MIT Open Access Articles

Information Causality, Szemerédi-Trotter and Algebraic Variants of CHSH

Information Causality, Szemerédi-Trotter and algebraic variants of CHSH

Mohammad Bavarian*
MIT

Peter W. Shor†
MIT

Abstract

In this paper, we consider the following family of two prover one-round games. In the $\mathsf{CHSH_q}$ game, two parties are given $x, y \in \mathbb{F}_q$ uniformly at random, and each must produce an output $a, b \in \mathbb{F}_q$ without communicating with the other. The players' objective is to maximize the probability that their outputs satisfy $a + b = xy$ in $\mathbb{F}_q$. This game was introduced by Buhrman and Massar [7] as a large alphabet generalization of the CHSH game—which is one of the most well-studied two-prover games in quantum information theory, and which has a large number of applications to quantum cryptography and quantum complexity. Our main contributions in this paper are the first asymptotic and explicit bounds on the entangled and classical values of $\mathsf{CHSH_q}$, and the realization of a rather surprising connection between $\mathsf{CHSH_q}$ and geometric incidence theory. On the way to these results, we also resolve a problem of Pawłowski and Winter [25] about pairwise independent Information Causality, which, beside being interesting on its own, gives as an application a short proof of our upper bound for the entangled value of $\mathsf{CHSH_q}$.

1. Introduction

In this work, we study a certain family of two prover one-round games. The study of multiprover one-round games (from now on, simply referred to as games) began in the late 20th century in the context of multiprover interactive proof systems in computer science [2], and also in the context of the Bell inequalities in physics [1] with the topic continuing to be of significant interest in both computer science and quantum physics to this day (see for example [5, 8, 12, 18, 28]). The particular family of games we shall study was first introduced by Buhrman and Massar [7] nearly a decade ago. It is defined as follows.

Definition 1.1. Let q be a prime, or a prime power, and $\mathbb{F}_q$ the unique field of size q. In the $\mathsf{CHSH_q}$ game, two non-communicating parties Alice and Bob are each given an input x and y from $\mathbb{F}_q$ chosen uniformly at random. Their objective is to maximize the probability that their outputs $a, b \in \mathbb{F}_q$ satisfy $a + b = xy$.
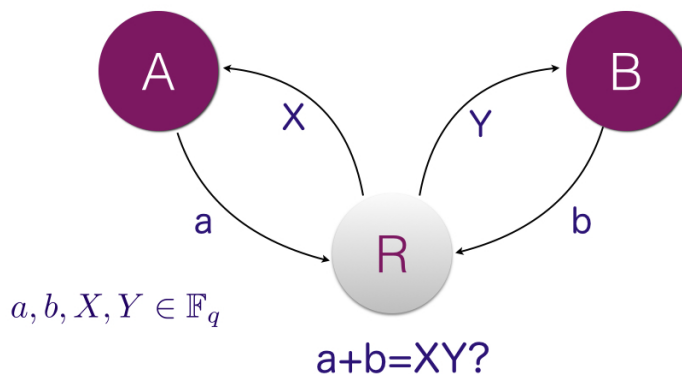
Fig. 1: The $\mathsf{CHSH_q}$ game.

The celebrated CHSH game, named after its inventors Clauser, Horne, Shimnoy and Holt [9], is the case $q = 2$ of the above definition. It is arguably the most well-studied game in quantum information theory [34], and has many applications in the study of entanglement ([6, 34]) and also in quantum cryptography [13] and quantum complexity [29]. Given the major role CHSH plays in many aspects of quantum information theory, it has been of great interest to find well-structured asymptotic generalizations of CHSH, as this could have much impact in the study of non-locality in general, and in the above applications of CHSH in particular. In this paper we focus on Buhrman and Massar's generalization, described in Definition 1.1, since we expect that the algebraic form of $\mathsf{CHSH_q}$ would lead to a interesting and useful structure for this family of games. In fact one of the main results of our work is the realization of a strong connection between the $\mathsf{CHSH_q}$ game and some remarkable mathematical results in incidence geometry and arithmetic combinatorics. This surprising connection combined with our other results further supports the intuition about the rich structure of these games.

Despite the simple form of this family of games and our precise understanding of the case $q = 2$, it turns out that analyzing $\mathsf{CHSH_q}$ beyond the $q = 2$ case is a rather difficult task. This difficulty is not restricted to analyzing $\mathsf{CHSH_q}$; it is actually an instance of a more general phenomenon, and is essentially shared with any game with $q \geqslant 3$. The main issue here is that we do not know a large alphabet generalization of the foundational result of Tsirelson on SDP characterization of the entangled value of XOR games (which are a subclass of games with $q = 2$).[1] This result of Tsirelson, combined with the tools of convex analysis such as complementary slackness and strong duality, gives a powerful path toward analyzing the entangled value and the optimal strategies for XOR games. The unavailability of the above powerful tool has resulted in a scarcity of results for analyzing the games in the case $q \neq 2$— which is regarded as one of the central challenges in the study of non-local games (see [6, 14]). Indeed, a major goal of this work is to expand on the set of examples and tools available for analyzing games beyond the relatively well-understood

---

[1] See Definition 2.1 for a precise definition of XOR games. The games we consider here are perhaps the most natural higher alphabet generalization of XOR games as the referee's acceptance predicate only depends on the sum over $\mathbb{F}_q$ of player's outputs. Remarkably, such a simple generalization from $q = 2$ to say $q = 3$ seem to make a substantial difference.

case of XOR games, which we do in the context of studying $\mathsf{CHSH_q}$. We note that our results do not go far on addressing the fundamental problems regarding the complexity of entangled two prover non-XOR games. However, we believe that for tackling this fundamental problem, a certain amount of preparatory work in the form of establishment of new tools and examples is a definite prerequisite. We hope that our work constitutes an advance in the foundation necessary for tackling the aforementioned fundamental problems.

Results. For a game G, we denote by $\omega(G)$ and $\omega^*(G)$ the maximum winning probability of classical and quantum strategies, respectively. These are usually referred to as the classical and entangled values, in short. Recall that since the quantum strategies contain the classical ones as a subset, it is clear that $\omega(G) \leqslant \omega^*(G)$.

Regarding the entangled value of $\mathsf{CHSH_q}$, we give two different proofs of the following theorem, which generalizes the well-known upper bound of Tsirelson [31] for the original CHSH game.

Theorem 1.2. *For any prime or prime power q we have*

$$\omega^*(\mathsf{CHSH_q}) \leqslant \frac{1}{q} + \frac{q-1}{q}\frac{1}{\sqrt{q}}\,.$$

Despite advances due to several researchers [7, 15, 21, 33] in analyzing the value of $\mathsf{CHSH_q}$ games, prior to our work there was no result, even in conjectured form, known for the asymptotic behavior of the classical and entangled value of $\mathsf{CHSH_q}$. Even for small values of q, most results, with the exception of the original $\frac{1}{3} + \frac{2}{3}\frac{1}{\sqrt{3}}$ upper bound of Buhrman and Massar for $q = 3$, were obtained using numerical methods. Thus, our work is the first to obtain asymptotic results on these games.

One interesting fact about the bound in Theorem 1.2 is the striking similarity of the bound $1/q + (q-1)/(q\sqrt{q})$ here, with the influential tight upper bound of Tsirelson [31] of $1/2 + 1/2\sqrt{2}$ for CHSH. This striking resemblance gives rise to the natural question of asymptotic (or exact) tightness of the bound in Theorem 1.2. Although we cannot answer the above questions in full, we provide some answers which clarify the situation to some extent, and highlight some of the relevant issues.

Theorem 1.3. *There exists a universal constant $\epsilon_0 > 0$ such that for any prime p and $k \geqslant 1$ we have*

$$\omega(\mathsf{CHSH_q}) = \begin{cases} \Omega(q^{-\frac{1}{2}}) & for \quad q = p^{2k} \\ O(q^{-\frac{1}{2}-\epsilon_0}) & for \quad q = p^{2k-1}\,. \end{cases}$$

To prove this theorem, we adopt a new view of $\mathsf{CHSH_q}$. The main insight here is the following.

Fact 1.4. *A classical strategy for $\mathsf{CHSH_q}$ is in direct correspondence with a configuration of q non-vertical lines and q points in $\mathbb{F}_q^2$, with no two lines having the same slope, and no two points lying on the same vertical line. Given such a configuration of lines and points, the winning probability of the corresponding strategy for $\mathsf{CHSH_q}$ is proportional to the number of point-line incidences.*

The correspondence in Fact 1.4 allows us access to some powerful results in arithmetic combinatorics where questions related to the incidences of collections of points and lines over finite fields have seen much progress recently. Most relevant to our problem is the celebrated finite field Szemerédi-Trotter theorem of Bourgain, Katz and Tao [4] which states that, under a certain size restriction

3

satisfied in our case, the number of incidences between a collection of points P and a collection of lines L is at most of the size $|P|^{\frac{3}{4}-\epsilon_0}|L|^{\frac{3}{4}-\epsilon_0}$ for some $\epsilon_0 > 0$.[2] This result combined with Fact 1.4 is essentially sufficient to prove the upper bound in Theorem 1.3.

In fact, the relation between $\mathsf{CHSH_q}$ and the finite field Szemerédi-Trotter theorem is closer than it might first appear to be. As we show in Section 5, understanding the classical value of $\mathsf{CHSH_q}$ is in some sense equivalent to the finite field Szemerédi-Trotter theorem with the appropriate parameters. We show this by proving that the restrictions on the points and lines in Fact 1.4, which is crucial in order to translate the geometric configuration to a legal $\mathsf{CHSH_q}$ strategy, can actually be relaxed without losing much in the bounds.

Going back to the quantum and classical values of $\mathsf{CHSH_q}$, it is important to notice that our classical lower bound in Theorem 1.3 shows that there is no asymptotic separation between the quantum and classical values of the game for $q = p^{2k}$, while the classical upper bound of $O(q^{-1/2-\epsilon_0})$ leaves open the possibility of such a separation in the setting of $q = p^{2k-1}$. Hence the most obvious gap in our bounds is captured by the following problem.

Problem 1.5 (Open). *Does there exists an infinite family of $q = p^{2k-1}$ such that $\omega^*(\mathsf{CHSH_q}) = \Omega(q^{-\frac{1}{2}})$, or some $\delta > 0$ and an infinite family of $q = p^{2k-1}$ such that $\omega^*(\mathsf{CHSH_q}) = O(q^{-\frac{1}{2}-\delta})$?*[3]

Given the geometric picture in Fact 1.4, we observe that our main open problem, Problem 1.5, is related to a question of Kempe and Kasher [19] about the security of Bourgain's two-source extractor [3, 27] in the presence of quantum memory. The main point is that Bourgain's extractor consists of two main ingredients: The first is a crucial preprocessing step, which (roughly-speaking) makes sure the two sources are in generic position with respect to each other. The second is an application of a Hadamard extractor on the two preprocessed sources. It was shown by Kasher and Kempe that the bare Hadamard extractor remains secure in the presence of quantum adversaries. Hence, the missing part in the analysis of Bourgain's extractor in the presence of quantum memory is the first step, analysis of which heavily relies on Szemerédi-Trotter theorem on finite fields. Thus, the core of both Kasher and Kempe's question and that of ours seem to be the extent to which finite-field Szemerédi-Trotter theorem can be (or fails to be) extended to the quantum setting.

1.1. Techniques

Let us start by giving more detail on our two different proofs of Theorem 1.2. A common aspect of both these methods is that they avoid a direct analysis of the norms of associated game operators. Instead, they take a novel indirect approach via reductions. In order to rule out a certain winning probability p for G, we show that the ability to win instances of G with probability greater than p would allow us to achieve a winning probability p' for a more generic game G', one which we already know to be impossible.

Both our methods for proving the upper bound on the entangled value of $\mathsf{CHSH_q}$ game work by a reduction to another generic result: the first method uses a reduction to a large alphabet variant

---

[2]The original proof of Bourgain *et al.* did not show an explicit bound on ϵ_0 however work in recent years have obtained explicit bounds of the form $\epsilon_0 > \frac{1}{700}$. However, the truth is believed to be much better than what is proved in these works.

[3]It is possible that the answer to both questions above are positive. In fact this would be the best possible outcome from the point of view of applications.

4

of the result of Linden et al. on quantum and classical strategies for certain distributed tasks [22].[4] This approach has the advantage of being self-contained and quite simple. The main idea here is to analyze a slightly different variant of $\mathsf{CHSH_q}$ game, called $\mathsf{CHSH_q^{dist}}$, in which two parties receive $(\alpha, \gamma) \in \mathbb{F}_q^2$ and $(\beta, \delta) \in \mathbb{F}_q^2$, and their objective is to produce outputs a and b satisfying $a + b = (\alpha + \beta)(\gamma + \delta)$.

The other approach is by a reduction to a new form of the principle of information causality due to Pawłowski and Winter [25], further generalized in this paper. A benefit of this approach is that executing the approach naturally leads us to an open problem of Pawłowski and Winter [25] which we resolve in Appendix B. Given the amount of attention paid to information causality principle in the foundations of quantum physics in recent years, this result is certainly significant on its own right. What makes this result even more attractive is that using it we can give as an application a very short proof of Theorem 1.2.

To discuss our result about *pairwise independent information causality*, it is best to first recall the standard scenario for information causality (IC) [24].

Definition 1.6 (IC). In an information causality game, Alice is given an input $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ from a known distribution π, and Bob an index $b \in [N]$. After making a measurement on her system, Alice sends a message $\alpha \in \Sigma$ to Bob. After receiving α from Alice, Bob makes a measurement on his system producing an output $Z \in \Lambda$. Alice and Bob's goal is to maximize the quantity $IC(A, B) = \sum_{i=1}^N I(X_i; Z | b = i)$.[5]

The main idea behind the *principle of information causality* is that assuming a certain form of independence among Alice's inputs (i.e. $\{X_i\}_{i=1}^N$), there is a stringent limit to the amount of correlation, as quantified by $IC(A, B)$, that the two parties can create by limited communication—even given arbitrarily entanglement between Alice and Bob. Our main result about the information causality game is as follows.

Theorem 1.7 (Pairwise Independent IC). *Consider an information causality game as in Definition 1.6. Assume Alice's input* $\mathbf{X} = (X_1, X_2, \ldots, X_N)$ *is drawn from an unbiased (i.e. with uniform marginals) pairwise independent distribution. Then we have*

$$IC(A, B) = \sum_{i=1}^N I(X_i; Z | b = i) = O_{|\Sigma|, |\Lambda|}(1),$$

where $O_{|\Sigma|, |\Lambda|}(1)$ *is a quantity depending only on the sizes of the alphabets of Alice's message to Bob and Bob's output, and not* N.[6]

The original setting of information causality from [24] is the case where $\{X_i\}_{i=1}^N$ are fully independent. Pawłowksi and Winter in [25] strengthened the original information causality by showing that

---

[4]Linden et al. prove their result only about $\mathbb{F}_2$-output games. We generalize this result to larger alphabets only as much as we need for our particular application. A full generalization of all aspects of the result of Linden et al. could be interesting but is beyond the scope of this article.

[5]A moment of reflection shows that the distribution of Bob's input $b \in [N]$ does not play any role here. Hence, it can be taken to be uniform over $[N]$ for simplicity.

[6]In most applications of information causality type theorems, including ours, the exact dependence on $|\Sigma|$ and $|\Lambda|$ is not important. The bound here is linear in $|\Sigma||\Lambda|$.

a similar bound holds even if the full independence condition is *relaxed* to pairwise independence, under the restriction that both Z and α have only two outcomes. They posed as an open problem to extend their result to larger alphabets. As it turns out such a theorem is precisely what we need to prove Theorem 1.2 by our information theoretic approach. We prove this Theorem in Appendix B and use it in Section 4.1 to prove Theorem 1.2.

Finally we end this part with a technical remark about Theorem 1.7.

Remark 1.8. (i) There is a crucial difference between $I(X_1 \ldots X_N; Z | b = i)$ and $\sum_{i=1}^{N} I(X_i; Z | b = i)$. It follows from the data processing inequality that the former is bounded by $O_{|\Sigma|}(1)$ with no need for any independence assumption. On the other hand, taking $X_1 = X_2 = \ldots = X_N = \text{Bernoulli}(\frac{1}{2})$, we see that $IC(A, B)$ could be as large as $\Omega(N)$ in this case. Thus, some independence assumption is necessary for Theorem 1.7 type results. (ii) The original information causality result of [24] can be proved by an appropriate use of data processing inequality and the chain rule applied to $I(X_1 X_2 \ldots X_N; \alpha) = O_{|\Sigma|}(1)$. Similar approaches run into trouble in the pairwise independence setting because of the correlations between X_i's.

1.2. Prior work

Buhrman and Massar were the first to study $\mathsf{CHSH_q}$ for $q \neq 2$ obtaining the upper bound of $1/3 + 2/3\sqrt{3}$ for $q = 3$ using information theoretic methods (different from the ones used here). Although the algebraic view of CHSH is in retrospect more or less clear, this was not the language originally used to describe the CHSH game. Hence, the contribution of Buhrman and Massar was to both realize this view, and to tackle the next interesting case after the original CHSH, which was the case of $q = 3$. In the same work, they mentioned that their method seemed not to work for higher values of q. After the work of Buhrman and Massar, the problem was attacked for small values of q by Ji et al. [15] and then by Liang, Lim and Deng [21] who, through a mix of numerical work and analytic insights, obtained several upper and lower bounds for quantum and classical value of the games for q's up to 13. Since the approach in the above line of work is mostly numerical, it is hard to infer much about the asymptotic questions of interest from the bounds there. The only work prior to ours to obtain some general results about $\mathsf{CHSH_q}$ is [33]. There, Wang proved various interesting results including a large alphabet generalization of a result of van Dam [11], on the collapse of communication complexity in the presence of perfect $\mathsf{CHSH_q}$ oracle boxes (this is a natural higher alphabet analogue of the well-known Popescu-Rohlich box [26]). He also approached the problem of analyzing the value of $\mathsf{CHSH_q}$ using the principle of information causality; however, the arguments there did not seem to provide explicit bounds. We discuss the similarities and the differences between our approach and Wang's in more detail in Section 3. This provides some intuition on why basic information causality of [24] seems insufficient to achieve the bound in Theorem 1.2 showing the advantage offered by the strengthened version in Theorem 1.7.

Before finishing the discussion of the prior work, let us take some time to elaborate more on the work of Liang et al. [21]. The upper bounds on their work is based on numerically solving the SDP hierarchies from [23]. Their method of lower bound (described on the bottom of page 4 of their paper) is a variant of the following natural heuristic: One starts with an arbitrary choice of Alice's and Bob's strategies and a joint state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$ in a candidate dimension d. Then one runs the following three-step iterative procedure: We first optimize for Bob's strategy, given the state ψ and Alice's strategy. Then, we optimize Alice's strategy fixing Bob's strategy and the state ψ. Finally, we optimize for ψ while fixing everything else. The first two steps are SDP computations, and the

latter is simply a maximum eigenvalue computation. The above process is repeated until a (near) local optimum is reached.[7]

1.3. Organization of the paper

To gain a better understanding of the main ideas behind the information theoretic approach in analyzing the two prover games, we recommend the reader to review the exposition of the approach as presented in Section 3. Although the arguments in Section 3 are not essential for proving the main results, it may provide much intuition and make the steps in various arguments more transparent. The most technical part of the paper is Appendix B which contains the proof of the open problem of Pawłowski and Winter. This section can be skipped if one is only interested in the application of Information Causality type principle to two-prover games rather than the proof of such results. Also for the sake of brevity, the second proof of Theorem 1.2 based on reduction to distributed non-local computation and the proof of Lemma 5.3, which requires more familiarity with projective spaces, are transferred from the main body to the appendices in this version.

2. Preliminaries

In this work, q will denote a prime power unless otherwise noted. We let $[n]$ denote the set of positive integers in the range 1 to n. By $O_{\alpha,\beta}(\cdot)$ we mean a quantity which is bounded by a universal constant depending only on the parameters α and β. In parts of the paper, we require a basic knowledge of information theory. A brief introduction to relevant notions from this area is presented in Appendix A.

By a game we always mean a two prover one-round refereed game unless otherwise specified. For our purposes, we only need the definition of following subclass of games.

Definition 2.1. An XOR-game G is specified by the following data: a set of question Q_1 for Alice and a set of question Q_2 for Bob, a distribution π over $Q_1 \times Q_2$, and matrix V with entries from $\mathbb{F}_2$ of size $Q_1 \times Q_2$. Given $q_1 \in Q_1$ to Alice, and $q_2 \in Q_2$ to Bob with probability $\pi(q_1, q_2)$, Alice and Bob must each produce $a, b \in \mathbb{F}_2$ respectively. They succeed if $a + b = V(q_1, q_2)$. If in the above setting one replaces the additive group of $\mathbb{F}_2$ with the additive group of $\mathbb{F}_q$, the resulting class of games will be called q-XOR games.

For q-XOR games, it is convenient to define the *bias* of a strategy as an alternative way of quantifying the winning probability.

Definition 2.2. For any probability $0 \leqslant p_{win} \leqslant 1$ corresponding to winning probability of some strategy for a q-XOR game, we define the bias parameter of the strategy to be

$$E \stackrel{\text{def}}{=} (q p_{win} - 1)/(q - 1).$$

With this definition we have $p_{win} = \frac{1}{q} + \frac{q-1}{q} E$. Note the bias parameter quantifies by how much a strategy outperforms the trivial random strategy which for q-XOR games achieves the expected winning probability of $1/q$. Hence, we have $E \geqslant 0$ for the optimal strategy of any q-XOR game.

---

[7] Let us note that an anonymous reviewer, based on their own personal experiments, has raised some doubts regarding the validity of the lower bounds for $q = 3$ in Liang et al. We believe that in this case Liang et al. result is correct, as prior to becoming aware of their result we independently obtained their lower bound for $q = 3$ via different methods.

Remark 2.3. We note that the notion of bias can also be defined in the same way in the more general setting of unique games with alphabet size q. However, we shall not need to work in this general setting for the purposes of this work and we do not require the definition of what a unique game is. However, the reader familiar with the terminology should note that all q-XOR games, and in particular $\mathsf{CHSH_q}$ games, are unique games.

3. Overview of Information Causality approach

Here we discuss the basic method that allows one to use information causality type theorems to prove bounds on the value of $\mathsf{CHSH_q}$ game. Since more precise arguments with better bounds are provided in Section 4 and Appendix C, we omit some of the details. For simplicity, we work in the fully independent setting of the information causality game; this has the added benefit of clarifying why we need the strengthened version of the information causality principle, Theorem 1.7, in order to prove Theorem 1.2. In this section, for the ease of notation, we use a slightly different convention from other sections by taking Alice's input to be indexed from 0 to $N-1$, as opposed to 1 to N.

The idea is best captured in the noiseless setting: assume that we have a (quantum or whatever) strategy that allows us to win the $\mathsf{CHSH_q}$ game with probability 1. Let us see that this leads to some unlikely consequences.
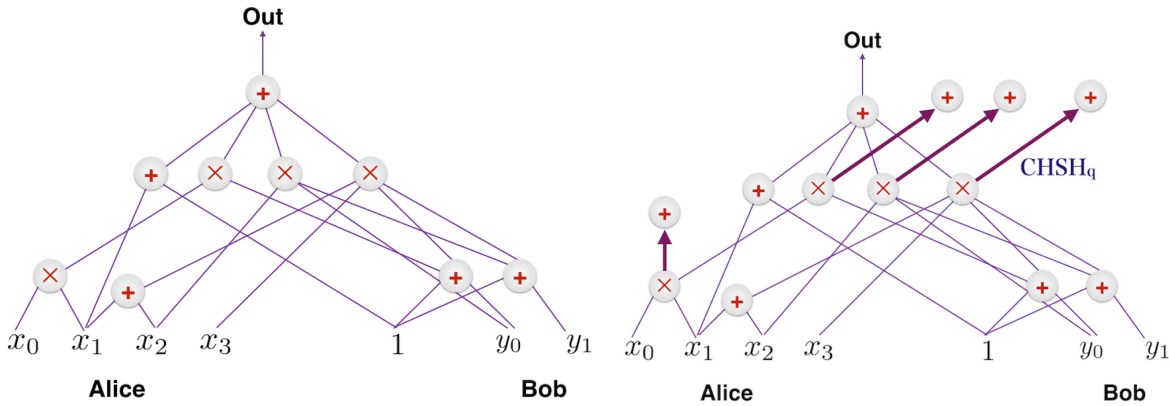

Fig. 2: The players' strategies for $\mathsf{CHSH_q}$ can be used to *eliminate* $\times$ gates from a distributed circuit, replacing them with $+$ gates—at the cost of some noise.

Let $q = 2$ for simplicity. Assume Alice is given $X_0, X_1, X_2, X_3 \in \mathbb{F}_2$ uniformly at random. Bob is given an index pointing to one of the Alice's inputs in the form of $b = b_0 b_1$ with $b_i \in \{0, 1\}$. The main observation is that the desired output of Bob, which is X_b, can be written as a polynomial over $\mathbb{F}_2$ in terms of X_i's and b_j's. Let us denote the desired output of Bob by $g_2(\cdot) = X_b$. In general g_k denote the polynomial that Alice and Bob want to compute in the "k$^{\text{th}}$ level" of the game, i.e., the setting where we have $N = 2^k - 1$ and $b = b_0 b_1 \ldots b_{k-1}$. Observe that

$$g_2(X_0, X_1, X_2, X_3, b_0, b_1) = b_0 b_1 X_3 + (1 - b_0) b_1 X_2 + b_0 (1 - b_1) X_1 + (1 - b_0)(1 - b_1) X_0 \,.$$

From the above, it is clear what Alice and Bob should do: given (say) the first monomial, $b_0 b_1 X_3$, Alice and Bob could use their perfect CHSH strategy to produce $u_3 + v_3 = b_0 b_1 X_3$. Doing the same

thing for all other monomials, they essentially dispose of all multiplications and are left with

$$g_2(X_0, X_1, X_2, X_3, b_0, b_1) = \sum_{i=0}^{3} u_i + \sum_{i=0}^{3} v_i \,.$$

Hence, if Alice sends $\sum_{i=0}^{3} u_i$ to Bob, using just one bit of communication, Alice and Bob manage to compute $g_2(\cdot)$ successfully. It is clear that the above perfect strategy can be extended in the same manner to arbitrary "level k," i.e. when Alice is given $\{X_i\}_{i=0}^{2^k-1}$ and Bob an index $b = b_0 b_1 \ldots b_{k-1}$, with no need for more than 1-bit of communication. Hence, it seems that there must be some information theoretic impossibility here, as one bit of communication should not be sufficient to allow two players to compute the rather complicated address function $g_k(\cdot)$ for large k. This is exactly the main idea of the information theoretic approach for Theorem 1.2.

To prove Theorem 1.2 we need a more careful version of the above argument in order to derive a contradiction even when Alice and Bob only share a *noisy strategy* for CHSH or CHSH$_q$. Let us give some more detail: imagine Alice and Bob now only have a strategy that allows them to win with probability $w < 1$. Instead of being as generous as before, Alice and Bob should now try to minimize the number of times they use their strategy for CHSH to transform a multiplication into an addition. Naively, if there exists a circuit with m multiplication gates for computing the polynomial $g_k(\cdot)$, one would expect the winning probability of Alice and Bob to scale like w^m. As one increases k, if Alice's and Bob's probability of success in computing $g_k(\cdot)$ deteriorates slowly enough compared to the rate of increase in the information complexity of computing $g_k(\cdot)$, this gives us our desired contradiction.

The above illustrates the basic strategy for proving upper bounds on winning probabilities of games, in our case CHSH$_q$, using information theoretic methods. The above analysis can be improved in several ways; however even with all improvements we are unable to prove Theorem 1.2 using just basic information causality result of [24]. The main difficulty here stems from the situation at the first level, and this difficulty in our view is one of the reasons why the method of Wang [33] and our own initial method based on the basic Information Causality theorem [24] seemed incapable of reproducing the bounds in Theorem 1.2.[8] To demonstrate this, assume $q = 3$. Let $X_0, X_1, X_2, b_0 \in \mathbb{F}_3$ and let $h_1(\cdot)$ be the polynomial over $\mathbb{F}_3$ which Alice and Bob would like to compute at level 1. We have

$$
\begin{aligned}
h_1(X_0, X_1, X_2, b_0) &= (1 - b_0^2)X_0 + (1 - (b_0 - 1)^2)X_1 + (1 - (b_0 + 1)^2)X_2 \\
&= X_0 + (X_2 - X_1)b_0 - (X_0 + X_1 + X_2)b_0^2 \,.
\end{aligned}
$$

We see that the number of multiplications we have at level 1 for $\mathbb{F}_3$ is 2; on the other hand, over $\mathbb{F}_2$ we have $g_1(X_0, X_2, b_0) = X_0 + (X_0 + X_1)b_0$. The increase in the number of multiplication gates per level is the main problem which forces us to use Theorem 1.7 as opposed to the fully independent version. To see the advantage of Theorem 1.7, suppose that we choose Alice's input uniformly at random from the subspace $X_0 + X_1 + X_2 = 0$. Notice that although X_i's are now correlated, they

---

[8]The other major difference between our work and that of Wang in our view is the use of *Regularization Lemma* 4.2 which greatly simplifies our analysis by reducing the number of parameters we have to keep track of in Proposition 4.4 to the bare minimum.

are still pairwise independent. The main advantage of taking X_i's from subspace $X_0 + X_1 + X_2 = 0$ is that this allows to save one multiplication in the evaluation of h_1 since

$$h_1(X_0, X_1, X_2, b_0) = X_0 + (X_2 - X_1)b_0$$

in this case. The above method indeed can be extended to higher values of k to prove Theorem 1.7. Essentially, the idea is to take Alice's input uniformly at random from an appropriately chosen subspace of the vector space (over $\mathbb{F}_3$) spanned by $\{X_i\}_{i=0}^{3^k-1}$. However, using the full strength of Theorem 1.7 we can actually give a simpler information theoretic proof of Theorem 1.2, as presented in Section 4.1.

4. Generalized Tsirelson bound for $\mathsf{CHSH_q}$

The aim of this section is to prove Theorem 1.2. We have two different proofs of this result. Here we just present the proof based on the pairwise independent Information Causality principle, Theorem 1.7. The second proof, which is based on a large alphabet generalization of result of Linden et al. [22], is presented in Appendix C. In both proofs of Theorem 1.2, we can substantially simplify the arguments by assuming the optimal strategy $\mathcal{P}$ for $\mathsf{CHSH_q}$ always produces different types of error with equal probability. This is formalized as follows.

Definition 4.1. A classical or quantum strategy for $\mathsf{CHSH_q}$ is called *regular* if the following holds:

$$\Pr_{a,b \leftarrow \mathcal{P}*}[a + b = xy + k | x, y] = \frac{1}{q} - \frac{E}{q} \qquad \forall\, k \in \mathbb{F}_q^*.$$

Here, a, b are the outputs of Alice and Bob's strategies given x, y as inputs, respectively. The symbol $a, b \leftarrow \mathcal{P}$ means that the outputs a, b of Alice and Bob's strategies are produced via the protocol $\mathcal{P}$ (given x, y as inputs). The symbol E as usual denotes the bias of the game defined as in Definition 2.2. In some occasions, for the ease of notation, we do not write out $a, b \leftarrow \mathcal{P}$ fully, and instead simply write $\mathcal{P}$ to represent the fact that the players follow the particular strategy $\mathcal{P}$ for producing their outputs.

It turns out that we can without loss of generality assume any protocol $\mathcal{P}$ for $\mathsf{CHSH_q}$ is regular which is the content of the following lemma.

Lemma 4.2 (Regularization Lemma). *Given any protocol $\mathcal{P}$ for $\mathsf{CHSH_q}$, there exists a generic method to obtain a regular protocol $\mathcal{P}^*$ from $\mathcal{P}$ without changing the winning probability.*

Proof. Given any quantum strategy $\mathcal{P}$ for $\mathsf{CHSH_q}$ game, define its regularized version $\mathcal{P}^*$ as follows: On inputs x and y, A and B use shared randomness to agree upon $\alpha, \beta \in \mathbb{F}_q^*$ and $\gamma, \delta \in \mathbb{F}_q$ uniformly at random. Then, they follow the original strategy on inputs $\tilde{x} = (\alpha x + \gamma)$, $\tilde{y} = (\beta y + \delta)$. Let $\tilde{a}$ and $\tilde{b}$ be their outputs here. Finally, A outputs $a = \frac{1}{\alpha\beta}(\tilde{a} - \delta\alpha x - \gamma\delta)$ and B outputs $\frac{1}{\alpha\beta}(\tilde{b} - \beta\gamma y)$. We show $\mathcal{P}^*$ satisfies the properties we wanted.

First, notice that $\mathcal{P}^*$ has the same winning probability as $\mathcal{P}$ since the input distribution remains uniform on $\mathbb{F}_q \times \mathbb{F}_q$, i.e. $\Pr_{\mathcal{P}*}[a + b = xy | x, y] = \Pr_{\mathcal{P},x',y'}[a' + b' = x'y'] = p_{win}$. It is not too hard to see that

$$\Pr_{a,b \leftarrow \mathcal{P}*}[a + b = xy + s | x, y] = \Pr_{a',b' \leftarrow \mathcal{P},x',y'}[a' + b' = x'y' + s\,\alpha\beta].$$

10

The key here is that quadruple (x', y', α, β) have a uniform product distribution over its domain $(\mathbb{F}_q)^2 \times (\mathbb{F}_q^*)^2$. For $s \in \mathbb{F}_q^*$, we have that $s\alpha\beta$ is uniformly distributed over $\mathbb{F}_q^*$. This and the fact that the input (x', y') is uniform on $\mathbb{F}_q \times \mathbb{F}_q$ imply the regularity property. $\qquad\square$

4.1. Reduction to pairwise independent information causality

Let m be a positive integer which will be a parameter taken to be sufficiently large in our proof. We want to instantiate Theorem 1.7 by a subcode of generalized Hadamard code over $\mathbb{F}_q$. Let us recall the definition of this code:

Definition 4.3. The generalized Hadamard code corresponds to an m-dimensional subspace of a $q^m - 1$ dimensional vector space over $\mathbb{F}_q$. Given a seed $\mathbf{Y} = (Y_1, Y_2, \ldots, Y_m) \in \mathbb{F}_q^m$ we have a coordinate per each $\xi \in \mathbb{F}_q^m \backslash \{0\}$ defined by

$$\mathrm{Had}_\xi = \xi_1 Y_1 + \xi_2 Y_2 + \ldots + \xi_m Y_m \,.$$

Hence, a codeword is a point in the above subspace and it is given by the list of $q^m - 1$ coordinates defined as above.

The overall plan is to let Alice's input be a random codeword from $\{\mathrm{Had}_\xi\}_{\xi \in \mathbb{F}_q^m \backslash \{0\}}$, and Bob's input to be some $\xi \in \mathbb{F}_q^m \backslash \{0\}$ which is an index to one of the coordinates of Alice's input. This, however, does not quite work as the generalized Hadamard codeword is not pairwise independent. To fix this issue, we instead take Alice's input to be a proper subset $U_m \subseteq \mathbb{F}_q^m \backslash \{0\}$ such that $\{\mathrm{Had}_\xi\}_{\xi \in U_m}$ is pairwise independent. Concretely, we take U_m to consist of $a \in \mathbb{F}_q^m \backslash \{0\}$ with their first non-zero coordinate equal to 1. [9] With this setup we have $n = \frac{q^m - 1}{q - 1}$, Bob's input is some $b \in U_m$ and Alice's input is a uniformly random word from $\{\mathrm{Had}_\xi\}_{\xi \in U_m}$.

Now we need a proposition.

Proposition 4.4. *Let $\mathcal{P}$ be a regular protocol for* $\mathsf{CHSH_q}$ *with bias E. Assume Alice is given* $(c_1, c_2, \ldots, c_m) \in \mathbb{F}_q^m$ *and Bob* $(d_1, d_2, \ldots, d_m) \in \mathbb{F}_q^m$. *Assume Alice and Bob use the protocol $\mathcal{P}$ once per input pair $\{(c_k, d_k)\}_{k=1}^m$ to produce $\{a_k\}_{k=1}^m$ and $\{b_k\}_{k=1}^m$, i.e. using $(a_k, b_k) \leftarrow \mathcal{P}(c_k, d_k)$. Let $Z = \sum_{i=1}^m a_k + \sum_{i=1}^m b_k$. We have*

$$\Pr\left[Z = \sum_{i=1}^m c_i d_i \right] = \frac{1}{q} + \frac{q-1}{q} E^m \,.$$

Also, for all $e \in \mathbb{F}_q^*$ we have

$$\Pr\left[Z = \sum_{i=1}^m c_i d_i + e \right] = \frac{1}{q} - \frac{1}{q} E^m \,.$$

Proof. The proof is by induction on m. For $m = 1$ this is clear. Assume the result for $m - 1$. Notice that for $Z = \sum_{i=1}^m c_i d_i$ to occur, it must be the case that the error in level $m - 1$, i.e.

---

[9]This is chosen such that the map $\pi : (\xi_1, \xi_2, \ldots, \xi_m) \mapsto (\xi_1 : \xi_2 : \ldots : \xi_m)$ injects onto $\mathbb{PF}_q^{m-1}$

$\sum_{i=1}^{m-1} c_i d_i - a_i - b_i$, exactly cancels out the error occurred in the last step, which is $a_m + b_m - c_m d_m$. Now by the assumptions,

$$Pr\left[Z = \sum_{i=1}^{m} c_i d_i\right] = \left(\frac{1}{q} + \frac{q-1}{q} E^{m-1}\right)\left(\frac{1}{q} + \frac{q-1}{q} E\right)$$
$$+ (q-1)\left(\frac{1}{q} - \frac{1}{q} E^{m-1}\right)\left(\frac{1}{q} - \frac{1}{q} E\right)$$
$$= \frac{1}{q} + \frac{q-1}{q} E^m .$$

This finishes the first claim of the theorem. The second claim of the theorem follows from the first one and the symmetry. $\qquad\square$

We apply this proposition to the setting where Alice is given $\{\text{Had}_\xi\}_{\xi \in U_m}$ and Bob $\xi^* \in U_m$. Alice and Bob want to compute $\text{Had}_{\xi^*} = \xi_1^* Y_1 + \xi_2^* Y_2 + \ldots + \xi_m^* Y_m$. Here, Y_i's are given to Alice as part of her inputs as any coordinate of the form $(0, 0, \ldots, 1, 0, \ldots, 0)$ is in U_m and ξ^* is precisely the input to Bob. Hence by following the above protocol, Alice ends up sending $\sum_{i=1}^{m} a_i$ to Bob as the message and Bob outputs $Z = \sum_{i=1}^{m} (a_i + b_i)$. The resulting output Z will satisfy the following:

$$I(X_\xi; Z | b = \xi) = \left(\frac{1}{q} + \frac{q-1}{q} E^m\right) \log_2(1 + (q-1)E^m)$$
$$+ \frac{q-1}{q^2}(1 - E^m) \log_2(1 - E^m) .$$

Notice that the calculation above was not too hard because the regularity guaranteed from Proposition 4.4 specifies the exact joint distribution of (X_ξ, Z). Now we take m large enough such that $E^m \ll \frac{1}{q}$. In this regime, it is easy to see that above expression is always larger than $\frac{E^{2m}}{\text{poly}(q)}$ for a fixed polynomial independent of m. Using $I(X_\xi; Z | b = \xi) \geqslant \frac{E^{2m}}{\text{poly}(q)}$ in Theorem 1.7, and noticing that $|U_m| = (q^m - 1)/(q-1)$, we see that $\frac{q^m - 1}{q-1} \frac{E^{2m}}{\text{poly}(q)} = O_q(1)$. For this to hold for arbitrarily large m, we must have

$$E \leqslant \frac{1}{\sqrt{q}},$$

which is our desired result.

5. Classical aspects of $\mathsf{CHSH_q}$ and point-line incidences

In this section, we present our results regarding the classical value of $\mathsf{CHSH_q}$. This includes Theorem 1.3 and various other results.

We begin by a short introduction to some notions from geometric incidence theory. Let $\Pi = \mathbb{F}^2$ be the plane over a field $\mathbb{F}$. For a collection of lines L and points P over a plane Π we define the set of incidences as

$$I(P, L) = \{(p, l) \in P \times L , \, p \in l\} .$$

A central question in geometric incidence theory is the following: Given $|P|$ and $|L|$, what can be said about the size $|I(P, L)|$? If $|P|$ and $|L|$ are of roughly the same size, it is hard to imagine a

configuration where every line in L would contain every point in P. Hence $|I(P, L)| \leqslant |P||L|$ seems a rather pessimistic upper bound. In fact, using the fact that there is at most one line through two distinct points and one point at the intersection of two distinct lines suffices to get a better upper bound of $|P|^{3/4}|L|^{3/4} + |P| + |L|$. As shown by Szemerédi and Trotter [30] the above bound can be improved when $\mathbb{F} = \mathbb{R}$ to $|P|^{2/3}|L|^{2/3} + |P| + |L|$. The proof of this result is very geometric and relies on localization techniques that do not work in the finite field settings. The situation over finite fields remained unclear until about a decade ago; finally Bourgain, Katz and Tao [4] used tools from arithmetic combinatorics to show an improved upper bound on $|I(P, L)|$ as long as the sets P and L are not too large. More specifically we have:

Theorem 5.1 (BKT). *For any $\delta > 0$, there exists some $\epsilon > 0$ such that for any prime field $\mathbb{F}_p$ and any collection of points P and lines L over $\mathbb{F}_p^2$ satisfying $p^\delta \leqslant |P|, |L| \leqslant p^{2-\delta}$, we have*

$$|I(P, L)| = O\left(|P|^{\frac{3}{4}-\epsilon}|L|^{\frac{3}{4}-\epsilon}\right).$$

Although the theorem of Bourgain et al. as stated above only holds for prime fields, it is not too hard to see that essentially the same argument goes through whenever $|P|, |L|$ are large compared to the proper subfields of $\mathbb{F}_q$. This is made explicit in the work of Jones [16, 17]. Although what is proved in [16, 17] is more general, we just need the following corollary.

Corollary 5.2 (Jones). *Let q be an odd power of a prime, and assume P and L are sets of points and lines in $\mathbb{F}_q^2$ of size $\Theta(q)$. There exists a universal constant $\epsilon > 0$ such that*

$$I(P, L) \leqslant q^{\frac{3}{2}-\epsilon}.$$

5.1. From point-line incidences to CHSH$_q$

Given the above results, Theorem 1.3 can be proved rather quickly. To see why, recall Fact 1.4 from the introduction where it was claimed that the winning probability of any classical strategy for CHSH$_q$ corresponds to the number of point-line incidences among q lines and q points in $\mathbb{F}_q^2$ under some restrictions on the points and lines. Hence, the result of Jones immediately implies the lower bound in Theorem 1.3. For the upper bound, we need the following lemma which allows to relax the restrictions on the points and lines in Fact 1.4.

Lemma 5.3. *Let P, L be a set of points and lines in $\mathbb{F}_q^2$ with $|P| = \Theta(q)$ and $|L| = \Theta(q)$. There exists a set of points P', and a set of lines L', satisfying the conditions of Fact 1.4 with $|P'| \leqslant |P|$ and $|L'| \leqslant |L|$ such that*
$$|I(P', L')| = \Omega\left(|I(P, L)|\right).$$

The main idea for proving this lemma is to start from the given configuration of P and L, and apply a random projective transformation to them. The next step is to remove the lines with the same slope, and the points on the same vertical line. This ensures that the remaining sets of points and lines, P' and L', satisfy the required condition. It is not too hard to show that this deletion process shrinks the number of incidences only by a constant factor in expectation finishing the proof of Lemma 5.3. Notice that one could not simply use affine transformations in the place of the projective ones as affine transformations preserve the direction of lines which is problematic if

13

the initial configuration of L has many parallel lines. A more detailed proof is given in Appendix D.

Before proving Theorem 1.3 we need to establish some useful notation.

Definition 5.4. Let $\mathbb{F}_q^2 = \mathbb{F}_q[z_1, z_2]$ be the plane. We denote by $\ell_{a,b}$ the line

$$\ell_{a,b} = \{(z_1, z_2) \in \mathbb{F}_q^2 : z_2 = az_1 - b\}.$$

Proof of Theorem 1.3 and Fact 1.4. The optimal classical strategies for $\mathsf{CHSH_q}$ are given by two functions $f, g : \mathbb{F}_q \to \mathbb{F}_q$ corresponding to Alice and Bob's strategies maximizing

$$|\{x, y \in \mathbb{F}_q \mid f(x) + g(y) = xy\}| \ .$$

Given $f : \mathbb{F}_q \to \mathbb{F}_q$ corresponding to Alice's strategy, let P be the collection of q points of the form $(x, f(x)) \in \mathbb{F}_q^2$. To Bob's strategy $g : \mathbb{F}_q \to \mathbb{F}_q$, we associate a collection L of q lines $\{\ell_{y,g(y)}\}$. Observe in this language any pair $(x, y) \in \mathbb{F}_q^2$ satisfying $\mathsf{CHSH_q}$ correspond to a point-line incidence. Hence, we have

$$f(x) + g(y) = xy \quad \Leftrightarrow \quad (x, f(x)) \in \ell_{y,g(y)} \, ,$$

which means that

$$\sum_{x,y \in \mathbb{F}_q} 1_{f(x)+g(y)=xy} = I(P, L).$$

The upper bound for $q = p^{2k-1}$ now follows from Corollary 5.2.

Now assume $q = p^{2k}$. Recall that in this case there exists a subfield $K \cong \mathbb{F}_{\sqrt{q}}$ such that $K \subset \mathbb{F}_q$. Let $P = \{(a, b) \in \mathbb{F}_q^2 : a, b \in K\}$ and $L = \{\ell_{c,d} : c, d \in K\}$. Notice that $|P| = |L| = q$, and $|I(P, L)| = q^{3/2}$. Combined with Lemma 5.3, this proves the lower bound. $\square$

This framework can also be used to give an improved lower bound for general q's.

Theorem 5.5. *There exists a strategy for $\mathsf{CHSH_q}$ achieving a winning probability $\Omega(q^{-2/3})$.*

Let us note that the above lower bound was somewhat counterintuitive to us at first. The point is that we expected that the function $(x, y) \mapsto xy$ to be in some sense *maximally psuedorandom* against the function $(a, b) \mapsto a + b$. Given this, it was reasonable to assume that the best classical strategy for $\mathsf{CHSH_q}$ would achieve a winning probability of $\widetilde{O}(q^{-1})$ which is up to polylogarithmic factor the same as that of a random strategy. The logarithmic advantages can be seen to be achievable using a simple balls-and-bins analysis by taking a random function $f : \mathbb{F}_q \to \mathbb{F}_q$ as Alice's strategy, and optimizing Bob's strategy $g : \mathbb{F}_q \to \mathbb{F}_q$ given that of Alice. In fact, numerical experiments which looked for locally optimal solutions confirmed the above intuition.[10] Despite all this, Theorem 5.5 states that much better lower bounds are achievable in general.

Proof of Theorem 5.5. This follows from Lemma 5.3 applied to the next proposition. $\square$

---

[10]More precisely, the algorithm used was the following: we start from two random strategies $f, g : \mathbb{F}_q \to \mathbb{F}_q$, and in every iteration we fix one of the functions and update the other one to the optimal strategy taking the other function as fixed.

Proposition 5.6. *For any finite field field $\mathbb{F}_q$, there exists a set of at most q lines and at most q points over $\mathbb{F}_q^2$ with $\Omega(q^{4/3})$ incidences.*

Proof. Let $q = p^s$. First we handle the case $s = 1$, then we give a construction for $s \geqslant 2$. Since we are concerned with an asymptotic statement, we can assume q is sufficiently large. As a result, we can safely ignore the ceiling and floor signs as they do not affect the asymptotic. For the prime case $s = 1$, the construction is very simple: let $P = [q^{1/3}] \times [q^{2/3}]$. Let L be the collection of lines $\ell_{c,d}$ of the form $y = cx + d$ with $c \in [q^{1/3}/2]$ and $d \in [q^{2/3}/2]$. It is clear that this achieves $I(P, L) = \Theta(q^{4/3})$ with $|P|, |L| \leqslant q$.

For the case $s \geqslant 2$, we choose our set P to be a product set, $P = A \times B$ where $A, B \subset \mathbb{F}_q$ are both subspaces. Let g be a primitive element of $\mathbb{F}_q$ so $\{1, g, g^2, \ldots, g^{s-1}\}$ form a basis of $\mathbb{F}_q$ as a vector space over $\mathbb{F}_p$.

Let $b \leqslant s$ be a positive integer, close to $2s/3$, to be specified later. Let $a = s - b$ (the condition $s \geqslant 2$ will turn out to be sufficient for $a \geqslant 1$ which we require). Define

$$A = \mathbb{F}_p + g\,\mathbb{F}_p + g^2\mathbb{F}_p + \ldots + g^{a-1}\mathbb{F}_p$$

and

$$B = \mathbb{F}_p + g\,\mathbb{F}_p + g^2\mathbb{F}_p + \ldots + g^{b-1}\mathbb{F}_p$$

and

$$C = \mathbb{F}_p + g\,\mathbb{F}_p + g^2\mathbb{F}_p + \ldots + g^{b-a}\mathbb{F}_p\,.$$

Notice that $|P| = |A||B| = q$.

Let $\ell_{c,d} \subset \mathbb{F}_q^2$ to be the line corresponding to $\{(x, y) \in \mathbb{F}_q^2 : y = cx + d\}$. Define

$$L = \{\ell_{c,d} : c \in C, d \in B\}\,.$$

Given this we can see

$$I(P, L) = |A||B||C| \quad, \quad |L| = |B||C|\,.$$

We want $|L| = O(q)$ while $I(P, L) = \Omega(q^{4/3})$. Since $|A||B| = q$, it suffices to choose b such that

$$|C| = p^{2b-s+1} = \Omega(p^{\frac{s}{3}}) \quad, \quad |B||C| = p^{3b-s+1} = O(p^s)\,.$$

Now if $s \bmod 3 = 2$ then

$$\mathbb{Z} \cap \left[\frac{2s}{3} - \frac{1}{2}, \frac{2s}{3} - \frac{1}{3}\right] \neq \varnothing.$$

Hence, we are done by taking b to be the integer in that interval. For, $s = 3k$, $s = 3k + 1$ we take $b = 2k$ and $b = 2k + 1$ respectively. In $s = 3k$ case, we have $|I(P, L) = p^{4k+1}$ and $|L| = p^{3k+1}$. The important thing is that although $|L|$ is larger than its desired size by a factor of p, we are also exceeding the desired $|I(P, L)|$ lower bound by a factor p. A moment of though reveals that choosing L' to be the subset of L of size p^{3k} with maximum number incidences will finish the proof in this case. The situation in $s = 3k + 2$ case is analogous: if we choose L' be the the subset of L of size p^{3k+1} with the maximum number of incident points from P that will finish the construction. $\qquad\square$

15

6. Concluding remarks

In this work, we initiated the study of $\mathsf{CHSH_q}$ in the asymptotic setting. We developed the theory of both quantum and classical values of this family of games, and outlined the connection to the problem of point-line incidences over finite fields. The fact that $\mathsf{CHSH_q}$ is a natural problem to consider in the study of non-XOR games (which is the original motivation of our work as well as Burhman and Massar's) while exhibiting intimate connections to above mathematical topics indicates that this problem deserves further investigation in the future. This is especially boosted by the fact that guaranteed progress can be made by using better numerical methods to investigate higher values of q, and also by attempting to quantize the results in arithmetic combinatorics. An investigation of the extent to which the results in additive and arithmetic combinatorics quantize could certainly have much further impact beyond the problems considered here.

One goal of our study was to further develop the techniques available for analyzing the entangled value of non-binary non-local games. We believe that by giving two rather different proofs of Theorem 1.2, we demonstrated the power of the indirect approach of analyzing non-local games.

Future directions. As discussed previously, Problem 1.5 remains the most clear open problem given the bounds proved here. As mentioned before, its resolutions is likely to also resolve to Kasher and Kempe's problem regarding the security of Bourgain's two-source extractor in the presence of entanglement[19]. We can think of two possible routes for resolving this problem: one is by trying to quantize the arguments in the paper of Bourgain, Katz and Tao [4], and the other is by investigating the SDP hierarchies of Navascués et al. [23] to see whether they could lead to any improvement to Theorem 1.2 or lead to tightness results via some rounding scheme. In the hierarchy approach, it might be useful to keep in mind the rounding scheme of Kempe et al. [20] (though their result seem more relevant when the game value is close to 1 which is not the case here). Currently, with some collaborators, we are pursuing the latter direction via the SDP hierarchies.

We finish by recounting the perhaps most intriguing (and rather open-ended) future direction. This is the question of the extent to which the relatively well-understood theory of XOR games extends to larger alphabets. A related question is to find a better explanation for the absence of any analogue of a large alphabet generalization of Tsirelson's theorem [32] for even slight variants of non-XOR games (say q-XOR games for $q = 3$) in the literature. A better understanding of the above issues would certainly constitute a major advance in our understanding of two prover games and non-locality in general.

References

[1] J. S. Bell. On the Einstein-Podolsky-Rosen paradox. *Physics*, 1(3), 1964.

[2] M. Ben-Or, S. Goldwasser, J. Kilian, and A. Wigderson. Multi-prover interactive proofs: How to remove intractability assumptions. In *Proceedings of the twentieth annual ACM symposium on Theory of computing(STOC)*, 1988.

[3] J. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

[4] J. Bourgain, N. Katz, and T. Tao. A sum-product estimate in finite fields, and applications. *Geometric & Functional Analysis*, 14(1), 2004.

[5] J. Briët and T. Vidick. Explicit lower and upper bounds on the entangled value of multiplayer XOR games. *Communications in Mathematical Physics*, 321(1):181–207, 2013.

[6] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner. Bell nonlocality. *quant-ph:1303.2849*, 2013.

[7] H. Buhrman and S. Massar. Causality and Tsirel'son bounds. *Physical Review A, 72(5), 052103*, 2005.

[8] H. Buhrman, O. Regev, G. Scarpa, and R. de Wolf. Near-optimal and explicit Bell inequality violations. In *IEEE 26th Annual Conference on Computational Complexity (CCC)*, pages 157–166, 2011.

[9] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt. Proposed experiment to test local hidden-variable theories. *Physical Review Letters*, 23(15):880–884, 1969.

[10] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley & Sons, 2012.

[11] W. van Dam. Implausible consequences of superstrong nonlocality. *Natural Computing*, 12(1), 2013.

[12] I. Dinur and D. Steurer. Analytical approach to parallel repetition. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing*, STOC, pages 624–633, 2014.

[13] A. K. Ekert. Quantum cryptography based on Bell's theorem. *Physical Review Letters*, 67(6):661–663, 1991.

[14] N. Gisin. Bell inequalities: Many questions, a few answers. *The Western Ontario Series in Philosophy of Science*, 73, 2009.

[15] S.-W. Ji, J. Lee, J. Lim, K. Nagata, and H.-W. Lee. Multisetting Bell inequality for qudits. *Physical Review A*, 78(5):052103, 2008.

[16] T. G. Jones. Explicit incidence bounds over general finite fields. *arXiv preprint arXiv:1009.3899*, 2010.

[17] T. G. Jones. New quantitative estimates on the incidence geometry and growth of finite sets. *PhD Thesis, University of Bristol. arXiv:1301.4853*, 2013.

[18] M. Junge and C. Palazuelos. Large violation of Bell inequalities with low entanglement. *Communications in Mathematical Physics*, 306(3):695–746, 2011.

[19] R. Kasher and J. Kempe. Two-source extractors secure against quantum adversaries. *Theory of Computing*, 8(21), 2012.

[20] J. Kempe, O. Regev, and B. Toner. Unique games with entangled provers are easy. In *49th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2008.

[21] Y.-C. Liang, C.-W. Lim, and D.-L. Deng. Reexamination of a multisetting Bell inequality for qudits. *Physical Review A*, 80(5), 2009.

[22] N. Linden, S. Popescu, A. J. Short, and A. Winter. Quantum nonlocality and beyond: limits from nonlocal computation. *Physical Review Letters*, 99(18):180502, 2007.

[23] M. Navascués, S. Pironio, and A. Acín. A convergent hierarchy of semidefinite programs characterizing the set of quantum correlations. *New Journal of Physics*, 10(7), 2008.

[24] M. Pawłowski, T. Paterek, D. Kaszlikowski, V. Scarani, A. Winter, and M. Zukowski. Information causality as a physical principle. *Nature*, 461, 2009.

[25] M. Pawłowski and A. Winter. Hyperbits: The information quasiparticles. *Physical Review A*, 85(2), 2012.

[26] S. Popescu and D. Rohrlich. Quantum nonlocality as an axiom. *Foundations of Physics*, 24(3):379–385, 1994.

[27] A. Rao. An exposition of Bourgains 2-source extractor. *Electronic Colloquium on Computational Complexity (ECCC)*, 14(034), 2007.

[28] R. Raz. A counterexample to strong parallel repetition. *SIAM Journal on Computing*, 40(3):771–777, 2011.

[29] B. W. Reichardt, F. Unger, and U. Vazirani. A classical leash for a quantum system: command of quantum systems via rigidity of CHSH games. In *Proceedings of the 4th conference on Innovations in Theoretical Computer Science (ITCS)*, 2013.

[30] E. Szemerédi and W. T. Trotter Jr. Extremal problems in discrete geometry. *Combinatorica*, 3(3-4):381–392, 1983.

[31] B. S. Tsirel'son. Quantum generalizations of Bell's inequality. *Letters in Mathematical Physics*, 4(2), 1980.

[32] B. S. Tsirel'son. Quantum analogues of the Bell inequalities. the case of two spatially separated domains. *Journal of Soviet Mathematics*, 36(4), 1987.

[33] G. Wang. Functional boxes, communication complexity and information causality. *arXiv preprint arXiv:1109.4988*, 2011.

[34] R. F. Werner and M. M. Wolf. Bell inequalities and entanglement. *Quantum Information and Computation (QIC)*, (3), 2001.

A. Information theory background

Here we give a brief account of some definition from Information Theory. We do not require any knowledge of quantum information theory as the proofs in Appendix B are all based on classical information theory. The only place where we use some quantum ideas (which we must as Theorem 1.7 is formulated in the quantum setting), is in Theorem B.1 which we borrow directly from Pawłowski and Winter [25].

For a random variables X over a domain $\mathcal{X}$, we define the entropy of X as

$$H(X) = \sum_{x \in \mathcal{X}} \Pr[X = x] \cdot \log\left(\frac{1}{\Pr[X = x]}\right).$$

The entropy of a random variable X conditioned on the value of random variable Y is the average (according to Y) entropy of random variables $X|Y = y$, i.e.

$$H(X|Y) = \sum_{y \in \mathcal{Y}} \Pr[Y = y] \cdot H(X|Y = y),$$

where $H(X|Y = y) = \sum_{x \in \mathcal{X}} \Pr[X = x|Y = y] \cdot \log\left(\frac{1}{\Pr[X=x|Y=y]}\right)$.

Remark A.1. The above definition is equivalent with the more concise definition of $H(X|Y) = H(XY) - H(Y)$.

Now we are ready to define and state the main property of *mutual information*.

Definition A.2 (Mutual Information). Given random variables X and Y we define their mutual information via

$$I(X;Y) = H(X) + H(Y) - H(XY) = H(X) - H(X|Y) = H(Y) - H(Y|X).$$

The main fact that we shall need regarding the mutual information is the following.

Proposition A.3 (Chain Rule). *For any random variables X, Y, Z, W we have*

$$I(X;YZ|W) = I(X;Y|W) + I(X;Z|Y,W).$$

For a more detailed introduction to information theory as well as the proof of the above claims, we recommend the early chapters of [10].

B. Pairwise independent Information Causality

In this section we prove Theorem 1.7 resolving the main open problem of [25]. Our starting point is the original theorem of Pawłowski and Winter which is the binary version of our result.

Theorem B.1 (PW). *Assume in the information causality game of Definition 1.6, Alice's input $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is drawn from an unbiased pairwise independent distribution. Assume Alice's message to Bob α and output of Bob Z have both only two outcomes. We have*

$$\sum_{i=0}^{N} I(X_i; Z|b = i) = O(1).$$

The main idea for proving this theorem is to reduce the sizes of alphabets of Bob's output and Alice's message (to Bob)—while keeping a control of the information theoretic quantities of interest. We do this in two steps: first, we relax the assumption on the alphabet size of Bob's output Z, and then we relax the assumption on the alphabet size of the message α.

Proposition B.2. *Assume in the information causality game of Definition 1.6, Alice's input* $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ *is drawn from an unbiased pairwise independent distribution. Assume Alice's message to Bob* α *has only two outcomes. Let* Λ *be the alphabet for Bob's output* Z. *Then,*

$$\sum_{i=0}^{N} I(X_i; Z|b=i) = O(|\Lambda|).$$

Note that what makes this alphabet reduction not so immediate is the fact that $I(X; Z_1 Z_2 \ldots Z_k) \geq m$ does not imply a lower bound on $\max_i \{I(X; Z_i)\}$. The idea of alphabet reduction is as follows: We imagine Bob, after doing his measurement and getting some output $Z \in \Lambda$, performs a post-processing on Z to produce a binary output Z_{bin}. This Z_{bin} will be only a function of Z and his input b, and will satisfy

$$I(X_i; Z_{bin}|b=i) \geq \frac{I(X_i; Z|b=i)}{|\Lambda|}.$$

This establishes the desired result since $\sum_{i=1}^{n} I(X_i; Z_{bin}|b=i)$ is at most a constant by Theorem B.1. In order, to define Z_{bin} as a function of Z and b to satisfy the above property, we need the following lemma.

Lemma B.3. *Let* X *and* Y *be two discrete random variables. Let* Y *take values in some domain* $\mathcal{B}$ *and assume* X *to be uniform on its domain* $\mathcal{A}$. *Then there exists a function* $f : \mathcal{B} \to \{0, 1\}$ *such that*

$$I(X; f(Y)) \geq \frac{I(X; Y)}{|\mathcal{B}|}.$$

Proof. Let s be the ratio of the mutual information between X and Y, and the entropy of X.

$$I(X; Y) = sH(X) \Rightarrow H(X|Y) = (1-s)H(X).$$

For any $j \in \mathcal{B}$, define $r_j = \Pr[Y = j]$ and $t_j = \frac{H(X|Y=j)}{H(X)}$. Notice that since $\sum_{j \in \mathcal{B}} r_j = 1$ we have

$$H(X|Y) = (1-s)H(X) = \sum_{j \in \mathcal{B}} r_j t_j H(X) = H(X) - \sum_{j \in \mathcal{B}} (1 - t_j) r_j H(X).$$

Hence, there exists $j^* \in \mathcal{B}$ such that $r_{j*}(1 - t_{j*}) \geq \frac{s}{|\mathcal{B}|}$. Define the function $f : \mathcal{B} \to \{0, 1\}$ to be everywhere zero expect at j^*. We have

$$H(X|f(Y)) = \Pr[f(Y) = 0] H(X|Y \neq j^*) + \Pr[f(Y) = 1] H(X|Y = j^*).$$

Since X is unbiased over its domain, the first term is bounded by $(1 - r_{j*})H(X)$. The second term is equal to $r_{j*} t_{j*} H(X)$. Adding these two terms and using the choice of j^* it follows that

$$H(X|f(Y)) \leq H(X) - r_{j*}(1 - t_j^*)H(X) \leq H(X)\left(1 - \frac{s}{|\mathcal{B}|}\right),$$

which using the definition mutual information is seen to be the desired result. $\square$

To prove Proposition B.2, for $i \in \Lambda$ we let $f_i : \Lambda \to \{0, 1\}$ be given by the above lemma applied to X_i and Z which is applicable since X_i is unbiased. We let $Z_{bin} = f_i(Z)$. By Theorem B.1 it follows that

$$\sum_{i=1}^{n} I(X_i; Z_{bin}|b = i) = O(1) \,.$$

On the other hand $I(X_i; Z|b = i) \leqslant I(X_i; Z_{bin}|b = i)|\Lambda|$ by the above lemma. Hence we are done with Proposition B.2. Finally, we prove Theorem 1.7 by a reduction to the above proposition.

Consider any protocol $\mathcal{C}$ for the information causality game in the setting of Theorem 1.2. From $\mathcal{C}$, we produce an auxiliary protocol $\mathcal{C}^*$ with only a binary message (from Alice to Bob). By applying Proposition B.2 to $\mathcal{C}^*$ we prove the result. In $\mathcal{C}^*$ the binary message sent from Alice to Bob is denoted α_{bin}, and Bob's output is denoted by $\widetilde{Z}$. $\mathcal{C}^*$ is defined as follows: Before even receiving their inputs, Alice and Bob use shared randomness to "guess" α which Alice was going to send to Bob in the original protocol $\mathcal{C}$. Let that guess be $\overline{\alpha} \in \Sigma$. Bob then proceeds with the original protocol treating $\overline{\alpha}$ as the message that Alice intended to send him in the original protocol $\mathcal{C}$. By following the original protocol, he produces a $\overline{Z} \in \Lambda$. Alice given her input makes the appropriate measurement on her system according to the recipe given in $\mathcal{C}$ producing some $\alpha \in \Sigma$. She checks if $\alpha = \overline{\alpha}$ in which case she sends $\alpha_{bin} = 1$ to Bob; otherwise she sends $\alpha_{bin} = 0$. If Bob receives $\alpha_{bin} = 1$, he sets $\widetilde{Z} = \overline{Z}$, and outputs $\widetilde{Z}\alpha_{bin}$. On the other hand, if he receives $\alpha_{bin} = 0$, he discards $\overline{Z}$, and produces a completely random $\widetilde{Z}$. He again outputs $\widetilde{Z}\alpha_{bin}$. The choice that Bob outputs α_{bin} as well as $\widetilde{Z}$ is for technical convenience. In any case, this only increases the output size of Bob in $\mathcal{C}^*$ from $|\Lambda|$ to $2|\Lambda|$, which only slightly affects the constant in $O(\cdot)$ in Theorem 1.7. Now let us fill the rest of the details.

Proof of Theorem 1.7. The setup as usual is the following: Alice's input $\mathbf{X} = (X_1, X_2, \ldots, X_n)$ is taken from an unbiased pairwise independent distribution. Given her input, Alice makes a measurement on her system and sends α to Bob where $\alpha \in \Sigma$. Bob, given his input $b = i \in [N]$ and α, outputs a $Z \in \Lambda$. Given any protocol $\mathcal{C}$ for the above, consider $\mathcal{C}^*$ which reduces the communication to binary according to the recipe described above: so the message would be α_{bin} and Bob's output $\widetilde{Z}\alpha_{bin}$.

Now by the chain rule,

$$I(X_i; \alpha_{bin}\widetilde{Z}|b = i) = I(X_i; \alpha_{bin}|b = i) + I(X_i; \widetilde{Z}|b = i, \alpha_{bin}) \,.$$

So we have

$$I(X_i; \alpha_{bin}\widetilde{Z}|b = i) \geqslant I(X_i; \widetilde{Z}|b = i, \alpha_{bin}) \,.$$

We know $\sum_{i=1}^{n} I(X_i; \alpha_{bin}\widetilde{Z}|b = i) = O(|\Lambda|)$ by Proposition B.2. Because in this situation Alice only communicates a bit α_{bin} to Bob who outputs $\widetilde{Z}\alpha_{bin}$ which has alphabet size $2|\Lambda|$.

Now we want to analyze the term $I(X_i; \widetilde{Z}|b = i, \alpha_{bin})$. First notice that the event $b = i$ is independent of the random variable α_{bin} because α_{bin} is produced entirely on Alice's side. Hence,

$$\Pr[\alpha_{bin} = 1] = \Pr[\alpha_{bin} = 1|b = i]$$

Now conditioned on $\alpha_{bin} = 0$, $\widetilde{Z}$ is chosen completely at random. Hence,

$$I(X_i; \widetilde{Z}|b = i, \alpha_{bin} = 0) = 0 \,.$$

This proves that

$$I(X_i; \widetilde{Z}|b = i, \alpha_{bin}) = \Pr[\alpha_{bin} = 1] I(X_i; \widetilde{Z}|b = i, \alpha_{bin} = 1).$$

Conditioning on $\alpha_{bin} = 1$ is independent of X_i as one can imagine that Alice first produces α using the original protocol, and then selects $\overline{\alpha} \in \Sigma$ uniformly at random. Since $\overline{\alpha}$ is completely independent of everything else we see that the event $\alpha_{bin} = 1$ is independent of X_i. The only effect of conditioning on the event $\alpha_{bin} = 1$ is that the joint distribution of $(X_i, \widetilde{Z})$ becomes the same as the joint distribution of (X_i, Z). This means that this conditioning reduces us to the case of original protocol $\mathcal{C}$ where $\alpha \in \Sigma$ was communicated in its entirety. Hence, we have

$$I(X_i; \widetilde{Z}|b = i, \alpha_{bin} = 1) = I(X_i; Z|b = i).$$

Noticing that $\Pr[\alpha_{bin} = 1] = \frac{1}{|\Sigma|}$ and using

$$\sum_{i=1}^{n} I(X_i; \alpha_{bin}\widetilde{Z}|b = i) = O(|\Lambda|),$$

we get our desired result

$$\sum_{i=0}^{N} I(X_i; Z|b = i) = O(|\Sigma||\Lambda|).$$

$\square$

C. Reduction to distributed nonlocal computation

Here we give an alternative proof of Theorem 1.2 by establishing a higher alphabet variant of powerful result of Linden et al. [22]. The main result of their work is that for a certain broad class of games, quantum and classical strategies are equivalent in their power. This class of games, called distributed non-local computation games, is defined as follows.

Definition C.1. Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be any function and $\mathcal{D}$ any distribution on $\mathbb{F}_2^n$. In the 2-party non-local computation of $(f, \mathcal{D})$, Alice receives $x \in \mathbb{F}_2^n$ uniformly at random and Bob receives $y \stackrel{\text{def}}{=} z - x$ where z is chosen according to $\mathcal{D}$. Alice and Bob succeed if their outputs $\alpha, \beta \in \mathbb{F}_2$ satisfy $\alpha + \beta = f(z)$.

Theorem C.2 (Linden et al.). *For any binary distributed computation problem S, given by $(f, \mathcal{D})$, the entangled value and the classical value of the game coincide to the best linear approximation of f.*

$$\omega^*(S) = \omega(S) = \max_{l \in \mathcal{L}} \Pr_{z \in \mathcal{D}}[f(z) = l(z)],$$

where $\mathcal{L} = \{l : \mathbb{F}_2^n \to \mathbb{F}_2 \mid l(x + y) = l(x) + l(y)\}$ is the set of all linear functions.

Now we can naturally define the distributed version of the $\mathsf{CHSH_q}$ game following the above recipe.

Definition C.3. In $\mathsf{CHSH_q^{dist}}$, Alice and Bob receive $(\alpha, \gamma) \in \mathbb{F}_q^2$ and $(\beta, \delta) \in \mathbb{F}_q^2$. Their objective is to produce outputs a and b satisfying $a + b = (\alpha + \beta)(\gamma + \delta)$.

Any strategy $\mathcal{P}$ for $\mathsf{CHSH_q}$ results in a natural strategy $\mathcal{P}^{\mathrm{dist}}$ for $\mathsf{CHSH_q^{dist}}$ as follows: assume Alice and Bob have a strategy $\mathcal{P}$ succeeding in $\mathsf{CHSH_q}$ game with probability $p_{win} = 1/q + E(q-1)/q$. In $\mathcal{P}^{\mathrm{dist}}$, first Alice and Bob use $\mathcal{P}$ on inputs (α, δ) to produce a_1 and b_1, and then use $\mathcal{P}$ for a second time to produce a_2 and b_2 for inputs (γ, β). Finally Alice outputs $a = a_1 + a_2 + \alpha\gamma$, and Bob outputs $b = b_1 + b_2 + \beta\delta$.

We have the following proposition about $\mathcal{P}^{\mathrm{dist}}$ which is straightforward to establish.

Proposition C.4. *Let $\mathcal{P}$ be a regular protocol for $\mathsf{CHSH_q}$ with bias E. Let $\mathcal{P}^{dist}$ be the resulting distributed protocol obtained through above procedure from $\mathcal{P}$. The winning probability of Alice and Bob in the protocol $\mathcal{P}^{\mathrm{dist}}$ has bias E^2, i.e. we have*

$$p_{win} = \frac{1}{q} + \frac{q-1}{q}E^2 \,.$$

Moreover, the resulting protocol for $\mathsf{CHSH_q^{dist}}$ is itself regular.

The fact that $\mathcal{P}^{\mathrm{dist}}$ is regular when $\mathcal{P}$ is regular because of the symmetry all elements of $\mathbb{F}_q^*$. To see the statement about the winning probability of $\mathcal{P}^{\mathrm{dist}}$, notice that $\mathcal{P}^*$ succeeds if and only if the type of errors produced in two uses of $\mathcal{P}$ in $\mathcal{P}^{\mathrm{dist}}$ give errors of opposite sign. Hence, the following calculation confirms the above proposition.

$$\left(\frac{1}{q} + \frac{q-1}{q}E\right)^2 + \sum_{k=1}^{q}\left(\frac{1}{q} - \frac{E}{q}\right)^2 = \frac{1}{q} + \frac{q-1}{q}E^2 \,.$$

Because of Proposition C.4, it suffices to prove an upper bound of $2/q - 1/q^2$ for the winning probability of regular strategies in $\mathsf{CHSH_q^{dist}}$ to finish the proof of Theorem 1.2. Notice that, as hinted before, this value is exactly the winning probability achieved by the trivial strategy in which both players just output 0. This can be seen to be the best linear approximation to the polynomial $f(x,y) = xy$ which is aligned with the philosophy of Linden et al. [22] on equivalence of quantum and classical players for the distributed version of the game.

Now we proceed to the proof of Theorem 1.2. For simplicity, throughout this proof we shall assume q is a prime. However, essentially the same argument works in general with some small modifications mentioned at the end of the section. First we need the following lemma.

Lemma C.5. *Let ω be a qth primitive root of unity. For any set of unit vector $u_x, v_y \in \mathbb{C}^n$ we have*

$$\left|\sum_{x,y}\omega^{-xy}\langle u_x, v_y\rangle\right| \leqslant q^{3/2} \,. \tag{1}$$

Proof. Since the discrete Fourier transform matrix $H_{x,y} = \frac{1}{\sqrt{q}}\omega^{-xy}$ is unitary it follows that for any $f : \mathbb{F}_q \to \mathbb{C}$ and $g : \mathbb{F}_q \to \mathbb{C}$ then

$$\left|\sum_{x,y}\omega^{-xy}f(x)g(y)\right| \leqslant \sqrt{q}\|f\|_2\|g\|_2 \,.$$

23

Here we used Cauchy-Schwarz inequality between $\hat{f}(y)$ and $g(y)$. To finish the proof of the lemma, we apply the above fact to the n coordinate functions $f_i, g_i : \mathbb{F}_q \to \mathbb{C}$ defined by $f_i(x) = \overline{u}_x(i)$ and $g_i(x) = v_y(i)$. Combining that with another Cauchy-Schwarz we get

$$\left| \sum_{x,y} \omega^{-xy} \langle u_x, v_y \rangle \right| = \left| \sum_{i,x,y} \omega^{-xy} f_i(x) g_i(y) \right|$$

$$\leq \sqrt{q} \sum_{i=1}^{n} \left(\sum_{x \in \mathbb{F}_q} |u_x(i)|^2 \right)^{\frac{1}{2}} \left(\sum_{y \in \mathbb{F}_q} |v_y(i)|^2 \right)^{\frac{1}{2}}$$

$$\leq \sqrt{q} \left(\sum_{i,x} |u_x(i)|^2 \right)^{\frac{1}{2}} \left(\sum_{i,y} |v_y(i)|^2 \right)^{\frac{1}{2}}$$

$$= q^{3/2}.$$

$\square$

We also need the following convenient notation for quantifying the probability of different types of error.

Notation C.6. We let $p_k = \Pr_{a,b \leftarrow \mathcal{P}^{\text{dist}}}[a + b = k + (\alpha + \beta)(\gamma + \delta)]$ according to $\mathcal{P}^{\text{dist}}$ (defined after definition C.3).

Proof of Theorem 1.2. Let ω be a primitive q^{th} root of unity. First notice that since $p_0 = \frac{1}{q} + \frac{q-1}{q} E^2$ and $p_k = \frac{1}{q} - \frac{1}{q} E^2$ for $k \in \mathbb{F}_q^*$, it follows that

$$\sum_{k=0}^{q-1} p_k \, \omega^k = E^2 + \left(\frac{1}{q} - \frac{E^2}{q} \right) \sum_{k=0}^{q-1} \omega^k = E^2. \tag{2}$$

Denote by $x = (\alpha, \gamma)$ the input of Alice, and by $y = (\beta, \delta)$ the input of Bob. Let P_x^a and Q_y^b be the measurement operators for Alice and Bob given inputs x and y. Define $U_x = \sum_{a=0}^{q-1} \omega^a P_x^a$ and $V_y = \sum_{b=0}^{q-1} \omega^b Q_y^b$. Since we have not assumed any bound on the dimension of Alice and Bob's system, we can assume the operators P_x^a and Q_y^b are projections. This implies that the operators U_x and V_y's are unitary. Now we have

$$E^2 = \sum_{k=0}^{q-1} p_k \omega^k = \sum_{k=0}^{q-1} \mathop{\mathbf{E}}_{x,y} \left[\sum_{a+b=k+xy} \langle \psi | P_x^a \otimes Q_y^b | \psi \rangle \right] \omega^k$$

$$= \langle \psi | \mathop{\mathbf{E}}_{x,y} \left[\sum_{a,b=0}^{q-1} \omega^{a+b-xy} P_x^a \otimes Q_y^b \right] | \psi \rangle$$

$$= \frac{1}{q^2} \sum_{x,y \in F_q^2} \omega^{-xy} \langle \psi | U_x \otimes V_y | \psi \rangle$$

$$= \frac{1}{q^2} \sum_{x,y \in F_q^2} \omega^{-xy} \langle u_x | v_y \rangle,$$

24

where $U_x^\dagger \otimes I|\psi\rangle = |u_x\rangle$ and $I \otimes V_y|\psi\rangle = |v_y\rangle$ are unit vectors. So Lemma C.5 implies $E \leqslant \frac{1}{\sqrt{q}}$ which establishes the desired result. $\qquad\square$

General prime powers. In the above discussion we assumed q was a prime. Here we present the slight modifications necessary in the more general case of $q = p^s$ with s not necessarily 1.

Suppose we have a function $\chi : \mathbb{F}_q \to \mathbb{C}$ with the following properties:

1. For all $x \in \mathbb{F}_q$ we have $|\chi(x)| = 1$.

2. $\chi(x + y) = \chi(x)\chi(y)$.

3. $\sum_{x \in \mathbb{F}_q} \chi(x) = 0$.

Now if we replace the function $x \mapsto \omega^x$ with $x \mapsto \chi(x)$ in our argument we claim that the argument goes through exactly the same as before. To check this, first note that the $q \times q$ matrix $M_{xy} = \frac{1}{\sqrt{q}}\chi(xy)$ is again unitary. This is because for any $y \neq y' \in \mathbb{F}_q^*$ we have

$$\sum_{x \in \mathbb{F}_q} \overline{\chi(xy)} \cdot \chi(xy') = \sum_{x \in \mathbb{F}_q} \chi(x(y' - y)) = 0,$$

where we used the second and third properties of χ. This establishes that the analogue of equation (1) holds.

Next note that the main property we used in equation (2) was the fact that $\sum_{k \in \mathbb{F}_q} \chi(k) = 0$ which again holds here. Similarly, from the fact that $|\chi(a)| = 1$ it follows that operators of the form $U_x = \sum_{a \in \mathbb{F}_q} \chi(a)P_x^a$ are again unitary. It is easy to check the rest of our calculations are similarly valid given properties 1-3 of χ.

Finally, we shall note that a function such as χ is easy to construct using an additive isomorphism between $\mathbb{F}_q$ and $\mathbb{F}_p^s$ where p is the characteristic of $q = p^s$. [11]

D. Projective transforms and the classical value of CHSH_q

The goal of this section is to give a proof of Lemma 5.3. To do so we need to introduce some basic concepts from projective geometry. The major advantages of working over the projective plane as opposed to the affine plane for our purposes here is that firstly, the duality between points and lines becomes quite transparent and clean over the projective plane, and secondly (and more importantly), it turns out that set of *projective transformations*, while still preserving the point-line incidence structure, is much larger and richer than the group of affine transformations.

Recall that the points of the projective plane $\mathbb{PF}_q^2$ are given by the triples $(x : y : z)$, with at least one coordinate non-zero, where we identify any two points $(x : y : z)$ and $(\lambda x : \lambda y : \lambda z)$ for $\lambda \in \mathbb{F}_q^*$. The lines in the projective plane are given by triples $(l : m : n)$ consisting of all point $(x : y : z)$ satisfying $lx + my + nz = 0$. Hence, $\mathbb{PF}_q^2$ consists of a total of $q^2 + q + 1$ points and lines, where each line contains $q + 1$ points, and similarly each point is contained in $q + 1$ lines. One can go

---

[11]A particularly common choice is $\omega^{\mathrm{Tr}(\cdot)}$ where ω is p^{th} root of unity and $\mathrm{Tr}(\cdot)$ denotes the trace function given by $\alpha \mapsto \alpha + \alpha^p + \ldots + \alpha^{p^{s-1}}$ which can be shown to be a map from $\mathbb{F}_q$ to $\mathbb{F}_p$.

from the projective plane to the affine plane by discarding all the points on the "line at infinity", which consists of points of the form $(a : b : 0)$. Any point remaining can then be put in the form $(a : b : 1)$ which corresponds to the point (a, b) of the affine plane. Two points lying on the same a vertical line in the affine plane are $(a : b : 1)$ and $(a : b' : 1)$. These lie on the projective line $(1 : 0 : -a)$. These projective lines all go through the point $(0 : 1 : 0)$, or the "vertical infinity".

Next we describe the concept of a projective transformation. Any 3×3 invertible matrix A induces a map on $\mathbb{PF}_q^2$ by its action on $\mathbb{F}_q^3$. Two matrices A and B induce the same action on $\mathbb{PF}_q^2$, if they are related to by $A = \lambda B$ by some $\lambda \in \mathbb{F}_q^*$. Each element of this equivalence class represents a projective transformation on the projective plane.

Now we are almost ready to prove Lemma 5.3. The main idea of the proof is to start from the given sets of points P and lines L with large $I(P, L)$ which possibly are far from satisfying the conditions of Fact 1.4 and apply a projective transformation. After, this we discard some of our points and lines so that the hypothesis of Fact 1.4 is satisfied. We argue that a good portion of our incidences would remain after the above deletion process finishing the proof. The details are as follows.

Proof of Lemma 5.3. Let P and L in $\mathbb{PF}_q^2$ be the set of points and lines given by the hypothesis. In the argument we shall operate under the assumption that $|P|, |L| \leq \frac{q}{2}$. Indeed, one can cut the sizes of P and L by any constant factor without losing much in $|I(P, L)|$ as follows: first we shrink $|P|$ by keeping the $\frac{q}{2}$ points with the most number of incident lines—which shrinks $|I(P, L)|$ only by a factor of $\frac{2|P|}{q} = O(1)$ factor—and after keeping those lines with the most number incident points (among the $\frac{q}{2}$ size points that we kept).

Now we apply a random projective transformation to $\mathbb{PF}_q^2$ which can be seen to be equivalent to the following operation: Take a line l from all $q^2 + q + 1$ lines in $\mathbb{PF}_q^2$ uniformly at random to be the new line at the infinity. Choose a random point on l to be the new vertical point at infinity (and a random points from the remaining q points on the line at the infinity to be the new horizontal point at infinity, but this latter point plays no role for us). Next we discard all but a single line from any subset of L that are parallel after this operation. Similarly, we discard all but one point from any set of points in P that are on the same vertical line after this operation. We also discard any point of P on the new line at infinity; this puts us back in the case of the affine plane $\mathbb{F}_q^2$. Let P' and L' be the set of points and lines in $\mathbb{F}_q^2$ obtained after the above operation. To finish the proof, it suffices to show that in expectation P', L' and $I(P', L')$ would remain within constant factors of their original sizes.

Let us first compute the probability that two given lines ℓ_1 and ℓ_2 have the same slope. If we choose a random line as ℓ_∞, two lines will have the same slope if they intersect the line ℓ_∞ at the same point. The probability of this event for a specific pair of lines is $1/(q+1)$ (assuming that neither is the line at infinity), as this happens exactly when ℓ_∞ intersects ℓ_1 at the same point that ℓ_2 intersects ℓ_1. We now compute the probability that a given line survives. The probability that it is not the line at infinity is $\frac{q^2+q}{q^2+q+1}$, and given that it is not the line at infinity, it is eliminated with probability at most $\frac{q-1}{2(q+1)}$, since there are $\frac{q-1}{2}$ other lines that could eliminate it by having the same slope. This gives a total probability that it is eliminated of $\frac{q(q+3)}{2(q^2+q+1)} < \frac{1}{2}$. This shows that the expected number of lines that survive is at least $\frac{q+1}{4}$.

We next need to analyze the effect of choosing the point at vertical infinity at random and throwing

out points lying on the same vertical line. The argument is analogous to the above. Again, the probability that a given two points lie on a vertical is $1/(q+1)$. The probability that a point is not on the line at infinity is $\frac{q^2}{q^2+q+1}$. By the same argument as before, we have that the expectation that a given point survives is at least $\frac{q^2(q+3)}{2(q^2+q+1)(q+1)}$. Thus, the expected number of points that survive is also $\Theta(q)$. Furthermore the process of throwing out points and the process of throwing out lines are independent. Thus, the probability that we keep a point-line incidence is at least $\Omega(1)$ which finishes the proof. $\qquad\square$