

Does ignorance of the whole imply ignorance of the parts?

— Large violations of non-contextuality in quantum theory

Thomas Vidick¹ and Stephanie Wehner²

¹Computer Science division, UC Berkeley, USA*

²Center for Quantum Technologies, National University of Singapore, 2 Science Drive 3, 117543 Singapore[†]

(Dated: May 5, 2011)

A central question in our understanding of the physical world is how our knowledge of the whole relates to our knowledge of the individual parts. One aspect of this question is the following: to what extent does ignorance about a whole preclude knowledge of at least one of its parts? Relying purely on classical intuition, one would certainly be inclined to conjecture that a strong ignorance of the whole cannot come without significant ignorance of at least one of its parts. Indeed, we show that this reasoning holds in any non-contextual hidden variable model (NC-HV). Curiously, however, such a conjecture is *false* in quantum theory: we provide an explicit example where a large ignorance about the whole can coexist with an almost perfect knowledge of each of its parts. More specifically, we provide a simple information-theoretic inequality satisfied in any NC-HV, but which can be *arbitrarily* violated by quantum mechanics. Our inequality has interesting implications for quantum cryptography.

In this note we examine the following seemingly innocent question: does one's ignorance about the whole necessarily imply ignorance about at least one of its parts? Given just a moments thought, the initial reaction is generally to give a positive answer. Surely, if one cannot know the whole, then one should be able to point to an unknown part. Classically, and more generally for any deterministic non-contextual hidden variable model, our intuition turns out to be correct: ignorance about the whole does indeed imply the existence of a specific part which is unknown, so that one can point to the source of one's ignorance. However, we will show that in a quantum world this intuition is flawed.

THE PROBLEM

Let us first explain our problem more formally. Consider two dits y_0 and $y_1 \in \{0, \dots, d-1\}$, where the string $y = y_0y_1$ plays the role of the whole, and y_0, y_1 are the individual parts. Let ρ_y denote an encoding of the string y into a classical or quantum state. In quantum theory, ρ_y is simply a density operator, and in a NC-HV model it is a preparation \mathcal{P}_y described by a probability distribution over hidden variables $\lambda \in \Lambda$. Let P_Y be a probability distribution over $\{0, \dots, d-1\}^2$, and imagine that with probability $P_Y(y)$ we are given the state ρ_y . The optimum probability of guessing y given its encoding ρ_y , which lies in a register E , can be written as

$$P_{\text{guess}}(Y|E) = \max_{\{\mathcal{M}\}} \sum_{y \in \{0, \dots, d-1\}^2} P_Y(y) p(y|\mathcal{M}, \mathcal{P}_y), \quad (1)$$

where $p(y|\mathcal{M}, \mathcal{P}_y)$ is the probability of obtaining outcome y when measuring the preparation \mathcal{P}_y with \mathcal{M} ,

and the maximization is taken over all d^2 -outcome measurements allowed in the theory. In the case of quantum theory, for example, the maximization is taken over POVMs $\mathcal{M} = \{M_y\}_y$ and $p(y|\mathcal{M}, \mathcal{P}_y) = \text{tr}(M_y \rho_y)$. The guessing probability is directly related to the conditional min-entropy $H_\infty(Y|E)$ through the equation [2]

$$H_\infty(Y|E) := -\log P_{\text{guess}}(Y|E). \quad (2)$$

This measure plays an important role in quantum cryptography and is the relevant measure of information in the single shot setting corresponding to our everyday experience, as opposed to the asymptotic setting captured by the von Neumann entropy. A closely related variant is the smooth min-entropy $H_\infty^\epsilon(Y|E)$ which can be thought of as being like $H_\infty(Y|E)$ except with some small error probability ϵ . The main question we are interested in can then be loosely phrased as:

How does $H_\infty(Y = Y_0Y_1|E)$ (ignorance about the whole) relate to $H_\infty(Y_C|EC)$, for $C \in \{0, 1\}$ (ignorance about the parts)?

Here the introduction of the additional random variable C is crucial, and it can be understood as a pointer to the part of Y about which there is large ignorance (given a large ignorance of the whole string Y); see Figure 1 for an illustration of this role. It is important to note that the choice of C should be consistent with the encoding prior to its definition. That is, whereas C may of course depend on Y_0, Y_1 and the encoding E , the reduced state on registers holding Y_0, Y_1 and E after tracing out C should remain the same. In particular, this condition states that C cannot be the result of a measurement causing disturbance to the encoding register; if we were allowed to destroy information in the encoding we would effectively alter the original situation.

*Electronic address: vidick@eecs.berkeley.edu

[†]Electronic address: wehner@nus.edu.sg

RESULTS

An inequality valid in any NC-HV model.

We first show that classically, or more generally in any non-contextual hidden variable model [22], ignorance about the whole really *does* imply ignorance about a part. More specifically, we show that for any random variable $Y = Y_0Y_1$ and side information E , there exists a random variable $C \in \{0, 1\}$ such that

$$H_\infty(Y_C|EC) \gtrsim \frac{H_\infty(Y_0Y_1|E)}{2}. \quad (3)$$

This inequality can be understood as an information-theoretic analogue of Bell inequalities to the question of non-contextuality. Classically, this inequality is known as the *min-entropy splitting inequality*, and plays an important role in the proof of security of some (classical) cryptographic primitives [3, 4]. The proof of (3) is a straightforward extension to the case of standard NC-HV models [7, 8] of a classical technique known as min-entropy splitting first introduced by Wullschleger [3], and we defer details to the appendix.

The fact that C is a random variable, rather than being deterministically chosen, is important, and an example will help clarify its role. Consider Y uniformly distributed over $\{0, \dots, d-1\}^2$ and $E = Y_0$ with probability $1/2$, and Y_1 with probability $1/2$. In this case it is easy to see that *both* Y_0 and Y_1 can be guessed from E with average success probability $1/2 + 1/(2d)$, so that $H_\infty(Y_0|E) = H_\infty(Y_1|E) \approx 1$, which is much less than $H_\infty(Y|E) \approx \log d$. However, define C as 0 if $E = Y_1$ and 1 if $E = Y_0$. Then it is clear that $H_\infty(Y_C|EC) = \log d$, as we are always asked to predict the variable about which we have no side information at all! In this case the random variable C “points to the unknown” by being correlated with the side information E , but is entirely consistent with our knowledge about the world: by tracing out C we recover the initial joint distribution on (Y, E) . This also highlights the important difference between the task we are considering and the well-studied random access codes [5, 6], in which the requirement is to be able to predict one of Y_0, Y_1 (adversarially chosen) from their encoding; for this task it has been demonstrated that there is virtually no asymptotic difference between classical and quantum encodings (see below for a discussion).

It is interesting to note that (3) still holds if we consider a somewhat “helpful” physical model in which in addition to the encoding one might learn a small number of “leaked” bits of information about Y . More specifically, if the NC-HV discloses m extra bits of information then it follows from the chain rule for the min-entropy (see appendix) that

$$H_\infty(Y_C|EC) \gtrsim \frac{H_\infty(Y_0Y_1|E)}{2} - m. \quad (4)$$

Violation in quantum theory. Our main result shows that (3) is violated in the strongest possible sense

by quantum theory. More specifically, we provide an explicit construction that demonstrates this violation: Let $Y = Y_0Y_1$ be uniformly distributed over $\{0, \dots, d-1\}^2$. Given $y = y_0y_1 \in \{0, \dots, d-1\}^2$, define its encoding $\rho_{y_0y_1}^E = |\Psi_y\rangle\langle\Psi_y|$ as

$$|\Psi_y\rangle := X_d^{y_0} Z_d^{y_1} |\Psi\rangle, \quad (5)$$

where X_d and Z_d are the generalized Pauli matrices and

$$|\Psi\rangle := \frac{1}{\sqrt{2\left(1 + \frac{1}{\sqrt{d}}\right)}} (|0\rangle + F|0\rangle), \quad (6)$$

with F being the matrix of the Fourier transform over \mathbb{Z}_d . Since we are only interested in showing a quantum violation, we will for simplicity always assume that d is prime [23]. The system YE is then described by the ccq-state

$$\rho_{Y_0Y_1E} = \frac{1}{d^2} \sum_{y_0, y_1} |y_0\rangle\langle y_0| \otimes |y_1\rangle\langle y_1| \otimes \rho_{y_0y_1}^E. \quad (7)$$

We first prove that $H_\infty(Y|E) = \log d$ for our choice of encoding. We then show the striking fact that, even though the encoding we defined gives very little information about the whole string Y , for any adversarially chosen random variable C (possibly correlated with our encoding) one can guess Y_C from its encoding ρ_E with essentially constant probability. More precisely, for any ccqc-state $\rho_{Y_0Y_1EC}$, with $C \in \{0, 1\}$, that satisfies the consistency relation $\text{tr}_C(\rho_{Y_0Y_1EC}) = \rho_{Y_0Y_1E}$, we have

$$H_\infty(Y_C|EC) \approx 1 \quad (8)$$

for *any* sufficiently large d . This shows that the inequality (3) can be violated arbitrarily (with d), giving a striking example of the malleability of quantum information. What’s more, it is not hard to show that this effect still holds even for H_∞^ε , for constant error ε , and a “helpful” physical model leaking $m \approx c \log d$ bits of information with $c < 1/2$. Hence, the violation of the inequality (3) has the appealing feature of being very robust. Indeed, for any number of bits m a NC-HV might leak in addition, we could find a d to ensure a violation.

PROOF OF THE QUANTUM VIOLATION

We now provide an outline of the proof that the encoding specified in (5) leads to a quantum violation of the splitting inequality (3); for completeness, we provide a more detailed derivation in the appendix. Our proof proceeds in three steps: first, by computing $H_\infty(Y|E)$ we show that the encoding does indeed not reveal much information about the whole. Second, we compute the

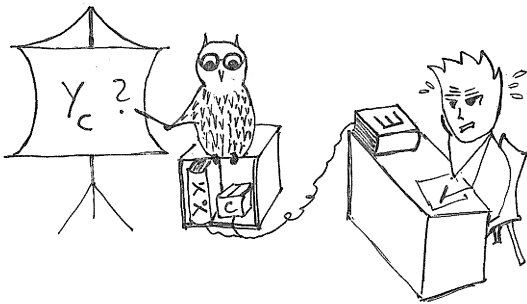


FIG. 1: Intuitively, one can also understand our result in terms of a game between Bob and a malicious challenger, the Owl. Imagine Bob is taking a philosophy class teaching him knowledge about Y , clearly chosen uniformly at random. Unfortunately, he never actually attended and had insufficient time to prepare for his exam. Luckily, however, he has been given some encoding E of the possible answers Y_0Y_1 , hastily prepared by his old friend Alice. When entering the room, he had to submit E for inspection to the challenger who knows Y_0, Y_1 as well as the encoding Alice might use. After inspection, the challenger may secretly keep a system C , possibly correlated with E , but such that the reduced system on Y_0, Y_1 and E looks untampered with. It is immediately obvious to the challenger that Bob must be ignorant about the whole of Y_0Y_1 . But can it always measure and point to a $C = c$ such that Bob is ignorant about Y_C ? That is, can it always detect Bob's ignorance by challenging him to output a single Y_C ? Classically, this is indeed possible: ignorance about the whole of Y_0Y_1 implies significant ignorance about one of the parts, Y_C . However, a quantum Bob could beat the Owl.

optimal measurements for extracting Y_0 and Y_1 on average, and show that these measurements perform equally well for any other prior distribution on Y . Finally, we show that even introducing an additional system C does not change one's ability to extract Y_C from the encoding.

Step 1: Very intuitively, ignorance about the whole string already follows from Holevo's theorem and the fact that we are trying to encode 2 dits into a d -dimensional quantum system. To see this more explicitly, recall that $H_\infty(Y|E) = \log d$ is equivalent to showing that $P_{\text{guess}}(Y|E) = 1/d$. From (1) we have that this guessing probability is given by the solution to the following semidefinite program (SDP)

$$\begin{aligned} & \text{maximize} && \frac{1}{d^2} \sum_{y_0, y_1} \text{tr}(M_{y_0 y_1} |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|) \\ & \text{subject to} && M_{y_0 y_1} \geq 0 \text{ for all } y_0, y_1, \\ & && \sum_{y_0, y_1} M_{y_0, y_1} = \mathbb{I}. \end{aligned}$$

The dual SDP is easily found to be

$$\begin{aligned} & \text{minimize} && \text{Tr}(Q) \\ & \text{subject to} && Q \geq \frac{1}{d^2} |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}| \text{ for all } y_0, y_1. \end{aligned}$$

Let v_{primal} and v_{dual} be the optimal values of the primal and dual respectively. By the property of weak duality, $v_{\text{dual}} \geq v_{\text{primal}}$ always holds. Hence, to prove our result, we only need to find a primal and dual solutions for which $v_{\text{primal}} = v_{\text{dual}} = 1/d$. It is easy to check that $\hat{Q} = \mathbb{I}/d^2$ is a dual solution with value

$v_{\text{dual}} = \text{tr}(\hat{Q}) = 1/d$. Similarly, consider the measurement $M_{y_0 y_1} = |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|/d$. Using Schur's lemma, one can directly verify that $\sum_{y_0, y_1} M_{y_0 y_1} = \mathbb{I}$, giving $v_{\text{primal}} = 1/d$. The claimed value of the conditional min-entropy follows.

Step 2: A similar argument, exploiting the symmetries in the encoding, can be used to show that

$$P_{\text{guess}}(Y_0|E) = P_{\text{guess}}(Y_1|E) = \frac{1}{2} + \frac{1}{2\sqrt{d}}. \quad (9)$$

The measurements that attain these values are given by the eigenbases of Z_d and X_d respectively.

As a remark to quantum information theorists, note that this means that our encoding doubles as a random access encoding of the string y into a d -dimensional quantum state ρ_y with probability (9) to recover y_0 or y_1 . For $d = 2$, such encodings have previously been considered in the realm of contextuality as a reinterpretation of the CHSH inequality [9, 10]. However, we note that this is *not* what is surprising here, as there exists an obvious classical random access encoding for 2 dits into a single dit (see discussion on C above), with recovery probability $1/2 + 1/(2d)$.

Simply computing (9) is hence insufficient for our purposes. Let us write $\{|y_0\rangle, y_0 \in \{0, \dots, d-1\}\}$ for the eigenbasis of Z_d , and note that its Fourier transform $\{|F|y_1\rangle, y_1 \in \{0, \dots, d-1\}\}$ is then the eigenbasis of X_d . Exploiting the symmetries in our problem, it is straightforward to verify that for all $y_0, y_1 \in \{0, \dots, d-1\}$

$$|\langle y_0 | \Psi_{y_0 y_1} \rangle|^2 = |\langle y_1 | F^\dagger | \Psi_{y_0 y_1} \rangle|^2 = \frac{1}{2} + \frac{1}{2\sqrt{d}}. \quad (10)$$

An important consequence of this is that for *any* other prior distribution $P_{y_0 y_1}$, measurement in the Z_d eigenbasis distinguishes the states

$$\sigma_{y_0} = \sum_{y_1} P_{y_0 y_1}(y_0, y_1) |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|, \quad (11)$$

with probability at least $1/2 + 1/(2\sqrt{d})$, even when the distribution is unknown. A similar argument can be made for the marginal states σ_{y_1} and measurement in the X_d eigenbasis.

Step 3: It now remains to show that, for any possible choice of an additional classical system C [24], one can still guess Y_C from the encoding with a good success probability: one cannot construct a C which would "point to the unknown". Note that we may express the joint state with any other system C as

$$\rho_{Y_0 Y_1 E C} = \frac{1}{d^2} \sum_{y_0 y_1} |y_0\rangle \langle y_0| \otimes |y_1\rangle \langle y_1| \otimes \rho_{y_0 y_1 c}^{EC}, \quad (12)$$

for some states $\rho_{y_0 y_1 c}^{EC}$ on registers E and C . Since the reduced state on Y_0, Y_1 and E should be the same for any C we have by the fact that Y_0 and Y_1 are classical that $\text{tr}_C(\rho_{y_0 y_1 c}^{EC}) = |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|$. Since $|\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|$ is a

pure state, this implies that $\rho_{y_0 y_1 c}^{EC} = |\Psi_{y_0 y_1}\rangle\langle\Psi_{y_0 y_1}| \otimes \sigma_{y_0 y_1}^C$. Now imagine that we were to perform some arbitrary measurement on C , whose outcome would supposedly point to an unknown substring. But this merely creates a different distribution $P_{y_0 y_1}$ over encoded strings, and we already know from the above that we can still succeed in retrieving either y_0 or y_1 with probability at least $1/2 + 1/(2\sqrt{d})$ by making a measurement in the X_d or Z_d basis respectively. Hence for large d we have a recovery probability of roughly $1/2$, implying

$$H_\infty(Y_0|EC=0) \approx H_\infty(Y_1|EC=1) \approx 1, \quad (13)$$

which is our main claim.

Note that the consistency condition, which states that our choice of C should be compatible with the original situation and should not affect the reduced state, is important, and makes our task non-trivial. As an example, consider our construction for $d = 2$. In that case the encoding states lie in the XZ -plane of the Bloch sphere. Imagine now that we measured the encoding register E in the eigenbasis of σ_y , and let the outcome be C . But for any measurement in the eigenbasis of σ_y we observe entirely random outcomes, and the post-measurement states trivially no longer carry any information about the encoded string. Indeed, any choice of C would do if we are allowed to destroy information in such a manner.

IMPLICATIONS FOR CRYPTOGRAPHY

Our result answers an interesting open question in quantum cryptography [11], namely whether min-entropy splitting can still be performed when conditioned on quantum instead of classical knowledge. This technique was used to deal with classical side information E in [4, 12]. Our example shows that quantum min-entropy splitting is impossible, even when we would be willing to accept subtracting a large error term on the r.h.s. of (3). This tells us that classical protocols that rely on such statements may become insecure in the presence of quantum side information, and highlights the importance of so-called min-entropy sampling results of [13] used in quantum cryptography [14] instead. It also indicates that contextuality may play a more important role in our understanding of the possibilities and limits of quantum cryptography than previously thought.

DISCUSSION

The first indication that something may be amiss when looking at knowledge from a quantum perspective was given by Schrödinger [15], who pointed out that one can have *knowledge* (not ignorance) about the whole, while still being ignorant about the parts [25]. Here, we tackled this problem from a very different direction, starting with the premise that one has ignorance about the whole.

Our results show that contextuality is responsible for much more significant effects than have previously been noted. In particular, it leads to arbitrarily large quantum violations of (3), which can be understood as a Bell-type inequality for non-contextuality. This is still true even for a somewhat “helpful” physical model, leaking additional bits of information. To our knowledge, this is the first *information-theoretic* inequality distinguishing NC-HV models from quantum theory. Our question and perspective are completely novel, and we hope that our observations will lead to an increased understanding of the role of contextuality. In this work, we have considered standard NC-HVs in which all HVs can be decomposed as convex combinations of extremal HVs which give deterministic outcomes for effects (see appendix). It is an interesting open question whether our results can be generalized to very general models that distinguish between measurement and preparation contextuality [1].

At the heart of our result lies the fact that contextuality allows for strong forms of complementarity in quantum mechanics (often conflated with uncertainty [16]), which intuitively is responsible for allowing the violation of (3). Typically, complementarity is discussed by considering examples of properties of a physical system that one may be able to determine individually, but which cannot all be learned at once. In spirit, this is similar to the notion of a random access encoding where we could determine either property Y_0 or Y_1 quite well, but not all of Y . However, as discussed above this can also be true classically, in a probabilistic sense. We would thus like to emphasize the novelty of our perspective, as we approach the problem from the other end, and first demonstrate the general result that in an NC-HV ignorance about the whole always implies ignorance about a part. We then show that in a quantum world, this principle is violated in the strongest possible sense, even with respect to an additional system C . One could think of this as a much more robust way of capturing the intuitive notion of complementarity [17].

Finally, it is an interesting open question whether our inequality can be experimentally verified. Note that this made difficult by the fact that our aim would be to test *ignorance* rather than knowledge. However, it is conceivable that such an experiment can be performed by building a larger cryptographic protocol whose security relies on being ignorant about one of the parts of a string Y created during that protocol [26]. A quantum violation could then be observed by breaking the security of the protocol, and exhibiting *knowledge* (rather than ignorance) about some information that could not have been obtained if the protocol was secure.

Acknowledgments

We thank Jonathan Oppenheim, Christian Schaffner, Tony Short, Robert Spekkens and CQT’s “non-local club” for useful comments. We are particularly grate-

ful to Tony Short for pointing out that our problem could more easily be explained by means of the game depicted in Figure 1. TV was supported by ARO Grant W911NF-09-1-0440 and NSF Grant CCF-0905626. SW was supported by the National Research Foundation, and

the Ministry of Education, Singapore. TV is grateful to CQT, Singapore, for hosting him while part of this work was done. SW is grateful for an invitation from the Mittag-Leffler Institute, Sweden, where part of this work was performed.

-
- [1] R. Spekkens, *Physical Review A* **71**, 052108 (2005).
- [2] R. König, R. Renner, and C. Schaffner, *IEEE Trans. Info.* **55** (2009).
- [3] J. Wullschleger, in *Advances in Cryptology — EURO-CRYPT '07* (Springer-Verlag, 2007), Lecture Notes in Computer Science.
- [4] I. B. Damgård, S. Fehr, R. Renner, L. Salvail, and C. Schaffner, in *Advances in Cryptology—CRYPTO '07* (Springer-Verlag, 2007), vol. 4622 of *Lecture Notes in Computer Science*, pp. 360–378.
- [5] A. Nayak, in *Proceedings of 40th IEEE FOCS* (1999), pp. 369–376.
- [6] A. Ambainis, A. Nayak, A. Ta-Shma, and U. Vazirani, in *Proceedings of 31st ACM STOC* (1999), pp. 376–383.
- [7] A. A. Klyachko, M. A. Can, S. Binicioğlu, and A. S. Shumsovsky, *Physical Review Letters* **101**, 020403 (2008).
- [8] A. Cabello, S. Severini, and A. Winter (2010), arXiv:1010.2163.
- [9] J. Clauser, M. Horne, A. Shimony, and R. Holt, *Phys. Rev. Lett.* **23**, 880 (1969).
- [10] R. Spekkens, D. Buzacott, A. Keehn, B. Toner, and G. Pryde, *Physical Review Letters* **102**, 010401 (2009).
- [11] C. Schaffner (2007), Personal communication.
- [12] I. B. Damgård, S. Fehr, L. Salvail, and C. Schaffner, in *Proceedings of 46th IEEE FOCS* (2005), pp. 449–458.
- [13] R. König and R. Renner (2007), arXiv:0712.4291.
- [14] R. König, S. Wehner, and J. Wullschleger (2009), arXiv:0906.1030.
- [15] E. Schrödinger, *Naturwissenschaften* **23**, 807,823,840 (1935).
- [16] J. Oppenheim and S. Wehner, *Science* **330**, 1072 (2010).
- [17] J. Oppenheim and S. Wehner (2010), in preparation.
- [18] A. J. Short and S. Wehner, *New Journal of Physics* **12**, 033023 (2010).
- [19] R. Renner, Ph.D. thesis, ETH Zurich (2005), quant-ph/0512258.
- [20] A. Bowie, *Introduction to German philosophy: from Kant to Habermas* (2003).
- [21] R. Q. Odendaal and A. R. Plastino, *Eur. J. of Physics* **31**, 193 (2010).
- [22] Intuitively, a non-contextual model is one in which the observable statistics of a particular measurement do not depend on the *context* in which the measurement is performed, and in particular on which other compatible measurements are possibly performed simultaneously. Here, and as is usual, we consider non-contextual models in which the measurements are composed of deterministic effects. We refer to the appendix for formal definitions.
- [23] Recall that there are infinitely many primes.
- [24] In principle, C could be arbitrary, but since we are only interested in the result of a 2-outcome measurement on C , we assume that C is indeed already classical and use the subscript C to denote that classical value.
- [25] The whole here being a maximally entangled state, and the parts being the individual (locally completely mixed) subsystems. See also [21] for an example.
- [26] For example, one could consider weak forms of oblivious transfer where one only demands security against the receiver.
- [27] A simple example in the classical setting — storing each of the halves with probability half — should convince the reader that this is necessary.
- [28] We will write a max instead of a sup everywhere, and assume that the dimension of the system E is finite.
- [29] Some of these variables may be hidden, in the sense that the operational description of a particular preparation does not necessarily fully determine the distribution on variables that describe it, but we will not need to make that distinction.

In this appendix, we provide a detailed derivation of our results. To this end, we first provide some more detailed background on the entropic quantities we use in Section A. In Section B we show that the splitting inequality (B8) is satisfied in any deterministic non-contextual hidden variable model (NC-HV model for short). This is a minor twist on the existing classical proof [4] due to Wullschleger [3]. Finally, in Section C, we proceed to prove our main result, that there exists a quantum encoding which strongly violates the splitting inequality (B8).

Appendix A: Entropy measures

Throughout, we will measure information in terms of the min-entropy, which is directly related to the *guessing probability* $P_{\text{guess}}(Y|E)$ [2], where Y is a classical string ranging in the set \mathcal{Y} and E an auxiliary system. It is defined as the maximum probability with which one can predict the whole string Y , given the system E . The maximization is over all possible observations, or measurements, on E ; these vary depending on the physical model (e.g. classical or quantum) under consideration.

Definition A.1. *Let Y be a classical random variable with distribution P_Y taking values in a set \mathcal{Y} , and $\{\mathcal{P}_y\}_{y \in \mathcal{Y}}$ any set of preparations on E . Then the maximum guessing probability of Y given E is defined as*

$$P_{\text{guess}}(Y|E) = \max_{\{\mathcal{M}_y\}} \sum_{y \in \mathcal{Y}} P_Y(y) p(y|\mathcal{P}_y, \mathcal{M}_y), \quad (\text{A1})$$

where the maximum is taken over all measurements $\mathcal{M} = \{\mathcal{M}_y\}$ allowed in the model.

For instance, in the case of quantum mechanics we simply have

$$P_{\text{guess}}(Y|E) := \max_{\substack{\{M_y\}_{y \in \mathcal{Y}} \\ \forall y M_y \geq 0 \\ \sum_y M_y = \mathbb{I}}} \sum_y P_Y(y) \text{tr}(M_y \rho_y^E), \quad (\text{A2})$$

where the maximization is taken over all POVMs [28] and ρ_y^E denotes the reduced state of the system on E , when $Y = y$. For *classical* side-information E this expression simplifies to

$$P_{\text{guess}}(Y|E) := \mathbb{E}_{e \leftarrow E} \left[\max_y P_{Y|E=e}(y) \right]. \quad (\text{A3})$$

In other words, for classical side information, the optimal guessing measurement is to simply output the y which is most likely given the classical value e .

For the case of classical and quantum theories it is known that the guessing probability directly relates to the conditional min-entropy [2]. Here, we follow the operational approach of [18], and define the conditional min-entropy for an arbitrary theory with classical Y as

$$H_{\infty}(Y|E) := -\log P_{\text{guess}}(Y|E). \quad (\text{A4})$$

For the case of quantum systems, the conditional min-entropy was first introduced by Renner [19] as a way to measure randomness conditioned on an adversary's knowledge. The min-entropy can also be defined when Y is quantum itself [19], but we will not need it here. In the quantum setting, we will also use a smoothed version of the quantum conditional min-entropy, defined for any $\varepsilon > 0$ as:

$$H_{\infty}^{\varepsilon}(Y|E) = \max_{\tilde{\rho}_{YE} \in \mathcal{B}_{\varepsilon}(\rho_{YE})} H_{\infty}(Y|E)_{\tilde{\rho}_{YE}}, \quad (\text{A5})$$

where the maximization is taken over all (subnormalized) states $\tilde{\rho}_{YE}$ within ε trace distance of ρ_{YE} . A similar definition could be made for arbitrary theories using the distance defined in [18], but we will not require it here.

The conditional min-entropy has a number of appealing properties, which for any NC-HV model essentially follow from its operational interpretation, and also hold in the quantum setting [19]. First of all consider the min-entropy of classical YZ conditioned on side information E . Clearly, since guessing Y and Z can only be more difficult than guessing Y alone, we have $P_{\text{guess}}(YZ|E) \leq P_{\text{guess}}(Y|E)$. Translated, this gives monotonicity of the min-entropy

$$H_{\infty}(YZ|E) \geq H_{\infty}(Y|E). \quad (\text{A6})$$

Similarly, to guess Y and Z from E one strategy would be to guess Z (in the worst case, choosing $Z = z$ with $z \in \mathcal{Z}$ taken from the uniform distribution) and then try to guess Y knowing Z . In terms of guessing probabilities, this means that $P_{\text{guess}}(YZ|E) \geq P_{\text{guess}}(Y|EZ)/|\mathcal{Z}|$. Translated, we obtain the chain rule

$$H_{\infty}(Y|EZ) \geq H_{\infty}(Y|E) - \log |\mathcal{Z}|. \quad (\text{A7})$$

A final property that will be important to us is that, as a direct consequence of (A4) we may also write the min-entropy as

$$H_{\infty}(Y|E) = \min_{\mathcal{M}} H_{\infty}(Y|\mathcal{M}(E)), \quad (\text{A8})$$

where the minimization is taken over all measurements \mathcal{M} , and $H_{\infty}(Y|\mathcal{M}(E))$ is the min-entropy conditioned on the classical information obtained by measuring E with \mathcal{M} .

Appendix B: A splitting inequality valid in any NC-HV model

Before turning to the proof of the generalized splitting inequality, let us briefly review what is meant by a non-contextual model. In any physical theory, we can imagine that a system is prepared according to some *preparation* \mathcal{P} , on which we later make *measurements* \mathcal{M} . Each such measurement can be viewed as a collection of elementary effects \mathbf{f} . The exact form of the effects depends on the model one considers. For example, in quantum theory the effects are simply given by POVM elements. A particularly useful effect is given by the so-called *unit effect* \mathbb{I} , corresponding to the identity in the quantum or classical setting. Hence to any effect one can associate a two-outcome measurement $\mathcal{M}_{\mathbf{f}} = \{\mathbf{f}, \mathbb{I} - \mathbf{f}\}$.

When discussing non-contextuality, this measurement is typically interpreted as a question one might pose to the underlying physical system and has two answers, “yes” for \mathbf{f} and “no” for $\mathbb{I} - \mathbf{f}$. We hence also refer to \mathbf{f} as a question. Of course, one might consider measurements that ask many questions simultaneously, that is, they consist of many individual effects. Two effects are called *compatible* if the corresponding questions can be answered simultaneously without causing disturbance to the underlying physical system, in the sense that we would obtain the same answers again were we to ask the same questions repeatedly.

A set of mutually compatible effects/questions is thereby called a *context*. For example, if \mathbf{f}_1 is compatible with \mathbf{f}_2 the set $\mathcal{C}_1 = \{\mathbf{f}_1, \mathbf{f}_2\}$ is called a *context*. Similarly, if \mathbf{f}_1 is compatible with \mathbf{f}_3 , then the set $\mathcal{C}_2 = \{\mathbf{f}_1, \mathbf{f}_3\}$ is also a context. Note, however, that in such a scenario it can still be that \mathbf{f}_2 and \mathbf{f}_3 are *not* compatible. That is, any effect can be part of multiple *distinct* contexts.

For each effect in a particular context, one can pose the question \mathbf{f} by making the measurement $\mathcal{M}_{\mathbf{f}}$ defined above. Informally, a model is called *non-contextual* if the answer to question \mathbf{f}_1 will always be the same in both contexts, whether $\mathcal{M}_{\mathbf{f}_2}$ or $\mathcal{M}_{\mathbf{f}_3}$ are performed simultaneously (which is possible by definition of being compatible). In our example this means that if were we to make measurement $\mathcal{M}_{\mathbf{f}_1}$ in context \mathcal{C}_1 , or context \mathcal{C}_2 , we would always obtain the same distribution on outcomes.

1. Classical theory

Recall that the phenomenon of min-entropy splitting guarantees that, if a string Y_0Y_1 has high min-entropy then there is a way to split it by introducing a binary random variable C such that the string Y_C has about half as much min-entropy as Y_0Y_1 . Classically, min-entropy splitting follows from the following statement.

Lemma B.1 ([4], Lemma 4.2). *Let $\varepsilon > 0$ and Y_0, Y_1 two random variables such that $H_\infty^\varepsilon(Y_0Y_1|E) \geq \alpha$, where E is classical. Then, there exists a binary random variable C such that $H_\infty^\varepsilon(Y_C|EC) \geq \alpha/2$.*

Using the chain rule (A7) and the monotonicity (A6) of the min-entropy one immediately obtains the statement of min-entropy splitting

$$H_\infty^\varepsilon P(Y_C|EC) \geq \alpha/2 - 1 - \log 1/\varepsilon' \quad (\text{B1})$$

2. Non-contextual hidden variable models

Typically, in a non-contextual hidden variable model it is assumed that a preparation \mathcal{P} is simply a distribution over hidden variables Λ , and a measurement then corresponds to “reading out” such hidden variables. Each outcome event $k \in K$ is associated with a corresponding effect \mathbf{f}_k , where intuitively \mathbf{f}_k “reads out” the hidden variables by mapping a certain subset of possible hidden variables to the outcome $k \in K$. In contrast, some works consider more generalized scenarios known as *ontological models* [1]. The main difference here is that these hidden variable models can locally model even contextual theories, but specify explicit conditions to make these generalized theories non-contextual again.

In this section we show that the splitting inequality holds in any standard *deterministic* NC-HV, which is the definition taken in most previous work, as in e.g. [7, 8]. We will, however, phrase our result in the general language of non-contextual models as introduced in [1], restricting our attention to those models which are deterministic.

a. Background

Very intuitively, a non-contextual *ontological model* for an operational theory associates intrinsic attributes to every physical system, which are supposed to exist independently of the particular context in which the system might be observed. These attributes are described by a set of hidden variables [29] $\lambda \in \Lambda$. Hence for us a hidden variable model consists of the following:

1. A set of hidden variables Λ .
2. For every preparation \mathcal{P} in the physical theory, a probability distribution $p(\lambda|\mathcal{P})$ over $\lambda \in \Lambda$.

3. For every ℓ -outcome measurement \mathcal{M} , and hidden variable $\lambda \in \Lambda$, a probability distribution $p(k|\lambda, \mathcal{M})$ over $k \in [\ell] := \{1, \dots, \ell\}$.

The model is indeed a model for the physical theory if it accurately predicts the outcome distribution of any measurement on any preparation, i.e. performing measurement \mathcal{M} on preparation \mathcal{P} produces outcome k with probability

$$p(k|\mathcal{P}, \mathcal{M}) = \sum_{\lambda} p(k|\lambda, \mathcal{M})p(\lambda|\mathcal{P}), \quad (\text{B2})$$

where for notational simplicity we assume that Λ is discrete.

Effects. We adopt the common notion that measurements are a collection of elementary effects. Here, an effect is a linear functional $\mathbf{f}_k : \Lambda \rightarrow [0, 1]$, mapping hidden variables to outcomes. As is common in the study of non-contextuality [8], we will consider only measurements which are a collection of deterministic effects $\mathbf{f}_k : \Lambda \rightarrow \{0, 1\}$. That is, we effectively work with a *deterministic* model. Much more general scenarios are certainly possible [1] but we will not consider them here. Note that a deterministic model does not mean that there is no more randomness: preparations are given as probability distributions over hidden variables and hence we generally do observe non-deterministic outcomes when measuring a preparation. Of particular importance is the unit effect \mathbb{I} (i.e. the identity), which obeys $\mathbb{I}(\lambda) = 1$ for all $\lambda \in \Lambda$. A measurement is thus a collection $\mathcal{M} := \{\mathbf{f}_k \mid \sum_k \mathbf{f}_k = \mathbb{I}\}$, where we usually index the effects by the outcome that they give in \mathcal{M} . We write the probability of obtaining the outcome k using measurement \mathcal{M} containing the effect \mathbf{f}_k as

$$p(k|\lambda, \mathbf{f}_k) := p(k|\lambda, \mathcal{M}) = \mathbf{f}_k(\lambda). \quad (\text{B3})$$

Note that with every effect, we can again associate a two-outcome measurement $\mathcal{M}_{\mathbf{f}} = \{\mathbf{f}, \mathbb{I} - \mathbf{f}\}$ where without loss of generality we label \mathbf{f} using the outcome ‘1’ and $\mathbb{I} - \mathbf{f}$ using the outcome ‘0’. When concerned with such a measurement $\mathcal{M}_{\mathbf{f}}$ we thus also use $p(1|\lambda, \mathbf{f})$ and $p(0|\lambda, \mathbb{I} - \mathbf{f})$ to denote the probabilities of obtaining outcomes ‘1’ and ‘0’ respectively.

Extensions. Often we wish to relate one physical system to another. For example, we may wish to perform an additional independent experiment such as flipping a coin. Given a system with a set of hidden variables Λ , we allow its extension to a second system in the following way: if Λ' is another set of hidden variables used to describe another physical system, then the combined system will have hidden variables $\Lambda \times \Lambda'$. For every preparation \mathcal{P} on the original system, we say that \mathcal{P}' is an extension of \mathcal{P} in the combined system if for every $\lambda \in \Lambda$

$$p(\lambda|\mathcal{P}) = \sum_{\lambda' \in \Lambda'} p'((\lambda, \lambda')|\mathcal{P}'). \quad (\text{B4})$$

A measurement \mathcal{M}' is similarly said to extend \mathcal{M} as long as

$$p(j|\lambda, \mathcal{M}) = \sum_{\lambda'} p'(j|\lambda, \lambda'), \mathcal{M}'. \quad (\text{B5})$$

Preparations. To study our problem, we will assume that there is an implicit prior distribution on preparations \mathcal{P} describing prior knowledge about the state of the system under consideration. More specifically, we will be concerned with encodings of a string y into preparations \mathcal{P}_y , where the probability $P_Y(y)$ of choosing the string y translates into a prior probability on the preparation as

$$p(\mathcal{P}_y) := P_Y(y). \quad (\text{B6})$$

b. Splitting inequality

We are now ready to generalize Lemma B.1 to any deterministic NC-HV model. The analogue of (B1) is then an easy corollary. Note that in this statement, the conditional min-entropy is understood as being defined through the guessing probability (A1) as in equation (A4). This assumes given a fixed distribution $P_{Y_0 Y_1}$ on the strings $y_0 y_1$, through which a prior distribution on the preparations $\mathcal{P}_{y_0 y_1}$ follows as explained at the end of Section B 2.

Theorem B.2. *Let a NC-HV model \mathfrak{M} be given, with corresponding set of hidden variables Λ . Let $Y = Y_0 Y_1$ be two classical random variables each taking values in a finite set \mathcal{Y} , and $\{\mathcal{P}_{y_0 y_1}\}_{(y_0, y_1) \in \mathcal{Y}^2}$ a corresponding fixed set of preparations on a register E such that*

$$H_\infty(Y_0 Y_1 | E) \geq \alpha. \quad (\text{B7})$$

Then there exists an extended model \mathfrak{M}' over the set of hidden variables $\Lambda' = \Lambda \times \{0_C, 1_C\}$, and a set of preparations $\mathcal{P}'_{y_0 y_1 c}$, for $c \in \{0, 1\}$, extending the $\mathcal{P}_{y_0 y_1}$ and such that

$$H_\infty(Y_C C | E) \geq \frac{\alpha}{2}. \quad (\text{B8})$$

Proof. Recall that we assume a prior distribution on the preparations given by $p(\mathcal{P}_{y_0 y_1}) := P_{Y_0 Y_1}(y_0 y_1)$. This lets us define the guessing probability, which by assumption is such that

$$2^{-\alpha} \geq P_{\text{guess}}(Y | E). \quad (\text{B9})$$

To rewrite the r.h.s. in terms of hidden variables, first of all note that given the prior distribution over preparations we can write the probability of a particular hidden variable $\lambda \in \Lambda$ as

$$p(\lambda) = \sum_{y_0 y_1} p(\mathcal{P}_{y_0 y_1}) p(\lambda | \mathcal{P}_{y_0 y_1}). \quad (\text{B10})$$

Fix a measurement $\mathcal{M} = \{\mathbf{f}_k\}_k$, where we indexed the effects by their outcome in the measurement. By definition, the probability of observing the outcome k when \mathcal{M} is performed on the preparation $\mathcal{P}_{y_0 y_1}$ is

$$p(k | \mathcal{P}_{y_0 y_1}, \mathcal{M}) = \sum_{\lambda} p(\lambda | \mathcal{P}_{y_0 y_1}) p(k | \lambda, \mathbf{f}_k), \quad (\text{B11})$$

The overall probability of observing the outcome k when \mathcal{M} is performed on the preparation \mathcal{P} corresponding to the mixture of the preparations $\mathcal{P}_{y_0 y_1}$ with associated probabilities $p(\mathcal{P}_{y_0 y_1})$ is then

$$p(k) := p(k | \mathcal{P}) = \sum_{y_0 y_1} p(\mathcal{P}_{y_0 y_1}) p(k | \mathcal{P}_{y_0 y_1}). \quad (\text{B12})$$

Note that by definition the hidden variables λ give deterministic outcomes under the measurement of any effect, and hence $p(k | \lambda, \mathbf{f}_k) = \lambda_{\mathbf{f}_k}$, where $\lambda_{\mathbf{f}_k}$ is 1 if the measurement $\{\mathbf{f}_k, \mathbb{I} - \mathbf{f}_k\}$ deterministically produces the outcome ' \mathbf{f}_k ' when performed on a system in state λ , and 0 otherwise. Using Bayes' rule twice we obtain

$$\begin{aligned} p(k) p(\mathcal{P}_{y_0 y_1} | k) &= p(\mathcal{P}_{y_0 y_1}) p(k | \mathcal{P}_{y_0 y_1}) \\ &= p(\mathcal{P}_{y_0 y_1}) \sum_{\lambda} p(\lambda | \mathcal{P}_{y_0 y_1}) p(k | \lambda, \mathbf{f}_k) \\ &= \sum_{\lambda} p(\lambda) p(k | \lambda, \mathbf{f}_k) p(\mathcal{P}_{y_0 y_1} | \lambda). \end{aligned} \quad (\text{B13})$$

Using (A3) and (A8) we obtain that for any measurement $\mathcal{M} = \{\mathbf{f}_k\}$, the guessing probability of $Y_0 Y_1$ is determined by the maximum posterior probability of any string $y_0 y_1$, conditioned on obtaining the outcome k when measuring \mathcal{P} with \mathcal{M} , so that (B9) implies

$$\begin{aligned} 2^{-\alpha} &\geq \sum_k p(k) \max_{y_0 y_1} p(\mathcal{P}_{y_0 y_1} | k) \\ &= \sum_{\lambda} p(\lambda) \max_{y_0 y_1} \left[\sum_k p(k | \lambda, \mathbf{f}_k) p(\mathcal{P}_{y_0 y_1} | \lambda) \right]. \end{aligned} \quad (\text{B14})$$

where in order to invert the summations over k and λ with the maximization we used the fact that for any k , there exists exactly one λ such that $\lambda_{\mathbf{f}_k} = 1$ and vice-versa, so that the summation over λ (resp. over k) which is after the max in the expressions above contains exactly one term. This is a consequence of the fact that the \mathbf{f}_k form a measurement, so that $\sum_k \mathbf{f}_k = 1$, together with the variables λ being deterministic, so that $p(k | \lambda, \mathbf{f}_k)$ can only be either 0 or 1.

We now need to define the additional single-bit random variable C , which is intuitively supposed to designate which of the two halves, y_0 or y_1 , the preparation $\mathcal{P}_{y_0 y_1}$ contains the least amount of information about, so that we can indeed lower-bound the min-entropy $H_\infty(Y_C C | \mathcal{P})$. For this we allow C to be correlated with the preparation $\mathcal{P}_{y_0 y_1}$.

In order to accommodate C , we extend the set of hidden variables Λ' as $\Lambda' \times \{0_C, 1_C\}$. Define $q_1 =$

$\sum_{\lambda} p(\lambda|\mathcal{P}_{y_0y_1})$, where the sum ranges over all λ such that $\sum_{y_0} p(\mathcal{P}_{y_0y_1}|\lambda) \geq 2^{-\alpha/2}$, and $q_0 = 1 - q_1$. Note that q_0 can be computed by the same summation, but now ranging over all λ such that $\sum_{y_0} p(\mathcal{P}_{y_0y_1}|\lambda) < 2^{-\alpha/2}$. Define two preparations as follows:

- $\mathcal{P}_{y_0y_11}$ is defined through the distribution

$$p((\lambda, 1_C)|\mathcal{P}_{y_0y_11}) = \begin{cases} p(\lambda|\mathcal{P}_{y_0y_1})/q_1 & \text{if } \sum_{y_0} p(\mathcal{P}_{y_0y_1}|\lambda) \geq 2^{-\alpha/2} \\ 0 & \text{otherwise} \end{cases} \quad (\text{B15})$$

and $p((\lambda, 0_C)|\mathcal{P}_{y_0y_11}) = 0$ for every λ .

- $\mathcal{P}_{y_0y_10}$ is defined analogously by

$$p((\lambda, 0_C)|\mathcal{P}_{y_0y_10}) = \begin{cases} p(\lambda|\mathcal{P}_{y_0y_1})/q_0 & \text{if } \sum_{y_0} p(\mathcal{P}_{y_0y_1}|\lambda) < 2^{-\alpha/2} \\ 0 & \text{otherwise} \end{cases} \quad (\text{B16})$$

and $p((\lambda, 1_C)|\mathcal{P}_{y_0y_10}) = 0$ for every λ .

Finally, we define the preparation $\mathcal{P}_{y_0y_1C}$ as the mixture of $\mathcal{P}_{y_0y_11}$ with probability q_1 , and of $\mathcal{P}_{y_0y_10}$ with probability q_0 . Note that the preparation $\mathcal{P}_{y_0y_1C}$ is indeed an extension of $\mathcal{P}_{y_0y_1}$ in the new theory, as $p((\lambda, 0_C)|\mathcal{P}_{y_0y_1C}) + p((\lambda, 1_C)|\mathcal{P}_{y_0y_1C}) = p(\lambda|\mathcal{P}_{y_0y_1})$. Finally, we update the prior on preparations by setting $p(\mathcal{P}_{y_0y_11}) = q_1 p(\mathcal{P}_{y_0y_1})$ and $p(\mathcal{P}_{y_0y_10}) = q_0 p(\mathcal{P}_{y_0y_1})$, so that

$$p(\mathcal{P}_{y_0y_1}) = p(\mathcal{P}_{y_0y_10}) + p(\mathcal{P}_{y_0y_11}). \quad (\text{B17})$$

One can check that with these definitions, whenever $\sum_{y_0} p(\mathcal{P}_{y_0y_1}|\lambda) \geq 2^{-\alpha/2}$ we have, using Bayes' rule twice,

$$\begin{aligned} & p(\mathcal{P}_{y_0y_11}|\lambda, 1_C) \\ &= \frac{p((\lambda, 1_C)|\mathcal{P}_{y_0y_11})p(\mathcal{P}_{y_0y_11})}{p((\lambda, 1_C))} \\ &= \frac{(p(\lambda|\mathcal{P}_{y_0y_1})/q_1) \cdot (q_1 p(\mathcal{P}_{y_0y_1}))}{p(\lambda)} \\ &= p(\mathcal{P}_{y_0y_11}|\lambda) \end{aligned} \quad (\text{B18})$$

and 0 otherwise, where for the second equality we used $p(\lambda) = p((\lambda, 1_C))$ for all those λ such that $p((\lambda, 1_C)|\mathcal{P}_{y_0y_11})$ is not zero.

From this point on, our proof follows very closely the classical proof of Lemma B.1. By definition, for every y_1 and every λ , we have that

$$\sum_{y_0} p(\mathcal{P}_{y_0y_10}|\lambda, 0_C) < 2^{-\alpha/2} \quad (\text{B19})$$

It does not seem possible to similarly bound $\sum_{y_1} p(\mathcal{P}_{y_0y_11}|\lambda, 1_C)$, but it is not necessary either, as we do not have access to this quantity directly.

Rather, let $\mathcal{M} = \{\mathbf{f}_k\}_k$ be any $2d$ -outcome measurement; as in (B14) we need to bound

$$\sum_{\lambda} p(\lambda) \max_{y_0} \left[\sum_k p(k|\lambda, 1_C, \mathbf{f}_k) \sum_{y_1} p(\mathcal{P}_{y_0y_11}|\lambda, 1_C) \right]$$

Note that by definition, $\sum_{y_0} p(\mathcal{P}_{y_0y_11}|\lambda, 1_C)$ is either 0 or at least $2^{-\alpha/2}$, so that for all y_0, y_1 and λ , we have the trivial bound

$$\begin{aligned} & p(\mathcal{P}_{y_0y_11}|\lambda, 1_C) \\ & \leq \max_{y_0y_1} p(\mathcal{P}_{y_0y_11}|\lambda, 1_C) 2^{\alpha/2} \sum_{y_0} p(\mathcal{P}_{y_0y_11}|\lambda, 1_C) \\ & = \max_{y_0y_1} p(\mathcal{P}_{y_0y_1}|\lambda) 2^{\alpha/2} \sum_{y_0} p(\mathcal{P}_{y_0y_11}|\lambda, 1_C) \end{aligned} \quad (\text{B20})$$

where for the last equality we used (B18). Summing this equation over all y_1 and combining it with (B14) lets us bound (B20) by $2^{-\alpha} 2^{\alpha/2} \cdot 1 = 2^{-\alpha/2}$. This bound together with (B19) proves the theorem. \square

The fact that min-entropy splitting holds in any NC-HV model now follows as a corollary from Theorem B.2 and the fact that the chain rule (A7) and monotonicity (A6) of the min-entropy also hold for NC-HV models.

Corollary B.3. *Let a NC-HV model \mathfrak{M} be given, with corresponding set of hidden variables Λ . Let $Y = Y_0Y_1$ be two classical random variables each taking values in a finite set \mathcal{Y} , and $\{\mathcal{P}_{y_0y_1}\}_{(y_0, y_1) \in \mathcal{Y}^2}$ a corresponding fixed set of preparations on a register E . Then there exists an extended model \mathfrak{M}' over the set of hidden variables $\Lambda' = \Lambda \times \{0_C, 1_C\}$, and a set of preparations $\mathcal{P}_{y_0y_1c}$, for $c \in \{0, 1\}$, extending the $\mathcal{P}_{y_0y_1}$ such that*

$$H_{\infty}(Y_C|EC) \geq \frac{H_{\infty}(Y_0Y_1|E)}{2} - 1. \quad (\text{B21})$$

To see that this equality is robust is now again an immediate consequence of the chain rule (A7) and monotonicity property (A6), which tell us that when we obtain some additional classical information $A = a$ with $a \in \mathcal{A}$ we have

$$H_{\infty}(Y_C|EAC) \geq H_{\infty}(Y_C A|EC) - \log |\mathcal{A}| \quad (\text{B22})$$

$$\geq H_{\infty}(Y_C|EC) - \log |\mathcal{A}| \quad (\text{B23})$$

$$\geq H_{\infty}(Y_0Y_1|E)/2 - \log |\mathcal{A}| - 1. \quad (\text{B24})$$

That is, a secretly helpful NC-HV leaking a small number $m = \log |\mathcal{A}|$ bits of additional information does not decrease the min-entropy by more than $\log |\mathcal{A}|$ bits.

Appendix C: Splitting is violated by quantum mechanics

We are now ready to show that the splitting inequality (B8) is violated by quantum mechanics in a very strong sense. To this end, we first construct a particular quantum encoding of two dits into one qudit.

1. The encoding

Consider the encoding $E : \{0, \dots, d-1\}^{\times 2} \rightarrow \mathbb{C}^d$ given by

$$E(y_0, y_1) = |\Psi_{y_0 y_1}\rangle := X_d^{y_0} Z_d^{y_1} |\Psi\rangle, \quad (\text{C1})$$

where X_d and Z_d are the generalized Pauli matrices given by their actions on an orthonormal basis $\{|y_0\rangle, y_0 \in \{0, \dots, d-1\}\}$

$$X_d |y_0\rangle = |y_0 + 1 \pmod{d}\rangle, \quad (\text{C2})$$

$$Z_d |y_0\rangle = \omega^{y_0} |y_0\rangle, \quad (\text{C3})$$

with $\omega = \exp(2\pi i/d)$, and

$$|\Psi\rangle := \frac{1}{\sqrt{2\left(1 + \frac{1}{\sqrt{d}}\right)}} (|0\rangle + F|0\rangle), \quad (\text{C4})$$

with F denoting the Quantum Fourier transform operator over \mathbb{Z}_d . Note that $X_d = FZ_dF^\dagger$. We also refer to the eigenbasis of Z_d as the *computational basis* and the eigenbasis of X_d as the *Fourier basis*. Below, it will be convenient to note that Z_d acts as the cyclic shift operator in the eigenbasis of X_d , and vice versa. Throughout, we will assume that d is prime.

Imagine a source that chooses $y_0, y_1 \in \mathcal{Y} := \{0, \dots, d-1\}$ uniformly at random and emits $|\Psi_{y_0 y_1}\rangle$, corresponding to the ccq-state

$$\begin{aligned} & \rho_{Y_0 Y_1 E} \\ & := \frac{1}{d^2} \sum_{y_0, y_1} \underbrace{|y_0\rangle\langle y_0|}_{Y_0} \otimes \underbrace{|y_1\rangle\langle y_1|}_{Y_1} \otimes \underbrace{|\Psi_{y_0 y_1}\rangle\langle\Psi_{y_0 y_1}|}_E. \end{aligned} \quad (\text{C5})$$

Throughout, we will consider the probability that we guess $Y_0 Y_1$ or the individual entries Y_0 and Y_1 given the register E . We begin by showing that for our specific encoding the probability of guessing *both* entries Y_0, Y_1 is small.

Lemma C.1. *For the ccq-state $\rho_{Y_0 Y_1 E}$ given by (C5)*

$$P_{\text{guess}}(Y_0 Y_1 | E) = \frac{1}{d}. \quad (\text{C6})$$

Proof. Computing the probability of guessing both bits is equivalent to solving the semidefinite program (SDP)

$$\begin{aligned} & \text{maximize} \quad \frac{1}{d^2} \sum_{y_0, y_1} \text{tr}(M_{y_0 y_1} |\Psi_{y_0 y_1}\rangle\langle\Psi_{y_0 y_1}|) \\ & \text{subject to} \quad M_{y_0 y_1} \geq 0 \text{ for all } y_0, y_1, \\ & \quad \quad \quad \sum_{y_0, y_1} M_{y_0, y_1} = \mathbb{I}. \end{aligned}$$

The dual SDP is easily found to be

$$\begin{aligned} & \text{minimize} \quad \text{Tr}(Q) \\ & \text{subject to} \quad Q \geq \frac{1}{d^2} |\Psi_{y_0 y_1}\rangle\langle\Psi_{y_0 y_1}| \text{ for all } y_0, y_1. \end{aligned}$$

Let v_{primal} and v_{dual} be the optimal values of the primal and dual respectively. Note that by weak duality we have $v_{\text{dual}} \geq v_{\text{primal}}$. Since $|\Psi_{y_0 y_1}\rangle\langle\Psi_{y_0 y_1}|$ is a pure state, $\hat{Q} = \mathbb{I}/d^2$ is a feasible dual solution with value $\text{tr}(\hat{Q}) = 1/d$.

We now show that \hat{Q} is in fact optimal, by constructing a solution to the primal that achieves the same value. Let $M_{y_0 y_1} = |\Psi_{y_0 y_1}\rangle\langle\Psi_{y_0 y_1}|/d$. Clearly, $M_{y_0 y_1} \geq 0$ for all y_0 and y_1 , and by Schur's lemma we have

$$\sum_{y_0, y_1} M_{y_0 y_1} = \frac{1}{d} \sum_{y_0, y_1} X_d^{y_0} Z_d^{y_1} |\Psi\rangle\langle\Psi| (X_d^{y_0} Z_d^{y_1})^\dagger \quad (\text{C7})$$

$$= \mathbb{I}. \quad (\text{C8})$$

Hence, our choice of operators is a feasible primal solution with primal value $1/d$ which concludes our claim. \square

We now show that the probability of retrieving any of the individual entries Y_0 and Y_1 is nevertheless quite large. To this end, let us first establish the following simple lemma.

Lemma C.2. *For the encoding defined in (C1) we have for all $y_0, y_1 \in \{0, \dots, d-1\}$*

$$|\langle y_0 | \Psi_{y_0 y_1} \rangle|^2 = \frac{1}{2} + \frac{1}{2\sqrt{d}}, \quad (\text{C9})$$

$$|\langle y_1 | F^\dagger | \Psi_{y_0 y_1} \rangle|^2 = \frac{1}{2} + \frac{1}{2\sqrt{d}}. \quad (\text{C10})$$

Proof. First of all, note that for all y_0 and y_1

$$\langle y_0 | \Psi_{y_0 y_1} \rangle = \langle 0 | (X_d^{y_0})^\dagger | \Psi_{y_0 y_1} \rangle \quad (\text{C11})$$

$$= \langle 0 | Z_d^{y_1} | \Psi \rangle = \langle 0 | \Psi \rangle, \quad (\text{C12})$$

where we have used the fact that $Z_d |0\rangle = |0\rangle$. Similarly, we have

$$\langle y_1 | F^\dagger | \Psi_{y_0 y_1} \rangle = \omega^{-y_1 a} \langle y_1 | F^\dagger Z_d^{y_1} | \Psi \rangle \quad (\text{C13})$$

$$= \omega^{-y_1 a} \langle 0 | F^\dagger (Z_d^{y_1})^\dagger Z_d^{y_1} | \Psi \rangle \quad (\text{C14})$$

$$= \omega^{-y_1 a} \langle 0 | F^\dagger | \Psi \rangle \quad (\text{C15})$$

$$= \omega^{-y_1 a} \langle 0 | \Psi \rangle, \quad (\text{C16})$$

with $\omega = \exp(2\pi i/d)$, where the first equality follows from the fact that $F|y_1\rangle$ is an eigenvector of X_d , and the last equality by noting that $F|\Psi\rangle = |\Psi\rangle$. It thus remains to compute

$$\langle 0 | \Psi \rangle = \frac{1}{\sqrt{2\left(1 + \frac{1}{\sqrt{d}}\right)}} (\langle 0 | 0 \rangle + \langle 0 | F | 0 \rangle) \quad (\text{C17})$$

$$= \frac{1}{\sqrt{2}} \sqrt{1 + \frac{1}{\sqrt{d}}}, \quad (\text{C18})$$

from which our claim follows. \square

It is now straightforward to compute the maximum probabilities that we retrieve Y_0 and Y_1 correctly.

Lemma C.3. For the cq-states $\rho_{Y_0 E}$ and $\rho_{Y_1 E}$ given by the reduced states of (C5) we have

$$P_{\text{guess}}(Y_0|E) = P_{\text{guess}}(Y_1|E) = \frac{1}{2} + \frac{1}{2\sqrt{d}}. \quad (\text{C19})$$

Proof. We first show our claim for $P_{\text{guess}}(Y_0|E)$. Consider the state corresponding to an encoding of y_0 given by $\sigma_{y_0} := \frac{1}{d} \sum_{y_1} |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|$. As before, we can express the winning probability as an SDP with primal

$$\begin{aligned} & \text{maximize} && \frac{1}{d} \sum_{y_0} \text{tr}(M_{y_0} \sigma_{y_0}) \\ & \text{subject to} && M_{y_0} \geq 0 \text{ for all } y_0, \\ & && \sum_{y_0} M_{y_0} = \mathbb{I}. \end{aligned}$$

We now show that without loss of generality, the optimal measurement has an extremely simple form. First of all note that $[\sigma_{y_0}, Z_d^a] = 0$ for all a and y_0 since

$$\sigma_{y_0} = \frac{1}{d} \sum_{y_1} Z_d^{y_1} |\Psi_{y_0 0}\rangle \langle \Psi_{y_0 0}| (Z_d^{y_1})^\dagger. \quad (\text{C20})$$

Hence, if $\{\hat{M}_{y_0}\}_{y_0}$ is an optimal solution then so is the measurement given by $\hat{M}_{y_0} = \frac{1}{d} \sum_a Z_d^a M_{y_0} (Z_d^a)^\dagger$. Thus without loss of generality we may assume that the optimal measurement operators are diagonal in the computational basis. Now consider the largest term corresponding to \hat{M}_{max} and σ_{max} such that

$$\text{tr}(\hat{M}_{\text{max}} \sigma_{\text{max}}) \geq \text{tr}(\hat{M}_{y_0} \sigma_{y_0}) \quad (\text{C21})$$

for all y_0 . Since all measurement operators are Hermitian, we can expand $\hat{M}_{\text{max}} = \sum_j \lambda_j |j\rangle \langle j|$ in its eigenbasis. We may now in turn consider the element $|j\rangle \langle j|$ which has the largest overlap with σ_{max} . That is, choose

$$m = \text{argmax}_j \langle j | \sigma_{\text{max}} | j \rangle, \quad (\text{C22})$$

that is, $\langle m | \sigma_{\text{max}} | m \rangle \geq \langle j | \sigma_{\text{max}} | j \rangle$ for all j . Clearly, we have that

$$P_{\text{guess}}(Y_0|E) \leq \langle m | \sigma_{\text{max}} | m \rangle. \quad (\text{C23})$$

It remains to prove that this inequality is tight. Without loss of generality assume that $\sigma_{\text{max}} = \sigma_0$, any other case will follow by a simple relabeling. Note that by Lemma C.2 we have

$$\langle y_0 | \sigma_0 | y_0 \rangle \leq \langle 0 | \sigma_0 | 0 \rangle = \frac{1}{2} + \frac{1}{2\sqrt{d}}, \quad (\text{C24})$$

for all y_0 and thus we choose $m = 0$ in (C22). Note that by construction we have $\sigma_{y_0} = X_d^{y_0} \sigma_0 (X_d^{y_0})^\dagger$, and hence $\langle y_0 | \sigma_{y_0} | y_0 \rangle = \langle 0 | \sigma_0 | 0 \rangle$. Thus for the measurement in the computational basis given by $M_{y_0} = |y_0\rangle \langle y_0|$, the inequality (C23) is tight which together with (C24) gives our claim. The case of retrieving Y_1 is exactly analogous, with the roles of X_d and Z_d interchanged. \square

2. Min-entropy splitting

We are ready to show that the min-entropy splitting inequality (B8) is violated for the ccq-state given in (C5).

Theorem C.4. For the ccq-state given in (C5), we have that for any ccqc state $\rho_{Y_0 Y_1 EC}$ with $\dim(C) = 2$ satisfying $\text{tr}_C(\rho_{Y_0 Y_1 EC}) = \rho_{Y_0 Y_1 E}$,

$$P_{\text{guess}}(Y_c | EC = c) \geq \frac{1}{2} + \frac{1}{2\sqrt{d}} \quad (\text{C25})$$

for all $c \in \{0, 1\}$.

Proof. Note that we may express

$$\rho_{Y_0 Y_1 EC} = \frac{1}{d^2} \sum_{y_0 y_1} |y_0\rangle \langle y_0| \otimes |y_1\rangle \langle y_1| \otimes \rho_{y_0 y_1 c}^{EC}. \quad (\text{C26})$$

We now first note that by the reduced trace condition and the fact that Y_0 and Y_1 are classical we must have that $\text{tr}_C(\rho_{y_0 y_1 c}^{EC}) = |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|$. Since $|\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|$ is a pure state, this implies that $\rho_{y_0 y_1 c}^{EC} = |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}| \otimes \sigma_{y_0 y_1}^C$. Since C is classical which we can express $\sigma_{y_0 y_1}^C$ without loss of generality in the computational basis as

$$\sigma_{y_0 y_1}^C = q_{y_0 y_1} |0\rangle \langle 0| + (1 - q_{y_0 y_1}) |1\rangle \langle 1|, \quad (\text{C27})$$

for some arbitrary distribution $\{q_{y_0 y_1}, 1 - q_{y_0 y_1}\}$.

Let us now consider how well we can compute $P_{\text{guess}}(Y_0 | EC = 0)$; the case of $C = 1$ is analogous. First of all, note that the state obtained from $\rho_{Y_0 Y_1 EC}$ after we measured C in the computational basis and obtained outcome $C = 0$, followed by tracing out C is given by

$$\begin{aligned} & \rho_{Y_0 Y_1 E} \\ &= \frac{1}{\hat{q}_0} \sum_{y_0 y_1} \tilde{q}_{y_0} \tilde{q}_{y_1 | y_0} |y_0\rangle \langle y_0| \otimes |y_1\rangle \langle y_1| \otimes |\Psi_{y_0 y_1}\rangle \langle \Psi_{y_0 y_1}|, \end{aligned} \quad (\text{C28})$$

where $\hat{q}_0 = \sum_{y_0 y_1} \tilde{q}_{y_0 y_1}$ and $\tilde{q}_{y_0 y_1} = (1/d^2) q_{y_0} q_{y_1 | y_0}$. The states we wish to distinguish are thus given by

$$\sigma_{y_0 | c=0} := \frac{1}{\sum_{y_1} \tilde{q}_{y_1 | y_0}} \sum_{y_1} \tilde{q}_{y_1 | y_0} \rho_{y_0 y_1}, \quad (\text{C29})$$

Note that from Lemma C.2 we have that for all y_0

$$\langle y_0 | \sigma_{y_0 | c=0} | y_0 \rangle = \frac{1}{2} + \frac{1}{2\sqrt{d}}. \quad (\text{C30})$$

Hence, for the measurement in the computational basis we succeed with probability at least $1/2 + 1/(2\sqrt{d})$, independent of the distributions $\{q_{y_0 y_1}, 1 - q_{y_0 y_1}\}$. Again, by exchanging the roles of X_d and Z_d the same probability can be achieved using a measurement in the Fourier basis, which proves the theorem. \square

In terms of min-entropy, we thus have that $H_\infty(Y_0Y_1|E) = \log d$ but for all C we have $H_\infty(Y_C|EC) \approx 1$! This effect is still observed for the ε -smooth min-entropy for small ε , since

$$H_\infty^\varepsilon(Y_0Y_1|E) \geq H_\infty(Y_0Y_1|E) = \log d, \quad (\text{C31})$$

and

$$-\log(P_{\text{guess}}(Y_C|EC) - \varepsilon) \geq H_\infty^\varepsilon(Y_C|EC). \quad (\text{C32})$$