On the Nonconvexity of Dimensionally Limited Quantum and Classical Correlations

John Matthew Donohue[1,*] and Elie Wolfe[2,†]

[1] *Institute for Quantum Computing and Department of Physics & Astrophysics, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1*

[2] *Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada, N2L 2Y5*

Quantum theory is known to be non-local in the sense that separated parties (who do not signal to each other in any way) can perform measurements on a shared quantum state to obtain correlated multipartite probability distributions which cannot be achieved if the parties share only classical randomness. Our emphasis here is that the set of distributions compatible with sharing quantum states of some fixed dimension is neither convex nor a superset of the classical distributions when the quantum dimension is sufficiently constrained. We begin to clarify the relationship between those quantum distributions associated with some dimensional constraint and those classical distributions associated with limited shared randomness. We raise several foundational questions regarding quantum correlations in finite Hilbert spaces, including: Is there a quantum dimensional advantage in terms of simulating classical randomness? (Yes, sometimes.) Are quantum-constrained correlations ever convex for a fixed dimension? (Yes, sometimes.) Can there exist scenarios where quantum correlations are never convex for any finite dimension? (Maybe.)

A poignant illustration of the non-classicality of quantum mechanics is evidenced by the predictions of quantum theory in Bell scenarios. A Bell scenario consists of space-like separated parties who cannot communicate whatsoever but who have access to some previously-prepared shared state which they each individually measure. If the shared state is quantum, then the conditional probability distribution resulting from the simultaneous measurements of all the parties can be outside the set of distributions achievable by sharing classical randomness, i.e. incompatible with any local hidden variable model (LHVM) [1, 2]. For brevity, we shall refere to a conditional probability distribution as a "box", per the device-independent formalism [3–9]; the parties' choices of measurement are abstracted to inputs, and their measurement results are abstracted to outputs, as illustrated in Fig. 1. A box outputs $a, b, c..$ when given inputs $x, y, z..$ according to $p[abc...|xyz...]$; specifying a distribution in terms of such conditional probabilities equivalently *defines* a box.

From a strictly operational perspective, the internal mechanism of the box is hidden from the parties; the classical or quantum nature of the shared preparation is not specified, although information about the predicating resource can sometimes be inferred from the probabilities. Indeed, the study of non-locality is the characterizing and discriminating between two sets of boxes: boxes predicated on sharing quantum states and boxes predicated on sharing classical randomness [1, 2]. Non-local boxes are those which cannot be implemented classically, and are highly valued as the resources which drive device-independent quantum communication and quantum cryptography [10–12].

For any given Bell scenario the set of LHVM-compatible boxes is known as the local polytope. The (finite) facets of the local polytope are given by Bell inequalities [1]. A box is defined as non-local if and only if it violates one of the Bell inequalities. By contrast, we refer to

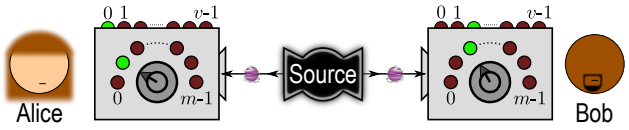


Figure 1. A schematic of the scenario we consider. n parties ($n = 2$ shown in this example) make synchronized measurements on systems prepared by a common source. Each party has m input (measurement setting) choices and v output (measurement result) possibilities. For $n = 2$, the parties are conventionally named Alice and Bob. Inputting x and y will return outputs a and b for Alice and Bob respectively with probability $p[ab|xy]$.

the set of boxes implementable by sharing any quantum state as the quantum elliptope. This is to indicate that it is convex but not describable in terms of a finite set of linear inequalities [13, see 1, Fig. 4]. To enforce no-signalling, i.e. to ensure that no information about the measurement choices of one party should be inferrable by the measurement results of another, the parties are assigned distinct Hilbert spaces[1] [18, 19]. Boxes inside the quantum elliptope thus have conditional probabilities of the form

$$p[abc...|xyz...] = \mathrm{Tr}\Big[\rho\,\Big(\hat{A}_{a|x}\otimes\hat{B}_{b|y}\otimes\hat{C}_{c|z}\,...\Big)\Big], \qquad (1)$$

where the dimensions of ρ and the local measurement operators $\hat{A}_{a|x}, \hat{B}_{b|y}$,... are unconstrained.

The operational boundary conditions delineating the quantum elliptope are notoriously hard to pin down, although various approximations to the quantum elliptope have recently become available [20–22]. Some properties

[1] *Technically* the measurement operators of the distinct parties need only commute with each other to form a genuine quantum multipartite implementation; the relegation of each party to a distinct Hilbert space is just a convenient trick to ensure appropriate commutativity, with apparently no loss of generality [14–17].

of the quantum elliptope, however, are unquestionable: (i) it is convex, (ii) the local polytope lies strictly inside it, (iii) all deterministic boxes reside on the boundary of the elliptope, and (iv) no box in the elliptope can be used to send a signal between parties [1]. The the first two properties are lost, however, when the dimension of the local Hilbert spaces is constrained [23]. Foundationally, the non-classical nature of quantum mechanics is richer in detail if we compare quantum preparations to classical preparations at finite sizes, as opposed to only in the asymptotic limits of the quantum elliptope and local polytope [24]. The aim of this Letter is to explore this finer characterization.

Although dimensionally constrained quantum correlations have already garnered considerable attention, most prior works have considered the parties as additionally having access to unlimited shared randomness. Examples include determining the degree to which Bell inequalities can be violated [25–27], assessing the security of a quantum cryptographic protocol [28–30], and testing if a given probability distribution is achievable [31–33]. The background assumption of unlimited shared randomness imposes convexity automatically, and obscures the fine-grained quantum-to-classical comparison which we seek. In contrast, our approach takes care to consider *purely* quantum systems and systems with limited shared randomness. Some important work has already been done in this framework. For example, Ref. [34] recently re-analyzed prepare-and-measure signalling scenarios without the assumption of unlimited shared randomness made in Ref. [35].

The minimal classical or quantum dimension required to achieve some unconditional joint probability distribution - limiting each party to a single input - has recently been equated with the distribution's non-negative or positive semi-definite rank, respectively [36, 37], and an arbitrary strong quantum dimensional advantage has been noted [38]. For the usual multiple-inputs Bell scenarios, the questions of minimizing the quantum dimension required required to implement a given box has recently been considered in Ref. [39], though see also Refs. [40, 41]. In this Letter we focus on the related foundational question of what minimal dimension is required for convexity. For classical correlations the degree of shared randomness necessary for convexity is surely finite, while for quantum correlations we show that the situation is much more complex.

Pál and Vértesi [23] notably called attention to the nonconvexity of some dimensionally constrained quantum correlations. We resolve the open question they posed, and extend the discussion to multipartite and multichotomous measurements. We show moreover that sharing merely qubits leads to non-convex correlations in *any* non-trivial Bell scenario. We also prove that, in some scenarios, quantum correlations are guaranteed to be convex if the underlying Hilbert space dimension is sufficiently large.

The operational implications of sharing dimensionally constrained quantum states, perhaps needless to say, has practical application. The local Hilbert space dimensions are often limited in real physical systems; for example spin-s particles have $d = 2s + 1$. Even in systems with an infinity of possible local states, practical quantum communication protocols typically only address a finite subset of states from each system. Our qualitative results herein regarding nonconvexity, and our quantitative results such as in Fig. 2, are novel dimension witnesses, and can be used for self-testing commercial equipment.

NOTATION AND FUNDAMENTAL AXIOM

For analytical clarity, in this Letter we consider exclusively symmetric Bell scenarios, where each of n space-like separated parties choose from among m inputs (measurement settings) and, in response, observe among v possible outcomes (measurement results), as illustrated in Fig. 1. We specify a symmetric Bell scenario by the index (n-m-v). The (n-2-2) scenarios, for example, are those for which every party has access to two binary-outcome observables. (2-2-2) is the familiar CHSH non-locality scenario [42, 43], and is the most fundamental scenario exemplifying our main results.

As is conventional, for (2-m-v) we shall refer to the two parties as Alice and Bob. We use x and y to indicate Alice and Bob's respective apparatus choices (box inputs), and use a and b to indicate their respective measurement outcomes (box outputs), starting all indices from 0. We always presume $v \geq 2$, as correlations between the n parties would be meaningless otherwise.

To avoid over-specifying a no-signalling box, we employ a box parameterization scheme which uses only the minimum number of conditional probabilities which allow implicit the full distribution to be reconstructed under the assumptions of normalization and no-signalling. One such parameterization (inspired by Acín *et al.* [44]) is based on reserving-as-implicit all probabilities involving the outcome 0. Since $\sum_{a=0}^{v-1} p[a|x] = 1$, one can implicitly define $p[0|x] \equiv 1 - \sum_{a=1}^{v-1} p[a|x]$. Similarly, since $\sum_{b=0}^{v-1} p[ab|xy] = p[a|x]$, one defines $p[a, 0 \mid x, y] \equiv p[a|x] - \sum_{b=1}^{v-1} p[ab|xy]$.

In this parameterization scheme, each of the $\binom{n}{k}m^k$ possible k-partite input tuples one might condition upon has $(v-1)^k$ freely specifiable output probabilities. Consequently, any no-signalling box is specified in terms of precisely

$$\mathcal{F} = \sum_{k=1}^{n} \binom{n}{k} m^k (v-1)^k = (m(v-1)+1)^n - 1 \qquad (2)$$

total explicit parameters, regardless of the choice of parameterization scheme [1, 45].

For further simplification, we shall herein only consider quantum measurement scenarios for which the local Hilbert spaces of every party have the same dimension, d. We denote the set of boxes achievable with such a dimensional constraint as $\mathcal{Q}_d^{(n\text{-}m\text{-}v)}$. In this notation, therefore, the quantum elliptope is identically $\mathcal{Q}_\infty^{(n\text{-}m\text{-}v)}$. We indicate $\mathcal{Q}_d^{(n\text{-}m\text{-}v)}$ via the shorthand $\mathcal{Q}_d$ when the scenario $(n\text{-}m\text{-}v)$ is clear from context.

It will also be important for us to consider the set of boxes achievable by sharing only a classical random variable λ (as opposed to sharing a quantum state ρ), as seen in Fig. 3(b). In parallel with the notion of constrained local Hilbert space dimension, we shall consider LHVM correlations where the dimension of the shared classical randomness, $|\lambda|$, might analogously be constrained, i.e. when the parties share no more than $\log_2|\lambda|$ classical bits in common. For example, if the parties share the outcome of rolling two distinguishable dice, then $|\lambda| = 36$. The set of classical boxes achieved using constrained shared randomness is denoted $\mathcal{L}_{|\lambda|}$.

Boxes predicated on classical randomness are identically mixtures of product distributions[2], i.e. $\mathcal{L}_{|\lambda|}$ is the set of boxes for which

$$p[ab...|xy...] = \sum_{\lambda=0}^{|\lambda|-1} p[\lambda]p[a|x\lambda]p[b|y\lambda]... \tag{3}$$

and thus $\mathcal{L}_\infty$ may be thought of as the convex hull of $\mathcal{L}_1$, where $\mathcal{L}_1$ is the set of boxes achievable without actually sharing any randomness. Indeed, $\mathcal{L}_\infty = \text{ConvexHull}\big[\mathcal{L}_1\big]$ is corollary of Fine's theorem [46–48].

While $\mathcal{L}_\infty = \text{ConvexHull}\big[\mathcal{L}_1\big]$ follows from Eq. (3), one can nevertheless span the local polytope without requiring infinite shared randomness. We define the minimum amount of shared randomness which allows every possible local box to be implemented as $|\lambda^\star|$:

Definition $|\lambda^\star|$: The Classical Carathéodory Number $|\lambda^\star|$ is the minimum amount of classical randomness that must be shared in order to span the local polytope. Formally,

$$|\lambda^\star| \equiv \min|\lambda| \text{ s.t. } \mathcal{L}_\infty^{(n\text{-}m\text{-}v)} = \mathcal{L}_{|\lambda|}^{(n\text{-}m\text{-}v)} \tag{4}$$

where $|\lambda^\star|$ intrinsically depends implicitly on $(n\text{-}m\text{-}v)$. We call $|\lambda^\star|$ is the classical Carathéodory number [49, 50] of $\mathcal{L}_1$ because every $P \in \text{ConvexHull}\big[\mathcal{L}_1\big]$ can be decomposed as a convex mixture of at-most $|\lambda^\star|$ boxes from $\mathcal{L}_1$, whereas $|\lambda^\star| - 1$ would not be adequate.

Details regarding the determination of $|\lambda^\star|$, including the result that $|\lambda^\star| = 4$ for the (2-2-2) scenario, can be found in the Supplementary Online Materials.

[2] The notion of sharing classical randomness can be generalized to random variables with "memory", but our analysis hews to the simplest model; see Ref. [1, Sec. II.G] for more details.

Note that boxes arising from measurements on separable quantum states are equivalent to boxes arising from shared classical randomness and vice versa. Critically, all classical boxes with degree of shared classical randomness $|\lambda|$ are subsumed by the set of boxes implemented quantumly by sharing (separable) qudits with local Hilbert space dimension $d \geq |\lambda|$. Formally,

Axiom (1) : $\mathcal{L}_{|\lambda|} \subseteq (\mathcal{Q} : \text{sep})_{d\geq|\lambda|}$, i.e. any box that can be constructed by sharing a classical dit can also be composed by sharing qudits.

Corollary (1a) : $\mathcal{L}_\infty \subseteq \mathcal{Q}_{d\geq|\lambda^\star|}$.

Corollary (1b) : If $\mathcal{L}_\infty \nsubseteq \mathcal{Q}_d$, then $\mathcal{Q}_d$ is not convex.

Proof. Every $P \in \mathcal{L}_{|\lambda|}$ can be mapped to a $P' \in (\mathcal{Q} : \text{sep})_{d=|\lambda|}$ by the following construction on Eq. (1): $\rho \to \sum_\lambda^{|\lambda|} p[\lambda](|\lambda\rangle\langle\lambda|)^{\otimes n}$, and $\hat{A}_{a|x} \to \sum_\lambda^{|\lambda|} p[a|x\lambda]|\lambda\rangle\langle\lambda|$ etc. Thus, for example, any box in $\mathcal{L}_1$ can be implemented quantumly by sharing a product state. The corollaries follow since $\mathcal{L}_\infty = \mathcal{L}_{|\lambda^\star|} = \text{ConvexHull}\big[\mathcal{L}_1\big]$. □

We conjecture that the shared quantum state must have local dimension $d \geq |\lambda^\star|$ in order to contain the local polytope, i.e. the converse of **Cor**. (1a), that $\mathcal{L}_\infty \nsubseteq \mathcal{Q}_{d<|\lambda^\star|}$. For that matter, we conjecture that the converse of **Ax**. (1) is also true, but this is a fundamental unproven open question.

Conjecture I : $\mathcal{L}_{|\lambda|\leq|\lambda^\star|} \nsubseteq \mathcal{Q}_{d<|\lambda|}$, i.e. the set of boxes which may be realized by sharing (possibly entangled) qudits of local Hilbert space dimension d is conjectured to *not* contain the set of boxes constructable by sharing a classical random variable of dimension $|\lambda| \leq |\lambda^\star|$ if $d < |\lambda|$.

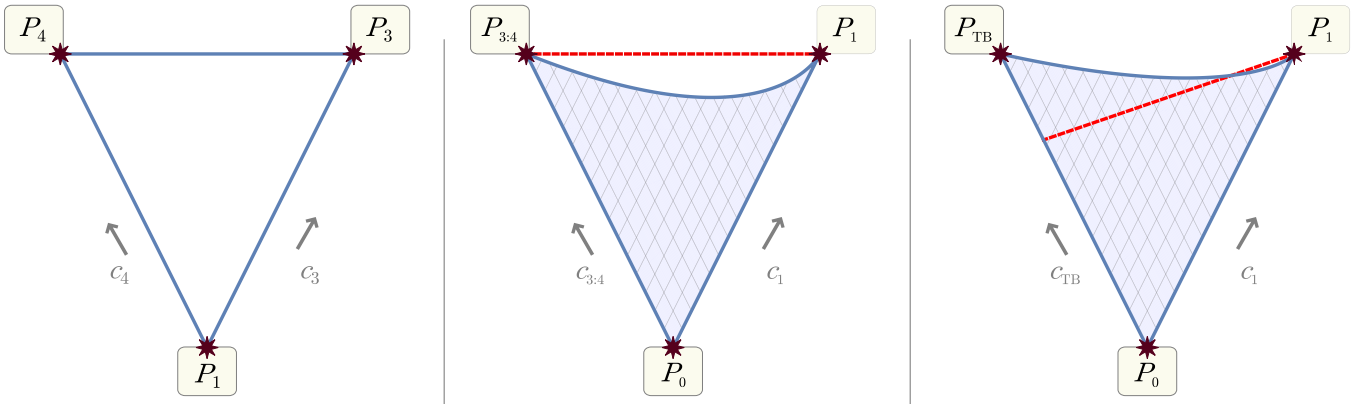
Note that **Conj**. I would imply the nonconvexity of $\mathcal{Q}_d$ whenever $d < |\lambda^\star|$, per **Cor**. (1b). One can use entropic measures to infer a trivial lower bound on the smallest d for which $\mathcal{L}_{|\lambda|} \subseteq \mathcal{Q}_d$ by comparing the largest Mutual Information which can be mediated through the given $|\lambda|$ to the maximum Total Correlation Capacity of quantum states of dimension d [51, Eq. (16)].

Eventually we will consider boxes achievable by sharing both a dimensionally-constrained quantum state and some classical randomness; see Fig. 3(c). We denote the set of boxes achievable using quantum systems of dimension d with the assistance of shared randomness of dimension $|\lambda|$ as $\mathcal{Q}_d + \mathcal{L}_{|\lambda|}$.

EXAMPLES OF QUANTUM NONCONVEXITY

Proposition (2) : $\mathcal{Q}_1$ is not convex.

Proof. In this trivial example there effectively isn't any shared resource at all, and the only randomness is local noise. Indeed, if $d = 1$ then the only possible quantum state that Alice and Bob can "share" is the one-dimensional identity, i.e. $|0\rangle\langle 0|_A \otimes |0\rangle\langle 0|_B$. No matter



(a) If $\widetilde{P} = c_1P_1 + c_3P_3 + c_4P_4$, then $\widetilde{P} \in \mathcal{L}_\infty$ iff $c_4 \leq 1 - c_3$, and $\widetilde{P} \in \mathcal{Q}_2$ iff $\{c_1 \text{ or } c_3 \text{ or } c_4\} = 0$.

(b) If $\widetilde{P} = c_0P_0 + c_1P_1 + c_{3:4}P_{3:4}$, then $\widetilde{P} \in \mathcal{L}_\infty$ iff $c_{3:4} \leq 1 - c_1$, and $\widetilde{P} \in \mathcal{Q}_2$ iff $c_{3:4} \leq (1 - c_1)^{3/2}$.

(c) If $\widetilde{P} = c_0P_0 + c_1P_1 + c_{\rm TB}P_{\rm TB}$, then $\widetilde{P} \in \mathcal{L}_\infty$ iff $c_{\rm TB} \leq 2^{-1/2}(1 - c_1)$, and $\widetilde{P} \in \mathcal{Q}_2$ iff $c_{\rm TB} \leq (1 - c_1)^{5/4}$.

Figure 2. Three examples of the nonconvexity of $\mathcal{Q}_2$. Each triangle represents the set of possible convex combination of its vertex boxes, ie. positively weighted mixtures such that $\sum_i c_i = 1$, where c_i represents the weight of P_i in the mixture. The named boxes referenced in these figures are defined in Table I. To emphasize the notion of convex combination we have plotted the regions as equilateral triangles; nevertheless any pair of edges should be thought of as two independent axes such that the third weight is fixed by normalization. Although in each triangle all three vertex boxes are achievable with shared-qubits preparations, only a fraction of possible convex combinations are also similarly achievable, as indicated by the shaded region. Indeed, in (a) only the zero-area edges of the triangle are qubit-achievable. The local polytope $\mathcal{L}_\infty$ contains the entire triangular regions of (a) and (b), since the vertex boxes of those triangle are all local, but in (c) only the region below the dashed red line is consistent with classical shared randomness. The essential feature of all three triangles is that the shaded subset representing $\mathcal{Q}_2$ is not convex. Incidentally, all three triangle further illustrate that $\mathcal{L}_\infty \not\subseteq \mathcal{Q}_2$. The $\mathcal{Q}_2$ region in (c) was obtained via computational maximization; $(1 - c_1)^{5/4}$ was later noticed to precisely coincide with the numeric boundary. Consistent with Ref. [23], the qubit-accessible region shrinks if one restricts to projective measurements. The edges where $c_1 \neq 0$ in (a), for example, are not in $\mathcal{Q}_2$-PVM, and in (c) we find that $\widetilde{P} \in \mathcal{Q}_2$-PVM only if $c_{\rm TB} \lesssim (1 - c_1)^{3/2}$. Thus the use of POVMs can be certified in a device-independent way given a dimension promise, akin the the results of Ref. [52].

how Alice and Bob choose their POVM elements, their joint probability distributions will always factorize to a product distribution, $p[ab|xy] = \mathrm{Tr}\big[\hat{A}_{a|x}\big]\,\mathrm{Tr}\big[\hat{B}_{b|y}\big] = p[a|x] \times p[b|y]$.

Now, suppose Alice and Bob both have access to a uniformly-distributed variable $0 \leq \lambda \leq v - 1$, and let their marginal probabilistic dependencies on λ be such that $p[a|x\lambda] \to \delta[a = \lambda]$ and $p[b|y\lambda] \to \delta[b = \lambda]$ regardless of x or y. The resulting box $P \in \mathcal{L}_v$ is such that $p[\lambda\lambda|xy] = p[\lambda|x] = p[\lambda|y] = v^{-1}$ where P does not factorize, i.e. $p[\lambda\lambda|xy] \neq p[\lambda|x] \times p[\lambda|y]$. As such, $P \in \mathcal{L}_\infty$ yet $P \notin \mathcal{Q}_1$. By **Cor.** (1b), then, $\mathcal{Q}_1$ is not convex. □

Proposition (3) : $\mathcal{Q}_2$ is not convex (for all $m \geq 2$). For example, in the (2-2-2) scenario, the set of boxes that can be achieved by sharing qubits is nonconvex.

Proof. We considered maximally general two-qubit preparations and measurement schemes, and found that convex combinations of various boxes in $\mathcal{Q}_2$ were not themselves achievable with qubits; see Fig. 2 for illustrative examples. Details of our state and measurements parametrization, along with additional examples of nonconvexity in a *Mathematica*™ notebook, may be found in the Supplementary Online Materials. □

Proposition (4) : $\big(\mathcal{L}_\infty \setminus \mathcal{L}_{|\lambda|=d}\big) \cap \mathcal{Q}_d \neq \emptyset$, i.e. there exist local boxes which classically require shared-randomness of dimension at least $|\lambda'|$, but which ad-

	$\langle A_0\rangle$	$\langle A_1\rangle$	$\langle B_0\rangle$	$\langle B_1\rangle$	$\langle A_0B_0\rangle$	$\langle A_1B_0\rangle$	$\langle A_0B_1\rangle$	$\langle A_1B_1\rangle$
P_0	0	0	0	0	0	0	0	0
P_1	1	1	1	1	1	1	1	1
P_2	-1	-1	-1	-1	1	1	1	1
P_3	1	-1	1	-1	1	-1	-1	1
P_4	-1	1	-1	1	1	-1	-1	1
$P_{3:4}$	0	0	0	0	1	-1	-1	1
$P_{1:4}$	0	0	0	0	1	0	0	1
$P_{\rm TB}$	0	0	0	0	$2^{-1/2}$	$2^{-1/2}$	$2^{-1/2}$	$-2^{-1/2}$

Table I. A list of bipartite conditional probability distributions, or boxes, in the (2-2-2) scenario. Per convention [1], we parameterize binary-output boxes in terms of outcome biases, ie. $\langle A_x\rangle = p[a{=}1|x] - p[a{=}0|x]$ and $\langle A_xB_y\rangle = p[a{=}b|xy] - p[a{\neq}b|xy]$. Note that the five boxes P_0 through P_4 are product distributions, achievable even absent any shared randomness, ie. $|\lambda| = 1$. $P_{i:j}$, by contrast, indicates the equally-weighted mixture of boxes P_i through P_j, requires non-trivial shared randomness, ie. $|\lambda| \geq 2$. Only $P_{\rm TB}$ is nonlocal; it is the quantum box which achieves the Tsirelson bound [43, 53]. Every box in this table is achievable with qubits, but many mixtures of these boxes *cannot* be achieved using qubits, per Fig. 2, hence illustrating the nonconvexity of $\mathcal{Q}_2$.

mit quantum "shortcuts" through implementations using quantum systems of smaller dimension $d < |\lambda'|$.

Proof. Consider the local box $P_{1:4}$ from Table I. If Alice and Bob choose the same input then their outputs are perfectly correlated, but if they choose different inputs then their outputs are unrelated. To achieve $P_{1:4}$ classically requires flipping two coins: When Alice or Bob input 0 their output is given by the first coin flip. The input 1, however, returns the second coin flip. Thus $P_{1:4}$ requires $|\lambda| \geq 4$, i.e. $P_{1:4} \in \mathcal{L}_\infty \setminus \mathcal{L}_3$. Alternatively, $P_{1:4}$ is quantumly achievable by sharing the entangled pure state $\frac{|00\rangle+|11\rangle}{\sqrt{2}}$ and letting $\hat{A}_{1|0} - \hat{A}_{0|0} = \hat{B}_{1|0} - \hat{B}_{0|0} = \hat{\sigma}_Z$ and $\hat{A}_{1|1} - \hat{A}_{0|1} = \hat{B}_{1|1} - \hat{B}_{0|1} = \hat{\sigma}_X$. So, $P_{1:4} \in \mathcal{Q}_2$. □

Although a box which requires $|\lambda| \geq 4$ is inside $\mathcal{Q}_2$, nevertheless $\mathcal{L}_4 \not\subseteq \mathcal{Q}_2$. Indeed, the empty interior of Fig. 2(a) shows that $\mathcal{L}_3 \not\subseteq \mathcal{Q}_2$, as is expected per **Conj**. I. Thus sharing quantum preparations enables not only nonlocality, ie. boxes forbidden classically, but also *super*-locality: quantum measurement schemes can occasionally reproduce boxes of higher corresponding shared randomness dimension, effectively simulating $|\lambda| > d$. See Zhang [38, Sec. 4.1] for a thorough discussion of both the existence and extent of super-locality in the special case (2-1-v). Interestingly, super-locality can occur even in the absence of entanglement [54].

AREN'T QUANTUM CORRELATIONS CONVEX?

The quantum elliptope, i.e. $\mathcal{Q}_\infty$, is convex. The convexity of the quantum elliptope, i.e. $\mathcal{Q}_\infty$, is well established; see for example Refs. [55, Sec. 5C] and [18, Sec. 5]; proof may also be given in terms of properties of C*-algebras [17, 56]. We have shown in **Prop**. (3), however, that for finite local Hilbert space dimension $\mathcal{Q}_d$ is sometimes not convex. The nonconvexity of $\mathcal{Q}_d$ was also demonstrated by Pál and Vértesi [23]. There is no contradiction between the convexity of $\mathcal{Q}_\infty$ and the nonconvexity of $\mathcal{Q}_d$; rather, convexification of quantum boxes requires either classical shared randomness or, equivalently, comes at the expense of increasing the local Hilbert space dimension.

The reason quantum boxes cannot be mixed without increasing Hilbert space dimension is because the measurements are local. The local nature of the measurement operators means that the composite POVM element associated with some global input and output $\hat{M}_{ab...|xy...} = \hat{A}_{a|x} \otimes \hat{B}_{b|y} \otimes ...$ is necessarily a product operator. Mixtures of product operators, sometimes known as separable superoperators [57, 58], are not generally product operators.

On the other hand, access to classical randomness allows for the quantum preparations and measurements to co-depend on a shared classical hidden variable. Indeed, any combination of N qudit-based boxes can be implemented by sharing a single qudit, *so long as the qudit is prepared according to a classical variable* λ, and the variable λ remains accessible to the measurements, i.e. per Fig. 3(c). By sharing supplementary classical randomness with $|\lambda| \geq N$ the quantum measurements can adequately reconfigure themselves to the appropriate basis depending on which qudit is prepared.

Explicitly, we imagine the quantum preparations and measurements to co-depend deterministically on λ, i.e. $p[\rho^{(\gamma)}|\lambda] = p[\hat{A}^{(\gamma)}_{a|x}|\lambda] = ... = \delta[\gamma - \lambda]$. By this construction, the resulting hybrid quantum-classical boxes in $\mathcal{L}_{|\lambda|} + \mathcal{Q}_d$ yields

$$p[ab...|xy...] = \sum_{\lambda=0}^{|\lambda|-1} p[\lambda] \, \mathrm{Tr}[\rho^{(\lambda)} \, \hat{A}^{(\lambda)}_{a|x} \otimes \hat{B}^{(\lambda)}_{b|y} \, ...] \qquad (5)$$

If we insist on recasting the convexification into a purely quantum implementation, then we can embed the shared randomness into the quantum state by constructing what's known as a classical-quantum (*cq*-) state [41, 59] which would have dimension $d \times |\lambda|$ instead of d. The cq-state is block diagonal in the larger Hilbert space.

Axiom (5) : $\mathcal{L}_{|\lambda|} + \mathcal{Q}_d \subseteq \mathcal{Q}_{d' \geq d \times |\lambda|}$, i.e. any hybrid quantum-classic box can be thought of as a purely-quantum box predicated on a sufficiently larger local Hilbert space dimension $d' = d \times |\lambda|$.

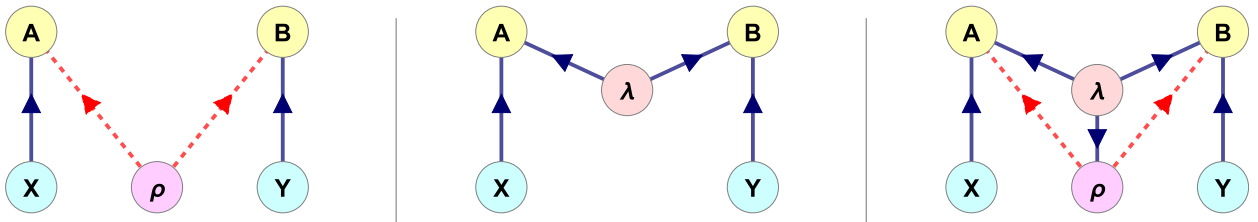
Proof. To quote Baez [60]: "Roughly speaking, if we have a physical system whose states are either states of $\rho^{(i)}$ OR states of $\rho^{(j)}$, its Hilbert space will be the direct sum $\mathcal{H}^{(i)} \oplus \mathcal{H}^{(j)}$." The direct sum of N Hilbert spaces each of dimension d can be embedded in a single Hilbert space of dimension $d \times N$ [61]. See the Supplementary Online Materials for further details. □

Granting that $\mathcal{Q}_d$ is presumed to be nonconvex, we can then ask how much supplementary shared randomness is required so that $\mathcal{L}_{|\lambda|} + \mathcal{Q}_d$ will be convex. Equivalently, what's the worst-case number of different $P \in \mathcal{Q}_d$ that must be combined in order to simulate *any* box which is expressible as a mixture of boxes from $\mathcal{Q}_d$? We rephrase that question into a definition.

Definition $|\lambda^\star_{\mathcal{Q}}|_d$: The Quantum Carathéodory Number $|\lambda^\star_{\mathcal{Q}}|^{(n\text{-}m\text{-}v)}_d$ is the minimum amount of supplementary classical correlation that must be supplied for $\mathcal{Q}^{(n\text{-}m\text{-}v)}_d$ to become completely convexified. Formally,

$$\begin{aligned} |\lambda^\star_{\mathcal{Q}}|_d \equiv \min \ |\lambda| \ \text{ such that} \\ \mathcal{L}_{|\lambda|} + \mathcal{Q}^{(n\text{-}m\text{-}v)}_d = \mathrm{ConvexHull}\big[\mathcal{Q}_d\big]. \end{aligned} \qquad (6)$$

We call $|\lambda^\star_{\mathcal{Q}}|_d$ the quantum Carathéodory number because Eq. (6) dictates that every $P \in \mathrm{ConvexHull}\big[\mathcal{Q}_d\big]$ can be decomposed as a convex mixture of at-most $|\lambda^\star_{\mathcal{Q}}|_d$ boxes from $\mathcal{Q}_d$, whereas $|\lambda^\star_{\mathcal{Q}}|_d - 1$ would not be adequate.



(a) Sharing a multipartite quantum state. **(b)** Sharing only classical randomness. **(c)** Sharing both resources.

Figure 3. Various causal structures associated with two space-like separated parties who make synchronized measurements on systems prepared by a common source, see Fig. 1. The joint distribution on the observed outcomes differs depending on the nature of the shared resource. If the preparation is quantum, such as in (a), then $p[ab|xy] = \mathrm{Tr}\big[\rho\, \hat{A}_{a|x} \otimes \hat{B}_{b|y}\big]$, per Eq. (1). If the preparation is classical, such as in (b), then $p[ab|xy] = \sum_\lambda p[\lambda] p[a|x\lambda] p[b|y\lambda]$, per Eq. (3). If the shared system is both classical and quantum, such as in (c), then $p[ab|xy] = \sum_\lambda p[\lambda]\, \mathrm{Tr}\big[\rho^{(\lambda)}\, \hat{A}^{(\lambda)}_{a|x} \otimes \hat{B}^{(\lambda)}_{b|y}\big]$, per Eq. (5). Hybrid preparations can equivalently be considered cq-states [41, 59].

Although $|\lambda^\star_{\mathcal{Q}}|_d$ may depend on d, there is a way to upper-bound the quantum Carathéodory number independently of d.

Theorem (6) : $\quad |\lambda^\star_{\mathcal{Q}}|_d \leq (m(v-1)+1)^n - 1\,.$

Proof. A 1929 theorem of Werner Fenchel [62, 63] states that any point within the convex hull of a not-necessarily-convex $\mathcal{F}$-dimensional closed and pathwise-connected set can be decomposed as a convex mixture of at-most $\mathcal{F}$ points in the set; Fenchel's theorem is a strengthening of Carathéodory's theorem [50] for the special case of indivisible sets such as pathwise-connected ones. Now, the *statistical space dimension* of no-signalling boxes in a symmetric Bell scenarios is $\mathcal{F}^{(n\text{-}m\text{-}v)} = (m(v-1)+1)^n - 1$ [1, 45], per Eq. (2). Consequently, $\mathcal{Q}_d \subseteq \mathcal{NS}$ is a pathwise-connected closed compactum of dimension $\mathcal{F}$; pathwise-connectedness follows from the continuous parameterization of both states and measurements and the Intermediate Value Theorem. □

For example, any box in $\mathrm{ConvexHull}\big[\mathcal{Q}_d^{(2\text{-}2\text{-}2)}\big]$ can be expressed as the convex combination of at-most eight qudit-based boxes. The actual quantum Carathéodory number may potentially be much lower than this upper bound, and is likely to have an explicit dependence on d. All we've established from **Prop**. (3) is that $|\lambda^\star_{\mathcal{Q}}|_2 \neq 1$. A tighter upper bound for $|\lambda^\star_{\mathcal{Q}}|_d$ than is given in **Thm**. (6) is desideratum for future research.

Importantly, the finiteness of $|\lambda^\star_{\mathcal{Q}}|_d$ can be used to guarantee the convexity of certain $\mathcal{Q}_d^{(n\text{-}m\text{-}v)}$ for finite d.

Theorem (7) : If for a given Bell scenario it so happens that $\mathcal{Q}_\infty = \mathrm{ConvexHull}\big[\mathcal{Q}_d\big]$, i.e. all extremal quantum distributions are achievable with qudits of dimension d, then $\mathcal{Q}_{d'}$ is convex if $d' \geq d \times |\lambda^\star_{\mathcal{Q}}|_d$.

Corollary (7a) : For (n-2-2) scenarios, $\mathcal{Q}_{d'}$ is convex if $d' \geq 2\,(3^n - 1)$, or assuming the upper bound of **Thm**. (6) is not tight, convexity is certain whenever $d' \geq |\lambda^\star_{\mathcal{Q}}|_d$.

Proof. By the definition of the quantum Carathéodory number, $\mathrm{ConvexHull}\big[\mathcal{Q}_d\big] = \mathcal{L}_{|\lambda^\star_{\mathcal{Q}}|_d} + \mathcal{Q}_d$. By **Ax**. (5) we can embed the supplementary classical correlations into the shared quantum state such that $\mathrm{ConvexHull}\big[\mathcal{Q}_d\big] \subseteq \mathcal{Q}_{d'}$ whenever $d' \geq d \times |\lambda^\star_{\mathcal{Q}}|_d$. However, the promise of $\mathcal{Q}_\infty = \mathrm{ConvexHull}\big[\mathcal{Q}_d\big]$ means that $\mathcal{Q}_\infty \subseteq \mathcal{Q}_{d'}$, from which it follows that $\mathcal{Q}_\infty = \mathcal{Q}_{d'}$. Thus, $\mathcal{Q}_{d'}$ is guaranteed to be convex.

The corollary is a consequence of Masanes' theorem [64, 65], which states that (projective) measurements on merely shared qubits are capable of achieving all extremal quantum distributions for scenarios involving two binary measurements per party [66]. Masanes' theorem is often cited when noting that the maximum violation of Bell inequalities for such scenarios can be computed by maximizing over qubit-based boxes. □

It is not clear if **Cor**. (7a) can be extended to more general scenarios. Consider (2-m-2) scenarios with $m \geq 3$, famous for the Bell inequality I_{3322} when $m = 3$ [67–69]. Pál and Vértesi [67] found that I_{3322} is apparently evermore-violated as d is increased. Granting the supposed *strict* hierarchy of $\mathcal{Q}_d$ for such scenarios, however, does not rule out $\mathcal{Q}_d$ being convex for large enough d; the strictness merely prevents using **Thm**. (7) to upper bound what constitutes "large enough". It is an open question if $\mathcal{Q}_d^{(2\text{-}3\text{-}2)}$ is convex for *any* finite d.

On the other hand there are additional Bell scenario for which the maximum violation of every Bell inequality is might be achieved for finite d. An example is the (2-2-3) scenario, famous for the CGMLP inequality [70]. There is numerical evidence that (2-2-m) scenarios achieve maximum nonlocality at $d \to m$ [21, 71]. The maximum violation of every Bell inequality, however, does not necessarily imply that all quantum extremal distributions have been achieved, so it is not certain that **Thm**. (7) applies. See Ref. [1, Sec. III.B] for further details.

Conjecture II : We conjecture that nonconvexity is nothing more than an artifact of not spanning the local polytope, i.e. that if $\mathcal{L}_{|\lambda^\star|} \subseteq \mathcal{L}_{|\lambda|} + \mathcal{Q}_d$ then $\mathcal{L}_{|\lambda|} + \mathcal{Q}_d$ is convex.

In stating **Conj**. II we used the equivalence $\mathcal{L}_\infty = \mathcal{L}_{|\lambda^\star|}$

per Eq. (4). Recall that the converse of **Conj**. II is obviously true: if a set of boxes – quantum, classical, or hybrid – does not contain the local polytope, then the set is not convex per **Cor**. (1b).

Conj. II amounts to speculating that $|\lambda^\star_{\mathcal{Q}}|_d = \min\ |\lambda|$ such that $\mathcal{L}_{|\lambda^\star|} \subseteq \mathcal{L}_{|\lambda|} + \mathcal{Q}_d$. If true, this would replace **Thm**. (6) with the claim $|\lambda^\star_{\mathcal{Q}}|_d \leq \lceil |\lambda^\star|/d \rceil$. This follows from $\mathcal{L}_d \subseteq \mathcal{Q}_d$ per **Ax**. (1) and then by $\mathcal{L}_{|\lambda^\star|} \subseteq \mathcal{L}_{\lceil |\lambda^\star|/d \rceil} + \mathcal{L}_d$, which merely notes that a random variable can be always be decomposed into multiple constituent parts. Equivalently, any integer $[1, N]$ can be mapped injectively to an ordered tuple $\{[1, M], [1, \lceil N/M \rceil]\}$.

To be clear, it is an open question if $\mathcal{Q}_d$ is *ever* convex for finite d in scenarios where **Thm**. (7) does not apply. **Conj**. II not only speculates the affirmative but effectively proposes a minimal value for what that finite d might be.

Finally, note that an implication of **Conj**. I and **Conj**. II combined is that $\mathcal{Q}_d$ should be convex if and only if $d \geq |\lambda^\star|$.

DISCUSSION

Questions remain even when still considering the (2-2-2) scenario. We've established that $\mathcal{L}_2^{(2\text{-}2\text{-}2)} \not\subseteq \mathcal{Q}_2$, but would qutrits be able to span the local polytope, or would $d = 4$ be required?[3] The insufficiency of qutrits is speculated in **Conj**. I, but this should certainly be investigated further.

Furthermore, although from **Cor**. (7a) is is clear that $\mathcal{Q}_{16}^{(2\text{-}2\text{-}2)}$ is convex, there's still a large gap between the not-convex result of $\mathcal{Q}_2$ and the yes-convex result of $\mathcal{Q}_{16}$. When exactly does convexity "kick in"? Could it be as early as $d = 4$? The convexity of $\mathcal{Q}_4^{(2\text{-}2\text{-}2)}$ is speculated in **Conj**. II, but this too should certainly be investigated further. For general scenarios where one presumes that **Thm**. (7) does not apply, such as (2-3-2) [67], we have noted it is a completely open question if $\mathcal{Q}_d$ is *ever* convex for finite d.

Most generally, given three descriptive elements: 1) an operational description of some Bell scenario such as (n-m-v), 2) the local Hilbert space dimension limit of d, and 3) the dimension limit of any shared classical randomness $|\lambda|$, are the resulting correlations $\mathcal{L}_{|\lambda|} + \mathcal{Q}_d$ convex? Such fundamental questions remained generally unanswered despite the broad results of **Thms**. (6) and (7). The purely classical regime is considered further, however, in the Supplementary Online Materials.

We have evidenced that nonconvexity should be expected in all non-locality scenarios with sufficiently restrictive constraints on the local Hilbert space dimensions. As many physical systems have implicit bounds on their local Hilbert space dimensions, it is all the more important to anticipate nonconvexity of quantum correlations in common practical quantum information-theoretic protocols. Purely quantum systems with appropriately small local Hilbert space dimensions are "forbidden" from accessing certain classical distributions. This promise can be exploited, for example as a device-authenticating security-check in quantum cryptographic implementations.

The nonconvexity of quantum correlations is no less relevant, practically, than other no-go results pertaining to finite Hilbert space dimensions. Quantifying the *genuine* boundaries of finite-dimensionally-generated quantum correlations, as opposed to the convex hull of such correlations [32, 33], is therefore an important question for future research. The apparent relationship between local Hilbert space dimension and degree of classical randomness is intriguing in this regard [72–74]. Perhaps classical information theory tools for considering limited shared randomness [75] can be adapted and applied to finite dimensional quantum systems. Fuller quantitative operational characterizations of finite dimension quantum systems are valuable desiderata for both foundational and practical agendas.

Acknowledgments We are grateful to Joshua Combes, Tobias Fritz, Ravi Kunjwal, Corsin Pfister, Marco Quintino, Sandu Popescu, Matt Pusey, Kevin Resch, Vincent Russo, and Rob Spekkens for valuable discussions, and to Matty Hoban for alerting us to a serious flaw in an early version of this manuscript. J. M. D. is grateful for support from the Natural Sciences and Engineering Research Council of Canada. Research at Perimeter Institute is supported by the Government of Canada through Industry Canada and by the Province of Ontario through the Ministry of Economic Development and Innovation.

[3] $|\lambda^\star|^{(2\text{-}2\text{-}2)} = 4$ is proven in the Supplementary Online Materials.

* jdonohue@uwaterloo.ca
† ewolfe@perimeterinstitute.ca

[1] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, "Bell Nonlocality," Rev. Mod. Phys. **86**, 419 (2014).
[2] S. Popescu, "Nonlocality Beyond Quantum Mechanics," Nat. Phys. **10**, 264 (2014).
[3] J. Barrett, N. Linden, S. Massar, S. Pironio, S. Popescu, and D. Roberts, "Nonlocal correlations as an information-theoretic resource," Phys. Rev. A **71**, 022101 (2005).
[4] J. Barrett and S. Pironio, "Popescu-Rohrlich Correlations as a Unit of Nonlocality," Phys. Rev. Lett. **95**, 140401 (2005).
[5] N. S. Jones and L. Masanes, "Interconversion of nonlocal correlations," Phys. Rev. A **72**, 052312 (2005).
[6] B. Jean-Daniel, B. Cyril, B. Nicolas, G. Nicolas, and L. Yeong-Cherng, "A Framework for the Study of Symmetric Full-Correlation Bell-Like Inequalities," J. Phys. A **45**, 125301 (2012).

[7] C. Miniatura, L. Kwek, M. Ducloy, B. Grémaud, B. Englert, L. Cugliandolo, and A. Ekert, *Lecture Notes of the Les Houches Summer School in Singapore: July 2009* (OUP Oxford, 2011); V. Scarani, "Quantum Information: Primitive Notions and Quantum Correlations," arXiv:0910.4222 (2009).

[8] V. Scarani, "The Device-Independent Outlook on Quantum Physics," Acta Physica Slovaca **62**, 347 (2012).

[9] J.-D. Bancal, *On the Device-Independent Approach to Quantum Physics* (Springer International Publishing, 2014).

[10] A. Acín, S. Massar, and S. Pironio, "Efficient Quantum Key Distribution Secure against No-Signalling Eavesdroppers," New J. Phys. **8**, 126 (2006).

[11] L. Masanes, S. Pironio, and A. Acín, "Secure Device-Independent Quantum Key Distribution with Causally Independent Measurement Devices," Nat. Comm. **2**, 238+ (2011).

[12] A. Ekert and R. Renner, "The Ultimate Physical Limits of Privacy," Nature **507**, 443 (2014).

[13] D. Avis and T. Ito, "Comparison of two bounds of the quantum correlation set," in *1st Inter. Conf. on Quant. Nano & Micro Tech.* (IEEE, 2007).

[14] V. B. Scholz and R. F. Werner, "Tsirelson's Problem," arXiv:0812.4305 (2008).

[15] M. Junge, M. Navascués, C. Palazuelos, D. Perez-Garcia, V. B. Scholz, and R. F. Werner, "Connes's Embedding Problem and Tsirelson's Problem," J. Math. Phys. **52**, 012102 (2011).

[16] M. Navascués, T. Cooney, D. Pérez-García, and N. Villanueva, "A Physical Approach to Tsirelson's Problem," Found. Phys. **42**, 985 (2012).

[17] T. Fritz, "Tsirelson's Problem and Kirchberg's Conjecture," Rev. Math. Phys. **24**, 1250012 (2012).

[18] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, "A Combinatorial Approach to Nonlocality and Contextuality," Comm. Math. Phys. **334**, 533 (2015).

[19] M. Navascués, Y. Guryanova, M. J. Hoban, and A. Acín, "Almost Quantum Correlations," Nat. Commun. **6**, 6288 (2015).

[20] M. Navascués, S. Pironio, and A. Acín, "Bounding the Set of Quantum Correlations," Phys. Rev. Lett. **98**, 010401 (2007).

[21] M. Navascués, S. Pironio, and A. Acín, "A Convergent Hierarchy of Semidefinite Programs Characterizing the Set of Quantum Correlations," New J. Phys. **10**, 073013 (2008).

[22] S. Pironio, M. Navascués, and A. Acín, "Convergent Relaxations of Polynomial Optimization Problems with Noncommuting Variables," SIAM J. Optim. **20**, 2157 (2010).

[23] K. F. Pál and T. Vértesi, "Concavity of the set of quantum probabilities for any given dimension," Phys. Rev. A **80**, 042114 (2009).

[24] G. Chiribella and X. Yuan, "Bridging the gap between general probabilistic theories and the device-independent framework for nonlocality and contextuality," arXiv:1504.02395 (2015).

[25] D. Pérez-García, M. Wolf, C. Palazuelos, I. Villanueva, and M. Junge, "Unbounded violation of tripartite bell inequalities," Comm. Math. Phys. **279**, 455 (2008).

[26] M. Junge, C. Palazuelos, D. Pérez-García, I. Villanueva, and M. M. Wolf, "Operator space theory: A natural framework for Bell inequalities," Phys. Rev. Lett. **104**, 170405 (2010).

[27] M. Epping, H. Kampermann, and D. Bruß, "Designing Bell inequalities from a Tsirelson bound," Phys. Rev. Lett. **111**, 240404 (2013).

[28] M. Zukowski and Č. Brukner, "Bell's theorem for general N-qubit states," Phys. Rev. Lett. **88**, 210401 (2002).

[29] I. Devetak and A. Winter, "Distillation of secret key and entanglement from quantum states," Proc. Roy. Soc. A **461**, 207 (2005).

[30] B. Kraus, N. Gisin, and R. Renner, "Lower and upper bounds on the secret-key rate for quantum key distribution protocols using one-way classical communication," Phys. Rev. Lett. **95**, 080501 (2005).

[31] N. Brunner, S. Pironio, A. Acin, N. Gisin, A. A. Méthot, and V. Scarani, "Testing the dimension of Hilbert spaces," Phys. Rev. Lett. **100**, 210503 (2008).

[32] M. Navascués, G. de la Torre, and T. Vértesi, "Characterization of quantum correlations with local dimension constraints and its device-independent applications," Phys. Rev. X **4**, 011011 (2014).

[33] M. Navascués and T. Vértesi, "Bounding the set of finite dimensional quantum correlations," arXiv:1412.0924 (2014).

[34] J. Bowles, M. T. Quintino, and N. Brunner, "Certifying the dimension of classical and quantum systems in a prepare-and-measure scenario with independent devices," Phys. Rev. Lett. **112**, 140407 (2014).

[35] R. Gallego, N. Brunner, C. Hadley, and A. Acín, "Device-independent tests of classical and quantum dimensions," Phys. Rev. Lett. **105**, 230501 (2010).

[36] R. Jain, Y. Shi, Z. Wei, and S. Zhang, "Efficient protocols for generating bipartite classical distributions and quantum states," IEEE Trans. Info. Theo. **59**, 5171 (2013).

[37] H. Fawzi, J. Gouveia, P. A. Parrilo, R. Z. Robinson, and R. R. Thomas, "Positive semidefinite rank," arXiv:1407.4095 (2014).

[38] S. Zhang, "Quantum strategic game theory," in *Proc. 3rd Innov. Theo. Comput. Sci. - ITCS '12* (ACM, New York, NY, USA, 2012) pp. 39–59, 429122.

[39] J. Sikora, A. Varvitsiotis, and Z. Wei, "On the minimum dimension of a Hilbert space needed to generate a quantum correlation," arXiv:1507.00213 (2015).

[40] N. Harrigan, T. Rudolph, and S. Aaronson, "Representing probabilistic data via ontological models," arXiv:0709.1149 (2007).

[41] S. Wehner, M. Christandl, and A. C. Doherty, "Lower bound on the dimension of a quantum system given measured data," Phys. Rev. A **78**, 062112 (2008).

[42] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, "Proposed Experiment to Test Local Hidden-Variable Theories," Phys. Rev. Lett. **23**, 880 (1969).

[43] E. Wolfe and S. F. Yelin, "Quantum Bounds for Inequalities Involving Marginal Expectation Values," Phys. Rev. A **86**, 012123 (2012).

[44] A. Acín, R. Gill, and N. Gisin, "Optimal bell tests do not require maximally entangled states," Phys. Rev. Lett. **95**, 210402 (2005).

[45] S. Pironio, "Lifting Bell Inequalities," J. Math. Phys. **46**, 062112 (2005).

[46] A. Fine, "Hidden variables, joint probability, and the bell inequalities," Phys. Rev. Lett. **48**, 291 (1982).

[47] R. Spekkens, "The status of determinism in proofs of the impossibility of a noncontextual model of quantum theory," Found. Phys. **44**, 1125 (2014).

[48] R. Kunjwal, "Fine's theorem, noncontextuality, and correlations in specker's scenario," Phys. Rev. A **91**, 022108 (2015).

[49] I. Bárány, "A generalization of Carathéodory's theorem," Discrete Mathematics **40**, 141 (1982).

[50] E. W. Weisstein, "Carathéodory's fundamental theorem," (2015), see also mathoverflow.net/q/77379.

[51] B. Groisman, S. Popescu, and A. Winter, "Quantum, classical, and total amount of correlations in a quantum state," Phys. Rev. A **72**, 032317 (2005).

[52] T. Vértesi and E. Bene, "Two-qubit bell inequality for which positive operator-valued measurements are relevant," Phys. Rev. A **82**, 062115 (2010).

[53] B. S. Cirel'son, "Quantum Generalizations of Bell's Inequality," Lett. Math. Phys. **4**, 93 (1980).

[54] G. K. Tong, Y. Zheng, Y. Zhen, K. Chen, J.-D. Bancal, and V. Scarani, "Entanglement witness of quantum system with trusted dimension," Unpublished.

[55] R. F. Werner and M. M. Wolf, "All-multipartite bell-correlation inequalities for two dichotomic observables per site," Phys. Rev. A **64**, 032112 (2001).

[56] N. Ozawa, "Tsirelson's problem and asymptotically commuting unitary matrices," J. Math. Phys. **54**, 032202 (2013).

[57] S. Bandyopadhyay and M. Nathanson, "Tight bounds on the distinguishability of quantum states under separable measurements," Phys. Rev. A **88**, 052313 (2013).

[58] S. Bandyopadhyay, A. Cosentino, N. Johnston, V. Russo, J. Watrous, and N. Yu, "Limitations on separable measurements by convex optimization," arXiv:1408.6981 (2014).

[59] M. Piani, P. Horodecki, and R. Horodecki, "No-local-broadcasting theorem for multipartite quantum correlations," Phys. Rev. Lett. **100**, 090502 (2008).

[60] J. Baez, "Tensor product and direct sum," See also ProofWiki:Hilbert_Space_Direct_Sum.

[61] J. K. Hunter and B. Nachtergaele, *Applied Analysis* (World Scientific, 2001).

[62] O. Hanner and H. Rådström, "A generalization of a theorem of Fenchel," Proc. Amer. Math. Soc. **2**, 589 (1951).

[63] I. Bárány and R. Karasev, "Notes about the Carathéodory number," Disc. Comp. Geom. **48**, 783 (2012).

[64] L. Masanes, "Extremal quantum correlations for N parties with two dichotomic observables per site," quant-ph/0512100 (2005).

[65] L. Masanes, "Asymptotic Violation of Bell Inequalities and Distillability," Phys. Rev. Lett. **97**, 050503 (2006).

[66] T. Fritz, "Polyhedral duality in bell scenarios with two binary observables," J. Math. Phys. **53**, 072202 (2012), 10.1063/1.4734586.

[67] K. F. Pál and T. Vértesi, "Maximal Violation of a Bipartite Three-Setting, Two-Outcome Bell Inequality using Infinite-Dimensional Quantum Systems," Phys. Rev. A **82**, 022116 (2010).

[68] Y.-C. Liang, T. Vértesi, and N. Brunner, "Semi-device-independent bounds on entanglement," Phys. Rev. A **83**, 022108 (2011).

[69] C. Palazuelos and Z. Yin, "Large bipartite Bell violations with dichotomic measurements," arXiv:1504.05769 (2015).

[70] D. Collins, N. Gisin, N. Linden, S. Massar, and S. Popescu, "Bell Inequalities for Arbitrarily High-Dimensional Systems," Phys. Rev. Lett. **88**, 040404 (2002).

[71] S. Zohren and R. D. Gill, "Maximal Violation of the Collins-Gisin-Linden-Massar-Popescu Inequality for Infinite Dimensional States," Phys. Rev. Lett. **100**, 120406 (2008).

[72] A. Acín, S. Massar, and S. Pironio, "Randomness versus nonlocality and entanglement," Phys. Rev. Lett. **108**, 100402 (2012).

[73] A. A. Methot and V. Scarani, "An anomaly of nonlocality," Quant. Info. Comp. **7**, 157 (2007).

[74] A. Plastino, G. Bellomo, and A. R. Plastino, "Quantum state space-dimension as a quantum resource," arXiv:1505.05455 (2015).

[75] R. Chaves, L. Luft, and D. Gross, "Causal structures from entropic information: geometry and novel scenarios," New J. Phys. **16**, 043001 (2014).

[76] E. W. Weisstein, "Stirling number of the second kind," (2015), see also math.stackexchange.com/a/264814.

[77] M. Horodecki, P. Horodecki, and R. Horodecki, "Separability of n-particle mixed states: necessary and sufficient conditions in terms of linear maps," Phys. Lett. A **283**, 1 (2001).

[78] R. Lockhart, "Optimal ensemble length of mixed separable states," J. of Math. Phys. **41**, 6766 (2000).

[79] A. Acín, A. Andrianov, L. Costa, E. Jané, J. I. Latorre, and R. Tarrach, "Generalized Schmidt decomposition and classification of three-quantum-bit states," Phys. Rev. Lett. **85**, 1560 (2000).

[80] D. Bruß, "Characterizing Entanglement," J. Math. Phys. **43**, 4237 (2002).

[81] L. Amico, A. Osterloh, and V. Vedral, "Entanglement in Many-Body Systems," Rev. Mod. Phys. **80**, 517 (2008).

[82] O. Gühne and G. Tóth, "Entanglement Detection," Phys. Rep. **474**, 1 (2009).

[83] T. Heinosaari, D. Reitzner, and P. Stano, "Notes on joint measurability of quantum observables," Found. Phys. **38**, 1133 (2008).

Bounding the Minimum Classical Shared Randomness which Spans the Local Polytope

In this section we establish some upper and lower bounds on $|\lambda^\star|$, as defined per Eq. (4) in the main text. We are grateful to Matt Pusey of the Perimeter Institute for Theoretical Physics for suggesting most of these proofs. The related question of how to find an explicit classical implementation of a box while minimizing $|\lambda|$ ("ontological compression") is considered in Ref. [40].

Proposition (8) : $|\lambda^\star| \leq v^{m(n-1)}$. In particular for the (2-2-2) scenario $|\lambda^\star| \leq 4$.

Proof. Let $n-1$ parties have all their measurements depend deterministically on λ, and only the n'th party depending probabilistically on λ. There are precisely $v^{m(n-1)}$ possible deterministic distributions among the $n-1$ parties, so without loss of generality is suffices to take $|\lambda| = v^{m(n-1)}$. In this manner every conceivable classical correlation is possible. □

Proposition (9) : $|\lambda^\star| \leq (m(v-1)+1)^n - 1$.

Proof. The proof is analogous to the proof of **Thm**. (6). $\mathcal{L}_1 \subseteq \mathcal{NS}$ is a pathwise-connected closed compactum of dimension $\mathcal{F}^{(n\text{-}m\text{-}v)} = (m(v-1)+1)^n - 1$ per Eq. (2). Fenchel's theorem [62, 63] then implies that the Carathéodory number of any box $\mathcal{L}_1$ is less than or equal to $\mathcal{F}$; i.e. any $P \in \text{ConvexHull}\big[\mathcal{L}_1\big]$ can be decomposed into a convex combination of at most $\mathcal{F}$ boxes from $\mathcal{L}_1$. $|\lambda^\star|$ is identically that Carathéodory of $\mathcal{L}_1$. □

Proposition (10) : $|\lambda^\star| \geq (m(v-1)+1)^{n-1} + \sum_{k=2}^{n}\sum_{j=2}^{k}\binom{n-1}{k-1}\binom{m}{j}S_2[k,j](j-1)!(v-1)^k$, where S_2 refers to Stirling number of the second kind [76]. For $n=2$, therefore, $|\lambda^\star| \geq \frac{m(m-1)(v-1)^2}{2} + m(v-1)+1$ and in particular for the (2-2-2) scenario $|\lambda^\star| \geq 4$.

Proof. Consider a box which is as random as possible while still satisfying

$$p[a+b+c... = 0 \mod v \mid xyz] = 1 \quad \text{iff} \quad x = y = z... \tag{A.1}$$

i.e. the outputs of the parties always (modularly) sum to zero when the inputs are aligned, but this perfect correlation is not detected whenever the inputs are not all aligned. One can then, given $n-1$ of the outputs, perfectly determine the remaining one as $a_x = -b_x - c_x ... \mod v$. The instances of perfect correlation enforce that there is no local noise, and thus that every party's output depends *deterministically* on the random variable. The degrees of freedom in the outputs of the first $n-1$ parties, given by Eq. (2) as $(m(v-1)+1)^{n-1}-1$, are all able to be set independently and yet must be determined solely by the shared random variable. We must give the shared variable an alphabet size equal to the number of degrees of freedom *plus one*, to account for normalization. Thus, $|\lambda^\star| \geq (m(v-1)+1)^{n-1}$.

This loose lower bound can be strengthened to the expression in **Prop**. (10) by counting the remaining degrees of freedom which include the last party's outcome. There are $\binom{n-1}{k-1}$ ways to choose a k-partite context involving the last party. We then need to assign inputs to the parties; as the case where all inputs are equal is already specified by the definition of the box, we only need to consider distributing $j>1$ distinct inputs among the k parties. There are $\binom{m}{j}$ ways to choose which j distinct inputs to distribute. The Stirling number $S_2[k,j]$ gives the number of ways k objects can be divided into j (indistinguishable) partitions. Each partition is assigned one input however, so we must consider the various permutations of mapping inputs to partitions. To avoid overspecifying the distribution we take the last party (and any other parties in the partition including the last party) to be associated with largest of the j selected inputs. Thus we only consider $(j-1)!$ permutations of how the smaller inputs can be assigned to the remaining partitions. Finally we specify the outputs, avoiding output 0 per the parameterization scheme discussed prior to Eq. (2). Thus there are $v-1$ distinct outputs considered for each of the k parties considered. □

For pedagogical clarity we demonstrate how to obtain $|\lambda^\star| \geq 7$ in this fashion for the (2-3-2) scenario, famous for the Bell inequality I_{3322} [67–69]. The classical box in (2-3-2) which satisfies $b_x{=}a_x$ has six degrees of freedom, namely $p[a_1{=}1], p[a_1{=}1], p[a_2{=}1], p[a_1{=}1, b_1{=}1], p[a_0{=}1, b_2{=}1], p[a_1{=}1, b_2{=}1]$. Trivially $p[a_x{=}0] = 1 - p[a_x{=}1]$ and $p[a_x{=}1, b_y{=}0] = p[a_x{=}1] - p[a_x{=}1, b_y{=}1]$ etc. Note that $p[b_y{=}i]$ is fixed as equal to $p[a_y{=}i]$. Furthermore, $p[a_x{=}i, b_y{=}j] = p[a_y{=}j, b_x{=}i]$, so we only ever enumerate $p[a_x{=}i, b_y{=}j]$ where $x<y$. Adding one to the six degrees of freedom yields $|\lambda^\star| \geq 7$.

A Complimentary Fundamental Axiom

Axiom (11) : $(\mathcal{Q}:\text{sep})_d \subseteq \mathcal{L}_{|\lambda|\geq d^n}$, i.e. all quantum correlations generated by separable states are also classically achievable given enough shared randomness, where enough means $|\lambda| \geq d^n$.

Corollary (11a) : Entanglement is required for non-locality.

Proof. Recall that by definition separable states can be written as $\rho_{\text{sep}} = \sum_i c^{(i)} \rho_A^{(i)} \otimes \rho_B^{(i)} \ldots$, and therefore satisfy $\text{Tr}\big[\rho_{\text{sep}} \left(\hat{A}_{a|x} \otimes \hat{B}_{b|y} \ldots\right)\big] = \sum_{i=1}^{i_{\max}} c^{(i)} \text{Tr}\big[\rho_A^{(i)} \hat{A}_{a|x}\big] \text{Tr}\big[\rho_B^{(i)} \hat{B}_{b|y}\big] \ldots$. Any separable state with local Hilbert space dimension d can be decomposed into a mixture of no-more-than d^n product states [77 Def. 6, 78 Thm. 2] so we can replace $i_{\max}$ with d^n without loss of generality. One can therefore map every $P \in (\mathcal{Q}:\text{sep})_d$ to a $P' \in \mathcal{L}_{|\lambda|=d^n}$ by the following construction on Eq. (3): $\lambda \to i$, $p[\lambda] \to c^{(i)}$, $p[a|x\lambda] \to \text{Tr}\big[\rho_A^{(i)} \hat{A}_{a|x}\big]$, etc. □

Tong *et al.* [54] have shown that merely $|\lambda| \geq d$ is *not* sufficient to classically simulate the correlations which result from separable states, i.e. that separable states can still manifest super-locality. In particuar, they show that the box $p[ab|xy] = \frac{2+(-1)^{a+b+xy}\sqrt{2}}{8}$ requires $|\lambda| > 2$, but can nevertheless be achieved by measurements on separable qubits.

Details of the Parameterization of Shared-Qubits Boxes in the (2-2-2) Scenario

We consider qubit-based boxes with explicit representations in terms of two-qubit states of arbitrary entanglement and general two-outcome POVMs. We need to consider exclusively pure states, per Ref. [39, Lemma 1]. As all pure states are equivalent to their Schmidt-decomposed form under local unitary transformations [79–82], it is thus sufficient to consider the state

$$|\psi\rangle = \cos\left(\frac{\alpha}{2}\right)|00\rangle + \sin\left(\frac{\alpha}{2}\right)|11\rangle \ , \ \alpha \in (0,\pi), \tag{A.2}$$

by folding the local degrees of freedom into the measurement operators.

Adapting the notation of [83], we express a general binary 0/1 outcome POVM element as

$$\hat{A}_{a|x} = \frac{1}{2}\left[\left(1+(-1)^{1-a}\kappa_{A_x}\right)\mathbb{1} + (-1)^{1-a}\eta_{A_x}\left(\vec{n}_{A_x}\cdot\vec{\sigma}\right)\right], \tag{A.3}$$

where $\vec{\sigma} = (\hat{\sigma}_X, \hat{\sigma}_Y, \hat{\sigma}_Z)$ is a vector of Pauli matrices and $\vec{n}_{A_x} = (\sin\theta_{A_x}\cos\phi_{A_x}, \sin\theta_{A_x}\sin\phi_{A_x}, \cos\theta_{A_x})$ is a unit vector defining a direction in the Bloch sphere in spherical coordinates. Bob's POVM elements are defined similarly. To ensure positivity of the POVM elements corresponding to both outputs, the following conditions must be met:

$$\forall_{A_x} : \eta_{A_x} - 1 \leq \kappa_{A_x} \leq 1 - \eta_{A_x}. \tag{A.4}$$

Note that Eq. (A.4) implies $0 \leq \eta_{k_x} \leq 1$. If both bounds in Eq. (A.4) are simultaneously saturated then $\kappa_{k_x} = 0$ and $\eta_{k_x} = 1$, and Eq. (A.3) represents a projection-valued measurement (PVM).

As we are concerned with two-outcome POVMs, it is conventional to parameterize boxes in terms of bias of the measurement outcome, i.e.

$$\langle A_x\rangle = \langle\hat{A}_{1|x}\rangle - \langle\hat{A}_{0|x}\rangle \tag{A.5}$$

$$\langle A_x B_y\rangle = \langle\hat{A}_{1|x}\hat{B}_{1|y}\rangle - \langle\hat{A}_{0|x}\hat{B}_{1|y}\rangle - \langle\hat{A}_{1|x}\hat{B}_{0|y}\rangle + \langle\hat{A}_{0|x}\hat{B}_{0|y}\rangle. \tag{A.6}$$

The four marginal and four joint biases (for all x and y options) parameterize the eight-dimensional conditional probability space for the (2-2-2) scenario, and relate to the 17 "backend" parameters of the state and measurements as

$$\langle A_x\rangle = \eta_{A_x}\cos(\alpha)\cos(\theta_{A_x}) + \kappa_{A_x} \ , \text{ and} \tag{A.7}$$

$$\begin{aligned}\langle A_x B_y\rangle =& \eta_{A_x}\eta_{B_y}\cos(\phi_{A_x}+\phi_{B_y})\sin(\alpha)\sin(\theta_{A_x})\sin(\theta_{B_y}) + \eta_{A_x}\eta_{B_y}\cos(\theta_{A_x})\cos(\theta_{B_y}) \\ &+ \eta_{A_x}\kappa_{B_y}\cos(\alpha)\cos(\theta_{A_x}) + \eta_{B_y}\kappa_{A_x}\cos(\alpha)\cos(\theta_{B_y}) + \kappa_{A_x}\kappa_{B_y}\,.\end{aligned} \tag{A.8}$$

Convexly Combining Quantum Boxes via Direct Sum of Hilbert Spaces

Suppose we wish to take the convex combination of N qudit-based multipartite boxes, i.e.

$$\widetilde{P} = \sum_{i=0}^{N-1} c_i P_i \quad \text{where} \quad P_i \in \mathcal{Q}_d \,. \tag{A.9}$$

Typically[4] $\widetilde{P} \notin \mathcal{Q}_d$. This can be understood as follows:

Index the local measurement operators of each of the N boxes being combined into $\widetilde{P}$ by $\hat{A}^{(i)}_{a|x}, \hat{B}^{(i)}_{b|y}$ etc. Now imagine – although entirely unjustified – that all the N boxes P_i are predicated on sharing the *same* composite quantum state, i.e. $\forall_i\, \rho_i = \rho_0$. In order to reproduce the marginal probabilities of $\widetilde{P}$ with a single quantum box (without supplementary shared randomness), we should take $\widetilde{\rho} = \rho_0$ and define the new measurement operators $\widetilde{\hat{A}_{a|x}} = \sum_{i=0}^{N-1} c_i\, \hat{A}^{(i)}_{a|x}$. This satisfies the requirement that $\widetilde{p}[a|x] = \sum_{i=0}^{N-1} c_i\, p_i[a|x]$ etc.

Unfortunately, choosing measurement operators to satisfy the single-partite marginal probabilities does not extend to satisfaction of the required bipartite joint probabilities. Following Eq. (1) we find that

$$p[ab|xy] = \mathrm{Tr}\big[\widetilde{\rho}\, \widetilde{\hat{A}_{a|x}} \otimes \widetilde{\hat{B}_{b|y}}\big] = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} c_i\, c_j\, \mathrm{Tr}\big[\rho_0\, \hat{A}^{(i)}_{a|x} \otimes \hat{B}^{(j)}_{b|y}\big] \tag{A.10}$$

which does not remotely match $\widetilde{p}[ab|xy] = \sum_i c_i\, p_i[ab|xy] = \sum_i c_i\, \mathrm{Tr}\big[\rho_0\, \hat{A}^{(i)}_{a|x} \otimes \hat{B}^{(i)}_{b|y}\big]$. Therefore quantum boxes predicated on different sets of local measurement operators cannot be combined without increasing the Hilbert space dimension[5]; as per **Ax**. (5), however, $\widetilde{P} \in \mathcal{Q}_{d\times N}$.

Imagine that each P_i is implemented in $\mathcal{H}^{(i)} = \big(\mathbb{C}^d\big)^{\otimes n}$ using the (distinct!) quantum states $\rho^{(i)}$ and local measurement operators $\hat{A}^{(i)}_x$, $\hat{B}^{(i)}_y$, etc. Then, to implement $\widetilde{P}$ using a single quantum box, one may take the direct sum of the component Hilbert spaces such that

$$\widetilde{\mathcal{H}} = \bigoplus_{i=0}^{N-1} \mathcal{H}^{(i)}\,, \quad \widetilde{\rho} = \bigoplus_{i=0}^{N-1} c_i\, \rho_i\,, \quad \widetilde{\hat{A}_{a|x}} = \bigoplus_{i=0}^{N-1} \hat{A}^{(i)}_{a|x}, \tag{A.11}$$

and so on. In the direct sum per Eq. (A.11), the $\mathcal{H}^{(i)}$ are orthogonal closed subspaces of $\widetilde{\mathcal{H}}$. $\widetilde{\mathcal{H}} \subset \big(\mathbb{C}^{d\times n}\big)^{\otimes N}$ in the sense that all $\widetilde{\rho} \in \widetilde{\mathcal{H}}$ are block-diagonal in $\big(\mathbb{C}^{d\times n}\big)^{\otimes N}$.

To explain how Eq. (A.11) automatically satisfies Eq. (A.9), and why the new local Hilbert space dimension in Eq. (A.11) can be thought of as $d \times N$, consider how $\widetilde{P}$ is implemented in $\Big(\big(\mathbb{C}^d\big)^{\otimes N}\Big)^{\otimes n}$. The idea is to assign an ancilla $\mathbb{C}^N$ to every party, and to make the new shared state simultaneously diagonalized in all the ancillas. Thus

$$\widetilde{\rho} = \sum_{i=0}^{N-1} c_i\, \rho_i \otimes |i\rangle\langle i|_A \otimes |i\rangle\langle i|_B \otimes \ldots \quad \text{and} \quad \widetilde{\hat{A}_{a|x}} = \sum_{i=0}^{N-1} \hat{A}^{(i)}_{a|x} \otimes |i\rangle\langle i|_A \quad \text{and} \quad \widetilde{\hat{B}_{b|y}} = \sum_{j=0}^{N-1} \hat{B}^{(j)}_{b|y} \otimes |j\rangle\langle j|_B, \tag{A.12}$$

and so on. Indeed, Eq. (A.12) amounts to the *definition* of the direct sum in Eq. (A.11). Thus the parties' new measurement operators now live in distinct Hilbert spaces N times larger than originally, and the new shared state is a cq-state [41, 59], namely a block-diagonal composition of the component original states in the convex combination per Eq. (A.9).

What we find is that *direct summation of Hilbert spaces is identically convexification* in the sense of Eq. (5). Since $\bigoplus^N \bigotimes^n \mathbb{C}^d \subset \bigotimes^N \bigotimes^n \mathbb{C}^d$, we have proven that $\mathcal{L}_N + \mathcal{Q}_d \subseteq \mathcal{Q}_{d\times N}$.

---

[4] The major exception is when $\mathcal{Q}_d$ is convex, as then $\widetilde{P} \in \mathcal{Q}_d$ by definition. Convexity might be "rare" though, see **Thm**. (7).

[5] Suppose we artificially consider global measurements of the form $\widehat{A''B''} = \frac{\hat{A}\otimes\hat{B} + \hat{A}'\otimes\hat{B}'}{2}$. Such not-local-but-still-separable measurements lack physical meaning, but they can be interpreted as representing the convex combination of boxes acting on the same shared state with different local, i.e. product state, measurements. Such artificial global measurements are known as *separable superoperators*, and have been studied elsewhere in the context of state discrimination [57, 58].