Contextuality in Entanglement-assisted One-shot Classical Communication

Shiv Akshar Yadavalli[1] and Ravi Kunjwal[2]

[1]Department of Physics, Duke University, Durham, North Carolina, USA 27708

[2]Centre for Quantum Information and Communication, Ecole polytechnique de Bruxelles, CP 165, Université libre de Bruxelles, 1050 Brussels, Belgium

October 13, 2022

We consider the problem of entanglement-assisted one-shot classical communication. In the zero-error regime, entanglement can increase the one-shot zero-error capacity of a family of classical channels following the strategy of Cubitt *et al.*, Phys. Rev. Lett. 104, 230503 (2010). This strategy uses the Kochen-Specker theorem which is applicable only to projective measurements. As such, in the regime of noisy states and/or measurements, this strategy cannot increase the capacity. To accommodate generically noisy situations, we examine the one-shot success probability of sending a fixed number of classical messages. We show that preparation contextuality powers the quantum advantage in this task, increasing the one-shot success probability beyond its classical maximum. Our treatment extends beyond Cubitt *et al.* and includes, for example, the experimentally implemented protocol of Prevedel *et al.*, Phys. Rev. Lett. 106, 110505 (2011). We then show a mapping between this communication task and a corresponding nonlocal game. This mapping generalizes the connection with pseudotelepathy games previously noted in the zero-error case. Finally, after motivating a constraint we term *context-independent guessing*, we show that contextuality witnessed by noise-robust noncontextuality inequalities obtained in R. Kunjwal, Quantum 4, 219 (2020), is sufficient for enhancing the one-shot success probability. This provides an operational meaning to these inequalities and the associated hypergraph invariant, the weighted max-predictability, introduced in R. Kunjwal, Quantum 3, 184 (2019). Our results show that the task of entanglement-assisted one-shot classical communication provides a fertile playground to study the interplay of the Kochen-Specker theorem, Spekkens contextuality, and Bell nonlocality.

Shiv Akshar Yadavalli: sy215@duke.edu
Ravi Kunjwal: ravi.kunjwal@ulb.be

Contents

1 Introduction

The problem of identifying the resources responsible for a quantum advantage over classical strategies in quantum information and computation is key to unlocking the potential of quantum technologies. Often, such resources are taken to be theory-dependent features like entanglement, coherence, incompatibility, or perhaps the exponential scaling of Hilbert space dimension with the number of quantum systems at hand. The nonclassicality witnessed by Bell violations [1, 2] makes it possible to identify a source of quantum advantage that can be assessed in a theory-independent fashion, relying only on empirical data

rather than internal features of the theory that generated the data. In contrast to the case of Bell nonlocality, Kochen-Specker (KS) contextuality [3], a notion of nonclassicality mathematically similar to Bell nonlocality, hasn't been as widely adopted as a theory-independent witness of quantum advantage. This is despite the existence of theoretical results on its relevance for quantum information and computation [4–7]. One reason for this is that it isn't robust to noise, unlike Bell nonlocality, making its experimental testability a matter of controversy [8, 9]

Recently, much work has been devoted to making contextuality a notion of nonclassicality that relies on empirical data without making assumptions about the representation of measurements (concerning, in particular, their sharpness [10–12]) in the theory generating the data [13, 14].[1] This noise-robust notion of contextuality due to Spekkens [13] has been shown to underlie several quantum information tasks such as parity-oblivious multiplexing, quantum random access codes, state discrimination, communication complexity, anomalous weak values, and state-dependent cloning [19–25]. These applications of Spekkens contextuality, though, have no counterpart in terms of KS-contextuality, leaving a gap in our understanding of how advantages from KS-contextuality can be turned into noise-robust advantages premised on Spekkens contextuality. This is in line with the spirit of Ref. [17], where the first noise-robust noncontextuality inequality, inspired by the Kochen-Specker theorem, was derived. Since the approach to noise-robust noncontextuality inequalities generalizes the KS paradigm by removing restrictions like projective measurements [17], it behooves us to ask if, and in what precise form, the advantages that derive from KS-contextuality persist when one considers noise-robust contextuality à la Spekkens [13, 26]. In this paper, we take the first steps in this research program using tools from previously proposed hypergraph frameworks [10, 27].

We consider the problem of one-shot classical com-

---

[1] Experimental testability of generalized contextuality – in particular, the need for tomographic completeness in order to verify operational equivalences – has been addressed in several recent papers [14–16] and we refer the interested reader to these for a discussion of such issues and how they are handled in the framework. Note, however, that if one assumes a quantum description of any experiment (as opposed to a general probabilistic theory – GPT – description), then verification of operational equivalences via tomography is straightforward as there is no ambiguity about the dimension of a well-characterized system's Hilbert space, hence about the minimum number of tomographically complete preparations/measurements needed to establish operational equivalences. This move – assuming a quantum description – is the natural one when considering applications of contextuality in quantum information as opposed to its foundational implications for quantum theory (which necessitates a broader framework like GPTs). On the other hand, even if one does assume a quantum description, Kochen-Specker contextuality cannot do justice to the case where this description involves non-projective measurements for reasons that have been extensively discussed elsewhere [10, 13, 17, 18].

munication where it has been shown that, assisted by entanglement, KS-contextuality provides an increase in the one-shot zero-error capacity of classical channels based on the KS theorem [3, 6]. We study a relaxation of this problem to one of enhancing the one-shot success probability of sending a fixed number of classical messages assisted by entanglement. Previous work [28, 29] has studied the one-shot success probability in the case of classical channels (unrelated to the KS theorem [3]) which do not admit an enhancement of zero-error channel capacity à la Cubitt *et al.* [6]. In contrast, we here study the one-shot success probability for the general case that includes, in particular, channels which *do* admit an enhancement of the one-shot zero-error channel capacity. A schematic of the task of entanglement-assisted one-shot classical communication is outlined in Fig. 1.

Our results can be summarized as follows:

- We show that preparation noncontextuality [13] relative to Bob's share of the system character-izes the classical upper bound on the task of en-hancing the one-shot success probability of a clas-sical channel assisted by entanglement.[2] Hence, preparation contextuality drives the quantum ad-vantage in the general task.

- We then show a mapping between the one-shot communication task and a corresponding non-local game: preparation contextuality powers an advantage in the first task if and only if Bell nonlocality powers an advantage in the sec-ond.[3] This generalizes the connection between entanglement-assisted one-shot zero-error capac-ity and nonlocal (pseudotelepathy) games noted in Ref. [6].

- We motivate a constraint on the communication task that we term *context-independent guessing* in a situation where the receiver (Bob) has no knowledge of (or doesn't trust) the exact chan-nel probabilities but knows only (or trusts only) the channel hypergraph. We then prove that, for some classical channels (including the one studied in Ref. [6]), the contextuality witnessed by a hypergraph invariant – the weighted max-predictability – implies an enhancement of the

---

[2]This does not require that the channel be based on the KS theorem in the sense of Cubitt *et al.* [6].

[3]Note that while the first task, by definition, requires com-munication via a classical channel—hence timelike separation, where the channel output (to the receiver) is within the fu-ture lightcone of the channel input (from the sender)—the sec-ond task, by definition, forbids all communication by requiring spacelike separation, where the sender and receiver are causally disconnected during each run of the nonlocal game. However, in both situations—timelike or spacelike separation—the shared common-cause resource doesn't result in signalling, i.e., we as-sume that this resource is described by a non-signalling theory (*e.g.*, shared entanglement in quantum theory) [30].
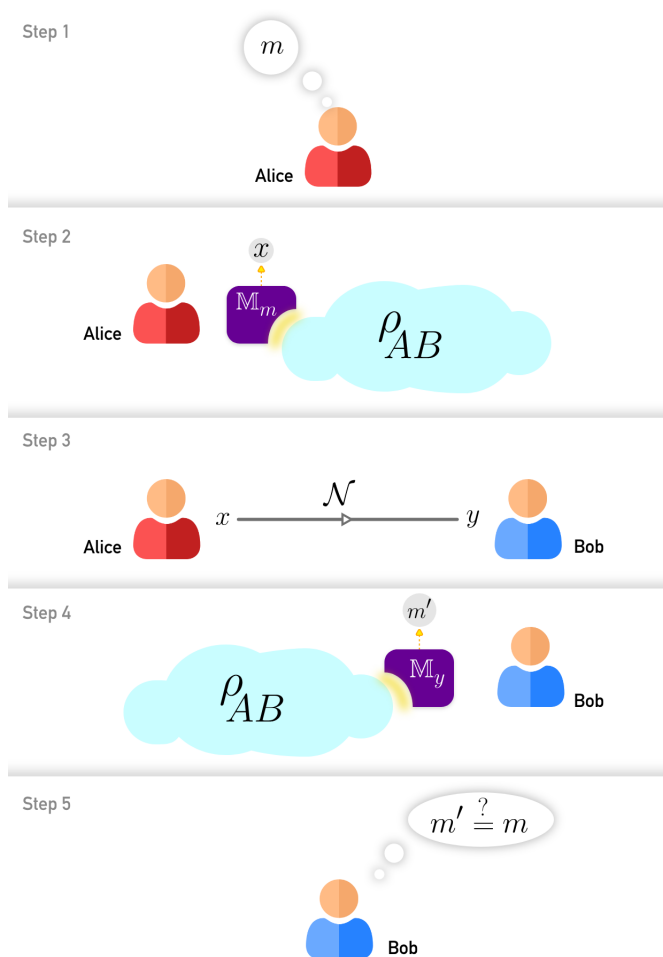
Figure 1: A schematic of the general protocol described in detail further on in Section 3.1. Alice and Bob are connected via a classical channel $\mathcal{N}$ and they share an entangled state ρ_{AB}. Once Alice decides to send a message m in Step 1, she encodes this message in her measurement choice in Step 2 and obtains outcome x. In Step 3, this outcome serves as the channel input and Bob obtains channel output y with probability $\mathcal{N}(y|x)$. Based on y, in Step 4, Bob measures his quantum system and, after some classical post-processing, obtains his guess m' for Alice's message m. They succeed if $m' = m$ (Step 5).

one-shot success probability in the communica-tion task. This makes direct connection with the formalism of Ref. [27], where weighted max-predictability provides an upper bound on the strength of source-measurement correlations un-der the assumption of noncontextuality. We thus provide an operational meaning to the violation of noise-robust noncontextuality inequalities in Ref. [27]: namely, such violations power the en-hancement of one-shot success probability of clas-sical communication assisted by entanglement.

The structure of the paper is as follows: In Sec-tion 2, we define some preliminary notions from the theory of one-shot classical communication as well as contextuality. In Section 3, we describe the general

protocol for entanglement-assisted one-shot classical communication that provides a unified description of protocols such as those of Refs. [6] and [28]. In Section 4, we discuss the resources that play a role in the quantum advantage in the communication task, including preparation contextuality and Bell nonlocality. In Section 5, we dive deep into the connection between the role of preparation contextuality in our communication task and the role of Bell nonlocality in a corresponding nonlocal game, proving some general relationships between them. In Section 6, we look at the problem of one-shot communication of a single bit through classical channels with complete confusability graphs, in particular the classical channel of Ref. [28]. In Section 7, we study the case of classical channels based on the KS theorem à la Ref. [6]. We conclude with a discussion in Section 8, mentioning some open problems and opportunities for future work.

2 Preliminaries

2.1 Classical Channels

Consider a discrete and memoryless classical channel $\mathcal{N}$ (*e.g.*, Fig. 2). Let X denote the set of input symbols of $\mathcal{N}$ and Y denote the set of output symbols so that $\{\mathcal{N}(y|x)\}_{x \in X, y \in Y}$ denotes the channel probabilities satisfying: $\mathcal{N}(y|x) \geq 0$ for all $x \in X, y \in Y$ and $\sum_{y \in Y} \mathcal{N}(y|x) = 1$ for all $x \in X$. Further, we denote by $Y_x \subseteq Y$ the set of output symbols that have a non-zero probability of occurrence when the input symbol is $x \in X$, i.e., the *support* of x, given by $Y_x \equiv \{y \in Y | \mathcal{N}(y|x) > 0\}$ for $x \in X$. Similarly, $X_y \subseteq X$ denotes the set of input symbols that yield a non-zero probability of occurrence for the output symbol $y \in Y$, i.e., the *support* of y, given by $X_y \equiv \{x \in X | \mathcal{N}(y|x) > 0\}$ for all $y \in Y$.

To the classical channel $\mathcal{N}$, we associate the channel hypergraph $H(\mathcal{N})$: vertices of $H(\mathcal{N})$ denote the input symbols $x \in X$ and hyperedges denote the output symbols $y \in Y$, such that each hyperedge representing $y \in Y$ contains the input symbols in X_y. Any two input symbols $x, x' \in X$ are said to be *confusable* when they share a hyperedge in $H(\mathcal{N})$, i.e., $Y_x \cap Y_{x'} \neq \varnothing$. The *confusability graph* $G(\mathcal{N})$ of the channel is given by the orthogonality graph of $H(\mathcal{N})$, i.e., its vertices are given by X and any two vertices in X are connected by an edge if and only if they are confusable.

Given the classical channel $\mathcal{N}$, Alice and Bob choose an *encoding* of the messages (say, $[q] \equiv \{m\}_{m=1}^q$) that Alice (the sender) wants to send to Bob (the receiver) through the channel. An *encoding* is a collection of mutually disjoint subsets of X.

More concretely, $\{X^{(m)}\}_{m=1}^q$ (where $X^{(m)} \subseteq X$ for all $m \in [q]$) is an encoding of the set of q messages in $[q]$ if and only if $X^{(m)} \cap X^{(m')} = \varnothing$ for all distinct $m, m' \in [q]$.[4]

A *zero-error code* is a set of input symbols $X_0 \subseteq X$ that are mutually non-confusable, i.e., no two symbols in this set can map to the same output symbol when fed into the channel $\mathcal{N}$. Hence, an encoding $\{X^{(m)}\}_{m=1}^q$ of the messages in $[q]$ is said to admit a zero-error code if and only if there exists a non-empty set

$$X_0 \subseteq \{x | x \in X^{(m)}, m \in [q]\}$$

such that $\forall x, x' \in X_0$, $Y_x \cap Y_{x'} = \varnothing$. The *one-shot zero-error capacity* of a classical channel is the number of messages that can be sent without error with one use of the channel, i.e., the cardinality of the largest zero-error code it admits. This is given by the independence number $\alpha(G(\mathcal{N}))$ of $G(\mathcal{N})$, namely, the cardinality of the largest set of vertices that share no edge in $G(\mathcal{N})$. Note that $\mathcal{N}$ does *not* admit a nontrivial zero-error code (i.e., with $q \geq 2$) if and only if $G(\mathcal{N})$ is a complete graph, i.e., $\alpha(G(\mathcal{N})) = 1$. Further, for any encoding $\{X^{(m)}\}_{m=1}^q$ that does admit a zero-error code using $\mathcal{N}$, we necessarily have $q \leq \alpha(G(\mathcal{N}))$. We illustrate the above notions in Fig. 2 with a simple example of a classical channel that was studied in Ref. [28].

In this paper we will consider the *one-shot success probability* for sending messages in a given encoding $\{X^{(m)}\}_{m=1}^q$ of q messages, where $q > \alpha(G(\mathcal{N}))$. Such an encoding does not admit a zero-error code, although the classical channel $\mathcal{N}$ may admit (smaller) encodings with zero-error codes, i.e., it may be that $\alpha(G(\mathcal{N})) > 1$. Of particular interest in this paper is a family of classical channels that we term *Kochen-Specker (KS) channels*. A KS channel is defined as any classical channel whose channel hypergraph satisfies the property of *KS-uncolourability* [27][31], i.e., it is impossible to assign a $\{0, 1\}$-valuation to the vertices such that the assignments in each hyperedge add up to 1.[6] A KS-uncolourable hypergraph is said to admit a *KS set* if it is possible to associate its vertices to projectors on a Hilbert space and hyperedges to projector-valued measures (PVMs), i.e., the projectors in any hyperedge are mutually orthogonal and sum up to the identity. Any set of such projectors for a KS-uncolourable hypergraph is called a *KS set*. We will consider two paradigmatic examples of KS channels, drawing upon Refs. [6, 28], only one of which [6] admits a KS set.

---

[4]Hence, given any input symbol $x^{(m)} \in X^{(m)}$, the message m can be uniquely inferred from $x^{(m)}$ since $X^{(m)} \cap X^{(m')} = \varnothing$ for all $m' \neq m$. In the interest of efficiency of encoding, we shall only consider encodings where each $X^{(m)}$ ($m \in [q]$) is a clique in $G(\mathcal{N})$.[5]

[6]What this means in terms of channel confusability is that it is impossible to pick a set of vertices with one vertex from each hyperedge such that all the vertices in the set are mutually non-confusable.
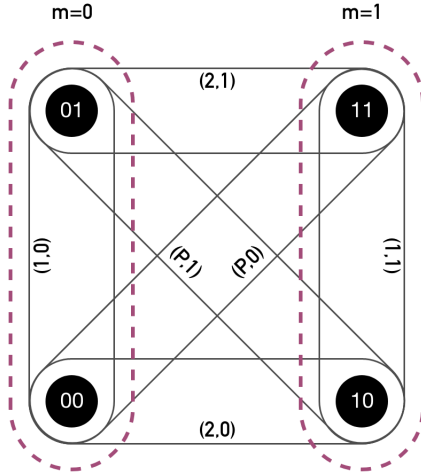
Figure 2: The channel hypergraph of the classical channel studied in Ref. [28] with the encoding indicated by dashed edges. Here $X = \{00, 01, 10, 11\}$ and $Y = \{(1,0), (1,1), (2,0), (2,1), (P,0), (P,1)\}$. The support of input 01, for example, is given by $Y_{01} = \{(1,0), (2,1), (P,1)\} \subseteq Y$ and the support of output $(P,0)$, for example, is given by $X_{(P,0)} = \{00, 11\} \subseteq X$. The channel probabilities are given by $\mathcal{N}(y|x) = 1/3$ for all $y \in Y_x$ and 0 otherwise for all $x \in X$. The one-bit encoding of $m \in \mathbf{Msg} = \{0, 1\}$ is given by $X^{(0)} = \{00, 01\} \subseteq X$ and $X^{(1)} = \{10, 11\} \subseteq X$. This encoding $\{X^{(0)}, X^{(1)}\}$ does not admit a zero-error code. In fact, a zero-error code doesn't exist for this channel hypergraph since it does not admit even a pair of inputs that are mutually non-confusable, i.e., do not share an edge.

2.2 Contextuality

We will be interested in the twin notions of preparation and measurement noncontextuality following Spekkens [13]. In a general operational theory, a preparation procedure consists of a source setting, $\mathbb{S}$, that prepares an ensemble of possible preparations indexed by source outcome $\mathrm{s} \in V_{\mathbb{S}}$, each with probability $p(\mathrm{s}|\mathbb{S})$. We denote the ensemble for source setting $\mathbb{S}$ by $\{(p(\mathrm{s}|\mathbb{S}), [\mathrm{s}|\mathbb{S}])\}_{\mathrm{s} \in V_{\mathbb{S}}}$. Formally, we refer to the (abstract) device implementing this preparation procedure as a multisource. A measurement procedure consists of a measurement setting $\mathbb{M}$ that yields one of possible outcomes indexed by $\mathrm{m} \in V_{\mathbb{M}}$ with probability $p(\mathrm{m}|\mathbb{M}, \mathbb{S}, \mathrm{s})$ when a system prepared according to $[\mathrm{s}|\mathbb{S}]$ is input to the measurement device. Formally, we refer to the (abstract) device implementing this measurement procedures as a multimeter. Together, the combination of source setting $\mathbb{S}$ and a measurement setting $\mathbb{M}$ yields conditional joint probability distribution given by $p(\mathrm{m}, \mathrm{s}|\mathbb{M}, \mathbb{S}) = p(\mathrm{m}|\mathbb{M}, \mathbb{S}, \mathrm{s})p(\mathrm{s}|\mathbb{S})$. (See Fig. 3.)

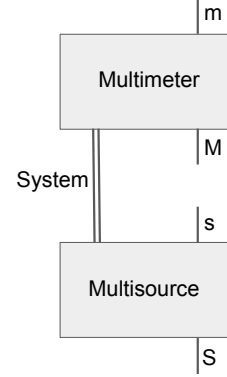Two source settings $\mathbb{S}$ and $\mathbb{S}'$ are said to be opera-

Figure 3: A generic prepare-and-measure setup. Classical inputs and outputs are indicated by single lines and the quantum/postquantum system passing from the multisource to the multimeter is indicated by double lines.

tionally equivalent if

$$\forall [\mathrm{m}|\mathbb{M}] :$$
$$\sum_{\mathrm{s}} p(\mathrm{m}, \mathrm{s}|\mathbb{M}, \mathbb{S}) = \sum_{\mathrm{s}'} p(\mathrm{m}, \mathrm{s}'|\mathbb{M}, \mathbb{S}'). \quad (1)$$

We will denote this operational equivalence by $\mathbb{S} \simeq \mathbb{S}'$. Two measurement events $[\mathrm{m}|\mathbb{M}]$ and $[\mathrm{m}'|\mathbb{M}']$ are said to be operationally equivalent if

$$\forall [\mathrm{s}|\mathbb{S}] : p(\mathrm{m}, \mathrm{s}|\mathbb{M}, \mathbb{S}) = p(\mathrm{m}', \mathrm{s}|\mathbb{M}', \mathbb{S}), \quad (2)$$

and we denote this by $[\mathrm{m}|\mathbb{M}] \simeq [\mathrm{m}'|\mathbb{M}']$.

In keeping with the original definition of generalized noncontextuality [13], and its further development in subsequent work [10], the notion of operational equivalence—for preparations and for measurements—is evaluated relative to *all* possible measurement and preparation events (respectively) in the operational theory governing the experiment of Fig. 3.

An ontological model of the operational theory consists of ontic states $\lambda \in \Lambda$ that are sampled by a preparation $[\mathrm{s}|\mathbb{S}]$ according to some probability distribution $\mu(\lambda|\mathbb{S}, \mathrm{s})$, so that $\mu(\lambda, \mathrm{s}|\mathbb{S}) = \mu(\lambda|\mathbb{S}, \mathrm{s})p(\mathrm{s}|\mathbb{S})$. Any measurement device responds to the input of an ontic state λ according to some probability, $\xi(\mathrm{m}|\mathbb{M}, \lambda)$, called a response function. The ontological model reproduces the operational statistics as follows:

$$p(\mathrm{m}, \mathrm{s}|\mathbb{M}, \mathbb{S}) = \sum_{\lambda} \xi(\mathrm{m}|\mathbb{M}, \lambda)\mu(\lambda, \mathrm{s}|\mathbb{S}). \quad (3)$$

The assumption of preparation noncontextuality entails the following implication:

$$\mathbb{S} \simeq \mathbb{S}' \Rightarrow \sum_{\mathrm{s}} \mu(\lambda, \mathrm{s}|\mathbb{S}) = \sum_{\mathrm{s}'} \mu(\lambda, \mathrm{s}'|\mathbb{S}'), \forall \lambda \in \Lambda,$$
$$\text{i.e., } \mu(\lambda|\mathbb{S}) = \mu(\lambda|\mathbb{S}'), \forall \lambda \in \Lambda. \quad (4)$$

The assumption of measurement noncontextuality entails the following implication:

$$[\mathbb{m}|\mathbb{M}] \simeq [\mathbb{m}'|\mathbb{M}'] \Rightarrow \xi(\mathbb{m}|\mathbb{M}, \lambda) = \xi(\mathbb{m}'|\mathbb{M}', \lambda), \forall \lambda \in \Lambda. \tag{5}$$

A failure of the joint assumption of preparation and measurement noncontextuality is then said to be a demonstration of contextuality.[7]

3 General protocol for one-shot classical communication assisted by a common-cause resource

3.1 One-shot success probability

We want to consider the situation where Alice and Bob have access to a shared common-cause resource such as quantum entanglement (but possibly also more general nonclassical common-cause resources [30, 33]) which they can use to enhance the one-shot success probability of sending messages through $\mathcal{N}$. This situation of entanglement-assisted one-shot classical communication has been studied previously [6, 28, 29]. We will below take the nonclassical common-cause resource to be quantum entanglement for ease of presentation but the ideas extend to post-quantum theories – corresponding, in general, to some convex subset of the set of no-signalling correlations – in a straightforward way.

The task at hand is the following: Alice and Bob share a classical channel $\mathcal{N}$ together with some bipartite quantum system in an entangled state ρ_{AB}. Alice wants to send messages from the set **Msg** according to a probability distribution $\{p(m)\}_{m \in \mathbf{Msg}}$. To do this for some encoding $\{X^{(m)}\}_{m \in \mathbf{Msg}}$ defined on $\mathcal{N}$, Alice implements a POVM $\mathbb{M}_m^{(\mathrm{A})} \equiv \{E_x^{(m)}\}_{x \in X}$ on her part, ρ_A, of the shared state ρ_{AB} given by $\rho_A = \mathrm{Tr}_B \rho_{AB}$. She obtains outcome $x \in X$ with probability $p(x|m) = \mathrm{Tr}(E_x^{(m)} \rho_A)$. She inputs x into the classical channel which yields output $y \in Y$ with probability $\mathcal{N}(y|x)$. Using the output y and his part of ρ_{AB}, Bob needs to figure out a strategy that will let him infer Alice's choice of measurement (hence the message m) with the maximum success probability, i.e., for every set of $\{m, x, y\}$ we want to maximize the probability $p(m' = m|y, m, x)$ that Bob's guess for the message, denoted m', agrees with the message Alice sent.

On receiving the channel output y, Bob implements some measurements, say $\{\mathbb{M}_v \equiv \{E_z^{(v)}\}_{z \in O}\}_{v \in V}$, on his part of the shared quantum system according to

some probability distribution, say $\{p(v|y)\}_{v \in V}$, that depends on y. The measurement outcome $z \in O$ occurs with probability $p(z|v, m, x) = \mathrm{Tr}(E_z^{(v)} \rho_x^{(m)})$ for the choice of POVM $\mathbb{M}_v$. Here $\rho_x^{(m)}$ is the state on Bob's side that Alice "steers" to when she obtains outcome x for measurement m with probability $p(x|m)$. Overall, Bob implements the effective POVM $\mathbb{M}'_y \equiv \sum_v p(v|y) \mathbb{M}_v$ given by the set of POVM elements $\{\sum_v p(v|y) E_z^{(v)}\}_{z \in O}$, where outcome z occurs with probability $p(z|y, m, x) \equiv \sum_v p(v|y) p(z|v, m, x)$. Bob's guess for the message, m', will then be a function of z, y, i.e., $m' = g(z, y)$, where g is a function from $O \times Y$ to **Msg**. We have a successful decoding when $m' = m$. This effectively defines the overall measurement $\mathbb{M}_y^{(\mathrm{B})}$ with outcomes $m' \in$ **Msg**, obtained by classical post-processing of the outcomes $z \in O$ of $\mathbb{M}'_y$, according to $p(m'|y, m, x) \equiv \sum_z \delta_{m', g(z, y)} p(z|y, m, x)$. The probability of a correct guess, $m' = m$, is then $p(m' = m|y, m, x) = \sum_z \delta_{m, g(z, y)} p(z|y, m, x)$.

The overall success probability is thus given by

$$S = \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(m'|y, m, x) \delta_{m, m'}. \tag{6}$$

In situations where Bob can do this perfectly (the zero-error regime), i.e., $S = 1$, we obtain an exact simulation of a noiseless classical channel **Id** with the same input and output alphabet **Msg** with channel probabilities $\mathbf{Id}(m'|m) = \delta_{m, m'}$ by using the noisy classical channel $\mathcal{N}$ and (potentially) some shared common-cause resource. A schematic of the protocol is provided in Fig. 1.

3.2 KS channels that admit KS sets: Cubitt *et al.* strategy

For some KS channels admitting KS sets (e.g., Fig. 4), the Cubitt *et al.* strategy [6] corresponds to choosing **Msg** $\equiv \{m\}_{m=1}^q$, $V \equiv Y$, $p(v|y) \equiv \delta_{v, y}$, $O \equiv X$ for all $v \in V$, (hence) $z \equiv x' \in X$, and $g(z, y) = g(x') = m'$, where m' is the unique message such that $x' \in X^{(m')}$. Thus, $p(m'|y, m, x) = \sum_{x'} \delta_{m', g(x')} p(x'|y, m, x)$. With these choices, we have

$$S = \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(m'|y, m, x) \delta_{m, m'}$$
$$= \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{x' \in X^{(m)}} p(x'|y, m, x). \tag{7}$$

---

[7]The assumption of noncontextuality is an instance of the Leibnizian idea of the ontological identity of operational indiscernibles [32]: operationally equivalent experimental procedures admit ontologically equivalent representations under this assumption.

As shown in Ref. [6], such channels admit an enhancement of their one-shot zero-error classical capacity in the presence of shared entanglement, i.e., their one-shot success probability $S = 1$ for some $q > \alpha(G(\mathcal{N}))$. We will discuss these channels in more detail in Section 7.

3.3 KS channels that do not admit KS sets

In general, a KS channel may not admit any KS set and in that case the strategy of Cubitt *et al.* [6] does not apply. An example of a KS channel that does not admit a KS set was studied for its one-shot success probability by Prevedel *et al.* [28]. This example fits within the general protocol described in Section 3.1 and will be discussed in Section 6.

4 Quantum advantage in one-shot classical communication

4.1 Classical one-shot success probability

Classically, the only information that Bob has about Alice's measurement and its outcome, i.e., m and x, is mediated by the output of the channel, y. We therefore have $p(m'|y, m, x) = p(m'|y)$, i.e., m' is conditionally independent of m and x, given y. Shared randomness does not help because it only amounts to a convex mixture of deterministic classical strategies (indexed by, say, $c \in \mathcal{C}$ according to some probability distribution $\{p(c)\}_{c \in \mathcal{C}}$ and no such convex mixture can do better than the best deterministic classical strategy. The classical one-shot success probability is therefore given by

$$
S_{\mathrm{Cl}} = \sum_{c \in \mathcal{C}} p(c) \left(\sum_m p(m) \sum_x p(x|m, c) \sum_y \mathcal{N}(y|x) \right.
$$
$$
\left. \sum_{m'} p(m'|y, c) \delta_{m,m'} \right), \tag{8}
$$

with the tight upper bound

$$
S_{\mathrm{Cl}}^{\max} \equiv \max_{c \in \mathcal{C}} \left(\sum_m p(m) \sum_x p(x|m, c) \sum_y \mathcal{N}(y|x) \right.
$$
$$
\left. \sum_{m'} p(m'|y, c) \delta_{m,m'} \right), \tag{9}
$$

so that

$$
S_{\mathrm{Cl}} \leq S_{\mathrm{Cl}}^{\max}. \tag{10}
$$

4.2 Preparation contextuality drives the quantum advantage

In the protocol we have described, the following operational equivalences hold: $\sum_x p(x|m_1)\rho_x^{(m_1)} =$ $\sum_x p(x|m_2)\rho_x^{(m_2)} \equiv \rho_B$ for all pairs of distinct messages $m_1, m_2 \in \mathbf{Msg}$.[8] This follows from the fact that the common-cause correlations shared between Alice and Bob must be nonsignalling: Alice's choice of POVM $\mathbb{M}_m$ encoding the message m ($m \in \mathbf{Msg}$) steers Bob's system to the ensemble of states $\{(p(x|m), \rho_x^{(m)})\}_{x \in X^{(m)}}$; however, on coarse-graining, the reduced state on Bob's side, $\rho_B = \mathrm{Tr}_A \rho_{AB}$, is the same for all choices of m, and thus the common-cause correlations cannot be used by Bob to infer m.

Preparation noncontextuality then entails that

$$
\sum_s p(x|m_1)p(\lambda|m_1, x) = \sum_x p(x|m_2)p(\lambda|m_2, x)
$$

$$
\sum_x p(x|m_1)p(\lambda|m_1, x) = \sum_x p(x|m_2)p(\lambda|m_2, x)
$$

for all $\lambda \in \Lambda$, for all $m_1, m_2 \in \mathbf{Msg}$. That is, $p(\lambda|m_1) = p(\lambda|m_2) \equiv p(\lambda)$ for all $\lambda \in \Lambda$, where Λ is the ontic state space of Bob's system.

Given the prior distribution $\{p(m)\}_{m \in \mathbf{Msg}}$ and the channel probabilities $\{\mathcal{N}(y|x)\}_{y,x}$, we obtain, under the assumption of preparation noncontextuality for Bob's system, the following expression for the one-shot success probability and the preparation noncontextual upper bound on it:

$$
S_{\mathrm{PNC}}
$$
$$
= \sum_\lambda p(\lambda) \left(\sum_m p(m) \sum_x p(x|m, \lambda) \sum_y \mathcal{N}(y|x) \right.
$$
$$
\left. \sum_{m'} p(m'|y, \lambda) \delta_{m,m'} \right)
$$
$$
\leq \max_\lambda \left(\sum_m p(m) \sum_x p(x|m, \lambda) \sum_y \mathcal{N}(y|x) \right.
$$
$$
\left. \sum_{m'} p(m'|y, \lambda) \delta_{m,m'} \right) \equiv S_{\mathrm{PNC}}^{\max}. \tag{11}
$$

[8]Although we are using quantum notation here, operationally, this prepare-and-measure setup on Bob's system (see Fig. 3) can be viewed as a multisource with settings $\mathbb{S} = m \in \mathbf{Msg}$ and outcomes $s = x \in X$ that occur with probability $p(s|\mathbb{S}) = p(x|m)$. The operational equivalences can then be expressed as in Eq. 1, i.e., $\forall [\mathbb{m}|\mathbb{M}]$, $\sum_{s_1} p(\mathbb{m}, s_1|\mathbb{M}, \mathbb{S}_1) = \sum_{s_2} p(\mathbb{m}, s_2|\mathbb{M}, \mathbb{S}_2)$, once we compute the probability $p(\mathbb{m}|\mathbb{M}, \mathbb{S}, s) = \mathrm{Tr}(\rho_s^{(\mathbb{S})} E_\mathbb{m}^{(\mathbb{M})})$ for all effects $E_\mathbb{m}^{(\mathbb{M})}$ representing measurement events $[\mathbb{m}|\mathbb{M}]$. That is, $\sum_x p(x|m_1)\rho_x^{(m_1)} = \sum_x p(x|m_2)\rho_x^{(m_2)}$ translates to $\sum_x p(x|m_1)\mathrm{Tr}(\rho_x^{(m_1)} E_\mathbb{m}^{(\mathbb{M})}) = \sum_x p(x|m_2)\mathrm{Tr}(\rho_x^{(m_2)} E_\mathbb{m}^{(\mathbb{M})})$ for all $[\mathbb{m}|\mathbb{M}]$. The multimeter has settings $\mathbb{M}$ and outcomes $\mathbb{m}$ that range over the set of all measurement events in the operational theory. Note, however, that as long as we assume the shared common-cause resource is non-signalling (as is the case with entangled states in quantum theory), we do not need to explicitly verify these operational equivalences by varying over all measurement events: these equivalences are implied by the non-signalling nature of the shared common-cause resource.

To see how this comes about, note that the one-shot success probability,

$$S = \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(m'|y,m,x)\delta_{m,m'},$$

when expressed in terms of an ontological model for Bob's system, requires that $p(m'|y,m,x) = \sum_\lambda p(m'|y,\lambda)p(\lambda|m,x)$. We can then write a joint probability distribution $p(x,\lambda|m) = p(\lambda|m,x)p(x|m)$, which can be rewritten as $p(x,\lambda|m) = p(x|m,\lambda)p(\lambda|m)$. Recalling that preparation noncontextuality requires $p(\lambda|m) = p(\lambda)$ for all m, we obtain the expression for S_{PNC}. Hence, we have: $S_{\mathrm{PNC}}^{\max} = S_{\mathrm{Cl}}^{\max}$.

We now argue that the upper bound $S_{\mathrm{PNC}}^{\max}$ can be saturated by a preparation noncontextual ontological model. Such a model (achieving $S_{\mathrm{PNC}} = S_{\mathrm{PNC}}^{\max}$) must necessarily have $p(\lambda|m) = p(\lambda) = \delta_{\lambda,\lambda_{\max}}$ for all $m \in$ **Msg**, where $\lambda_{\max}$ is the[9] ontic state of Bob's system ($\lambda_{\max} \in \Lambda$) that satisfies

$$S_{\mathrm{PNC}}^{\max} = \sum_m p(m) \Bigg(\sum_x p(x|m,\lambda_{\max}) \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(m'|y,\lambda_{\max})\delta_{m,m'} \Bigg) \qquad (12)$$

Thus, $p(\lambda|m) = \sum_x p(x|m)p(\lambda|m,x) = \delta_{\lambda,\lambda_{\max}}$ implies that $p(\lambda|m,x) = \delta_{\lambda,\lambda_{\max}}$ for all m,x. We must then have $p(m'|y,m,x) = \sum_\lambda p(m'|y,\lambda)p(\lambda|m,x) = \sum_\lambda p(m'|y,\lambda)\delta_{\lambda,\lambda_{\max}} = p(m'|y,\lambda_{\max})$, i.e., the statistics of y does not change in response to variations in m and x but is directly determined by the ontic state that is deterministically sampled by *every* preparation procedure. In fact, the response functions cannot even deviate from the best deterministic classical strategy. This preparation noncontextual ontological model, therefore, trivially reproduces the operational equivalence required on Bob's system, i.e., the operational equivalence between all the coarse-grainings of preparation ensembles induced by Alice's measurements. It also achieves $S = S_{\mathrm{PNC}}^{\max}$ by fixing the response functions on Bob's side to mimic the best deterministic classical strategy.[10]

[9]For our purposes, we can take $\lambda_{\max}$ to be unique without any loss of generality.

[10]Hence, the ontological model can only simulate an operational theory that has just one equivalence class of preparations and, furthermore, associates outcomes to its measurements deterministically. As such, in the presence of other empirical facts that an operational theory might present (such as the simple fact that Bob's system can be prepared in operationally inequivalent ways), this preparation noncontextual model will fail to reproduce predictions of the theory that go beyond the required operational equivalence between preparation procedures. Generically, therefore, any non-trivial (at least in the

Hence, we have that $S \leq S_{\mathrm{Cl}}^{\max}$ *is* a preparation noncontextuality inequality and any quantum advantage in this communication task witnesses preparation contextuality.

4.3 The shared state must violate a Bell inequality under the local measurements used in the communication protocol

In a quantum implementation of the communication protocol, shared entanglement between Alice and Bob is crucial for there to be an advantage over the classical one-shot success probability. However, entanglement alone is not enough: the entanglement must be such that it enables a Bell inequality violation relative to the local measurements that Alice and Bob implement. To see this, consider a nonlocal game that uses the same resources—shared entanglement and local measurements—as the communication protocol but under spacelike separation (hence no classical channel): Alice and Bob implement their local measurements labelled by m and y, respectively, and obtain their respective outcomes x and m' with a joint probability $p(x,m'|m,y)$ and the joint statistics thus collected admits a locally causal model, i.e.,

$$p(x,m'|m,y) = \sum_{\omega \in \Omega} p(x|m,\omega)p(m'|y,\omega)p(\omega), \quad (13)$$

where ω denotes the shared ontic state sampled from the ontic state space Ω of the bipartite system Alice and Bob share. (Note that this is, in general, different from the ontic state space Λ for Bob's system alone.) In such a case, it is straightforward to see that the achievable success probability is no better than the best deterministic classical strategy. Firstly,

$$S = \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(m'|y,m,x)\delta_{m,m'}$$
$$= \sum_m p(m) \sum_x \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(x|m)p(m'|y,m,x)\delta_{m,m'}$$
$$= \sum_m p(m) \sum_x \sum_y \mathcal{N}(y|x) \sum_{m'} p(x,m'|m,y)\delta_{m,m'},$$
$$(14)$$

sense of admitting operationally inequivalent preparation procedures) preparation noncontextual ontological model will only achieve $S < S_{\mathrm{PNC}}^{\max}$.

which can be now be expressed as

$$S = \sum_m p(m) \sum_x \sum_y \mathcal{N}(y|x) \sum_{m'} \left(\sum_{\omega \in \Omega} p(x|m,\omega) \right.$$
$$\left. p(m'|y,\omega) p(\omega) \right) \delta_{m,m'}$$
$$= \sum_{\omega \in \Omega} p(\omega) \left(\sum_m p(m) \sum_x p(x|m,\omega) \sum_y \mathcal{N}(y|x) \right.$$
$$\left. \sum_{m'} p(m'|y,\omega) \delta_{m,m'} \right)$$
$$\leq S_{\max}^{\mathrm{Cl}}. \tag{15}$$

Hence, Bell nonlocality of the shared entangled state relative to the measurements carried out by Alice and Bob is a necessary condition for a quantum advantage in this communication task.[11]

5 Preparation contextuality vis-à-vis Bell nonlocality: the connection with nonlocal games

It is known that any bipartite proof of Bell nonlocality can be turned into a proof of preparation contextuality on each wing of the Bell experiment, i.e., Bell nonlocality implies preparation contextuality on both wings of the Bell experiment. It is easiest to see this in the contrapositive: that is, the existence of a preparation noncontextual ontological model on any wing of the Bell experiment implies the existence of a locally causal model for the Bell experiment. We provide an explicit argument in Appendix A.[12]

For the task of enhancing the one-shot success probability of a classical channel, preparation contextuality and Bell nonlocality are even more intimately related than the general situation above. As we just showed in Section 4.3, $S > S_{\mathrm{Cl}}^{\max}$ implies Bell nonlocality of the joint statistics $\{p(x,m'|m,y)\}_{x,m',m,y}$. This allows us to state the following proposition:

Proposition 1. *For every classical channel $\mathcal{N}$ that admits an enhancement of the one-shot success probability driven by preparation contextuality, i.e., $S > S_{\mathrm{Cl}}^{\max}$, there exists a nonlocal game which can be won with a better-than-classical success probability by the same entangled state and local measurements which enable an advantage in the communication task. By construction, we also have the converse: an advantage in this nonlocal game would imply an advantage in the communication task.*

Indeed, if this were not the case (i.e., no such nonlocal game existed) then the enhancement of the one-shot success probability couldn't have been exhibited because the shared correlations between Alice and Bob would then be Bell-local. Hence, the problem of one-shot classical communication assisted by nonsignalling correlations characterizes a family of Bell scenarios where a proof of preparation contextuality on one wing implies a proof of Bell nonlocality between the two wings. This provides further insight into the conditions under which preparation contextuality for a single system can be said to imply Bell nonlocality for its appropriate bipartite extensions, in line with previous work where a certain type of preparation contextuality was shown to imply Bell nonlocality relative to a bipartite extension [36]. [13]

The explicit construction of a nonlocal game instantiating Proposition 1, however, would depend on the properties of the channel $\mathcal{N}$. We know that at least in the case of the Cubitt *et al.* protocol under ideal conditions, these nonlocal games correspond to pseudo-telepathy (PT) games inspired by the KS theorem [5, 6, 39]. In the case of the Prevedel *et al.* example [28], the associated nonlocal game is essentially the well-known CHSH game [2, 40]. It is an open question whether there exists a generic construction of a nonlocal game, instantiating Proposition 1, that always works starting from any channel $\mathcal{N}$.

We *can*, however, provide a fairly general construction of nonlocal games starting from a family of classical channels, thus instantiating Proposition 1. This general construction, in particular, reproduces as special cases the examples studied in Refs. [6, 28]. It is inspired by the pseudotelepathy (PT) game discussed in Ref. [6], allowing, however, the case where the quantum strategy is imperfect and where neither Alice nor Bob might have access to a KS set. We define this mapping from the communication task to a nonlocal game below.

---

[11] Note, however, that – contrary to the counterfactual Bell scenario we just considered – the measurement choice y in the communication protocol of interest is not free and is determined probabilistically by Alice's measurement outcome x. That is, the protocol requires a wiring of Alice's system with Bob's using the classical channel, something distinct from a Bell scenario.

[12] On the other hand, in the simplest possible scenario capable of exhibiting preparation contextuality (with two tomographically complete binary measurements and four preparations), it has been shown that the existence of a preparation noncontextual ontological model is *equivalent* to the existence of a locally causal model in any bipartite extension of the one-party scenario to a CHSH scenario [34]. Ref. [34] also noted that Barrett was the first to show the implication from preparation noncontextuality to local causality in any bipartite extension of a given one-party scenario. This has also been observed in Ref. [35]

[13] In the case of bipartite pure entangled states of Schmidt rank greater than two (such as the two-ququart maximally entangled state used in Ref. [6]), it has been shown that preparation contextuality of the reduced state on Bob's system, in conjunction with Alice's ability to remotely steer Bob's system to arbitrary preparation ensembles using entanglement [37], implies that the entangled state can exhibit Bell nonlocality [38]. So, at least in the case of such pure entangled states, the implication from preparation contextuality to Bell nonlocality that we consider in this paper also follows from Ref. [38].

The family of classical channels for which our construction works satisfies two properties for any channel $\mathcal{N}$ in the family: first, its channel hypergraph $H(\mathcal{N})$ is k-regular (i.e., every vertex appears in k hyperedges) for some positive integer k, and second, the channel probabilities are *entirely* fixed by the combinatorial structure of the channel, i.e., $\mathcal{N}(y|x) = \frac{1}{|Y_x|}\delta(y \in Y_x)$, where $\delta(a \in A)$ defines an indicator function for membership in set A, taking value 1 if $a \in A$ and 0 otherwise. Channels satisfying the first property will be called *k-regular* channels and those satisfying the second property will be called *output-uniform* channels. Hence, the classical channels we consider below will be k-regular and output-uniform classical channels. Further, we will assume that Alice's choice of the message to send in a particular run is uniformly random, i.e., $p(m) = \frac{1}{|\mathbf{Msg}|}$ for all $m \in \mathbf{Msg}$. All this amounts to the following expression for the one-shot success probability:

$$S = \frac{1}{|\mathbf{Msg}|}\frac{1}{k}\sum_{m,m',x,y \in Y_x}\delta_{m,m'}p(x,m'|m,y). \quad (16)$$

The corresponding nonlocal game is specified by the following: Alice receives questions $m \in \mathbf{Msg}$ and replies with answers $x \in X$; Bob receives questions $y \in Y$ and replies with answers $m' \in \mathbf{Msg}$; the conditional joint probability distributions of interest, therefore, are given by $\{p(x,m'|m,y)\}_{x,m',m,y}$; the Referee sends them questions m,y according to the probability distribution $p(m,y) = p(m)p(y) = \frac{1}{|\mathbf{Msg}|}\frac{1}{|Y|}$; in order to win the game, Alice and Bob must produce outputs x, m' (respectively) such that the condition $V(x,m',m,y) = 1$ is satisfied, where

$$V(x,m',m,y) \equiv \begin{cases} 1, & \text{if } y \notin Y_x \text{ (i.e., } \mathcal{N}(y|x) = 0), \\ \delta_{m,m'}, & \text{if } y \in Y_x \text{ (i.e., } \mathcal{N}(y|x) > 0). \end{cases}$$
(17)

The probability of winning the game is then given by

$$S_{\text{Bell}} \equiv \frac{1}{|\mathbf{Msg}|}\frac{1}{|Y|}\sum_{m,m',x,y}p(x,m'|m,y)V(x,m',m,y)$$
(18)

$$= \frac{1}{|\mathbf{Msg}|}\frac{1}{|Y|}\sum_{m,m',x,y \in Y_x}\delta_{m,m'}p(x,m'|m,y)$$

$$+ \frac{1}{|\mathbf{Msg}|}\frac{1}{|Y|}\sum_{m,m',x,y \notin Y_x}p(x,m'|m,y). \quad (19)$$

Note that this mapping relies only on combinatorial properties of $\mathcal{N}$, namely, its channel hypergraph $H(\mathcal{N})$, and is a straightforward generalization of the connection between the one-shot classical communication protocol and pseudo-telepathy games note in Proposition 3 of Ref. [6]. The connection of the protocol of Ref. [28] with the CHSH game, for example, falls under this generalization.

We are now ready to prove the following theorem:

Theorem 1. *Consider a k-regular output-uniform classical channel $\mathcal{N}$, i.e., $\mathcal{N}(y|x) = \frac{1}{k}\delta(y \in Y_x)$. Then, for any bipartite entangled state ρ_{AB} shared between Alice and Bob, Alice's local measurements $\{\mathbb{M}_m^{(A)} \equiv \{E_x^{(m)}\}_{x \in X^{(m)}}\}_{m \in \mathbf{Msg}}$, and Bob's local measurements $\{\mathbb{M}_y^{(B)} \equiv \{E_{m'}^{(y)}\}_{m' \in \mathbf{Msg}}\}_{y \in Y}$ — all denoted by the triplet $(\rho_{AB}, \{\mathbb{M}_m^{(A)}\}_{m \in \mathbf{Msg}}, \{\mathbb{M}_y^{(B)}\}_{y \in Y})$ — the following are equivalent:*

1. *The triplet $(\rho_{AB}, \{\mathbb{M}_m^{(A)}\}_{m \in \mathbf{Msg}}, \{\mathbb{M}_y^{(B)}\}_{y \in Y})$ provides a quantum advantage in the task of one-shot classical communication over $\mathcal{N}$, i.e., $S > S_{\text{Cl}}^{\max}$,*

2. *The triplet $(\rho_{AB}, \{\mathbb{M}_m^{(A)}\}_{m \in \mathbf{Msg}}, \{\mathbb{M}_y^{(B)}\}_{y \in Y})$ provides a quantum advantage in the corresponding nonlocal game (cf. Eq. (17)), i.e., $S_{\text{Bell}} > S_{\text{local}}^{\max}$, where $S_{\text{local}}^{\max}$ is the Bell-local bound given by*

$$S_{\text{local}}^{\max} \equiv \max_{p(x,m'|m,y) \in \mathcal{L}} S_{\text{Bell}}, \quad (20)$$

$\mathcal{L}$ being the set of Bell-local probability distributions. We use "$p(x,m'|m,y) \in \mathcal{L}$" as shorthand for membership of the full probability vector $(p(x,m'|m,y))_{x,m',m,y}$ in the set of Bell-local probability vectors [40].

Further, we have that $S = 1 \Leftrightarrow S_{\text{Bell}} = 1$.[14]

Proof. See Appendix B. $\square$

Theorem 1 provides us a way to characterize a family of classical channels $\mathcal{N}$ for which the one-shot success probability can be enhanced by nonsignalling correlations: namely, all k-regular and output-uniform classical channels for which there is a gap between the classical and the nonsignalling value of the nonlocal game defined by them following the recipe we have just outlined, *cf.* Eq. (17).

It is worth emphasizing here the physical distinction between the two tasks – the one-shot communication task and the corresponding nonlocal game – we have considered in this section. In the communication task, Alice and Bob must necessarily be timelike separated, but in the nonlocal game, they must necessarily be spacelike separated. In the absence of spacelike separation in the communication task, it is inaccurate to state that Bell nonlocality drives the quantum advantage in the task: to be sure, the states and measurements that drive the quantum advantage in the communication task can also drive the quantum advantage in the nonlocal game, but the two tasks correspond to fundamentally different physical situations. Contextuality, for this reason, is the more natural notion of nonclassicality to appeal to as the driver of quantum advantage in the communication task.

---

[14]Note that this recovers the special case of pseudotelepathy games considered in Proposition 3 of Ref. [6].

6 One-shot success probability of communicating a single bit

The problem of communicating a single bit through a noisy classical channel has been previously studied in Refs. [28, 29]. Note that the classical value of the one-shot success probability of communicating a single bit (i.e., one out of two messages) in this problem is strictly less than 1 if and only if the confusability graph of the channel $\mathcal{N}$ is a complete graph, i.e., $\alpha(G(\mathcal{N})) = 1$. Hence, it is only for such channels that the possibility of enhancing their one-shot success probability of sending a single bit using shared entanglement exists: for all other channels (with $\alpha(G(\mathcal{N})) \geq 2$), a single bit can always be sent with zero error classically. In the rest of this section, therefore, we will only consider output-uniform channels with a confusability graph that is complete. These channels are a special case of Kochen-Specker (KS) channels, namely, those where it is impossible to pick even a *pair* of non-confusable vertices from distinct hyperedges.[15] Such KS channels obviously do not admit any KS sets since any set of projectors associated with their vertices will necessarily have to be pairwise commuting, i.e., there would be no incompatibility between the projectors. Our general protocol applies to such channels and here we will consider one particular example, the one studied in Ref. [28], as a paradigmatic case and show how it fits within our framework.

The Prevedel *et al.* protocol

The classical channel considered by Prevedel *et al.* [28] is specified as follows: the input alphabet is the set of two-bit strings, $X = \{00, 01, 10, 11\}$, and the output alphabet is a set of trit-bit strings $Y = \{(1, 0), (2, 0), (P, 0), (1, 1), (2, 1), (P, 1)\}$; the channel hypergraph therefore consists of 4 vertices (labelled by $x = b_1 b_2 \in X$) with all possible two-vertex hyperedges, i.e., 6 hyperedges labelled by $y = (t, b) \in Y$; each hyperedge $y = (t, b)$ uniquely identifies a pair of inputs $x \in X_y$ sharing one of three properties: the value of the first bit (i.e., $y = (1, b_1)$), the value of the second bit (i.e., $y = (2, b_2)$), or the parity of the two bits (i.e., $y = (P, b_1 \oplus b_2)$); so, for example, the output $y = (1, 0)$ identifies the pair of inputs $X_y = \{00, 01\}$ in its support; the channel is, therefore, 3-regular and output-uniform with channel probabilities $\mathcal{N}(y|x) = \frac{1}{3}\delta(y \in Y_x)$. The channel hypergraph is illustrated in Fig. 2.

The general protocol of Section 3 takes the following

---

[15]Recall that a KS channel is defined by a channel hypergraph where it is impossible to pick a *set* of vertices, one vertex from *each* hyperedge, such that all the vertices in this set are mutually non-confusable.

form:

Msg $= \{m\}_{m=0}^1$, with the encoding $\{X^{(m)}\}_m$ given by $X_{m=0} \equiv \{00, 01\}$ and $X_{m=1} \equiv \{10, 11\}$. (21)

Alice carries out one of two possible measurements labelled by $m \in \{0, 1\}$, their outcomes labelled by $b_2 \in \{0, 1\}$. On obtaining outcome b_2 for measurement m, Alice inputs the two-bit string $x = mb_2$ to the channel. Bob possesses one of two possible binary measurements labelled by $v \in \{0, 1\}$ (their outcomes labelled by $z \in \{0, 1\}$) and must use the output y from the classical channel (which gives him some information about the possible inputs X_y) to decide his measurement strategy in order to infer the message Alice's message m. The full strategy is detailed below:

$v \in \{0, 1\}$,

$$p(v|y) = \begin{cases} \delta_{v,1}, & \text{for } y = (2, b_2), \\ \delta_{v,0}, & \text{for } y = (P, b_1 \oplus b_2), \\ \text{arbitrary for } y = (1, b_1). \end{cases}$$

$z \in \{0, 1\}$

$$g(z, y) = \begin{cases} b, & \text{for } y = (t, b) = (1, b_1), \\ b \oplus z, & \text{for } y = (t, b) \in \{(2, b_2), (P, b_1 \oplus b_2)\} \end{cases}$$
(22)

$$p(m' = m|z, y) = \delta_{g(z,y),m}. \tag{23}$$

Assuming $p(m) = \frac{1}{2}$, the expression for the success probability is given by

$$S = \frac{1}{3} + \frac{1}{6} \sum_{b_2, z, m, v} p(b_2, z|m, v)\delta_{b_2 \oplus z, mv}. \tag{24}$$

We refer to Appendix C for a complete derivation of the above expression. Now, the joint statistics $p(b_2, z|m, v)$ can be interpreted as arising from a Bell-CHSH scenario [2], i.e., a Bell scenario where each party has two possible binary-outcome measurements (m and v in this case), noting that the statistics is non-signalling. We can then define the success probability in such a CHSH game [40] as

$$S_{\text{CHSH}} \equiv \frac{1}{4} \sum_{b_2, z, m, v} p(b_2, z|m, v)\delta_{b_2 \oplus z, mv}, \tag{25}$$

so that

$$S = \frac{1}{3} + \frac{2}{3}S_{\text{CHSH}}. \tag{26}$$

We have

1. Classically:

$$\sum_{b_2, z, m, v} p(b_2, z|m, v)\delta_{b_2 \oplus z, mv} \leq 3, \tag{27}$$

so that $S = S_{\text{Cl}} \leq \frac{5}{6} \approx 0.833$, and

2. Quantumly:

$$\sum_{b_2, z, m, v} p(b_2, z | m, v) \delta_{b_2 \oplus z, mv} \leq 2 + \sqrt{2}, \quad (28)$$

so that $S = S_Q \leq \frac{1}{3} + \frac{2+\sqrt{2}}{6} \approx 0.902$.

And, of course, on allowing arbitrary nonsignalling correlations, a PR-box [40, 41] can achieve $S = S_{PR} = 1$.

How is the CHSH game related to the nonlocal game corresponding to the Prevedel *et al.* protocol that one would obtain following Theorem 1? In Appendix D, we show that, in fact, this nonlocal game is essentially the CHSH game rewritten in such a way that Alice has two inputs while Bob has six.

7 One-shot success probability of classical communication via general Kochen-Specker (KS) channels

7.1 Channel hypergraph, encoding, and context-independent guessing (CIG): the Cubitt *et al.* strategy

In this section we will go beyond channels with complete confusability graphs (for which one-shot zero-error communication is impossible) and consider general Kochen-Specker (KS) channels. Of particular interest will be KS channels that admit KS sets: for some of these channels it is possible to achieve an enhancement of the one-shot zero-error capacity using entanglement, e.g., in Ref. [6], Cubitt *et al.* showed that one can use a classical channel based on Peres's 24-ray two-qubit KS set [42] which also underlies the Peres-Mermin proof of KS-contextuality [43, 44].

We will focus on the one-shot success probability that can be achieved using the Cubitt *et al.* [6] strategy when Bob only assumes the structure of the channel hypergraph, $H(\mathcal{N})$, and his knowledge of the encoding Alice uses but makes no assumptions about the exact channel probabilities $\{\mathcal{N}(y|x)\}_{x \in X, y \in Y}$. This could, for example, happen when Alice and Bob trust the channel hypergraph but they do not trust the channel probabilities of the classical channel given by some provider.

Recall that the one-shot success probability following the Cubitt *et al.* strategy is given by

$$S = \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{m'} p(m'|y, m, x) \delta_{m, m'}$$
$$= \sum_m p(m) \sum_x p(x|m) \sum_y \mathcal{N}(y|x)$$
$$\sum_{x' \in X^{(m)}} p(x'|y, m, x). \quad (29)$$

Bob uses his knowledge of the channel hypergraph, $H(\mathcal{N})$, and the output received from the channel, y, along with any nonsignalling correlations shared with Alice, to make a guess $x' \in X_y$ for Alice's input x. Our assumption that Bob is oblivious of the channel probabilities means that the probability with which Bob makes his guess, $p(x'|y, m, x)$, should be independent of the particular $y \in Y_{x'}$ ("context" of the guess x') that Bob receives. It can depend only on the support of x' via an indicator function, i.e., $p(x'|y, m, x) = p(x'|m, x)\delta(y \in Y_{x'})$, where $\delta(y \in Y_{x'}) = 1$ if $y \in Y_{x'}$ and 0 otherwise. Overall, we have

for any $x' \in X$:
$$p(x'|y_1, m, x) = p(x'|y_2, m, x) \equiv p(x'|m, x),$$
$$\forall y_1, y_2 \in Y_{x'}, \forall x \in X^{(m)}, \forall m \in \mathbf{Msg}. \quad (30)$$

We term this condition *context-independent guessing* (CIG). We will see that the quantum strategy of the Cubitt *et al.* protocol [6] satisfies this constraint and this fact allows us to invoke the assumption of measurement noncontextuality in addition to preparation noncontextuality in placing a noncontextual upper bound on the one-shot success probability. This will in turn allow us to analyze the Cubitt *et al.* construction and the critical role of the KS theorem in it in the light of generalized noncontextuality à la Spekkens [13]. More concretely, we will see that a non-trivial upper bound on the classical one-shot success probability in this protocol can be characterized by a hypergraph invariant – the weighted max-predictability [18] – following the approach of Ref. [27].

On the other hand, note that classically we have $p(x'|y, x, m) = p(x'|y)$ for all $x \in X^{(m)}, m \in \mathbf{Msg}$ and the CIG constraint of Eq. (30) (in a classical strategy) then requires that $p(x'|y) = p(x')\delta(y \in Y_{x'})$. That is,

for any $x' \in X$: $p(x'|y_1, x, m) = p(x'|y_2, x, m) \equiv p(x')$,
$$\forall y_1, y_2 \in Y_{x'}, \forall x \in X^{(m)}, \forall m \in \mathbf{Msg}. \quad (31)$$

Indeed, in a classical strategy for this communication task, any assignment $p : X \to [0, 1]$ respecting the context-independence property defines what is usually called a (general) probabilistic model when the channel hypergraph is viewed as a contextuality scenario [27, 31].

7.2 One-shot success probability of a KS channel under the CIG constraint: contextuality and quantum advantage

As our working example, we will consider the same classical channel considered by Cubitt *et al.*, cf. Fig. 4. This channel admits a KS set, i.e., a set of projectors on a 4-dimensional quantum system, each projector associated with a vertex in the channel hyper-
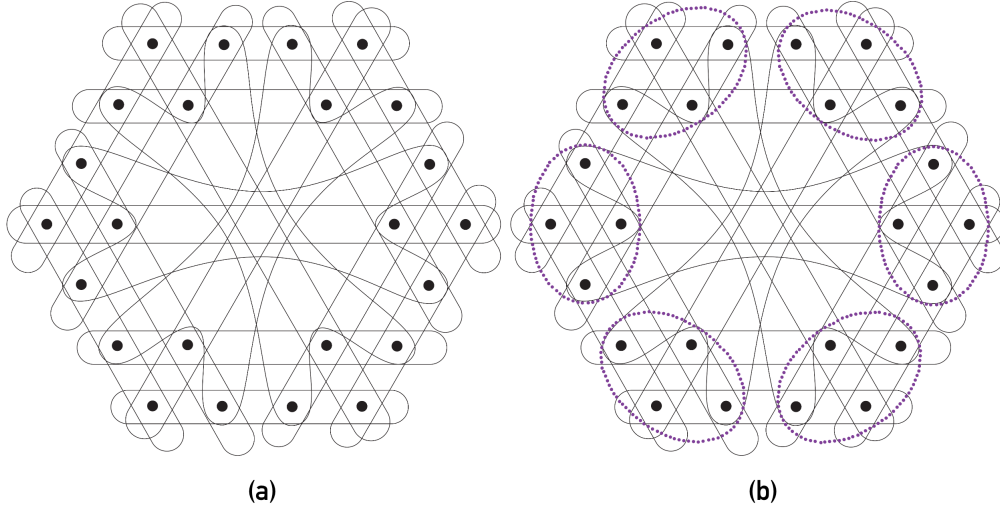
Figure 4: (a) The channel hypergraph considered in Ref. [6], and (b) the encoding used, highlighted with dotted hyperedges. Each message, corresponding to a dotted loop, contains four mutually confusable input symbols.

graph such that each hyperedge constitutes a projective measurement.

7.2.1 The one-shot success probability with context-independent guessing

The one-shot success probability is given by

$$
\begin{aligned}
S &= \sum_m p(m) \sum_x p(x|m) \\
&\quad \sum_y \mathcal{N}(y|x) \sum_{x' \in X^{(m)}} p(x'|y, m, x) \\
&= \sum_m p(m) \sum_x p(x|m) \\
&\quad \sum_y \left(\mathcal{N}(y|x) \sum_{x' \in X^{(m)}} p(x'|m, x) \delta(y \in Y_{x'}) \right) \\
&\quad \text{(using the CIG constraint)} \\
&= \sum_m p(m) \sum_x p(x|m) \\
&\quad \sum_{x' \in X^{(m)}} \left(\sum_y \mathcal{N}(y|x) \delta(y \in Y_{x'}) p(x'|m, x) \right).
\end{aligned}
\tag{32}
$$

For any two symbols $x, x' \in X$, we quantify the confusability of x with respect to x' via the function

$$
\eta(x, x') \equiv \sum_y \mathcal{N}(y|x) \delta(y \in Y_{x'}).
\tag{33}
$$

This is the probability that, for input x, the channel $\mathcal{N}$ yields an output that could also arise from the input x'. Obviously, $\eta(x, x) = 1$ for all $x \in X$.

Noting that $p(x|m)p(x'|m, x) = p(x, x'|m)$, we have

$$
\begin{aligned}
S &= \sum_m p(m) \sum_{x, x' \in X^{(m)}} p(x, x'|m) \eta(x, x') \tag{34} \\
&= \sum_m p(m) \sum_{x, x' \in X^{(m)}} \delta_{x, x'} p(x, x'|m) \\
&\quad + \sum_m p(m) \sum_{x, x' \in X^{(m)}} (1 - \delta_{x, x'}) p(x, x'|m) \eta(x, x'),
\end{aligned}
\tag{35}
$$

where we used the fact that $\eta(x, x) = 1$. Defining

$$
S_{\text{perf}} \equiv \sum_m p(m) \sum_{x, x' \in X^{(m)}} \delta_{x, x'} p(x, x'|m),
\tag{36}
$$

$$
S_{\text{imperf}} \equiv \sum_m p(m) \sum_{x, x' \in X^{(m)}} (1 - \delta_{x, x'}) p(x, x'|m) \eta(x, x'),
\tag{37}
$$

we have that $S = S_{\text{perf}} + S_{\text{imperf}}$. Here S_{perf} denotes the contribution to the success probability from the situation where Bob guesses Alice's input x to the channel exactly (i.e., $x' = x$, and therefore also infers m correctly) and S_{imperf} denotes the remaining contribution to the success probability from the situation where Bob doesn't guess x correctly (i.e., $x' \neq x$) but nevertheless infers m correctly from x' (i.e., $x' \in X^{(m)}$).

Recalling the source-measurement correlation function Corr studied in Ref. [27], we have

$$
\text{Corr} = S_{\text{perf}} \equiv \sum_m p(m) \sum_{x, x' \in X^{(m)}} \delta_{x, x'} p(x, x'|m).
\tag{38}
$$

In the context of Ref. [27], this correlation function captures how predictable the measurements corresponding to the hyperedges m can be made when

one varies over corresponding preparation ensembles (also labelled by m) that average to the same mixed state. In the ideal case of projective measurements from Peres' 24-vector KS set, this quantity is 1, since every measurement is perfectly predictable when the input state is picked from its orthonormal basis and the uniform mixture over input states from any such basis is the maximally mixed state. However, in a noncontextual ontological model, this quantity is a upper bounded by a hypergraph invariant [27] that will turn out to be relevant for the one-shot success probability in the following subsections.

We then have the following bounds on S:

$$\text{Corr} + \eta_{\min}(1 - \text{Corr})$$
$$\leq S = \text{Corr}$$
$$+ \sum_m p(m) \sum_{x,x' \in X^{(m)}} (1 - \delta_{x,x'}) p(x, x'|m) \eta(x, x')$$
$$\leq \text{Corr} + \eta_{\max}(1 - \text{Corr}), \qquad (39)$$

where

$$\eta_{\min} \equiv \min_m \min_{x \neq x' \in X^{(m)}} \eta(x, x'), \qquad (40)$$

$$\eta_{\max} \equiv \max_m \max_{x \neq x' \in X^{(m)}} \eta(x, x'). \qquad (41)$$

Here $\eta_{\min}$ is the minimum confusability between any two symbols in any message in the encoding. Similarly, $\eta_{\max}$ is the maximum confusability between any two symbols in any message in the encoding. We have $0 < \eta_{\min} \leq \eta_{\max} < 1$.

7.2.2 The assumption of measurement noncontextuality and how any noncontextual strategy satisfies context-independent guessing

In the Cubitt *et al.* strategy, it is assumed that both Alice and Bob have access to specific sets of measurements carved out of a KS set for the channel of Fig. 4. Alice's six measurements $\{\mathbb{M}_m^{(A)}\}_{m=1}^6$ correspond to the six dotted hyperedges in Fig. 4, while Bob's measurements $\{\mathbb{M}_y^{(B)}\}_{y=1}^{18}$ correspond to partitioning the 24 projectors into 18 hyperedges, denoted by solid hyperedges in Fig. 4.

In our treatment of the problem, since we want to allow more general choices of measurements, we make two relaxations:

1. Firstly, we do not restrict ourselves to projective measurements on Bob's side, i.e., we allow any set of positive operators (each operator associated to a vertex in the channel hypergraph) that satisfy the requirement that the (solid *and* dotted) hyperedges in Fig. 4 form complete measurements, and

2. Secondly, although in the Cubitt *et al.* protocol, Alice's measurements are carved out of the

same set of positive operators that Bob can implement, and this is clearly the optimal choice of measurements for Alice (yielding $S = 1$ for sending six messages, quantumly), we allow that, in general, Alice could associate some *other* measurements with the messages she wants to send even if the outcomes of such measurements do not satisfy the operational equivalences implicit in the channel hypergraph of Fig. 4. That is, the encoding strategy of Alice could use a set of positive operators for her measurements $\{\mathbb{M}_m^{(A)}\}_{m=1}^6$ that is completely different from the set of positive operators used by Bob for his measurements $\{\mathbb{M}_y^{(B)}\}_{y=1}^{18}$ in the decoding strategy.

We mention this to emphasize that our generalization of the Cubitt *et al.* strategy does not rely on identifying the measurement outcomes of Alice and Bob in the way they are identified in the optimal strategy and our use of the assumption of measurement noncontextuality is restricted to Bob's system, i.e., response functions of Bob's measurements. In particular, the positive operators that constitute Bob's measurements can, in principle, be reconfigured to define the six measurements $\{\mathbb{M}_m^{(B)}\}_{m=1}^6$ that it would be optimal for Alice to choose for her encoding measurements $\{\mathbb{M}_m^{(A)}\}_{m=1}^6$.

In a noncontextual ontological model of Bob's system, the response functions associated with the vertices (labelled by $x \in X$) respect measurement noncontextuality, i.e., for all x, y_1, y_2 such that $x \in X_{y_1} \cap X_{y_2} \neq \varnothing$,

$$\xi(x|y_1, \lambda) = \xi(x|y_2, \lambda) \equiv \xi(x|\lambda), \forall \lambda \in \Lambda. \qquad (42)$$

Now, even though Bob may not implement or have access to the six measurements $\{\mathbb{M}_m^{(B)}\}_{m=1}^6$ that would be optimal for Alice in the protocol, the operational equivalences implicit in the hypergraph of Fig. 4 indicate that the response functions for $\mathbb{M}_m^{(B)}$ on Bob's system must, under the assumption of measurement noncontextuality, also satisfy

$$\xi(x|m, \lambda) = \xi(x|\lambda), \forall \lambda \in \Lambda, \qquad (43)$$

for all $x \in X$ and all $m \in \mathbf{Msg}$.

Recalling that

$$p(x'|y, m, x) = \sum_\lambda \xi(x'|y, \lambda) \mu(\lambda|m, x), \qquad (44)$$

we have that $p(x'|y_1, m, x) = p(x'|y_2, m, x)$ for all x', y_1, y_2, m, x such that $x' \in X_{y_1} \cap X_{y_2}$ (equivalently, $y_1, y_2 \in Y_{x'}$), so that the CIG constraint is satisfied by any noncontextual strategy and the one-shot success probability takes the form of Eq. (34).

7.2.3 Upper bound on the one-shot success probability from preparation and measurement noncontextuality

We now proceed to upper bound the success probability under the assumption of noncontextuality, i.e.,

preparation and measurement noncontextuality, and obtain

$$S \leq \max_\lambda \sum_m p(m) \sum_{x,x' \in X^{(m)}} \xi(x'|m,\lambda)\mu(x|m,\lambda)\eta(x,x')$$

$$\equiv S_{\mathrm{NC}}^{\max}. \tag{45}$$

To see how this comes about starting from the expression for the one-shot success probability in Eq. (34), i.e.,

$$S = \sum_m p(m) \sum_{x,x' \in X^{(m)}} p(x,x'|m)\eta(x,x'),$$

first note that $p(x,x'|m) = \sum_\lambda \xi(x'|m,\lambda)\mu(\lambda,x|m)$, where $\mu(\lambda,x|m) = \mu(\lambda|m,x)p(x|m) = \mu(x|m,\lambda)\mu(\lambda|m)$. As in the general case of one-shot classical communication assisted by entanglement, given the operational equivalence of ensembles prepared on Bob's side by Alice's measurements $\mathbb{M}_m^{(\mathrm{A})}$ and the assumption of preparation noncontextuality, we have that $\mu(\lambda|m) = \nu(\lambda)$ for all m. We then have

$$S = \sum_m p(m)$$
$$\sum_{x,x' \in X^{(m)}} \sum_\lambda \xi(x'|m,\lambda)\mu(x|m,\lambda)\nu(\lambda)\eta(x,x')$$
$$= \sum_\lambda \nu(\lambda) \sum_m p(m)$$
$$\sum_{x,x' \in X^{(m)}} \xi(x'|m,\lambda)\mu(x|m,\lambda)\eta(x,x')$$
$$\leq S_{\mathrm{NC}}^{\max}. \tag{46}$$

7.2.4 Contextuality drives the quantum advantage

We show that the one-shot success probability achievable via any noncontextual strategy is no better than best classical strategy with context-independent guessing. For any extremal classical strategy $i \in \mathcal{I}$ ($\mathcal{I}$ being the set of extremal classical strategies satisfying CIG), the success probability is given by

$$S_{\mathrm{Cl}}(i)$$
$$= \sum_m p(m) \sum_x p_{\mathrm{A}}(x|m,i) \sum_y \mathcal{N}(y|x) \sum_{x' \in X^{(m)}} p_{\mathrm{B}}(x'|y,i)$$
$$= \sum_m p(m) \sum_x p_{\mathrm{A}}(x|m,i) \sum_y \mathcal{N}(y|x)$$
$$\sum_{x' \in X^{(m)}} p_{\mathrm{B}}(x'|i)\delta(x' \in X_y), \tag{47}$$

where $p_{\mathrm{B}}(x'|y,m,x,i) = p_{\mathrm{B}}(x'|y,i) = p_{\mathrm{B}}(x'|i)\delta(x' \in X_y)$, since Bob has no access to any information from Alice besides the shared variable i (denoting the strategy both of them agree to implement) and the channel output y, the latter specifying a confusable set X_y containing Alice's input x (and Bob's guess x'). An arbitrary classical strategy can then be represented

by a convex mixture of extremal classical strategies according to some probability distribution $\{p(i)\}_{i \in \mathcal{I}}$ and the classical success probability is then given by

$$S_{\mathrm{Cl(CIG)}} = \sum_i p(i) S_{\mathrm{Cl(CIG)}}(i) \leq \max_i S_{\mathrm{Cl(CIG)}}(i)$$
$$= \max_i \sum_m p(m) \sum_x p_{\mathrm{A}}(x|m,i) \sum_y \mathcal{N}(y|x)$$
$$\sum_{x' \in X^{(m)}} p_{\mathrm{B}}(x'|i)$$
$$\equiv S_{\mathrm{Cl(CIG)}}^{\max}. \tag{48}$$

Using the CIG constraint, we have $p_{\mathrm{B}}(x'|y,i) = p_{\mathrm{B}}(x'|i)\delta(y \in Y_{x'})$, where any extremal classical strategy i specifies a particular extremal probabilistic model on the channel hypergraph (viewed as a contextuality scenario [31]) in Fig. 4(a). This allows us to express the maximal classical success probability satisfying CIG as

$$S_{\mathrm{Cl(CIG)}}^{\max}$$
$$= \max_i \sum_m p(m) \sum_x p_{\mathrm{A}}(x|m,i) \sum_{x' \in X^{(m)}} \Bigg(p_{\mathrm{B}}(x'|i)$$
$$\sum_y \mathcal{N}(y|x)\delta(y \in Y_{x'}) \Bigg)$$
$$= \max_i \sum_m p(m) \sum_{x,x' \in X^{(m)}} p_{\mathrm{A}}(x|m,i)p_{\mathrm{B}}(x'|i)\eta(x,x')$$
$$\tag{49}$$

We therefore have

$$S_{\mathrm{NC}}^{\max} = S_{\mathrm{Cl(CIG)}}^{\max}. \tag{50}$$

Since any classical strategy is a convex mixture of extremal classical strategies, the upper bound $S_{\mathrm{Cl(CIG)}}^{\max}$ can always be achieved by a classical strategy, i.e., there exists an extremal classical strategy $i_* \in \mathcal{I}$ such that $S_{\mathrm{Cl(CIG)}}(i_*) = S_{\mathrm{Cl(CIG)}}^{\max}$. Similarly, the upper bound $S_{\mathrm{NC}}^{\max}$ can be saturated by a noncontextual strategy, albeit a very trivial one, following a similar reasoning as at the end of Section 4.2 (except that the extremal response functions here are indeterministic on account of KS-uncolourability). Thus, we have that contextuality also drives the quantum advantage in one-shot classical communication when Alice and Bob trust the channel hypergraph but make no assumptions about the channel probabilities.[16]

---

[16]Note that, under the CIG constraint, we no longer have the exact correspondence with nonlocal games exemplified by Theorem 1. This is because the connection between prepare-and-measure scenarios on Bob's system alone and Bell scenarios where Bob is one of the parties in a two-party Bell experiment breaks down when one imposes, besides preparation noncontextuality, the assumption of measurement noncontextuality on Bob's system (based on the operational equivalences be-

7.2.5 Contextuality witnessed by a hypergraph-invariant – the weighted max-predictability – is sufficient for a quantum advantage

In this section we point out an explicit connection between contextuality that can be witnessed via the noise-robust noncontextuality inequalities of Ref. [27] and the quantum advantage in the one-shot classical communication task with context-independent guessing. The full technical argument supporting the claims in this section is presented in Appendix E.

In Appendix E, we first show that

$$S \le S_{\mathrm{NC}}^{\max} \le \eta_{\max} + (1 - \eta_{\max})\beta(\Gamma, \{p(m)\}_m), \quad (51)$$

where

$$\beta(\Gamma, \{p(m)\}_m) \equiv \max_\lambda \sum_m p(m) \max_{x \in X^{(m)}} \xi(x|m, \lambda) \quad (52)$$

is the weighted max-predictability, a hypergraph invariant that was defined in Ref. [10] and studied extensively in Ref. [27], appearing in the upper bounds of noise-robust noncontextuality inequalities proposed in the frameworks of Refs. [10, 27]. Here, Γ is the hypergraph defined by the contextuality scenario from which the measurements of Alice and Bob are drawn in the ideal Cubitt *et al.* strategy [6], i.e., Fig. 4(b) including all the hyperedges (solid and dotted).

We then consider the special case $\eta_{\max} = \eta_{\min} = \eta$ and show that firstly, from Eq. (39), we have

$$S = \mathrm{Corr} + \eta(1 - \mathrm{Corr}) = \eta + \mathrm{Corr}(1 - \eta). \quad (53)$$

We then have that

$$\begin{aligned}
&\mathrm{Corr} > \beta(\Gamma, \{p(m)\}_m) \\
&\Leftrightarrow S > \eta + (1 - \eta)\beta(\Gamma, \{p(m)\}_m) \\
&\qquad \ge S_{\mathrm{NC}}^{\max} = S_{\mathrm{Cl(CIG)}}^{\max} \\
&\Rightarrow S > S_{\mathrm{NC}}^{\max}.
\end{aligned} \quad (54)$$

Now,

$$\mathrm{Corr} \le \beta(\Gamma, \{p(m)\}_m) \quad (55)$$

is an instance of a noise-robust noncontextuality inequality following the approach of Ref. [27], inspired by logical proofs of KS-contextuality [3, 17]. Hence, the contextuality witnessed by $\mathrm{Corr} > \beta(\Gamma, \{p(m)\}_m)$ is sufficient for a quantum advantage in this task when $\eta_{\max} = \eta_{\min}$.[17] Output-uniform channels with

k-regular hypergraphs have $\eta(x, x') = \frac{1}{k}\sum_y \delta(y \in Y_x \cap Y_{x'})$ and this quantity is independent of x, x' ($x \ne x'$) if and only if $|Y_x \cap Y_{x'}|$ is constant for all $x \ne x'$, i.e., the number of hyperedges shared by any two confusable vertices of the channel hypergraph is constant across all pairs of confusable vertices. In the channel hypergraph of Fig. 4, for example, we have $|Y_x \cap Y_{x'}| = 1$ for all confusable pairs of vertices x, x', so the classical channel studied in Ref. [6] (where $k = 3$ and $\mathcal{N}(y|x) = \frac{1}{3}$ for all $y \in Y_x, x \in X$) satisfies the condition required for $\mathrm{Corr} > \beta(\Gamma, \{p(m)\}_m)$ to imply a quantum advantage. The case $\mathrm{Corr} = 1$ corresponds to the situation studied in Ref. [6]. We have shown that this quantum advantage can persist even when $\mathrm{Corr} < 1$, i.e., in the regime of noisy measurements.

Another special case concerns the situation where $S_{\mathrm{imperf}} = 0$. In this situation too, the violation of $\mathrm{Corr} \le \beta(\Gamma, \{p(m)\}_m)$ implies a quantum advantage (cf. Appendix E). Using the set-up in Ref. [6], one can achieve $\mathrm{Corr} = 1$, i.e., zero-error communication.

We have thus provided an instance of an information-theoretic task where the noise-robust signatures of contextuality à la Refs. [17, 27] witness a quantum advantage in the task. This provides an operational meaning to noise-robust noncontextuality inequalities of the type in Eq. (55) that were proposed in Ref. [27].

7.3 KS basis sets and the ideal Cubitt *et al.* strategy

Can one use the strategy of Ref. [6] starting from *any* KS set of vectors?[18] The strategy requires not merely a KS set – namely, a set of vectors with orthogonality relations represented by a KS-uncolourable hypergraph – but, in fact, a *KS basis set*, i.e., a set of *disjoint* complete orthogonal bases $\mathcal{Z} \equiv \{\mathcal{B}_m\}_{m=1}^q$ such that it is impossible to pick a vector from each basis ensuring that no two are orthogonal. Clearly, the vectors in a KS basis set constitute a KS set. Denoting the vectors in basis $\mathcal{B}_m$ as $\{\psi_{mj}\}_{j=1}^d$, where d is the dimension of the Hilbert space spanned by the basis, we construct a classical channel $\mathcal{N}$ with inputs labelled by $\{(m, j)|m \in [q], j \in [d]\}$ ([N] denoting $\{1, 2, \dots, N\}$ for any positive integer N). The confusability graph, $G(\mathcal{N})$, of the channel is such that two inputs are confusable if and only if the corresponding vectors are orthogonal. The definition of a KS basis set then implies that $\alpha(G(\mathcal{N})) < q$ for any channel thus constructed from it.

As we have noted, the set of vectors appearing in any KS basis set form a KS set. However, it is not *a priori* obvious that, given a KS set, it is always possible to carve it up into a KS basis set $\{\mathcal{B}_m\}_{m=1}^q$

---

tween his local measurement events). This restricts the scope of response functions for Bob's measurements beyond anything required by local causality (under which no restriction on local response functions is imposed). The interested reader may look at a discussion of this point at the end of Section 2.7 in Ref. [10].

[17]Note that when $\eta_{\min} = \eta_{\max} \equiv \eta$, all pairs of distinct input symbols are equally confusable, i.e., $\eta(x, x') = \sum_y \mathcal{N}(y|x)\delta(y \in Y_{x'})$ is constant across all $x, x' \in X$.

[18]Recall that the general protocol of Section 3.1 does not rely on the existence of KS sets. It's only the particular strategy of Ref. [6] that makes use of them.

such that $q > \alpha(G(\mathcal{N}))$ for any channel constructed from it following the prescription of Cubitt *et al.* [6] mentioned above.[19] Hence, to answer whether the entanglement-assisted enhancement of the one-shot zero-error capacity achieved by the strategy of Ref. [6] carries through for any KS set, we need to settle the following question:

Does every KS set admit a KS basis set of size $q > \alpha(O(\Gamma))$? Here Γ is the contextuality scenario corresponding to the KS set and $O(\Gamma)$ is the orthogonality graph of Γ.[10, 27]

If the answer is in the affirmative, then the Cubitt *et al.* strategy can be used starting from arbitrary KS sets. If not, then there must exist a counter-example. Indeed, we can find such a counter-example and, therefore, the Cubitt *et al.* strategy is not applicable to arbitrary KS sets: it only works for KS sets that admit (disjoint) KS basis sets. Our counter-example comes from the Conway-Kochen 31-vector KS set, the smallest known KS set in dimension $d = 3$ [45]. We refer to Appendix F for details.

This raises the following important open problem: *Given an arbitrary KS set, what are the necessary and sufficient criteria for it to admit a KS basis set?*

8 Discussion and outlook

We have generalized and unified the protocols of Refs. [6, 28] in a broad framework for entanglement-assisted one-shot classical communication that should prove useful for future investigations. Our results bear witness to the role that noise-robust contextuality à la Spekkens [13] plays in this task. Indeed, the problem of entanglement-assisted one-shot classical communication provides a fertile ground to study the rich interplay between the Kochen-Specker theorem [3], Spekkens contextuality [13] and its hypergraph-theoretic formulations [10, 27], and nonlocal games. Several open questions and opportunities for future work arise:

1. Does there exist a generic construction of a nonlocal game, instantiating Proposition 1, for any channel $\mathcal{N}$ that admits an enhancement of its one-shot success probability? Even extending the family of channels for which such a construction exists beyond the case we have shown, i.e., the

family of output-uniform k-regular channels of Theorem 1, would constitute progress in this direction. Furthermore, even within this family of channels, an important question is to characterize those for which the corresponding nonlocal game admits a gap between classical and quantum/nonsignalling correlations. These channels would, in turn, admit an advantage in a corresponding one-shot communication task because of Theorem 1.

2. While the channels considered in Refs. [6, 28] are KS channels, it remains an open question whether more general channels (in particular, with KS-colourable channel hypergraphs) exhibit non-trivial advantages in enhancing the one-shot success probability using entanglement. Our general protocol in Section 3 does not specifically rely on the channel being KS-uncolourable. For example, a simple channel corresponding to a statistical proof of KS-contextuality is the one based on the KCBS construction on a qutrit [46]. It consists of 10 vertices, denoted $\{v_i, w_i\}_{i=1}^{5}$, and 5 hyperedges, denoted $\{v_i, w_i, v_{i+1}\}_{i=1}^{5}$ (addition modulo 5, so that $i + 1 = 1$ for $i = 5$), with $\alpha(G(\mathcal{N})) = 3$ (equal to its one-shot zero-error capacity); a natural question then arises: can entanglement be used to enhance the one-shot success probability of sending two bits using this channel? What would be the role of noise-robust contextuality à la Refs. [10, 18] in enabling such an enhancement? If not, can *any* channel with a KS-colourable hypergraph admit enhancement of its one-shot success probability?

3. The Cubitt *et al.* protocol [6] requires the existence of (disjoint) KS basis sets. Is it possible to modify this protocol to use KS sets which do not admit (disjoint) KS basis sets, e.g., the Conway-Kochen 31-vector KS set? Would such a modification still allow for the possibility of enhancing the one-shot zero-error capacity of a classical channel? Or would it, maybe, only allow for an enhancement of the one-shot success probability following the general protocol we discussed in Section 3? The existence of pseudotelepathy games based on KS sets [5] suggests that a quantum advantage in some corresponding one-shot communication task (following Theorem 1) should be possible. A related question is: what is the simplest scenario that admits enhancement of the one-shot zero-error capacity of a classical channel? Is the example studied in Ref. [6] the simplest one, or is it possible to further reduce, say, the size of the input alphabet or the dimension of the quantum system for which the enhancement is achieved? Of course, insofar as one uses KS sets to achieve this enhancement, this is also related to the smallest possible KS sets: in

---

[19]In Ref. [6], there is a claim that the existence of KS basis sets is "a corollary of the KS theorem". This is true if a KS basis set is allowed, in general, to contain bases that share vectors. However, for the Cubitt *et al.* construction to work, the bases in a *KS basis set* must be disjoint, i.e., no vectors are shared between bases: this is what allows Alice to encode her messages unambiguously in the outcomes of these measurement bases. Further, for an advantage, the number of these disjoint bases in a KS basis set must exceed the independence number of the confusability graph of the channel constructed from the KS basis set. Hence, in our definition of a *KS basis set*, we explicitly include the disjointness of bases, something Ref. [6] implicitly assumed.

dimension 3, it's been shown that the smallest KS set can have no fewer than 22 vectors [47, 48] (the 31-vector Conway-Kochen construction still being the smallest one known in 3 dimensions). Is there a smaller KS set (fewer than 24 vectors) than Peres's 24-vector set that also admits a KS basis set?

4. The enhancement of the one-shot success probability using the Cubitt *et al.* strategy relies on the orthogonality (compatibility) relations between projectors (projective measurements). Are there nontrivial examples of enhancement of the one-shot success probability that are only achievable with nonprojective measurements, perhaps inspired by joint measurability structures that lie outside the purview of projective measurements [49, 50]? We know that there exist such joint measurability structures, e.g., Specker's scenario, admitting proofs of contextuality [51, 52].

5. The connection of the one-shot communication task with preparation contextuality could also be leveraged to obtain bounds on inaccessible information in preparation contextual ontological models of quantum theory, following the ideas recently proposed in Ref. [53].

More generally, the problem of entanglement-assisted one-shot zero-error communication can be viewed as a channel simulation problem, i.e., using a noisy channel to simulate a noiseless channel in a one-shot setting using nonsignalling correlations [54]. The relaxation of it to the case of enhancing the one-shot success probability (which we have studied) can be viewed as using a noisy channel to simulate a less noisy channel using nonsignalling correlations, i.e., noise-attenuation of a classical channel using a nonclassical common-cause resource [33]. We have focussed in this paper on the interplay of this latter channel simulation problem with the contextuality of the system that the receiver (Bob) holds in the communication task. A worthwhile project here is a rigorous resource-theoretic account of this problem to better understand how various resources affect the simulation preorder over classical channels in this noisy setting [55]: whether perhaps the resource of LOSR-entanglement [56] is more appropriate than LOCC-entanglement when viewing the resource aspects of entanglement (and how this affects, for example, the usefulness of Tsirelson boxes vs. Hardy boxes [33, 56, 57] in this task), how the resource of noise-robust contextuality on one wing of a Bell experiment plays with bipartite nonlocality, and, more abstractly, the usefulness of a common-cause resource in simulating a direct-cause resource (e.g., the fact that entanglement can increase the one-shot zero-error capacity of a classical channel [6]). It would also be interesting to see if the contextuality witnesses we have considered in this paper turn out to be related to some

monotones for channel (non-)conversions in a resource theory of channel simulation.

References

[1] J. S. Bell, On the Einstein-Podolsky-Rosen paradox, Physics **1**, 195 (1964).

[2] 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).

[3] S. Kochen and E. P. Specker, The Problem of Hidden Variables in Quantum Mechanics, in *The logico-algebraic approach to quantum mechanics* (Springer, 1975) pp. 293–328.

[4] R. Renner and S. Wolf, Quantum pseudo-telepathy and the Kochen-Specker theorem, in *International Symposium on Information Theory, 2004. ISIT 2004. Proceedings.* (IEEE, 2004) pp. 322–322.

[5] G. Brassard, A. Broadbent, and A. Tapp, Quantum pseudo-telepathy, Foundations of Physics **35**, 1877 (2005).

[6] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, Improving Zero-Error Classical Com-

munication with Entanglement, Phys. Rev. Lett. **104**, 230503 (2010).

[7] M. Howard, J. Wallman, V. Veitch, and J. Emerson, Contextuality supplies the 'magic' for quantum computation, Nature **510**, 351 (2014).

[8] J. Barrett and A. Kent, Non-contextuality, finite precision measurement and the Kochen-Specker theorem, Studies in History and Philosophy of Science Part B: Studies in History and Philosophy of Modern Physics **35**, 151 (2004).

[9] A. Winter, What does an experimental test of quantum contextuality prove or disprove?, Journal of Physics A: Mathematical and Theoretical **47**, 424031 (2014).

[10] R. Kunjwal, Beyond the Cabello-Severini-Winter framework: Making sense of contextuality without sharpness of measurements, Quantum **3**, 184 (2019).

[11] A. Cabello, What do we learn about quantum theory from Kochen-Specker quantum contextuality?, PIRSA 17070034 (2017).

[12] G. Chiribella and X. Yuan, Measurement sharpness cuts nonlocality and contextuality in every physical theory, arXiv preprint arXiv:1404.3348 (2014).

[13] R. W. Spekkens, Contextuality for preparations, transformations, and unsharp measurements, Phys. Rev. A **71**, 052108 (2005).

[14] M. D. Mazurek, M. F. Pusey, R. Kunjwal, K. J. Resch, and R. W. Spekkens, An experimental test of noncontextuality without unphysical idealizations, Nature Communications **7**, 1 (2016).

[15] M. F. Pusey, L. Del Rio, and B. Meyer, Contextuality without access to a tomographically complete set, arXiv preprint arXiv:1904.08699 (2019).

[16] M. D. Mazurek, M. F. Pusey, K. J. Resch, and R. W. Spekkens, Experimentally bounding deviations from quantum theory in the landscape of generalized probabilistic theories, PRX Quantum **2**, 020302 (2021).

[17] R. Kunjwal and R. W. Spekkens, From the Kochen-Specker Theorem to Noncontextuality Inequalities without Assuming Determinism, Phys. Rev. Lett. **115**, 110403 (2015).

[18] R. Kunjwal and R. W. Spekkens, From statistical proofs of the Kochen-Specker theorem to noise-robust noncontextuality inequalities, Phys. Rev. A **97**, 052110 (2018).

[19] R. W. Spekkens, D. H. Buzacott, A. J. Keehn, B. Toner, and G. J. Pryde, Preparation Contextuality Powers Parity-Oblivious Multiplexing, Phys. Rev. Lett. **102**, 010401 (2009).

[20] A. Chailloux, I. Kerenidis, S. Kundu, and J. Sikora, Optimal bounds for parity-oblivious random access codes, New Journal of Physics **18**, 045003 (2016).

[21] D. Schmid and R. W. Spekkens, Contextual Advantage for State Discrimination, Phys. Rev. X **8**, 011015 (2018).

[22] D. Saha and A. Chaturvedi, Preparation contextuality as an essential feature underlying quantum communication advantage, Phys. Rev. A **100**, 022108 (2019).

[23] D. Saha, P. Horodecki, and M. Pawłowski, State independent contextuality advances one-way communication, New Journal of Physics **21**, 093057 (2019).

[24] R. Kunjwal, M. Lostaglio, and M. F. Pusey, Anomalous weak values and contextuality: Robustness, tightness, and imaginary parts, Phys. Rev. A **100**, 042116 (2019).

[25] M. Lostaglio and G. Senno, Contextual advantage for state-dependent cloning, Quantum **4**, 258 (2020).

[26] R. Kunjwal, Contextuality beyond the Kochen-Specker theorem, arXiv preprint arXiv:1612.07250 (2016).

[27] R. Kunjwal, Hypergraph framework for irreducible noncontextuality inequalities from logical proofs of the Kochen-Specker theorem, Quantum **4**, 219 (2020).

[28] R. Prevedel, Y. Lu, W. Matthews, R. Kaltenbaek, and K. J. Resch, Entanglement-Enhanced Classical Communication Over a Noisy Classical Channel, Phys. Rev. Lett. **106**, 110505 (2011).

[29] B. Hemenway, C. A. Miller, Y. Shi, and M. Wootters, Optimal entanglement-assisted one-shot classical communication, Phys. Rev. A **87**, 062301 (2013).

[30] J. Barrett, Information processing in generalized probabilistic theories, Phys. Rev. A **75**, 032304 (2007).

[31] A. Acín, T. Fritz, A. Leverrier, and A. B. Sainz, A Combinatorial Approach to Nonlocality and Contextuality, Communications in Mathematical Physics **334**, 533 (2015).

[32] R. W. Spekkens, The ontological identity of empirical indiscernibles: Leibniz's methodological principle and its significance in the work of Einstein, arXiv preprint arXiv:1909.04628 (2019).

[33] E. Wolfe, D. Schmid, A. B. Sainz, R. Kunjwal, and R. W. Spekkens, Quantifying Bell: the Resource Theory of Nonclassicality of Common-Cause Boxes, Quantum **4**, 280 (2020).

[34] M. F. Pusey, Robust preparation noncontextuality inequalities in the simplest scenario, Phys. Rev. A **98**, 022112 (2018).

[35] A. Tavakoli and R. Uola, Measurement incompatibility and steering are necessary and sufficient for operational contextuality, Phys. Rev. Research **2**, 013011 (2020).

[36] M. S. Leifer and O. J. E. Maroney, Maximally Epistemic Interpretations of the Quantum State

and Contextuality, Phys. Rev. Lett. **110**, 120401 (2013).

[37] L. P. Hughston, R. Jozsa, and W. K. Wootters, A complete classification of quantum ensembles having a given density matrix, Physics Letters A **183**, 14 (1993).

[38] M. Banik, S. S. Bhattacharya, S. K. Choudhary, A. Mukherjee, and A. Roy, Ontological models, preparation contextuality and nonlocality, Foundations of Physics **44**, 1230 (2014).

[39] P. Heywood and M. L. Redhead, Nonlocality and the Kochen-Specker paradox, Foundations of Physics **13**, 481 (1983).

[40] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, Rev. Mod. Phys. **86**, 419 (2014).

[41] S. Popescu and D. Rohrlich, Quantum nonlocality as an axiom, Foundations of Physics **24**, 379 (1994).

[42] A. Peres, Two simple proofs of the Kochen-Specker theorem, Journal of Physics A: Mathematical and General **24**, L175 (1991).

[43] A. Peres, Incompatible results of quantum measurements, Physics Letters A **151**, 107 (1990).

[44] N. D. Mermin, Hidden variables and the two theorems of John Bell, Rev. Mod. Phys. **65**, 803 (1993).

[45] A. Peres, *Quantum theory: concepts and methods*, Vol. 57 (Springer Science & Business Media, 2006).

[46] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumovsky, Simple Test for Hidden Variables in Spin-1 Systems, Phys. Rev. Lett. **101**, 020403 (2008).

[47] S. Uijlen and B. Westerbaan, A Kochen-Specker system has at least 22 vectors, New Generation Computing **34**, 3 (2016).

[48] F. Arends, A lower bound on the size of the smallest Kochen-Specker vector system, Master's thesis, Oxford University (2009).

[49] R. Kunjwal, C. Heunen, and T. Fritz, Quantum realization of arbitrary joint measurability structures, Phys. Rev. A **89**, 052126 (2014).

[50] N. Andrejic and R. Kunjwal, Joint measurability structures realizable with qubit measurements: Incompatibility via marginal surgery, Phys. Rev. Research **2**, 043147 (2020).

[51] R. Kunjwal and S. Ghosh, Minimal state-dependent proof of measurement contextuality for a qubit, Phys. Rev. A **89**, 042118 (2014).

[52] X. Zhan, E. G. Cavalcanti, J. Li, Z. Bian, Y. Zhang, H. M. Wiseman, and P. Xue, Experimental generalized contextuality with single-photon qubits, Optica **4**, 966 (2017).

[53] I. Marvian, Inaccessible information in probabilistic models of quantum systems, non-contextuality inequalities and noise thresholds for contextuality, arXiv preprint arXiv:2003.05984 (2020).

[54] T. S. Cubitt, D. Leung, W. Matthews, and A. Winter, Zero-error channel capacity and simulation assisted by non-local correlations, IEEE Transactions on Information Theory **57**, 5509 (2011).

[55] C. E. Shannon, A note on a partial ordering for communication channels, Information and control **1**, 390 (1958).

[56] D. Schmid, T. C. Fraser, R. Kunjwal, A. B. Sainz, E. Wolfe, and R. W. Spekkens, Understanding the interplay of entanglement and nonlocality: motivating and developing a new branch of entanglement theory, arXiv preprint arXiv:2004.09194 (2020).

[57] L. Hardy, Nonlocality for two particles without inequalities for almost all entangled states, Phys. Rev. Lett. **71**, 1665 (1993).

[58] A. Cabello, J. Estebaranz, and G. García-Alcaine, Bell-Kochen-Specker theorem: A proof with 18 vectors, Physics Letters A **212**, 183 (1996).

A Preparation noncontextuality on one wing of a bipartite Bell experiment implies local causality

We will use quantum notation below for ease of understanding, but the argument applies to all non-signalling general probabilistic theories (GPTs).

Consider a general bipartite Bell scenario where Alice's measurement settings are labelled by s, Bob's settings are labelled by t, and their respective outcomes are labelled by a and b. Their joint statistics is, therefore, given by $p(a,b|s,t) = \mathrm{Tr}(E_a^{(s)} \otimes E_b^{(t)} \rho_{\mathrm{AB}})$, where $\{E_a^{(s)}\}_a$ denotes the POVM associated with s, $\{E_b^{(t)}\}_b$ denotes the POVM associated with t, and ρ_{AB} is the entangled state shared between Alice and Bob. We now consider the prepare-and-measure experiment on Bob's side[20] that this Bell scenario induces: Bob's preparations are steered by Alice's measurements, i.e., every measurement outcome $E_a^{(s)}$ on Alice's side steers Bob's system to (an unnormalized state) $\sigma_{a|s} = \mathrm{Tr}_A(E_a^{(s)} \otimes I \rho_{\mathrm{AB}})$. However, no-signalling requires that Bob should not be able to infer Alice's measurement setting s by local interventions on his system alone, so that we have $\sum_a \sigma_{a|s} = \rho_{\mathrm{B}}$ for all measurement settings s that Alice can choose. Each s therefore labels a preparation ensemble $\{p(a|s), \rho_{a|s}\}_a$ on Bob's side such that $\sum_a p(a|s)\rho_{a|s} = \rho_{\mathrm{B}}$ for all s, where $p(a|s) = \mathrm{Tr}_B \sigma_{a|s}$

---

[20]The same argument goes through with the roles of Alice and Bob interchanged.

and $\rho_{a|s} = \frac{\sigma_{a|s}}{p(a|s)}$. Given this operational equivalence between the preparation ensembles on Bob's side, the assumption of preparation noncontextuality entails that any ontological model of Bob's system must satisfy $\sum_a p(a|s)p(\lambda|s,a) = p(\lambda)$ for all s. This can be rewritten as $\sum_a p(a|s,\lambda)p(\lambda|s) = p(\lambda)$ for all s, i.e., $p(\lambda|s) = p(\lambda)$ for all s. We then have, given Bob's measurement outcomes $E_b^{(t)}$, that

$$p(b|t,s,a) = \mathrm{Tr}_B(E_b^{(t)}\rho_{a|s}) = \sum_\lambda p(b|t,\lambda)p(\lambda|s,a), \tag{56}$$

and

$$
\begin{aligned}
&p(a,b|s,t)\\
=&p(a|s)p(b|t,s,a)\\
=&\sum_\lambda p(b|t,\lambda)p(a|s)p(\lambda|s,a)\\
=&\sum_\lambda p(b|t,\lambda)p(a|s,\lambda)p(\lambda|s)\\
=&\sum_\lambda p(b|t,\lambda)p(a|s,\lambda)p(\lambda), \tag{57}
\end{aligned}
$$

where the last equality follows from the assumption of preparation noncontextuality. Thus, the existence of a preparation noncontextual ontological model for Bob's system (or for Alice's system, by symmetry) implies the existence of locally causal ontological model for the bipartite Bell experiment. Note that

$$
\begin{aligned}
&p(a,b|s,t)\\
=&\mathrm{Tr}(E_a^{(s)} \otimes E_b^{(t)}\rho_{\mathrm{AB}})\\
=&\mathrm{Tr}(E_a^{(s)} \otimes I\rho_{\mathrm{AB}})\mathrm{Tr}_{\mathrm{B}}(E_b^{(t)}\rho_{a|s}), \tag{58}
\end{aligned}
$$

where $p(a|s) = \mathrm{Tr}(E_a^{(s)} \otimes I\rho_{\mathrm{AB}})$ and $p(b|t,s,a) = \mathrm{Tr}_{\mathrm{B}}(E_b^{(t)}\rho_{a|s})$.

B Proof of Theorem 1

We begin by noting that, following Eq. (15), we have

$$
\begin{aligned}
&S_{\mathrm{Cl}}^{\max}\\
=&\max_{p(x,m'|m,y)\in\mathcal{L}} S\\
=&\frac{1}{|\mathbf{Msg}|}\frac{1}{k}\max_{p(x,m'|m,y)\in\mathcal{L}}\sum_{m,m',x,y\in Y_x}\delta_{m,m'}p(x,m'|m,y). \tag{59}
\end{aligned}
$$

The Bell expression of Eq. (19) can be rewritten by making the substitution (recalling that $p(y) = \frac{1}{|Y|}$)

$$\sum_{y\notin Y_x} p(y) = 1 - \sum_{y\in Y_x} p(y) = 1 - \frac{k}{|Y|}, \tag{60}$$

and using the no-signalling condition

$$\sum_{m'} p(x,m'|m,y) = p(x|m), \text{ for all } y\in Y, \tag{61}$$

to express the second term of Eq. (19) as follows:

$$
\begin{aligned}
&\frac{1}{|\mathbf{Msg}|}\sum_{m,m',x,y\notin Y_x}\frac{1}{|Y|}p(x,m'|m,y)\\
=&\frac{1}{|\mathbf{Msg}|}\sum_{m,x,y\notin Y_x}\frac{1}{|Y|}p(x|m)\\
=&\frac{1}{|\mathbf{Msg}|}|\mathbf{Msg}|\left(1-\frac{k}{|Y|}\right)\\
=&1-\frac{k}{|Y|}. \tag{62}
\end{aligned}
$$

The Bell expression then becomes

$$
\begin{aligned}
&S_{\mathrm{Bell}}\\
=&1-\frac{k}{|Y|}+\frac{1}{|\mathbf{Msg}|}\frac{1}{|Y|}\sum_{m,m',x,y\in Y_x}\delta_{m,m'}p(x,m'|m,y)\\
=&1-\frac{k}{|Y|}+\frac{k}{|Y|}S, \tag{63}
\end{aligned}
$$

using Eq. (16). This means that the following holds:

$$S = 1 \Leftrightarrow S_{\mathrm{Bell}} = 1, \tag{64}$$

i.e., the one-shot zero-error communication occurs if and only if the corresponding nonlocal game is won with certainty. On the other hand, this also means that

$$
\begin{aligned}
S_{\mathrm{local}}^{\max} =& \max_{p(x,m'|m,y)\in\mathcal{L}} S_{\mathrm{Bell}}\\
=&1-\frac{k}{|Y|}+\frac{k}{|Y|}\max_{p(x,m'|m,y)\in\mathcal{L}} S\\
=&1-\frac{k}{|Y|}+\frac{k}{|Y|}S_{\mathrm{Cl}}^{\max}, \tag{65}
\end{aligned}
$$

so that we finally have

$$S > S_{\mathrm{Cl}}^{\max} \Leftrightarrow S_{\mathrm{Bell}} > S_{\mathrm{local}}^{\max}. \tag{66}$$

C The one-shot success probability in the Prevedel *et al.* protocol

The expression for the one-shot success probability in Eq. (24) can be obtained as follows: starting from the general expression for S, we have

$$
\begin{aligned}
S =& \sum_m p(m)\sum_x p(x|m)\sum_y \mathcal{N}(y|x)\sum_v p(v|y)\\
&\sum_z p(z|v,m,x)p(m'=m|z,y) \tag{67}\\
=& \sum_m p(m)\sum_x p(x|m)\sum_y \mathcal{N}(y|x)\sum_v p(v|y)\\
&\sum_z p(z|v,m,x)\delta_{g(z,y),m}. \tag{68}
\end{aligned}
$$

This then becomes

$$
\begin{aligned}
S = &\sum_m p(m) \sum_x p(x|m) \mathcal{N}(y=(1,b_1)|x=(b_1,b_2)) \\
&\sum_v p(v|y) \sum_z p(z|v,m,x)\delta_{b_1,m} \\
+&\sum_m p(m) \sum_x p(x|m) \mathcal{N}(y=(2,b_2)|x=(b_1,b_2)) \\
&\sum_v \delta_{v,1} \sum_z p(z|v,m,x)\delta_{b_2\oplus z,m} \\
+&\sum_m p(m) \sum_x p(x|m) \mathcal{N}(y=(P,b_1\oplus b_2)|x=(b_1,b_2)) \\
&\sum_v \delta_{v,0} \sum_z p(z|v,m,x)\delta_{b_1\oplus b_2\oplus z,m} \\
=&\sum_m p(m) \sum_{b_2} p(x=(m,b_2)|m) \\
&\mathcal{N}(y=(1,m)|x=(m,b_2)) \\
+&\sum_m p(m) \sum_x p(x|m) \mathcal{N}(y=(2,b_2)|x) \\
&\sum_z p(z|v=1,m,x)\delta_{b_2\oplus z,m} \\
+&\sum_m p(m) \sum_x p(x|m) \mathcal{N}(y=(P,b_1\oplus b_2)|x) \\
&\sum_z p(z|v=0,m,x)\delta_{b_1\oplus b_2\oplus z,m}.
\end{aligned}
\tag{69}
$$

Finally, for $p(m)=\frac{1}{2}$, we have

$$
\begin{aligned}
S =&\sum_m \frac{1}{2} \sum_{b_2} p(x=(m,b_2)|m)\frac{1}{3} \\
+&\sum_m \frac{1}{2} \sum_x p(x|m)\frac{1}{3} \sum_z p(z|v=1,m,x)\delta_{b_2\oplus z,m} \\
+&\sum_m \frac{1}{2} \sum_x p(x|m)\frac{1}{3} \sum_z p(z|v=0,m,x)\delta_{m\oplus b_2\oplus z,m} \\
=&\frac{1}{3}+\frac{1}{6}\sum_m \sum_x p(x|m) \sum_z p(z|v=1,m,x)\delta_{b_2\oplus z,m} \\
+&\frac{1}{6}\sum_m \sum_x p(x|m) \sum_z p(z|v=0,m,x)\delta_{b_2\oplus z,0} \\
=&\frac{1}{3}+\frac{1}{6}\sum_m \sum_x \sum_z p(x,z|m,v=1)\delta_{b_2\oplus z,m} \\
+&\frac{1}{6}\sum_m \sum_x \sum_z p(x,z|m,v=0)\delta_{b_2\oplus z,0} \\
=&\frac{1}{3}+\frac{1}{6}\sum_m \sum_{b_2} \sum_z p(b_2,z|m,v=1)\delta_{b_2\oplus z,m} \\
+&\frac{1}{6}\sum_m \sum_{b_2} \sum_z p(b_2,z|m,v=0)\delta_{b_2\oplus z,0} \\
=&\frac{1}{3}+\frac{1}{6}\sum_{b_2,z,m,v} p(b_2,z|m,v)\delta_{b_2\oplus z,mv}.
\end{aligned}
\tag{70}
$$

D The nonlocal game for the Prevedel et al. protocol following Theorem 1

The Prevedel *et al.* protocol [28] is evidently built around the CHSH game. How does this square with the general construction of a nonlocal game that we referred to in Theorem 1? We do a consistency check here. Following the recipe for constructing a nonlocal game starting from a classical channel $\mathcal{N}$, outlined in Section 5, we have the following expression for the probability of success in the nonlocal game:

$$
\begin{aligned}
S_{\text{Bell}} \\
=\frac{1}{12}&\sum_{m,m',x,y\notin Y_x} p(x,m'|m,y) \\
+\frac{1}{12}&\sum_{m,m',x,y\in Y_x} p(x,m'|m,y)\delta_{m,m'},
\end{aligned}
\tag{71}
$$

where we used $p(m,y)=p(m)p(y)=\frac{1}{2}\frac{1}{6}=\frac{1}{12}$.

The expression for the one-shot success probability, S, is, of course, given by (using $p(m)=\frac{1}{2}$ and $\mathcal{N}(y|x)=\frac{1}{3}\delta(y\in Y_x)$)

$$
S = \frac{1}{6}\sum_{m,m',x,y\in Y_x} p(x,m'|m,y)\delta_{m,m'}.
\tag{72}
$$

Hence, using the fact that $k=3$ and $|Y|=6$ in this example, and following Eq. (63), we have

$$
S_{\text{Bell}} = \frac{1}{2}+\frac{1}{2}S.
\tag{73}
$$

Recalling Eq. (26),

$$
S = \frac{1}{3}+\frac{2}{3}S_{\text{CHSH}},
\tag{74}
$$

and we therefore have

$$
S_{\text{Bell}} = \frac{2}{3}+\frac{1}{3}S_{\text{CHSH}}
\tag{75}
$$

for the nonlocal game defined according to Section 5 and used in Theorem 1. Hence, we have that the nonlocal game constructed from our general recipe is essentially the CHSH game, except that the two inputs on Bob's side are disguised as six inputs labelled by $y\in Y$ that are classically post-processed to obtain $v\in\{0,1\}$ and the output for each y is given by $m'\in\{0,1\}$ obtained by classically post-processing z,y (Fig. 1).

E Contextuality witnessed by the weighted max-predictability is sufficient for a quantum advantage

Recall the expression for S given in Eq. (39), i.e.,

$$
S = \text{Corr}+\sum_m p(m) \sum_{x,x'\in X^{(m)}} (1-\delta_{x,x'})p(x,x'|m)\eta(x,x'),
\tag{76}
$$

where, from Eq. (38), we have

$$\text{Corr} = \sum_m p(m) \sum_{x,x' \in X^{(m)}} \delta_{x,x'} p(x,x'|m). \quad (77)$$

From Eq. (46), we have, under the assumption of noncontextuality, that

$$S = \sum_\lambda \nu(\lambda) \sum_m p(m)$$
$$\sum_{x,x' \in X^{(m)}} \xi(x'|m,\lambda)\mu(x|m,\lambda)\eta(x,x'), \quad (78)$$

so that

$$\text{Corr}(\lambda) \equiv \sum_m p(m) \sum_{x,x' \in X^{(m)}} \delta_{x,x'}\xi(x'|m,\lambda)\mu(x|m,\lambda), \quad (79)$$

and

$$\text{Corr} = \sum_\lambda \text{Corr}(\lambda)\nu(\lambda) \quad (80)$$

We now proceed to upper bound $S_{\text{NC}}^{\max}$ in terms of a hypergraph invariant:

$$S_{\text{NC}}^{\max}$$
$$= \max_\lambda \Bigg(\text{Corr}(\lambda) + \sum_m p(m)$$
$$\sum_{x,x' \in X^{(m)}} \Big((1-\delta_{x,x'})\xi(x'|\mathbb{M}_m^{(A)},\lambda)$$
$$\mu(x|\mathbb{S}_m,\lambda)\eta(x,x') \Big) \Bigg)$$
$$\leq \max_\lambda \left(\text{Corr}(\lambda) + \eta_{\max}(1-\text{Corr}(\lambda)) \right)$$
$$= \max_\lambda \left(\eta_{\max} + \text{Corr}(\lambda)(1-\eta_{\max}) \right)$$
$$= \eta_{\max} + (1-\eta_{\max})\max_\lambda \text{Corr}(\lambda)$$
$$\leq \eta_{\max} + (1-\eta_{\max})\beta(\Gamma,\{p(m)\}_m), \quad (81)$$

where

$$\beta(\Gamma,\{p(m)\}_m) \equiv \max_\lambda \sum_m p(m) \max_{x \in X^{(m)}} \xi(x|m,\lambda) \quad (82)$$

is the weighted max-predictability [10, 27]. Hence, we have

$$S \leq S_{\text{NC}}^{\max} \leq \eta_{\max} + (1-\eta_{\max})\beta(\Gamma,\{p(m)\}_m). \quad (83)$$

A special case: We now consider the special case when $\eta_{\max} = \eta_{\min} = \eta$. Following Eq. (39), the lower and upper bounds on S coincide and we have

$$S = \text{Corr} + \eta(1-\text{Corr}) \quad (84)$$
$$= \eta + \text{Corr}(1-\eta). \quad (85)$$

The upper bound from noncontextuality becomes (cf. Eq. (81))

$$S_{\text{NC}}^{\max} = \eta + \max_\lambda \text{Corr}(\lambda)(1-\eta). \quad (86)$$

We then have, from noncontextuality, that

$$S \leq S_{\text{NC}}^{\max}$$
$$\leq \eta + (1-\eta)\beta(\Gamma,\{p(m)\}_m). \quad (87)$$

Together with Eq. (85), this gives us the following:

$$\text{Corr} > \beta(\Gamma,\{p(m)\}_m)$$
$$\Leftrightarrow S > \eta + (1-\eta)\beta(\Gamma,\{p(m)\}_m)$$
$$\geq S_{\text{NC}}^{\max} = S_{\text{Cl(CIG)}}^{\max}. \quad (88)$$

Recalling that

$$\text{Corr} \leq \beta(\Gamma,\{p(m)\}_m) \quad (89)$$

is a noise-robust noncontextuality inequality [27], we have that the contextuality witnessed by $\text{Corr} > \beta(\Gamma,\{p(m)\}_m)$ is sufficient for a quantum advantage when $\eta_{\max} = \eta_{\min}$.

Another special case: The sufficiency of $\text{Corr} > \beta(\Gamma,\{p(m)\}_m)$ for a quantum advantage also arises when $S_{\text{imperf}} = 0$. Then we have that $S = S_{\text{perf}} = \text{Corr}$ and $S_{\text{NC}}^{\max} \leq \beta(\Gamma,\{p(m)\}_m)$, so that the violation of $\text{Corr} \leq \beta(\Gamma,\{p(m)\}_m)$ implies the violation of $S \leq S_{\text{NC}}^{\max}$. Indeed, in the ideal quantum case considered by Cubitt *et al.*[6], we see that $\text{Corr} = 1$, maximally violating the noncontextuality inequality and achieving a success probability of 1.

F   Not every KS set admits a KS basis set: the Conway-Kochen 31-vector KS set

Consider the simplest known KS set in $d = 3$ dimensions, namely, the Conway-Kochen 31-vector KS set [45]. The 31 vectors are carved up into 17 complete orthogonal bases (with 3 vectors each) and 20 incomplete orthogonal bases (with 2 vectors each). The orthogonality graph has an independence number of 11 and the only disjoint basis sets of size greater than 11 are those of size 12 and 13. None of these disjoint basis sets forms a KS basis set, hence no quantum advantage over the unassisted one-shot zero-error capacity of 11 can be obtained via the methods of Ref. [6] for this construction. A remaining possibility is that, on adding the missing vectors in the 20 incomplete orthogonal bases to the KS set, the orthogonality relations between the resulting set of 51 vectors will perhaps allow for an advantage. We rule out this possibility as well: after including 20 additional vectors that render all the bases that appear in this KS set complete, we have a contextuality scenario represented by a hypergraph containing 51 vertices carved up into 37

(three-vertex) hyperedges. We check for any additional orthogonality relations arising from the newly introduced 20 vectors and find 4 additional incomplete orthogonal bases. On further completing these 4 bases by adding 4 more vectors, we find that there are, overall, 55 vertices (vectors) carved up into 41 hyperedges (complete orthogonal bases) and there are no additional orthogonality relations. The orthogonality graph associated with this extended contextuality scenario has an independence number of 25. Hence, for an advantage based on the strategy of Ref. [6], there must exist a KS basis set of size $q > 25$. However, the largest disjoint basis set is still of size 13 and it does not form a KS basis set. Hence, the strategy of Ref. [6] does not provide an advantage even (and especially) when extending Conway-Kochen 31-vector KS set to complete all incomplete orthogonal bases and include any additional orthogonality relations (leaving no incomplete orthogonal bases). This provides a counter-example to the question we posed, showing that the Cubitt *et al.* strategy doesn't work for arbitrary KS sets.

One might wonder why we bother "completing" the original 31-vector KS set to 55-vector KS set with no incomplete bases. We do this to rule out the possibility that something akin to the 18-vector KS set in 4 dimensions [58] is happening here: for that KS set, it's not possible to implement the Cubitt *et al.* protocol, but supplementing it with the remaining set of 6 vectors (out of Peres's 24-vector KS set [42] from which the 18-vector set is drawn) and taking into account the resulting additional orthogonality relations yields Peres's 24-vector KS set for which the Cubitt *et al.* protocol works. From our investigation, it is clear that for the 31-vector KS set, such a situation doesn't arise even after "completing" it.

We provide below a list of all the vectors and bases in the orginal as well as the "completed" KS set for the Conway-Kochen argument, so that the interested reader may verify our claims concerning this KS set.

F.1 Conway-Kochen 31-vector KS set

The 31 vectors (labelled from 1 to 31) are:

$1 : (-1, 2, 1), 2 : (-1, 2, 0), 3 : (0, 2, 1), 4 : (-1, 2, -1),$
$5 : (0, 2, 0), 6 : (1, 2, 1), 7 : (0, 2, -1), 8 : (1, 2, 0),$
$9 : (1, 2, -1), 10 : (0, 2, -2), 11 : (2, 2, 0), 12 : (2, 2, -2),$
$13 : (-1, 1, -2), 14 : (0, 1, -2), 15 : (-1, 0, -2),$
$16 : (0, 0, -2), 17 : (-1, -1, -2), 18 : (0, -1, -2),$
$19 : (0, -2, -2), 20 : (2, 1, -1), 21 : (2, 1, 0),$
$22 : (2, 0, -1), 23 : (2, 0, 0), 24 : (2, -1, -1),$
$25 : (2, -1, 0), 26 : (2, -2, 0), 27 : (2, 0, -2),$
$28 : (2, -2, -2), 29 : (2, 2, 2), 30 : (2, 0, 2),$
$31 : (2, -2, 2).$ (90)

The orthogonality relations of between these vec-

tors are the following (the first entry in each list is the vector with respect to which the remaining vectors in the list are orthogonal):

$[1; 12, 14, 21, 30], [2; 16, 20, 21], [3; 13, 14, 23],$
$[4; 18, 21, 27, 29], [5; 15, 16, 22, 23, 27, 30],$
$[6; 14, 25, 27, 31], [7; 17, 18, 23], [8; 16, 24, 25],$
$[9; 18, 25, 28, 30], [10; 19, 23, 24, 28, 29],$
$[11; 13, 16, 26, 28, 31], [12; 1, 17, 19, 26, 30],$
$[13; 3, 11, 22, 28], [14; 1, 3, 6, 23], [15; 5, 20, 22, 24],$
$[16; 2, 5, 8, 11, 21, 23, 25, 26], [17; 7, 12, 22, 26],$
$[18; 4, 7, 9, 23], [19; 10, 12, 20, 23, 31], [20; 2, 15, 19, 31],$
$[21; 1, 2, 4, 16], [22; 5, 13, 15, 17],$
$[23; 3, 5, 7, 10, 14, 16, 18, 19], [24; 8, 10, 15, 29],$
$[25; 6, 8, 9, 16], [26; 11, 12, 16, 17, 29],$
$[27; 4, 5, 6, 29, 30, 31], [28; 9, 10, 11, 13, 30],$
$[29; 4, 10, 24, 26, 27], [30; 1, 5, 9, 12, 27, 28],$
$[31; 6, 11, 19, 20, 27].$ (91)

The 17 complete orthogonal bases (vectors labelled as above) are:

$\{1, 12, 30\}, \{2, 16, 21\}, \{3, 14, 23\}, \{4, 27, 29\}, \{5, 15, 22\},$
$\{5, 16, 23\}, \{5, 27, 30\}, \{6, 27, 31\}, \{7, 18, 23\}, \{8, 16, 25\},$
$\{9, 28, 30\}, \{10, 19, 23\}, \{10, 24, 29\}, \{11, 13, 28\},$
$\{11, 16, 26\}, \{12, 17, 26\}, \{19, 20, 31\}.$ (92)

The 20 incomplete orthogonal bases are:

$\{1, 14\}, \{1, 21\}, \{2, 20\}, \{3, 13\}, \{4, 18\},$
$\{4, 21\}, \{6, 14\}, \{6, 25\}, \{7, 17\}, \{8, 24\},$
$\{9, 18\}, \{9, 25\}, \{10, 28\}, \{11, 31\}, \{12, 19\},$
$\{13, 22\}, \{15, 20\}, \{15, 24\}, \{17, 22\}, \{26, 29\}.$ (93)

F.2 The "completed" 55-vector KS set

The 20 vectors that complete the incomplete bases are:

$32 : (4, -4, -8), 33 : (-5, -1, 2), 34 : (-4, -4, 8),$
$35 : (-1, -2, -5), 36 : (5, 2, -1), 37 : (-1, -5, -2),$
$38 : (1, -5, 2), 39 : (-1, -2, 5), 40 : (-8, -4, -4),$
$41 : (5, -1, -2), 42 : (-5, 2, -1), 43 : (2, 5, -1),$
$44 : (-2, 1, -5), 45 : (1, -2, -5), 46 : (-2, -1, -5),$
$47 : (-5, -2, -1), 48 : (-5, 2, 1), 49 : (8, -4, 4),$
$50 : (-1, 2, -5), 51 : (-2, 5, 1).$ (94)

The resulting set of 17 newly complete bases is then:

$\{1, 14, 47\}, \{1, 21, 50\}, \{2, 20, 46\}, \{3, 13, 33\}, \{4, 18, 36\},$
$\{4, 21, 45\}, \{6, 14, 48\}, \{6, 25, 39\}, \{7, 17, 41\}, \{8, 24, 44\},$
$\{9, 18, 42\}, \{9, 25, 35\}, \{10, 28, 40\}, \{11, 31, 32\},$
$\{12, 19, 49\}, \{13, 22, 37\}, \{15, 20, 51\}, \{15, 24, 43\},$
$\{17, 22, 38\}, \{26, 29, 34\}.$ (95)

Taking into account possible extra orthogonality relations not captured by the set of 37 complete bases, it turns out that there are 4 additional incomplete bases in the set of 51 vectors above:

$$\{2, 40\}, \{3, 34\}, \{7, 32\}, \{8, 49\}. \qquad (96)$$

To complete these bases we add 4 more vectors to the 51-vector KS set:

$$52 : (20, 4, 8), 53 : (-20, 4, -8),$$
$$54 : (8, 4, -20), 55 : (8, -4, -20). \qquad (97)$$

The 4 additional newly complete bases are then

$$\{2, 40, 54\}, \{3, 34, 53\}, \{7, 32, 52\}, \{8, 49, 55\}. \qquad (98)$$

There are no new orthogonality relations in this completed set of 55 vectors carved up into 41 complete orthogonal bases.