\Big[\Phi_{A_1^L}\Big] = 1, \quad P_L^+ \Phi_{A_1^L} = \Phi_{A_1^L}, \\
& \operatorname{tr}_{O_1}\Big[\Phi_{A_1^L}\Big] = \frac{\mathbb{1}_{I_1}}{d_{ES}} \otimes \operatorname{tr}_{A_1}\Big[\Phi_{A_1^L}\Big]
\end{aligned}
\tag{44}
$$

However, $\Phi_{A_1^L} \in \mathrm{SEP}_L$ is still not a linear constraint and highly non-trivial to be enforced, as a full characterization of the separable set is NP-HARD [32]. Therefore, while a priori appearing more manageable, the above problem is still not in a form that can be tackled numerically. Fortunately, the symmetric constraint can also be exploited to obtain approximations of the symmetric separable states.

A.2 The SEP constraint

A.2.1 Approximation of SEP via the quantum de Finetti theorem

Ensuring $\Phi_{A_1^L} \in \mathrm{SEP}_L$ cannot be done exactly, so as it stands the problem is still numerically intractable. However, it is possible to define arbitrarily-precise outer approximations of the separable set through symmetric extensions and the quantum de Finetti theorem [15, 20]. Concretely, we can approximate, in principle to any desirable precision, a $\Phi_{A_1^L} \in \mathrm{SEP}_L$ by a marginal over a larger symmetric (and not necessarily separable) state $\Phi_{A_1^N}$, provided N is sufficiently large. For any finite N, this leads to an outer approximation of $\mathcal{T}$, which establishes a convergent hierarchy of outer approximations, such that for any $N \geq L$, we have

$$\omega_{d_\mathrm{E}}^{\boldsymbol{a}} \leq \cdots \leq \tilde{\omega}_{d_\mathrm{E}}^{\boldsymbol{a},N+1} \leq \tilde{\omega}_{d_\mathrm{E}}^{\boldsymbol{a},N} \leq \cdots \leq \tilde{\omega}_{d_\mathrm{E}}^{\boldsymbol{a},L},$$
$$\text{with} \quad \lim_{N\to\infty} \tilde{\omega}_{d_\mathrm{E}}^{\boldsymbol{a},N} = \omega_{d_\mathrm{E}}^{\boldsymbol{a}}. \tag{45}$$

More precisely, using the quantum de Finetti theorem we can establish asymptotic error bounds on this approximation [16, Cor.1]. In terms of the trace norm, this asymptotic error bound is given by

$$\left\| \Phi_{A_1^L} - \mathrm{tr}_{A_{L+1}^N}\left[\Phi_{A_1^N} \right] \right\|_1 \leq \frac{2L(L+(d_\mathrm{ES})^2+1)}{N+(d_\mathrm{ES})^2}. \tag{46}$$

As the symmetry requirement applies to both $\Phi_{A_1^L}$ and $\Phi_{A_1^N}$, this application of the quantum de Finetti theorem for approximating $\Phi_{A_1^L} \in \mathrm{SEP}_L$ is straightforward in the SDP in Eq. (15), only requiring the use of a larger state $\Phi_{A_1^N}$ as the optimization variable, and a suitable adjustment of the objective function using $(X^\mathrm{T} \otimes \mathbb{1}_{A_{L+1}^N})$.

A.2.2 Additional separability constraints

To improve the approximation, at an extra computational cost, we may also include additional separability constraints. In our case, we consider the constraints of positive partial transpose (PPT) on $\Phi_{A_1^N}$:

$$\Phi_{A_1^N}^{\mathrm{T}_\alpha} \geq 0, \quad \forall \text{ non-equivalent bipartitions } \alpha. \tag{47}$$

Here, "non-equivalent bipartitions" refers to the fact that it is unnecessary to include all bipartitions, as the state $\Phi_{A_1^N}$ is symmetric, and thus permutation invariant. Therefore, the bipartitions needed are α_k, where, e.g., we transpose only the first k subspaces, for $1 \leq k \leq \lfloor N/2 \rfloor$, with symmetry taking care of the remaining constraints.

The existence of entangled states which have positive partial transpose [36] means these constraints reduce the feasible convex set to that of PPT states, not the separable states. While entangled PPT states are not suitable solutions to the original problem, they are still adequate approximations and provide upper bounds for the actual optimal values. Inclusion of PPT constraints will perform at least as well as optimizing without them (i.e., the solution can only be improved), but convergence is significantly improved in certain cases [49] to $O(1/N^2)$ as opposed to $O(1/N)$ in the absence of these constraints. As shown in Table 1, for the $d_\mathrm{E} = 1$ case PPT constraints were sufficient to provide the exact analytical bounds.

A.2.3 SDP for SEP relaxation

By all of the above results, we arrive at the following SDP relaxation of the original problem:

Optimization Problem 4. *Final outer approximation (SDP).*

$$
\begin{aligned}
\textbf{\textit{Given:}} \quad & d_E,\ d_S,\ \rho_E^0,\ \rho_S^0,\ \{M_a\}_a,\ \boldsymbol{a} \\[4pt]
\textbf{\textit{Find:}} \quad & \tilde{\omega}_{d_E}^{\boldsymbol{a},N} := \max_{\Phi_{A_1^N}} \mathrm{tr}\left[(X^T \otimes \mathbb{1}_{A_{L+1}^N}) \Phi_{A_1^N} \right] \\[4pt]
\textbf{\textit{Subject to:}} \quad & \Phi_{A_1^N} \geq 0, \quad \mathrm{tr}\left[\Phi_{A_1^N} \right] = 1, \quad P_N^+ \Phi_{A_1^N} = \Phi_{A_1^N}, \\[4pt]
& \mathrm{tr}_{O_1}\left[\Phi_{A_1^N} \right] = \frac{\mathbb{1}_{I_1}}{d_{ES}} \otimes \mathrm{tr}_{A_1}\left[\Phi_{A_1^N} \right] \\[4pt]
& \Phi_{A_1^N}^{T_\alpha} \geq 0, \forall \alpha \in \mathfrak{A},
\end{aligned}
\tag{48}
$$

with $\mathfrak{A}$ denoting the set of non-equivalent bipartitions (App. A.2.2). The solutions of this SDP provide upper-bounds for the maxima $\omega_{d_\mathrm{E}}^{\boldsymbol{a}}$.

B Implementation

While in principle numerically accessible, the SDP in Prob. 4 still contains too many variables to allow for the computation of upper bounds of $\tilde{\omega}_{d_E}^{\boldsymbol{a},N}$, even for low-dimensional cases. This problem can be tackled by exploiting the fact that all appearing objects can be expressed within the symmetric subspace, as well as the (potential) sparsity of the problem.

In our case, a further simplification is possible by first noticing that both X and P_N^+ are real-valued. If we additionally consider a real-valued basis for the partial trace, we conclude that, for any feasible $\Phi_{\mathrm{A}_1^N}$, its entry-wise complex conjugate $\Phi_{\mathrm{A}_1^N}^*$ is also feasible, while providing the same value for the objective function. Therefore, we may perform the optimization using a real-valued $\Phi_{\mathrm{A}_1^N}$. Next, we explain how to exploit the symmetry property.

B.1 Symmetric representation

The symmetry constraint can be satisfied automatically by expressing $\Phi_{\mathrm{A}_1^N}$ directly in terms of a basis for the symmetric subspace. This allows for a great reduction in the number of variables and linear constraints in the SDP, which is required for its efficient numerical optimization. In the following, we describe how to construct this basis for the symmetric subspace, largely based on [34], and how to adapt the remaining constraints to act directly in this symmetric representation, based on a generalization of results provided in [62] for the qubit case.

We begin by defining a canonical basis for the symmetric subspace. With the convention $\mathbb{N} = \{1, 2, 3, \ldots\}$ and $\mathbb{N}_0 := \mathbb{N} \cup \{0\}$, for $d, n \in \mathbb{N}$, let

$$\mathfrak{T}_d^n := \left\{ (t_1, t_2, \ldots, t_d) \ \middle| \ t_i \in \mathbb{N}_0, \quad \sum_{i=1}^d t_i = n \right\} \tag{49}$$

be the set of weak integer compositions for n into exactly d parts, possibly of zero size, which we refer to as *types*. These types form a canonical labeling for the basis of the symmetric subspace. Now, defining $[d] := \{1, 2, \ldots, d\}$, let $\boldsymbol{u} \in [d]^n$, i.e., $\boldsymbol{u} = (u_1, \ldots, u_n)$ with $u_\ell \in [d]$, and $T(\boldsymbol{u}) = \boldsymbol{t}$ be the type of the vector $\boldsymbol{u}$, where t_i counts the number of instances where $u_\ell = i$ holds. As a concrete example, if $d = 6$ and $n = 8$, we might have:

$$\boldsymbol{u} = (1, 4, 1, 2, 3, 2, 2, 6) \qquad \longrightarrow \qquad T(\boldsymbol{u}) = (2, 3, 1, 1, 0, 1). \tag{50}$$

| In words, types count how many times a number $1, \ldots, d$ occurs in $\boldsymbol{u}$, and therefore are invariant under permutations, i.e., $T(\boldsymbol{u}) = T(P_\sigma \boldsymbol{u})$ for any permutation operator P_σ acting on entries of $\boldsymbol{u}$. Given a Hilbert space $\mathcal{H}$ with $d = \dim \mathcal{H}$, we may now define a basis for $\mathrm{Sym}_n(\mathcal{H})$ in terms of the types $\boldsymbol{t} \in \mathfrak{T}_d^n$ by constructing the non-normalized and normalized orthogonal basis vectors for the symmetric subspace [34], respectively, as

$$|\mathrm{Sym}(\boldsymbol{t})\rangle := \sum_{\boldsymbol{u}; T(\boldsymbol{u}) = \boldsymbol{t}} |u_1, u_2, \cdots, u_n\rangle, \qquad |\mathrm{sym}(\boldsymbol{t})\rangle := (\boldsymbol{t})^{-1/2} |\mathrm{Sym}(\boldsymbol{t})\rangle, \tag{51}$$

where $(\boldsymbol{t}) = \frac{(\sum_i t_i)!}{t_1! \cdot t_2! \cdots t_d!}$ denotes the multinomial coefficient for the normalization, and, for convenience in notation, we adopt the canonical basis $\{|u\rangle\}_{u=1}^d$ for each individual u_i. Equation (51) clearly defines a symmetric state, since each $\boldsymbol{u}$ occurs only once in the sum. Thus, for any permutation operator V_σ, we have $V_\sigma |\mathrm{sym}(\boldsymbol{t})\rangle = |\mathrm{sym}(\boldsymbol{t})\rangle$, and similarly for the non-normalized $|\mathrm{Sym}(\boldsymbol{t})\rangle$.

From Eq. (37), we know any symmetric separable state can be written as $\sum_i p_i |\phi_i\rangle\!\langle\phi_i|^{\otimes n}$. Therefore, by expressing $|\phi_i\rangle^{\otimes n}$ in terms of the symmetric basis in Eq. (51), we may write any symmetric operator by indexing the degrees of freedom of $\Phi_{\mathrm{A}_1^N}$ by the types $\boldsymbol{t}, \boldsymbol{t}'$. In our current problem, with a local dimension $d = (d_E d_S)^2 = d_{ES}^2$, the symmetric space has total dimension [34]

$$D_{\mathcal{S}} = \dim \mathrm{Sym}_N(\mathcal{H}) = \left(\!\! \left(\begin{array}{c} d_{ES}^2 \\ N \end{array} \right) \!\! \right) = \binom{N + d_{ES}^2 - 1}{N}, \tag{52}$$

with $\left(\!\!\left(\begin{smallmatrix} d \\ n \end{smallmatrix}\right)\!\!\right)$ the multiset notation, i.e., the number of ways of picking n elements out of d, with repetitions allowed, which corresponds to the cardinality of $\mathfrak{T}_d^n$. This is a significant reduction from the original $\dim \mathrm{A}_1^N = (d_{ES})^{2N}$.

For simplicity in notation, we consider the dependence of $D_{\mathcal{S}}$ on d_{ES} and N implicit in the following. To concretely exploit this reduction of the number of variables, i.e., in order to write $\Phi_{\mathrm{A}_1^N}$ in terms of a smaller matrix, we re-express the problem as follows. Let $\mathcal{S}_N := \mathbb{C}^{D_{\mathcal{S}} \otimes D_{\mathcal{S}}}$, and $\phi \in \mathcal{S}_N$. We specify elements $\phi_{\boldsymbol{t}, \boldsymbol{t}'}$, with $\boldsymbol{t}, \boldsymbol{t}' \in \mathfrak{T}_{d_{ES}^2}^N$, by a canonical orthonormal basis $\{|\boldsymbol{t}\rangle_{\mathcal{S}_N}\}_{\boldsymbol{t}}$. We define an isometry between the two representations

of the symmetric subspace, i.e., the normalized symmetric vectors in the full space A_1^N and the canonical basis for $\mathcal{S}_N$, as follows:

$$V_N = \sum_{\boldsymbol{t}} |\boldsymbol{t}\rangle_{\mathcal{S}_N} \langle \mathrm{sym}(\boldsymbol{t})|_{A_1^N} . \tag{53}$$

Note that this isometry acts as a projector in A_1^N, i.e., $V_N V_N^\dagger = \mathbb{1}_{\mathcal{S}_N}$ but $V_N^\dagger V_N \neq \mathbb{1}_{A_1^N}$. Therefore, if $P_N^+ \Phi_{A_1^N} = \Phi_{A_1^N}$, we may write it concisely as:

$$\Phi_{A_1^N} = V_N^\dagger \phi V_N = \sum_{\boldsymbol{t}, \boldsymbol{t}'} \phi_{\boldsymbol{t}, \boldsymbol{t}'} |\mathrm{sym}(\boldsymbol{t})\rangle\!\langle\mathrm{sym}(\boldsymbol{t}')|_{A_1^N} ,$$

$$\phi = V_N \Phi_{A_1^N} V_N^\dagger, \quad \text{and} \quad x = V_N (X^\mathrm{T} \otimes \mathbb{1}_{A_{L+1}^N}) V_N^\dagger . \tag{54}$$

so that the objective function, which acts entirely within the symmetric subspace, can be written as

$$\mathrm{tr}\left[(X^\mathrm{T} \otimes \mathbb{1}_{A_{L+1}^N}) \Phi_{A_1^N} \right] = \mathrm{tr}\left[(X^\mathrm{T} \otimes \mathbb{1}_{A_{L+1}^N}) V_N^\dagger \phi V_N \right] = \mathrm{tr}\left[V_N (X^\mathrm{T} \otimes \mathbb{1}_{A_{L+1}^N}) V_N^\dagger \phi \right] = \mathrm{tr}\left[x\phi \right] . \tag{55}$$

Positivity of $\Phi_{A_1^N}$ can be guaranteed by requiring $\phi \geq 0$ directly, as ϕ and $\Phi_{A_1^N}$ have the same non-zero eigenvalues. With the chosen normalization of C_U, we may also write $\mathrm{tr}\,[\phi] = 1$. In practice, it is sufficient to numerically compute x directly by means of Eq. (54), as computing the projection of $X^\mathrm{T} \otimes \mathbb{1}_{A_{L+1}^N}$ onto the symmetric subspace analytically can be cumbersome, and this computation only needs to be performed once before the numerical optimization.

To rewrite the partial trace constraints of the SDP directly in this symmetric basis, we must find a way to express partial traces in terms of the symmetric representation. In the following, we generalize some results known for qubits [62] to arbitrary local dimension. But first, recall that the constraint we wish to rewrite is given by

$$\mathrm{tr}_{O_1}\left[\Phi_{A_1^N}\right] = \frac{\mathbb{1}_{I_1}}{d_{\mathrm{ES}}} \otimes \mathrm{tr}_{A_1}\left[\Phi_{A_1^N}\right] . \tag{56}$$

Treating types as ordinary vectors, we may define addition between types with the same d, such that if $\boldsymbol{r} \in \mathfrak{T}_d^n$ and $\boldsymbol{s} \in \mathfrak{T}_d^m$, then $\boldsymbol{r} + \boldsymbol{s} = \boldsymbol{t} \in \mathfrak{T}_d^{n+m}$. This corresponds to the fact that $\mathcal{S}_{n+m} \subset \mathcal{S}_n \otimes \mathcal{S}_m$ [62]. Generalizing upon this idea, we may split an n-partite symmetric state into m parts of sizes k_i, with $\sum_{i=1}^m k_i = n$, through the decomposition

$$|\mathrm{Sym}(\boldsymbol{t})\rangle_{A_1^N} = \sum_{\boldsymbol{r}_1 + \cdots + \boldsymbol{r}_m = \boldsymbol{t}} |\mathrm{Sym}(\boldsymbol{r}_1)\rangle_{B_1} \cdots |\mathrm{Sym}(\boldsymbol{r}_m)\rangle_{B_m} , \tag{57}$$

where the sum is over all tuples $(\boldsymbol{r}_1, \ldots, \boldsymbol{r}_m) \in (\mathfrak{T}_d^{k_1} \times \cdots \times \mathfrak{T}_d^{k_m})$ satisfying $\boldsymbol{r}_1 + \cdots + \boldsymbol{r}_m = \boldsymbol{t}$. Here, $B_\ell := A_{b_\ell+1}^{b_\ell+k_\ell}$, with $b_\ell = \sum_{i=1}^{\ell-1} k_i$, corresponding to the subspace of the ℓ-th part of the decomposition. Importantly, this decomposition is performed in the full space with non-normalized vectors. For completion, the normalized version of the decomposition in Eq. (57) is given by

$$|\mathrm{sym}(\boldsymbol{t})\rangle_{A_1^N} = \sum_{\boldsymbol{r}_1 + \cdots + \boldsymbol{r}_m = \boldsymbol{t}} \left(\frac{(\boldsymbol{t})}{(\boldsymbol{r}_1) \cdots (\boldsymbol{r}_m)} \right)^{-1/2} |\mathrm{sym}(\boldsymbol{r}_1)\rangle_{B_1} \cdots |\mathrm{sym}(\boldsymbol{r}_m)\rangle_{B_m} . \tag{58}$$

However, working under such normalization is cumbersome and inefficient due to the several coefficients involved. Instead, we have chosen to normalize directly in terms of the original types $\boldsymbol{t}, \boldsymbol{t}'$, which makes normalization straightforward. To adapt the partial trace constraint, we first write the state $\Phi_{A_1^N}$ as in Eq. (22), and using Eq. (57) split the basis into two parts of sizes $(1, N-1)$, obtaining

$$\Phi_{A_1^N} = \sum_{\boldsymbol{t}, \boldsymbol{t}'} \sum_{\boldsymbol{r} + \boldsymbol{s} = \boldsymbol{t}} \sum_{\boldsymbol{r}' + \boldsymbol{s}' = \boldsymbol{t}'} \hat{\phi}_{\boldsymbol{t}, \boldsymbol{t}'} |\mathrm{Sym}(\boldsymbol{r})\rangle\!\langle\mathrm{Sym}(\boldsymbol{r}')|_{A_1} \otimes |\mathrm{Sym}(\boldsymbol{s})\rangle\!\langle\mathrm{Sym}(\boldsymbol{s}')|_{A_2^N} , \tag{59}$$

where the proper normalization is now taken care of by defining $\hat{\phi}_{\boldsymbol{t}, \boldsymbol{t}'} := \phi_{\boldsymbol{t}, \boldsymbol{t}'} ((\boldsymbol{t})(\boldsymbol{t}'))^{-1/2}$. Since $\boldsymbol{r}$ and $\boldsymbol{r}'$ are types for a single party, we have that $|\mathrm{Sym}(\boldsymbol{r})\rangle_{A_1}$ are simply vectors in the canonical basis $\{|u\rangle\}_{u=1}^d$. Thus, we will adopt the notation $|u(\boldsymbol{r})\rangle_{A_1} = |\mathrm{Sym}(\boldsymbol{r})\rangle_{A_1}$ in what follows. The tensor product form allows for the simplified application of the partial traces in Eq. (56). By linearity of the partial trace, we can treat each term of Eq. (59) separately, so that the terms on the left (Y_L) and right (Y_R) hand sides of Eq. (56) can be written as

$$Y_L := \mathrm{tr}_{O_1}\left[|u(\boldsymbol{r})\rangle\!\langle u(\boldsymbol{r}')|_{A_1} \otimes |\mathrm{Sym}(\boldsymbol{s})\rangle\!\langle\mathrm{Sym}(\boldsymbol{s}')|_{A_2^N} \right]$$

$$= \mathrm{tr}_{O_1}\left[|u(\boldsymbol{r})\rangle\!\langle u(\boldsymbol{r}')|_{A_1} \right] \otimes |\mathrm{Sym}(\boldsymbol{s})\rangle\!\langle\mathrm{Sym}(\boldsymbol{s}')|_{A_2^N} \tag{60}$$

$$Y_R := \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{tr}_{\mathrm{A}_1}\Big[|u(\boldsymbol{r})\rangle\!\langle u(\boldsymbol{r}')|_{\mathrm{A}_1} \otimes |\mathrm{Sym}(\boldsymbol{s})\rangle\!\langle \mathrm{Sym}(\boldsymbol{s}')|_{\mathrm{A}_2^N}\Big]$$

$$= \mathrm{tr}\left[|u(\boldsymbol{r})\rangle\!\langle u(\boldsymbol{r}')|_{\mathrm{A}_1}\right] \cdot \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes |\mathrm{Sym}(\boldsymbol{s})\rangle\!\langle \mathrm{Sym}(\boldsymbol{s}')|_{\mathrm{A}_2^N}. \tag{61}$$

A few simplifications are now evident. First, we observe that $\mathrm{tr}\left[|u(\boldsymbol{r})\rangle\!\langle u(\boldsymbol{r}')|_{\mathrm{A}_1}\right] = \delta_{\boldsymbol{r},\boldsymbol{r}'}$, where $\delta_{\boldsymbol{r},\boldsymbol{r}'}$ is the Kronecker delta. Secondly, we may split the states $|u(\boldsymbol{r})\rangle_{\mathrm{A}_1}$ into the local input and output spaces

$$|u(\boldsymbol{r})\rangle_{\mathrm{A}_1} = |i(\boldsymbol{r})\rangle_{\mathrm{I}_1} |o(\boldsymbol{r})\rangle_{\mathrm{O}_1}, \tag{62}$$

so that we obtain

$$\mathrm{tr}_{\mathrm{O}_1}\big[|u(\boldsymbol{r})\rangle\!\langle u(\boldsymbol{r}')|_{\mathrm{A}_1}\big] = \mathrm{tr}_{\mathrm{O}_1}\big[|i(\boldsymbol{r})\rangle\!\langle i(\boldsymbol{r}')|_{\mathrm{I}_1} \otimes |o(\boldsymbol{r})\rangle\!\langle o(\boldsymbol{r}')|_{\mathrm{O}_1}\big]$$

$$= |i(\boldsymbol{r})\rangle\!\langle i(\boldsymbol{r}')|_{\mathrm{I}_1} \cdot \mathrm{tr}\big[|o(\boldsymbol{r})\rangle\!\langle o(\boldsymbol{r}')|_{\mathrm{O}_1}\big]$$

$$= |i(\boldsymbol{r})\rangle\!\langle i(\boldsymbol{r}')|_{\mathrm{I}_1} \cdot \delta_{o(\boldsymbol{r}),o(\boldsymbol{r}')}. \tag{63}$$

Using the above results, and correcting for the missing normalizations, we can apply the isometry $\mathbb{1}_{\mathrm{I}_1} \otimes V_{N-1}$, so that Eqs. (60) and (61) become:

$$Y_L \to ((\boldsymbol{s})(\boldsymbol{s}'))^{1/2} \cdot \delta_{o(\boldsymbol{r}),o(\boldsymbol{r}')} \cdot |i(\boldsymbol{r})\rangle\!\langle i(\boldsymbol{r}')|_{\mathrm{I}_1} \otimes |\boldsymbol{s}\rangle\!\langle \boldsymbol{s}'|_{\mathcal{S}_{N-1}}$$

$$Y_R \to ((\boldsymbol{s})(\boldsymbol{s}'))^{1/2} \cdot \frac{\delta_{\boldsymbol{r},\boldsymbol{r}'}}{d_{\mathrm{ES}}} \cdot \sum_{i=1}^{d_{\mathrm{ES}}} |i\rangle\!\langle i|_{\mathrm{I}_1} \otimes |\boldsymbol{s}\rangle\!\langle \boldsymbol{s}'|_{\mathcal{S}_{N-1}}. \tag{64}$$

Here, we have written $\mathbb{1}_{\mathrm{I}_1} = \sum_{i=1}^{d_{\mathrm{ES}}} |i\rangle\!\langle i|_{\mathrm{I}_1}$ as to make the $\mathrm{I}_1 \otimes \mathcal{S}_{N-1}$ decomposition explicit in both expressions. By inserting Eq. (64) back into the sum of Eq. (59), we can appreciate the fact that Eq. (64) neatly separates each term of the sum as square matrices of size $d_{\mathrm{ES}} \times \left(\binom{d_{\mathrm{ES}}^2}{N-1}\right)$, written in terms of $|i\rangle\!\langle i'|_{\mathrm{I}_1} \otimes |\boldsymbol{s}\rangle\!\langle \boldsymbol{s}'|_{\mathcal{S}_{N-1}}$. The constraint of Eq. (56) then tells us we must sum over all $\boldsymbol{t}, \boldsymbol{t}'$ on both sides, where we can then apply the equality constraint element-wise by simply matching the resulting $(i, \boldsymbol{s}, i', \boldsymbol{s}')$ entries. As the extra normalization $((\boldsymbol{s})(\boldsymbol{s}'))^{1/2}$ of Eq. (64) is always equal between these elements, it cancels out in their element-wise equality constraint. Thus, it is sufficient to use the normalization of $\hat{\phi}_{\boldsymbol{t},\boldsymbol{t}'}$.

During implementation in software, the summations in Eq. (59) can be efficiently performed by taking into account the fact that valid decompositions of the form $\boldsymbol{t} = \boldsymbol{r} + \boldsymbol{s}$ are very restricted in number, and can be efficiently enumerated, grouped, filtered and counted. By assigning a tuple $(i, o, \boldsymbol{s})$ to each $\boldsymbol{t}$, the tracing over inputs and output spaces, and applications of the Kronecker deltas, can thus be efficiently computed. Therefore, the linear constraints between variables $\phi_{\boldsymbol{t},\boldsymbol{t}'}$ can be constructed by bucketing terms $\phi_{\boldsymbol{t},\boldsymbol{t}'}$ by matching $|i\rangle\!\langle i'|_{\mathrm{I}_1} \otimes |\boldsymbol{s}\rangle\!\langle \boldsymbol{s}'|_{\mathcal{S}_{N-1}}$, and normalization is simplified by attaching it to $\phi_{\boldsymbol{t},\boldsymbol{t}'}$ at the very end, i.e., using $\hat{\phi}_{\boldsymbol{t},\boldsymbol{t}'}$.

A similar strategy as above may be employed to define partial transpositions through the symmetric representation, as was done in [62] for the qubit case, in order to implement the PPT condition. However, for $d_{\mathrm{E}} = 2$ in our case, this approach would render the matrices and the number of constraints too large for a viable computation, once again, and therefore we did not pursue a generalization of this idea. The optimizations we have performed for $d_{\mathrm{E}} = 1$ involving PPT constraints, as shown in Table 1, used $\Phi_{\mathrm{A}_1^N}$ directly following Eq. (22), which was still tractable in this case.

While the above steps lead to a significantly smaller representation of the problem, it is not generally a sufficient reduction in variables and constraints for the SDP to be numerically tractable. For concreteness, in the smallest non-trivial case of $d_{\mathrm{E}} = d_{\mathrm{S}} = 2$, this gives for $N = 2, \ldots, 5$ a symmetric space of size $D_{\mathcal{S}} = 136,\ 816,\ 3\,876,\ 15\,504,\ \ldots$, such that the SDP would be written in terms of $(D_{\mathcal{S}})^2 = 18\,496,\ 665\,856,\ 15\,023\,376,\ 240\,374\,016, \ldots$ variables. These examples illustrate how quickly our problem can become numerically intractable, even with symmetry taken into account.

B.2 Sparse implementation of the SDP

In the following, we describe in detail the algorithm we have developed to obtain a sparse version of the SDP. This is an essential step in rendering the problem numerically tractable. The algorithm works by heuristically exploiting any existing sparsity of the original problem, in particular its objective function, in order to automatically construct a sparse outer approximation. In our particular case, the sparse SDP resulting from our algorithm was not only far simpler, it was in fact an exact sparse representation of the original problem. Below, we also discuss the general conditions required for this to be the case for our algorithm. We highlight that it

is potentially applicable to many other SDPs, providing at least an outer approximation, although there are no guarantees that a sparse simplification of the original problem will be obtained.

To see how a problem's sparsity can be exploited, consider a standard form SDP over $\mathbb{C}^{n \times n}$:

$$
\begin{aligned}
\textbf{Given:} \quad & F, \ (C^{(k)})_k, \ (b_k)_k \\
\textbf{Find:} \quad & \max_X \ \langle F, X \rangle \\
\textbf{Subject to:} \quad & \langle C^{(k)}, X \rangle = b_k, \quad k = 1, ..., m \\
& X \geq 0.
\end{aligned}
\tag{65}
$$

where $\langle A, B \rangle = \operatorname{tr}\left[A^\dagger B \right]$. We assume F is Hermitian so that $F_{ij} \neq 0$ implies $F_{ji} \neq 0$. Our goal is to exploit any sparsity of F, which defines the objective function, to simplify this problem by obtaining a suitable sparse version. This is achieved by finding a subset of variables X_{ij} which is self-sufficient to solve the problem, and this is done by keeping track of the indices (i, j) which can influence the objective function, directly or indirectly.

We begin by defining the initial *base sparsity* of the problem as

$$
\theta_{\text{base}}^0 := \left\{ \ (i, j) \ \mid \ F_{ij} \neq 0 \ \right\} \cup \{(i, i)\}_{i=1}^n,
\tag{66}
$$

i.e., the pairs of indices with non-zero entries in F, together with the indices in the diagonal. We include the diagonal indices by default in order to ensure $X \geq 0$ can be satisfied properly in the sparse version of the problem, as will be explained later. As $\langle F, X \rangle = \sum_{i,j} F_{ij}^* X_{ij}$, the set θ_{base}^0 corresponds to the entries of X which appear explicitly in the objective function, i.e., the optimization variables X_{ij} affecting its value, together with the diagonal entries. However, these are not the only variables relevant to the problem, as they also appear in linear constraints involving other variables not present in the objective, which may affect the objective function indirectly. Our next goal is to collect all of these indirect constraints to the objective function's variables. Defining the subset of constraints $C^{(k)}$ which relate to θ_{base}^0 as

$$
\mathcal{K}^0 := \bigcup_{(i,j) \in \theta_{\text{base}}^0} \left\{ \ k \ \middle| \ C_{ij}^{(k)} \neq 0 \ \right\},
\tag{67}
$$

we can define the initial *extended sparsity*

$$
\theta_{\text{ext}}^0 := \bigcup_{k \in \mathcal{K}^0} \left\{ \ (i, j) \ \middle| \ C_{ij}^{(k)} \neq 0 \ \right\}.
\tag{68}
$$

In words, this set includes the indices for any variables present in a linear constraint also involving the variables directly affecting the objective function. The extended sparsity θ_{ext}^0 specifies an initial guess for the minimum subset of variables $\{X_{ij} \mid (i, j) \in \theta_{\text{ext}}^0\}$ which should be considered in the SDP. The positivity constraint $X \geq 0$ could then be specified, in principle, in terms of a block-diagonal matrix which contains the variables specified by θ_{ext}^0 within its blocks. Positivity of this block-diagonal matrix, and in turn of X, could then be achieved by enforcing positivity for each block separately.

As positivity constraints require significant computational resources, ideally these blocks should be made as small as possible. In order to obtain the minimal block-diagonal form, we first use θ_{ext}^0 to construct an adjacency matrix A^0

$$
[A^0]_{ij} := \begin{cases} 1, & \text{if } (i, j) \in \theta_{\text{ext}}^0 \\ 0, & \text{else} \end{cases},
\tag{69}
$$

with which we construct an undirected graph $G(A^0)$ with n vertices, where each edge corresponds to a pair of indices (i, j). We can use this graph to obtain a permutation which leads to the minimal block-diagonal form, as each block corresponds to a disjoint subset of indices which are interrelated. This in turn, corresponds to the notion of *connected components* of a graph, i.e., its disjoint subgraphs, as shown in Fig. 7. Let $\{g_r\}_{r=1}^R$ be the R connected components of $G(A^0)$, so that $G(A^0) = \bigcup_{r=1}^R g_r$, and let ν_r be the set of indices of the vertices in g_r. We may use the ν_r to construct a permutation $\pi = (\nu_1 \ \nu_2 \ \cdots \ \nu_R)$, together with its associated permutation matrix P, such that $P A^0 P^{\mathrm{T}}$ is block-diagonal, with one block per connected component. E.g., from Fig. 7, we could use the permutation $\pi = [(5273)(4)(16)]$, where the ordering of the ν_r, or within each ν_r, are irrelevant. This permutation leads to the block-diagonal adjacency matrix shown in figure Fig. 8, on the left.

Therefore, if we take θ_{ext}^0 to be the sparsity pattern for X, we can construct the sparse matrix

$$
\widetilde{X} := P^{\mathrm{T}} \left(\bigoplus_{r=1}^R \widetilde{B}_r \right) P, \quad \widetilde{B}_r \geq 0 \ \forall r,
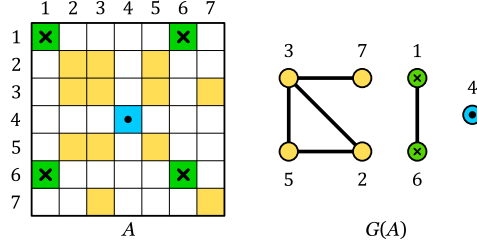\tag{70}
$$

Figure 7: An adjacency matrix A and its corresponding graph $G(A)$, with three connected components. Non-zero entries are depicted with the colors (and markers) of the component they belong to. Self-edges, corresponding to (i, i) entries, are omitted.

with sparse blocks of variables given by $\widetilde{B}_r$. However, imposing positivity for each block is not always possible in this case, as the blocks themselves have a sparse structure incompatible with positivity constraints. Concretely, there might be *no* $\widetilde{X} \geq 0$ matrix with sparsity structure given by $\widetilde{X}_{ij} = 0 \ \forall \ (i, j) \notin \theta^0_{\text{ext}}$.
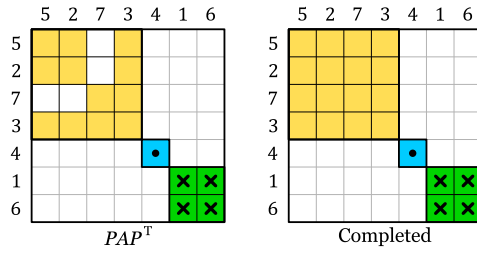
Figure 8: (Left) Under the permutation $\pi = [(5273)(4)(16)]$, the matrix A of Fig. 7 becomes block-diagonal, with blocks possibly containing "gaps" indicating that not all pairs of vertices share an edge for that component. (Right) The blocks can be completed by adding the missing entries (i.e., edges).

To address this in general, we instead assume all the blocks are dense. We can then define the initial *completed sparsity*

$$\theta^0_{\text{comp}} := \bigcup_{r=1}^{R} \{ \ (i, j) \in \nu_r \times \nu_r \ \}, \tag{71}$$

by including all missing indices (i, j) within each block. In graph theoretical terms, we turn each connected component into a complete graph, i.e., every pair of vertices shares an edge, and use the non-zero entries of the resulting adjacency matrix to define θ^0_{comp}; see Fig. 8, right panel. Denoting by B_r the dense block matrices, we can define the first sparse approximation of X as

$$\widetilde{X}^0 := P^{\text{T}} \left(\bigoplus_{r=1}^{R} B_r \right) P, \quad B_r \geq 0 \ \forall \ r, \tag{72}$$

such that $\widetilde{X}^0 \geq 0$ can be safely imposed through block-positivity, with no risk of rendering the SDP unfeasible.

The next step is to analyze how this sparse structure affects the other constraints in the SDP. These procedures expanded our initial base sparsity θ^0_{base} by including a larger set of indices (and therefore, variables) of the original dense X, resulting in θ^0_{comp}. However, these additional variables may themselves appear in other linear constraints, involving even further variables we have not yet taken into account. These inter-dependencies between variables lead to a cascade of new constraints and variables that must be considered, i.e., the additional indices included in θ^0_{comp} specify variables which appear in constraints which were previously ignored, as they did not involve variables specified by θ^0_{base}. Therefore, we repeat the above procedures until a stable and completely self-contained set of indices naturally emerges, i.e., we define a new base sparsity $\theta^1_{\text{base}} := \theta^0_{\text{comp}}$, and repeat the process $n + 1$ times until $\theta^{n+1}_{\text{base}} = \theta^n_{\text{base}}$, at which point we call this stable set the *effective sparsity* θ^* for the problem. In our particular applications, the algorithm converges within 2 to 4 iterations.

Note that, by construction, any variable X_{ij} for $(i, j) \in \theta^*$ appears in the objective function directly, or if not, affects the variables in the objective either directly (through linear constraints) or indirectly (through positivity constraints and, in turn, through further constraints, and so on, recursively). The goal of the above algorithm is to automatically discover a set of variables and constraints which is self-sufficient, given any objective function and linear constraints for an SDP.

While the algorithm generally provides an outer approximation, it can also lead to an exact sparse version of the original problem if the unused linear constraints are homogeneous, which is the case for the problem we consider. To understand why, we define the set of unused constraints as

$$\mathcal{K}_\perp^* := \left\{ k \ \middle| \ C_{ij}^{(k)} = 0, \ \forall \, (i,j) \in \theta^* \right\}. \tag{73}$$

If the unused constraints are all homogeneous, i.e., $b_k = 0$ for all $k \in \mathcal{K}_\perp^*$, then the original dense problem can be solved entirely within θ^*, without affecting the objective function, as any sparse solution can be made into a feasible solution to the dense problem. This can be understood as follows. Let X^* be a solution to the sparse problem, and $X_\perp^*$ a matrix in the complementary space $\theta_\perp^*$, containing all indices not appearing in θ^*. Any constraints discarded by the algorithm (i.e., the indices in $\mathcal{K}_\perp^*$) will act entirely on $\theta_\perp^*$. Then, as the objective function contains only variables within θ^*, we may write a dense solution $X = X^* + X_\perp^*$, such that

$$\mathrm{tr}\left[F^\dagger X\right] = \mathrm{tr}\left[F^\dagger X^*\right] + \mathrm{tr}\left[F^\dagger X_\perp^*\right] = \mathrm{tr}\left[F^\dagger X^*\right], \tag{74}$$

with $X^*, X_\perp^* \geq 0$, and such that all constraints $\langle C^{(k)}, X \rangle = b_k$ uncouple onto orthogonal spaces. Now, if the linear constraints involving $X_\perp^*$ are all homogeneous ($b_k = 0$ for all $k \in \mathcal{K}_\perp^*$), we may set $X_\perp^* = 0$ directly, while simultaneously satisfying all constraints in the dense problem, and without affecting the objective function. Therefore, any sparse solution X^* is also a valid solution to the dense problem, with the same value for the objective: the sparse problem is exact. Note that the diagonal indices needed to be included in θ_{base}^0 in order to ensure $X \geq 0$ is satisfied because of $X^* \geq 0$, as otherwise positivity of X could always be satisfied for any X^* by choosing arbitrarily large diagonal entries in $X_\perp^*$.

Despite the success of this algorithm in giving a sparse solution to our particular problem, in general, there is no guarantee that θ^* will be sparse. It may be the case that the algorithm eventually includes all entries of X, in which case an alternative approach must be used, e.g., stopping the algorithm before a stable sparsity pattern is reached (generally resulting in an outer approximation), using a more refined completion procedure (e.g., block-wise chordal completions [28, 74]), or exploiting additional structures of the original problem. Moreover, the sparsity of the original problem, and thus the efficacy of the above algorithm, relies heavily on the choice of basis in which the problem is represented. Fortunately, for our choice of parameters, the objective function and the resulting θ^* were sufficiently sparse for the SDP to be numerically tractable. Furthermore, as the partial trace constraints $\mathrm{tr}_{\mathrm{O}_1}\left[V_N^\dagger \phi V_N\right] = \frac{\mathbb{1}_{\mathrm{I}_1}}{d_{\mathrm{ES}}} \otimes \mathrm{tr}_{\mathrm{A}_1}\left[V_N^\dagger \phi V_N\right]$ are all homogeneous, this approach lead to an exact sparse representation of Eq. (15).

Finally, we note that other techniques exist to formulate sparse SDPs [28, 74], typically relying on the problem having an inherent "aggregate sparsity" (the joint sparse structure of F and all $C^{(k)}$ taken together), or being analytically amenable to a sparser representation by an appropriate change of variables. In our problem, no such structure is a priori apparent: only the objective is inherently sparse, with the linear constraints jointly involving all variables in a non-trivial manner. In such cases, our algorithm is then a suitable technique to obtain an "effective" sparsity focused on the objective function's inherent sparse structure, with the advantage of being more agnostic to the problem's overall structure.

C Maximum probability for closed systems

Here, we provide a proof for the analytical bounds for the $d_{\mathrm{E}} = 1$ case mentioned in Sec. 3. To this end, recall that, in the sequential measurement protocol, the probability of a sequence of outcomes can be written as

$$p(\boldsymbol{a}|d_{\mathrm{E}}) = p(a_1, \ldots, a_L|d_{\mathrm{E}}) = \mathrm{tr}\left[\mathcal{M}_{a_L} \circ \mathcal{U} \circ \cdots \circ \mathcal{M}_{a_1} \circ \mathcal{U}(\rho_{\mathrm{E}}^0 \otimes \rho_{\mathrm{S}}^0)\right], \tag{75}$$

with the measure-and-prepare map $\mathcal{M}_a(\rho_{ES}) = \mathrm{tr}_S[\rho_{ES} \cdot (\mathbb{1}_E \otimes E_S^a)] \otimes \rho_S^0$. If $d_{\mathrm{E}} = 1$, the unitary corresponds only to a local rotation on the system, and can be embedded in the state ρ_{S}^0 and measurements E_S^a. Thus, the maximum depends only on the state and measurements. To compute it, we first note that since both ρ_{S}^0 and $(E_S^a)_a$ are the same at each step and $\mathcal{M}_a$ is a measure and prepare operation, all outcomes are independent and identically distributed. Writing $q_a = \mathrm{tr}\left[\rho_{\mathrm{S}}^0 E_S^a\right]$ as the probability of outcome a, with $\sum_a q_a = 1$, and using n_a as the number of occurrences of a symbol a in $\boldsymbol{a}$, we can write the probability as:

$$p(\boldsymbol{a}|d_{\mathrm{E}} = 1) = \prod_{\ell=1}^L q_{a_\ell} = \prod_{a \in \mathcal{A}} q_a^{n_a}, \qquad \text{with} \qquad n_a \in \mathbb{N}, \quad \sum_{a \in \mathcal{A}} n_a = L, \quad q_a > 0, \text{ and } \sum_{a \in \mathcal{A}} q_a = 1. \tag{76}$$

For simplicity, we assume that $\boldsymbol{a}$ contains every symbol of $\mathcal{A}$ at least once, such that $n_a, q_a > 0$. Otherwise, we could simply assume a smaller $\mathcal{A}$ where this is the case. The maximum probability $\omega_1^{\boldsymbol{a}}$ can be found with standard techniques, such as Lagrange multipliers. Using the Lagrangian

$$\mathcal{L} = \prod_{a \in \mathcal{A}} q_a^{n_a} - \alpha \left(\sum_{a \in \mathcal{A}} q_a - 1 \right), \tag{77}$$

we calculate the partial derivatives for each q_a, and equate them to zero:

$$\frac{\partial \mathcal{L}}{\partial q_a} = \left(\frac{n_a}{q_a} \right) p - \alpha = 0, \ \forall a. \tag{78}$$

Since $q_a > 0$ for all a, $p > 0$, we may rewrite this as:

$$\frac{n_a}{q_a} = \frac{\alpha}{p} = \gamma \quad \rightarrow \quad q_a = \frac{n_a}{\gamma}, \ \forall a. \tag{79}$$

Summing over a and applying the constraints $\sum_a q_a = 1$ and $\sum_a n_a = L$, we obtain:

$$\sum_a q_a = \frac{1}{\gamma} \sum_a n_a \quad \rightarrow \quad \gamma = L. \tag{80}$$

Thus, the maximum value is

$$\omega_1^{\boldsymbol{a}} = \prod_{a \in \mathcal{A}} \left(\frac{n_a}{L} \right)^{n_a}, \qquad \text{for} \quad q_a = \frac{n_a}{L}, \ \forall a \tag{81}$$

Note that this solution is unique if $q_a > 0$ and $p > 0$. The only alternative solutions would require that $q_a = 0$ for some a, and thus $p = 0 < \omega_1^{\boldsymbol{a}}$. Therefore, this is indeed the global maximum.

We emphasize that this direct calculation of $\omega_1^{\boldsymbol{a}}$ crucially depends on the specific form of the instrument we employ. If the respective outcomes do *not* correspond to a $\mathcal{M}_a$ of measure-and-prepare form, the joint probabilities $p(\boldsymbol{a}|d_{\mathrm{E}})$ do not factorize like in Eq. (75), and an alternative approach must be used.

References

[1] L. Accardi, A. Frigerio, and J. T. Lewis. Quantum Stochastic Processes. *Publ. Rest. Inst. Math. Sci.*, 18:97–133, 1982. DOI: 10.2977/prims/1195184017.

[2] Akshay Agrawal, Robin Verschueren, Steven Diamond, and Stephen Boyd. A rewriting system for convex optimization problems. *J. Control. Decis*, 5(1):42–60, 2018. DOI: 10.1080/23307706.2017.1397554.

[3] S. Alipour, M. Mehboudi, and A. T. Rezakhani. Quantum metrology in open systems: Dissipative cramér-rao bound. *Phys. Rev. Lett.*, 112:120405, Mar 2014. DOI: 10.1103/PhysRevLett.112.120405.

[4] Mario Berta, Francesco Borderi, Omar Fawzi, and Volkher B Scholz. Semidefinite programming hierarchies for constrained bilinear optimization. *Math. Program.*, 194:781–829, 2022. DOI: 10.1007/s10107-021-01650-1.

[5] Stephen Boyd and Lieven Vandenberghe. *Convex optimization.* Cambridge university press, 2004. ISBN 9780521833783. DOI: 10.1017/CBO9780511804441. URL https://web.stanford.edu/~boyd/cvxbook/.

[6] V. B. Braginsky and F. Y. Khalili. *Quantum Measurement.* Cambridge University Press, 1992. DOI: 10.1017/CBO9780511622748.

[7] Heinz-Peter Breuer and Francesco Petruccione. *The Theory of Open Quantum Systems.* Oxford University Press, 2002. ISBN 978-0-198-52063-4. DOI: 10.1093/acprof:oso/9780199213900.001.0001.

[8] Heinz-Peter Breuer, Elsi-Mari Laine, Jyrki Piilo, and Bassano Vacchini. Colloquium: Non-markovian dynamics in open quantum systems. *Rev. Mod. Phys.*, 88:021002, Apr 2016. DOI: 10.1103/RevModPhys.88.021002.

[9] Nicolas Brunner, Miguel Navascués, and Tamás Vértesi. Dimension witnesses and quantum state discrimination. *Phys. Rev. Lett.*, 110:150501, Apr 2013. DOI: 10.1103/PhysRevLett.110.150501.

[10] Adrián A. Budini. Embedding non-Markovian quantum collisional models into bipartite Markovian dynamics. *Phys. Rev. A*, 88(3): 032115, September 2013. DOI: 10.1103/PhysRevA.88.032115.

[11] Costantino Budroni and Clive Emary. Temporal quantum correlations and Leggett-Garg inequalities in multilevel systems. *Phys. Rev. Lett.*, 113:050401, Jul 2014. DOI: 10.1103/PhysRevLett.113.050401.

[12] Costantino Budroni, Gabriel Fagundes, and

Matthias Kleinmann. Memory cost of temporal correlations. *New J. Phys.*, 21(9):093018, sep 2019. DOI: 10.1088/1367-2630/ab3cb4.

[13] Costantino Budroni, Giuseppe Vitagliano, and Mischa P Woods. Ticking-clock performance enhanced by nonclassical temporal correlations. *Phys. Rev. Research*, 3(3):033051, 2021. DOI: 10.1103/PhysRevResearch.3.033051.

[14] Paul Busch, Pekka J. Lahti, and Peter Mittelstaedt. *The Quantum Theory of Measurement*, volume 2 of *Lecture Notes in Physics Monographs*. Springer-Verlag Berlin Heidelberg, 2 edition, 1996. DOI: 10.1007/978-3-540-37205-9.

[15] Carlton M. Caves, Christopher A. Fuchs, and Rüdiger Schack. Unknown quantum states: The quantum de Finetti representation. *J. Math. Phys.*, 43(9):4537–4559, 2002. DOI: 10.1063/1.1494475.

[16] Giulio Chiribella. On quantum estimation, quantum cloning and finite quantum de finetti theorems. In Wim van Dam, Vivien M. Kendon, and Simone Severini, editors, *Theory of Quantum Computation, Communication, and Cryptography*, pages 9–25, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. DOI: 10.1007/978-3-642-18073-6_2.

[17] Giulio Chiribella, Giacomo Mauro D'Ariano, and Paolo Perinotti. Theoretical framework for quantum networks. *Phys. Rev. A*, 80:022339, Aug 2009. DOI: 10.1103/PhysRevA.80.022339.

[18] Giulio Chiribella, Giacomo Mauro D'Ariano, Paolo Perinotti, and Benoit Valiron. Quantum computations without definite causal structure. *Phys. Rev. A*, 88:022318, Aug 2013. DOI: 10.1103/PhysRevA.88.022318.

[19] Man-Duen Choi. Completely positive linear maps on complex matrices. *Linear Algebra Its Appl.*, 10(3):285–290, 1975. ISSN 0024-3795. DOI: 10.1016/0024-3795(75)90075-0.

[20] Matthias Christandl, Robert König, Graeme Mitchison, and Renato Renner. One-and-a-half quantum de Finetti theorems. *Commun. Math. Phys.*, 273(2):473–498, 2007. DOI: 10.1007/s00220-007-0189-3.

[21] Luis A. Correa, Mohammad Mehboudi, Gerardo Adesso, and Anna Sanpera. Individual quantum probes for optimal thermometry. *Phys. Rev. Lett.*, 114:220405, Jun 2015. DOI: 10.1103/PhysRevLett.114.220405.

[22] C. L. Degen, F. Reinhard, and P. Cappellaro. Quantum sensing. *Rev. Mod. Phys.*, 89:035002, Jul 2017. DOI: 10.1103/RevModPhys.89.035002.

[23] Steven Diamond and Stephen Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *J. Mach. Learn. Res*, 17(83):1–5, 2016. DOI: 10.5555/2946645.3007036. URL https://dl.acm.org/doi/10.5555/2946645.3007036.

[24] A. C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Distinguishing separable and entangled states. *Phys. Rev. Lett.*, 88:187904, Apr 2002. DOI: 10.1103/PhysRevLett.88.187904.

[25] Andrew C. Doherty, Pablo A. Parrilo, and Federico M. Spedalieri. Complete family of separability criteria. *Phys. Rev. A*, 69:022308, Feb 2004. DOI: 10.1103/PhysRevA.69.022308.

[26] Clive Emary, Neill Lambert, and Franco Nori. Leggett–Garg inequalities. *Rep. Prog. Phys.*, 77(1):016001, dec 2013. ISSN 0034-4885. DOI: 10.1088/0034-4885/77/1/016001.

[27] Tobias Fritz. Quantum correlations in the temporal Clauser–Horne–Shimony–Holt (CHSH) scenario. *New J. Phys.*, 12(8):083055, 2010. DOI: 10.1088/1367-2630/12/8/083055.

[28] Mituhiro Fukuda, Masakazu Kojima, Kazuo Murota, and Kazuhide Nakata. Exploiting sparsity in semidefinite programming via matrix completion I: General framework. *SIAM J. Optim.*, 11(3):647–674, 2001. DOI: 10.1137/S1052623400366218.

[29] Rodrigo Gallego, Nicolas Brunner, Christopher Hadley, and Antonio Acín. Device-independent tests of classical and quantum dimensions. *Phys. Rev. Lett.*, 105:230501, Nov 2010. DOI: 10.1103/PhysRevLett.105.230501.

[30] Christina Giarmatzi and Fabio Costa. Witnessing quantum memory in non-Markovian processes. *Quantum*, 5:440, April 2021. ISSN 2521-327X. DOI: 10.22331/q-2021-04-26-440.

[31] Otfried Gühne, Costantino Budroni, Adán Cabello, Matthias Kleinmann, and Jan-Åke Larsson. Bounding the quantum dimension with contextuality. *Phys. Rev. A*, 89:062107, Jun 2014. DOI: 10.1103/PhysRevA.89.062107.

[32] Leonid Gurvits. Classical deterministic complexity of edmonds' problem and quantum entanglement. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, page 10–19, New York, NY, USA, 2003. Association for Computing Machinery. ISBN 1581136749. DOI: 10.1145/780542.780545.

[33] Otfried Gühne and Géza Tóth. Entanglement detection. *Phys. Rep.*, 474(1):1–75, 2009. ISSN 0370-1573. DOI: 10.1016/j.physrep.2009.02.004.

[34] Aram W Harrow. The church of the symmetric subspace. *arXiv:1308.6595*, 2013. URL https://arxiv.org/abs/1308.6595.

[35] Jannik Hoffmann, Cornelia Spee, Otfried Gühne, and Costantino Budroni. Structure of temporal correlations of a qubit. *New J. Phys.*, 20(10):102001, oct 2018. DOI: 10.1088/1367-2630/aae87f.

[36] Michał Horodecki, Paweł Horodecki, and Ryszard Horodecki. Mixed-state entanglement and distillation: Is there a "bound" entanglement

in Nature? *Phys. Rev. Lett*, 80:5239–5242, Jun 1998. DOI: 10.1103/PhysRevLett.80.5239.

[37] A. Jamiołkowski. Linear transformations which preserve trace and positive semidefiniteness of operators. *Rep. Math. Phys.*, 3(4):275–278, 1972. ISSN 0034-4877. DOI: 10.1016/0034-4877(72)90011-0.

[38] Hyejung H. Jee, Carlo Sparaciari, Omar Fawzi, and Mario Berta. Quasi-Polynomial Time Algorithms for Free Quantum Games in Bounded Dimension. In Nikhil Bansal, Emanuela Merelli, and James Worrell, editors, *48th International Colloquium on Automata, Languages, and Programming (ICALP 2021)*, volume 198 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 82:1–82:20, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. ISBN 978-3-95977-195-5. DOI: 10.4230/LIPIcs.ICALP.2021.82.

[39] J. K. Korbicz, J. I. Cirac, and M. Lewenstein. Spin squeezing inequalities and entanglement of n qubit states. *Phys. Rev. Lett.*, 95:120502, Sep 2005. DOI: 10.1103/PhysRevLett.95.120502.

[40] A. J. Leggett. Realism and the physical world. *Rep. Prog. Phys.*, 71(2):022001, jan 2008. ISSN 0034-4885. DOI: 10.1088/0034-4885/71/2/022001.

[41] A. J. Leggett and Anupam Garg. Quantum mechanics versus macroscopic realism: Is the flux there when nobody looks? *Phys. Rev. Lett.*, 54(9):857–860, mar 1985. DOI: 10.1103/PhysRevLett.54.857.

[42] Göran Lindblad. Non-Markovian quantum stochastic processes and their entropy. *Comm. Math. Phys.*, 65(3):281–294, 1979. DOI: 10.1007/BF01197883.

[43] I. A. Luchnikov, S. V. Vintskevich, and S. N. Filippov. Dimension truncation for open quantum systems in terms of tensor networks, jan 2018. URL http://arxiv.org/abs/1801.07418. arXiv:1801.07418.

[44] I. A. Luchnikov, S. V. Vintskevich, H. Ouerdane, and S. N. Filippov. Simulation Complexity of Open Quantum Dynamics: Connection with Tensor Networks. *Phys. Rev. Lett.*, 122(16):160401, apr 2019. DOI: 10.1103/PhysRevLett.122.160401.

[45] I. A. Luchnikov, E. O. Kiktenko, M. A. Gavreev, H. Ouerdane, S. N. Filippov, and A. K. Fedorov. Probing non-Markovian quantum dynamics with data-driven analysis: Beyond "blackbox" machine-learning models. *Phys. Rev. Res.*, 4(4):043002, October 2022. DOI: 10.1103/PhysRevResearch.4.043002.

[46] Yuanyuan Mao, Cornelia Spee, Zhen-Peng Xu, and Otfried Gühne. Structure of dimension-bounded temporal correlations. *Phys. Rev. A*,

105:L020201, Feb 2022. DOI: 10.1103/PhysRevA.105.L020201.

[47] Mohammad Mehboudi, Anna Sanpera, and Luis A Correa. Thermometry in the quantum regime: recent theoretical progress. *Journal of Physics A: Mathematical and Theoretical*, 52(30):303001, jul 2019. DOI: 10.1088/1751-8121/ab2828.

[48] Simon Milz and Kavan Modi. Quantum stochastic processes and quantum non-Markovian phenomena. *PRX Quantum*, 2:030201, Jul 2021. DOI: 10.1103/PRXQuantum.2.030201.

[49] Miguel Navascués, Masaki Owari, and Martin B. Plenio. Power of symmetric extensions for entanglement detection. *Phys. Rev. A*, 80:052306, Nov 2009. DOI: 10.1103/PhysRevA.80.052306.

[50] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. Conic optimization via operator splitting and homogeneous self-dual embedding. *J. Optim. Theory Appl*, 169(3):1042–1068, June 2016. DOI: 10.1007/s10957-016-0892-3.

[51] Brendan O'Donoghue, Eric Chu, Neal Parikh, and Stephen Boyd. SCS: Splitting Conic Solver, version 3.2.2. https://github.com/cvxgrp/scs, November 2022.

[52] Ognyan Oreshkov, Fabio Costa, and Časlav Brukner. Quantum correlations with no causal order. *Nat. Commun.*, 3(1):1092, Oct 2012. DOI: 10.1038/ncomms2076.

[53] Asher Peres. Separability criterion for density matrices. *Phys. Rev. Lett.*, 77:1413–1415, Aug 1996. DOI: 10.1103/PhysRevLett.77.1413.

[54] Felix A. Pollock, César Rodríguez-Rosario, Thomas Frauenheim, Mauro Paternostro, and Kavan Modi. Non-Markovian quantum processes: Complete framework and efficient characterization. *Phys. Rev. A*, 97:012127, Jan 2018. DOI: 10.1103/PhysRevA.97.012127.

[55] Ángel Rivas and Susana F Huelga. *Open Quantum Systems: An Introduction*. Springer Berlin, Heidelberg, 2011. ISBN 978-3-642-23353-1. DOI: 10.1007/978-3-642-23354-8.

[56] Ángel Rivas, Susana F Huelga, and Martin B Plenio. Quantum non-markovianity: characterization, quantification and detection. *Rep. Prog. Phys.*, 77(9):094001, aug 2014. DOI: 10.1088/0034-4885/77/9/094001.

[57] Carlos Sabín, Angela White, Lucia Hackermuller, and Ivette Fuentes. Impurities as a quantum thermometer for a Bose-Einstein condensate. *Sci. Rep.*, 4(1):1–6, 2014. DOI: 10.1038/srep06436.

[58] Greg Schild and Clive Emary. Maximum violations of the quantum-witness equality. *Phys. Rev. A*, 92:032101, Sep 2015. DOI: 10.1103/PhysRevA.92.032101.

[59] Paul Skrzypczyk and Daniel Cavalcanti. *Semidefinite Programming in Quantum Information Science*. 2053-2563. IOP Publish-

ing, 2023. ISBN 978-0-7503-3343-6. DOI: 10.1088/978-0-7503-3343-6.

[60] Adel Sohbi, Damian Markham, Jaewan Kim, and Marco Túlio Quintino. Certifying dimension of quantum systems by sequential projective measurements. *Quantum*, 5:472, June 2021. ISSN 2521-327X. DOI: 10.22331/q-2021-06-10-472.

[61] Cornelia Spee, Costantino Budroni, and Otfried Gühne. Simulating extremal temporal correlations. *New J. Phys.*, 22(10):103037, oct 2020. DOI: 10.1088/1367-2630/abb899.

[62] John K. Stockton, J. M. Geremia, Andrew C. Doherty, and Hideo Mabuchi. Characterizing the entanglement of symmetric many-particle spin-$\frac{1}{2}$ systems. *Phys. Rev. A*, 67:022112, Feb 2003. DOI: 10.1103/PhysRevA.67.022112.

[63] D. Tamascelli, A. Smirne, S. F. Huelga, and M. B. Plenio. Nonperturbative Treatment of non-Markovian Dynamics of Open Quantum Systems. *Phys. Rev. Lett.*, 120(3):030402, January 2018. DOI: 10.1103/PhysRevLett.120.030402.

[64] Armin Tavakoli, Alejandro Pozas-Kerstjens, Peter Brown, and Mateus Araújo. Semidefinite programming relaxations for quantum correlations. 2023. URL https://arxiv.org/abs/2307.02551.

[65] Barbara M. Terhal. Bell inequalities and the separability criterion. *Phys. Lett. A*, 271(5):319–326, 2000. ISSN 0375-9601. DOI: 10.1016/S0375-9601(00)00401-1.

[66] Géza Tóth, Tobias Moroder, and Otfried Gühne. Evaluating convex roof entanglement measures. *Phys. Rev. Lett.*, 114:160501, Apr 2015. DOI: 10.1103/PhysRevLett.114.160501.

[67] Lucas B. Vieira and Costantino Budroni. Temporal correlations in the simplest measurement sequences. *Quantum*, 6:623, 2022. DOI: 10.22331/q-2022-01-18-623.

[68] Giuseppe Vitagliano and Costantino Budroni. Leggett-garg macrorealism and temporal correlations. *Phys. Rev. A*, 107:040101, Apr 2023. DOI: 10.1103/PhysRevA.107.040101.

[69] John Watrous. *The Theory of Quantum Information.* Cambridge University Press, 2018. DOI: 10.1017/9781316848142.

[70] Henry Wolkowicz, Romesh Saigal, and Lieven Vandenberghe. *Handbook of semidefinite programming: theory, algorithms, and applications*, volume 27. Springer Science & Business Media, 2012. DOI: 10.1007/978-1-4615-4381-7.

[71] Shibei Xue, Matthew R. James, Alireza Shabani, Valery Ugrinovskii, and Ian R. Petersen. Quantum filter for a class of non-Markovian quantum systems. In *54th IEEE Conference on Decision and Control (CDC)*, pages 7096–7100, December 2015. DOI: 10.1109/CDC.2015.7403338.

[72] Shibei Xue, Thien Nguyen, Matthew R. James, Alireza Shabani, Valery Ugrinovskii, and Ian R. Petersen. Modeling for Non-Markovian Quantum Systems. *IEEE Trans. Control Syst. Technol.*, 28 (6):2564–2571, November 2020. ISSN 1558-0865. DOI: 10.1109/TCST.2019.2935421.

[73] Xiao-Dong Yu, Timo Simnacher, H. Chau Nguyen, and Otfried Gühne. Quantum-inspired hierarchy for rank-constrained optimization. *PRX Quantum*, 3:010340, Mar 2022. DOI: 10.1103/PRXQuantum.3.010340.

[74] Yang Zheng, Giovanni Fantuzzi, and Antonis Papachristodoulou. Chordal and factor-width decompositions for scalable semidefinite and polynomial optimization. *Annu. Rev. Control*, 52:243–279, 2021. ISSN 1367-5788. DOI: 10.1016/j.arcontrol.2021.09.001.