Electronic Colloquium on Computational Complexity, Report No. 160 (2016)

Multiplayer parallel repetition for expander games

Irit Dinur [*][†] Prahladh Harsha [*][‡] Rakesh Venkat[‡] Henry Yuen[§]

October 26, 2016

Abstract

We investigate the value of parallel repetition of one-round games with any number of players $k \geq 2$. It has been an open question whether an analogue of Raz's Parallel Repetition Theorem holds for games with more than two players, i.e., whether the value of the repeated game decays exponentially with the number of repetitions. Verbitsky has shown, via a reduction to the density Hales-Jewett theorem, that the value of the repeated game must approach zero, as the number of repetitions increases. However, the rate of decay obtained in this way is extremely slow, and it is an open question whether the true rate is exponential as is the case for all two-player games.

Exponential decay bounds are known for several special cases of multi-player games, e.g., free games and anchored games. In this work, we identify a certain expansion property of the base game and show all games with this property satisfy an exponential decay parallel repetition bound. Free games and anchored games satisfy this expansion property, and thus our parallel repetition theorem reproduces all earlier exponential-decay bounds for multiplayer games. More generally, our parallel repetition bound applies to all multiplayer games that are *connected* in a certain sense.

We also describe a very simple game, called the GHZ game, that does *not* satisfy this connectivity property, and for which we do not know an exponential decay bound. We suspect that progress on bounding the value of this the parallel repetition of the GHZ game will lead to further progress on the general question.

1 Introduction and Results

We consider multi-player one-round games, and their parallel repetition. In a k-player game G, a referee chooses a k-tuple of questions $(x^1, \ldots, x^k)$ from some question distribution μ, and sends x^t to player t. Each player t gives an answer a^t that only depends on their question (i.e., they cannot communicate with each other). The referee evaluates the players' questions and answers according to some predicate $V((x^1, \ldots, x^k), (a^1, \ldots, a^k))$, and the players win if this predicate evaluates

[*]Research of the first and second author supported in part by an ISF-UGC grant number 1399/14.

[†]Weizmann Institute of Science. email: irit.dinur@weizmann.ac.il

[‡]Tata Institute of Fundamental Research. email: prahladh@tifr.res.in, rakesh09@gmail.com

[§]UC Berkeley. email: hyuen@cs.berkeley.edu

ISSN 1433-8092

to 1. The *value* of a game G, denoted by $\mathrm{val}(G)$, is the players' maximum success probability over all possible strategies they may use.

Here is a very natural operation on games, called *parallel repetition*: starting with a k-player game G, we can construct a new k-player game $G^{\otimes n}$, called the n-fold parallel repetition of G. In $G^{\otimes n}$, the referee will select n independent question tuples $(x_i^1, \ldots, x_i^t)$ from μ for each *coordinate* $i = 1, \ldots, n$, and send $(x_1^t, \ldots, x_n^t)$ to each player t. Each player has to respond with n answers, and they win this repeated game if their answers and questions for each coordinate i satisfy the original game predicate V. We call G the *base game* of the parallel repeated game $G^{\otimes n}$.

The central question we consider is how $\mathrm{val}(G^{\otimes n})$ depends on the base game G and the number of repetitions n. When G is a two-player game, the behavior of $\mathrm{val}(G^{\otimes n})$ has been extensively studied, especially due to its applications to probabilistically checkable proofs and hardness of approximation. The central result in this area is Raz's Parallel Repetition Theorem [Raz98], which states the following:

Theorem 1 (Two-player parallel repetition)**.** *Let G be a one-round two-player game with $\mathrm{val}(G) \leq 1 - \varepsilon$ for some $\varepsilon \in (0,1)$. Then for all $n \geq 0$,*

$$\mathrm{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\varepsilon^3 n}{\log|\mathscr{A}|}\right)$$

where $|\mathscr{A}|$ is the answer alphabet size of G and $c > 0$ is a universal constant.

In other words, for *nontrivial* two-player games G (i.e., games whose value is less than 1), $\mathrm{val}(G^{\otimes n})$ decays exponentially fast in n.

What about parallel repetition for games involving more than two players? It remains an intriguing open question whether Raz's Theorem can be extended to the multiplayer case. An early result of Verbitsky [Ver96] shows that for multiplayer games G with $\mathrm{val}(G) < 1$, the value of the repeated game $G^{\otimes n}$ must decay to 0 as n goes to infinity. However, the bound on the rate of decay is extremely weak: his result only shows that $\mathrm{val}(G^{\otimes n})$ is bounded by a function that is inversely proportional to the inverse Ackermann function of n [Pol12]! This poor rate of decay comes from its black-box usage of the density Hales-Jewett theorem from extremal combinatorics.

So far, Verbitsky's theorem is still the only result that gives a general parallel repetition bound for all multiplayer games. Exponential-decay parallel repetition bounds (à la Raz) for multiplayer games have been proven when there are additional assumptions on the game; for example, it has long been a folklore result that multiplayer *free* games satisfy an exponential-decay parallel repetition theorem [CWY15].[1] Recently, Bavarian, Vidick and Yuen [BVY15] studied a variant of parallel repetition (called "anchoring") where the base game G is first modified to an equivalent game $\widetilde{G}$ before being repeated in parallel, producing $\widetilde{G}^{\otimes n}$. They proved that the value of $\widetilde{G}^{\otimes n}$ is exponentially small in n when $\mathrm{val}(G) < 1$, and otherwise $\mathrm{val}(\widetilde{G}^{\otimes n}) = 1$.[2]

[1] A free game is one where each players' question is independent of all the other players'.

[2] In fact, Bavarian, Vidick and Yuen were motivated by the question of parallel repetition for *quantum* players; they showed that so-called "anchored" games satisfy quantum parallel repetition theorems.

We observe that the class of multiplayer games for which we have exponential-decay parallel repetition bounds (or, for that matter, any rate of decay better than inverse Ackermann!) all share a particular feature in common: when viewed as hypergraphs, the games all possess a certain "well-connectedness" property. For example, consider any two question tuples $x, \widehat{x}$ in the support of the question distribution μ of a free game. The question tuple $x = (x^1, \ldots, x^k)$ can be "locally morphed" to $\widehat{x} = (\widehat{x}^1, \ldots, \widehat{x}^k)$ via a sequence of question tuples $(\widehat{x}^1, \ldots, \widehat{x}^j, x^{j+1}, \ldots, x^k)$ for $j = 1, \ldots, k$, each of which remain in the support of μ. Furthermore, the anchoring transformations of [BVY15] can be understood as improving the connectivity properties of the base game before repetition. In this paper, we formalize this well-connectedness property as a type of *expansion* of the base game, and show that any connected multiplayer game has exponential-decay parallel repetition bounds. We associate with every base game G, a related graph H_G (see Definition 2) and show that if H_G is connected, then the value of the repeated game, $\mathrm{val}(G^{\otimes n})$, goes down exponentially in n, more precisely, for sufficiently large n,

$$\mathrm{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\varepsilon^5\lambda^2 n}{\log|\mathscr{A}|}\right),$$

where λ is the spectral gap of the Laplacian of the graph H_G and c is some universal constant (see Theorem 6 for an exact statement of the result). Thus, if the graph H_G is connected (i.e., $\lambda > 0$), we have an exponential decay in n. In the case of games G, wherein the associated graph H_G is not only connected but also expanding (i.e., λ is a constant), as is the case with free games, and anchored games, the rate of exponential decay is a function of alphabet size $|\mathscr{A}|$ of the base game G as in Raz's theorem.

Why care about games with more than two players? The notion of a game is an extremely basic notion, and it's use is pervasive in communication complexity, probabilistically checkable proofs (PCPs), etc. Whereas two-player games are already quite powerful and give us a lot, many problems are inherently higher-dimensional, i.e., would more naturally be cast as games with more than two players. The reason this is not commonly done is because we don't know how to analyze these creatures. For example, constraint-satisfaction-problems with arity k are naturally cast as a k-player game. They can be reduced to a two-player game in the same way that a hypergraph can be converted to a graph, but this reduction in dimensionality might be lossy. Indeed, it is empirically true that PCPs with 3 or more queries are much more powerful than 2-query PCPs, but what is the reason for this?

Furthermore, this sudden jump in difficulty in going from two-player problems to three or more players is encountered also when studying multiparty communication complexity, and seemingly because of the same technique limitations. While direct sum and direct product theorems are known for two-party communication complexity, nothing is known for the multiparty setting (in the so-called *number-on-forehead* model), and in fact making progress on this is connected to hard problems in circuit complexity.

We feel that the study of games with three or more players is a very important component in understanding such questions.

1.1 Notation

We establish some notational conventions, before stating our results formally.

For a k-player game G, we will let $\mathscr{X}^t$ denote the question alphabet for player t, and $\mathscr{X} = \mathscr{X}^1 \times \mathscr{X}^2 \times \cdots \times \mathscr{X}^k$ is the question alphabet for all the players together, underlying the question distribution μ. We will let $\mathscr{A}^t$ denote the answer alphabet for player t, and $\mathscr{A} = \mathscr{A}^1 \times \cdots \times \mathscr{A}^k$ to denote the answer alphabet for all the players together.

We will use superscripts to denote the players, and subscripts to denote the coordinate in parallel repetition. For example, x_i^t denotes the question received by player t in coordinate i. A single variable x can denote questions to the players in some coordinate clear from context, or a single coordinate game. We use x^{-t} to denote the questions to all but the t-th player in a single coordinate. When talking about multiple coordinates, we will use subscripts: x_{-i} denotes the questions to players in all but the i'th coordinate, and x_{-i}^t denotes the all questions to player t in the repeated game except for the i'th coordinate. To denote the question to player t in all coordinates, we use $x_{[n]}^t$.

We largely adopt the notational conventions from [Hol09] for probability distributions. We let capital letters denote random variables and lower case letters denote specific samples. We use P_X to denote the probability distribution of random variable X, and $\mathsf{P}_X(x)$ to denote the probability that $X = x$ for some value x. For multiple random variables, e.g., X, Y, Z, $\mathsf{P}_{XYZ}(x, y, z)$ denotes their joint distribution with respect to some probability space understood from context.

We use $\mathsf{P}_{Y|X=x}(y)$ to denote the conditional distribution $\mathsf{P}_{YX}(y, x)/\mathsf{P}_X(x)$, which is defined when $\mathsf{P}_X(x) > 0$. When conditioning on many variables, we usually use the shorthand $\mathsf{P}_{X|y,z}$ to denote the distribution $\mathsf{P}_{X|Y=y,Z=z}$. For an event W we let $\mathsf{P}_{XY|W}$ denote the distribution conditioned on W. We use the notation $\mathbb{E}_X f(x)$ and $\mathbb{E}_{\mathsf{P}_X} f(x)$ to denote the expectation $\sum_x \mathsf{P}_X(x) f(x)$.

Let P_{X_0} be a distribution over $\mathscr{X}$ and $\mathsf{P}_{X_1,Y}$ a joint distribution over $\mathscr{X} \times \mathcal{Y}$. Suppose for every x in the support of P_{X_0}, the conditional distribution $\mathsf{P}_{Y|X_1=x}$ defined over $\mathcal{Y}$ is well-defined. We then define the distribution $\mathsf{P}_{X_0}\mathsf{P}_{Y|X_1}$ over $\mathscr{X} \times \mathcal{Y}$ as

$$(\mathsf{P}_{X_0}\mathsf{P}_{Y|X_1})(x, y) := \mathsf{P}_{X_0}(x) \cdot \mathsf{P}_{Y|X_1=x}(y).$$

For two random variables X_0 and X_1 over the same set $\mathscr{X}$, we use

$$\|\mathsf{P}_{X_0} - \mathsf{P}_{X_1}\| := \frac{1}{2} \sum_{x \in \mathscr{X}} |\mathsf{P}_{X_0}(x) - \mathsf{P}_{X_1}(x)|,$$

to denote the total variation distance between P_{X_0} and P_{X_1}.

1.2 Our results

To define our class of connected and expanding games, we need the following notion of the $(k-1)$-*connection graph* of a game G. This graph, denoted H_G, has a vertex for every k-tuple of questions, and two k-tuples are connected by an edge if they agree on $(k-1)$ coordinates. A game is $(k-1)$-connected iff H_G is connected.

To further define our notion of expansion for a k-player game we need to take the weights of G into account when defining H_G. For this it is instructive to think of an intermediate bipartite graph $B_G = (\mathscr{X}', \mathscr{X}, E)$ as follows. The right hand vertices is simply $\mathscr{X}$, the set of all k-tuples of questions, and we endow these vertices with weights as given by G. The left hand vertices consists of all punctured k-tuples, which are k-tuples of questions where exactly one of the entries is replaced by a special $\star$ symbol. Connect each k-tuple of questions to all of the k ways to make it into a punctured k-tuple. Now, consider the distribution on punctured tuples obtained by selecting a random k-tuple from $\mathscr{X}$ according to the game distribution, and then puncturing it in a random location. The graph H_G is defined by selecting a random punctured tuple according to this distribution, and then selecting independently two k-tuples conditioned on this puncturing. Note that each completion is distributed exactly according to the original game distribution.

We now move to a completely explicit description consistent with the above. In what follows, $\mathsf{P}_X(x)$ denotes the probability of question tuple x under the question distribution μ, $\mathsf{P}_{X^t}(x^t)$ denotes the marginal probability of player t's question, and $P_{X^t|X^{-t}=x^{-t}}(x^t)$ denotes the same probability, conditioned on the other players having received x^{-t}.

Definition 2 (($k-1$)-connection graph of G)**.** *Let $G = (\mu, V)$ be a k-player game with question set $\mathscr{X} = \mathscr{X}^1 \times \cdots \mathscr{X}^k$. The $(k-1)$-**connection graph of** G is the weighted graph $H_G = (V_H, \rho)$ with vertex set $V_H = \mathscr{X}$ and weight function $\rho : \mathscr{X} \times \mathscr{X} \to [0,1]$, defined as follows: for every $x, x' \in \mathscr{X}$,*

$$\rho(x,x') = \begin{cases} \frac{1}{k}\,\mathsf{P}_X(x)\left[\sum_{t\in[k]} \mathsf{P}_{X^t|x^{-t}}(x^t)\right] & \text{if } x = x', \\ \frac{1}{k}\,\mathsf{P}_{X^{-t}}(x^{-t}) \cdot \mathsf{P}_{X^t|x^{-t}}(x^t) \cdot \mathsf{P}_{X^t|x^{-t}}(x'^t) & \text{if there exists } t \text{ s.t. } x^{-t} = x'^{-t} \text{ and } x^t \neq x'^t, \\ 0 & \text{otherwise.} \end{cases}$$

The weight function $\rho(x,x')$ can be viewed as the probability of generating the pair (x,x') according to the following random process: first, $x \in \mathscr{X}$ is sampled from the distribution P_X. Then, a coordinate $t \in [k]$ is chosen uniformly at random, and x' is sampled from the conditional distribution $\mathsf{P}_{X|x^{-t}}$ (that is, the distribution μ conditioned on x^{-t}).

Observe that ρ is symmetric, i.e., $\rho(x,x') = \rho(x',x)$. Furthermore, note that the weight on any given vertex is exactly:

$$\begin{aligned} \rho(x,\cdot) &= \sum_{x'} \rho(x,x') = \\ &= \mathsf{P}_X(x) \cdot \frac{1}{k} \sum_{t\in[k]} \mathsf{P}_{X^t|x^{-t}}(x^t) + \mathsf{P}_X(x) \cdot \frac{1}{k} \sum_{t\in[k]} \sum_{x'^t \neq x^t} \mathsf{P}_{X^t|x^{-t}}(x'^t) \\ &= \mathsf{P}_X(x). \end{aligned}$$

Therefore $\rho(\cdot,\cdot)$ is a probability distribution over $\mathscr{X} \times \mathscr{X}$.

Remark 3. Henceforth, when we talk about graph properties such as diameter, connectedness or expansion of H_G, we will do so only with respect to the vertices having non-zero weight.

We now recall the definition of a graph with a weight function ρ being a spectral expander:

Definition 4 (Normalized Laplacian)**.** *Let H be a weighted graph where $\rho(u,v) \leq 1$ is the weight between vertices u and v. The normalized Laplacian $L_H \in \mathbb{R}^{|V|\times|V|}$ of H is defined to be*

$$(L_H)_{u,v} = \begin{cases} 1 - \frac{\rho(u,v)}{\rho(v)} & \text{if } u = v \text{ and } \rho(v) \neq 0 \\ -\frac{\rho(u,v)}{\sqrt{\rho(u)\rho(v)}} & \text{if } \rho(u),\rho(v) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

where $\rho(u) = \sum_v \rho(u,v)$ and $\rho(v) = \sum_u \rho(u,v)$.

It is well-known that the second smallest eigenvalue of H is given by the following variational formula: for all $r \in \mathbb{N}$,

$$\lambda(H) = \inf_g \frac{\sum_{u,v} \rho(u,v)\|g(u) - g(v)\|^2}{\sum_u \rho(u)\|g(u) - \overline{g}\|^2} \tag{1}$$

where the infimum is over all vector-valued functions $g : V(H) \to \mathbb{R}^r$ defined on the vertices of H, and $\overline{g}$ is a vector in $\mathbb{R}^r$ where for each $i \in [r]$, $\overline{g}_i = \sum_u \rho(u)g(u)_i$.

Definition 5 (Expander graph)**.** *Let $\lambda \in (0,1)$. A graph H is a λ-expander if $\lambda(H) \geq \lambda$.*

Our main result is an exponential-decay parallel repetition bound for multiplayer games whose 1-connection graph is expanding:

Theorem 6 (Main theorem)**.** *Let $\epsilon, \lambda \in (0,1)$. Let G be a k-player game with* $\mathrm{val}(G) \leq 1 - \varepsilon$*. If the $(k-1)$-connection graph H_G is a λ-expander, then we have, for all $n \geq \frac{\log 4/\varepsilon}{\varepsilon^5\lambda^2}$:*

$$\mathrm{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\varepsilon^5\lambda^2 n}{\log|\mathscr{A}|}\right)$$

where $\mathscr{A} = \mathscr{A}^1 \times \cdots \times \mathscr{A}^k$ is the answer alphabet in G, and c is a universal constant.

By applying our main theorem to free games and anchored games, we recover existing exponential-decay parallel repetition results for multiplayer games [CWY15, BVY15]. We also get an exponential-decay lower bound for *connected games* – games whose $(k-1)$-connection graph is connected. We record these consequences in the following corollary:

Corollary 7. *Let G be a k-player game with* $\mathrm{val}(G) \leq 1 - \varepsilon$*, and let $n \geq \frac{\log 4/\varepsilon}{\varepsilon^5\lambda^2}$. If G is:*

1. *Free, i.e., $\mu(x) = \mu^1(x^1) \times \cdots \times \mu^k(x^k)$, then*

$$\mathrm{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\,\varepsilon^5\,n}{k^2\log|\mathscr{A}|}\right).$$

2. *α-Anchored (see Definition 13, and [BVY15] for more details), then*

$$\mathrm{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\,\alpha^{2k}\,\varepsilon^5\,n}{64\,k^2\log|\mathscr{A}|}\right).$$

3. *Connected, i.e., the $(k-1)$-connection graph is connected, then*

$$\mathrm{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\,\rho_{min}^2\,\varepsilon^5\,n}{\log|\mathscr{A}|}\right)$$

   where $\rho_{min} = \min_{u,v:\rho(u,v)>0} \rho(u,v)$. In particular, if the game G is such that μ is the uniform distribution over some set $S \subseteq \mathscr{X}$, then $\rho_{min} \geq (k|S|^2)^{-1}$.

where c is a universal constant.

The proof of Corollary 7 can be found in Appendix A.

Observe that our proof of exponential decay for games whose corresponding $(k-1)$-connection graph is connected proves a rate of exponential decay that is dependent on the size of the the base game G. It is conceivable that this rate of decay can be further improved to depend only on the alphabet size $|\mathscr{A}|$ of the base game and be independent of the size of the base game (as is the case in Raz's theorem for 2 player games). For games whose corresponding $(k-1)$-connection graph is expanding (as is the case with free games and anchoring games), we obtain a rate of exponential decay which is a function of only the base game's alphabet size.

A comment about fortified games. Bavarian, Vidick and Yuen also proved a parallel repetition bound for a special class of multiplayer games *fortified games* [BVY16] (a class of games introduced by Moshkovitz [Mos14]). However, we do not consider this a "true" exponential-decay parallel repetition bound, because it does not establish a decay bound of the form $\mathrm{val}(G^{\otimes n}) \leq \exp(-\beta n)$ for some constant β that depends on the game G, but is independent of n. Instead, it proves a decay bound that is exponential only for a small number of repetitions (depending on the base game). After this small number of repetitions, there are no guarantees about any further value decay (other than that promised by Verbitsky's theorem). Because we are interested in the asymptotic behavior of an n-repeated multiplayer game as n goes to infinity, we do not consider the parallel repetition of fortified games here.

A disconnected three-player game. It may seem that, given Corollary 7, we have established a general exponential-decay parallel repetition bound for *all* multiplayer games, albeit with some slightly annoying dependency on a quantity related to the minimum probability of any question from μ. Unfortunately, this is far from the case.

Here is a simple three-player game called the *GHZ game* whose parallel repetition resists analysis; the best decay bound we have comes from Verbitsky's theorem [Ver96]. The GHZ game is a three-player game[3] where the referee samples a question triple (x, y, z) uniformly at random from $\{(1,0,0), (0,1,0), (0,0,1), (1,1,1)\}$, and sends each bit of the triple to the corresponding player. The players respond with bits a, b, c respectively, and they win iff $x \wedge y \wedge z = a \oplus b \oplus c$. It is easy to

[3] The GHZ game comes from the study of non-locality in quantum physics; when the players use classical strategies, their maximum success probability is $\mathrm{val}(G) = 3/4$, but using quantum entanglement, the GHZ can be won with certainty [GHSZ90].

see that $\mathrm{val}(GHZ) = 3/4$ (achieved by the strategy where all players always output "0"). However, the best general bound we have on $\mathrm{val}(GHZ^{\otimes n})$ is the weak inverse-Ackermann decay given by Verbitsky's theorem.

Our main Theorem does not apply because the $(k-1)$-connection graph H_{GHZ} of the GHZ game is actually *disconnected*; no two question triples are connected via a single coordinate change. One necessary criterion for the $(k-1)$-connection graph to be connected is that, after fixing any subset of $(k-1)$ players' questions, the remaining player's question is yet undetermined. On the other hand, the players' questions in the GHZ game satisfy a linear relation (i.e. $x \oplus y \oplus z = 1$), and thus fixing two players' questions also fixes the third.

We believe that the strong correlations present in the GHZ question distribution represent the "hardest instance" of the multiplayer parallel repetition problem. Existing techniques from the two-player case (which we leverage in this paper) appear to be incapable of analyzing games with question distributions with such strong correlations. Thus we explicitly raise the open question of proving an exponential-decay parallel repetition bound for the GHZ game:

Conjecture 8 (GHZ parallel repetition)**.** There exists a constant $\beta > 0$ such that for all n,

$$\mathrm{val}(GHZ^{\otimes n}) \leq \exp(-\beta n).$$

Finally, we remark that this challenge of handling strongly correlated question distributions is reminiscent of the challenge of proving *direct sum* theorems for multiparty communication complexity in the *Number-on-Forehead* (NOF) model. There, each player sees every players' inputs but their own, so fixing $(k-1)$ out of k players' inputs will fix the remaining player's inputs. Proving direct sum results in NOF communication complexity has resisted progress for reasons that appear to be related to the multiplayer parallel repetition problem.

2 Proof of Theorem 6

2.1 Proof outline

Our proof starts out with arguments seen in previous proofs: if there is a great strategy for $G^{\otimes n}$, we can extract out an impossibly good one for G with value $\geq 1 - \varepsilon$, leading to a contradiction. However, in order to actually extract out a strategy for the single coordinate game, the players had to go through *correlated sampling* to sample some questions and answers in a subset of the n coordinates (altogether called a *dependency-breaking variable*) and then simulate $G^{\otimes n}$ on these sampled questions. In the two-player version, each player had enough information in hand to go about doing this (i.e. they only need knowledge of their own question). The main difficulty in extending this to three or more players is that the straightforward generalization the dependency-breaking variable cannot be "correlatedly sampled" without knowledge of some other player's question.

We avoid this roadblock by proving in Section 2.4 that if the $(k-1)$-connected graph H_G is connected , then the players can avoid correlated sampling altogether. In fact, they can sample an appropriate dependency-breaking variable from a *global* distribution that does not depend on any player's question.

2.2 Following Raz-Holenstein

Fix a k-player game $G = (\mu, V)$, with answer alphabet $\mathscr{A} = \mathscr{A}_1 \times \cdots \times \mathscr{A}_k$ and $\mathrm{val}(G) = 1 - \varepsilon$. Consider the n-fold parallel repetition $G^{\otimes n}$ and consider an optimal strategy $\{f^t : (\mathscr{X}^t)^{\otimes n} \to (\mathscr{A}^t)^{\otimes n}\}_{t \in [k]}$ for the k players.

For $i \in [n]$, let W_i denote the event that the players win coordinate i using this optimal strategy. Let $W = W_1 \wedge \cdots \wedge W_n$ denote the event that the players win all coordinates. For a set $S \subseteq [n]$, let $W_S = \wedge_{i \in S} W_i$. In the following, all probabilities are with respect to this optimal strategy.

Proposition 9. *Let* $\varepsilon > 0$*. Suppose that* $\log 1/\Pr(W) \leq \varepsilon n/16 - \log 4/\varepsilon$*. Then there exists a set* $S \subseteq [n]$ *of size at most* $t = \frac{8}{\varepsilon}(\log 4/\varepsilon + \log 1/\mathsf{P}(W))$ *such that*

$$\mathsf{P}_{i \notin S}(\neg W_i | W_S) \leq \varepsilon/2$$

where i *is chosen uniformly from* $[n] - S$*.*

Proof. Set $\delta = \varepsilon/8$. Let $W_{>1-\delta}$ denote the event that the players won more than $(1-\delta)n$ rounds. To show existence of such a set S, we will show that $\mathbb{E}_S \mathsf{P}(\neg W_i | W_S) \leq \varepsilon/2$, where C is a (multi)set of t independently chosen indices in $[n]$. This implies that there exists a particular set S such that $\mathsf{P}(\neg W_i | W_S) \leq \varepsilon/2$, which concludes the claim.

First we write, for a fixed S,

$$\mathsf{P}(\neg W_i | W_S) = \Pr(\neg W_i | W_S, W_{>1-\delta}) \mathsf{P}(W_{>1-\delta} | W_S) + \mathsf{P}(\neg W_i | W_S, \neg W_{>1-\delta}) \mathsf{P}(\neg W_{>1-\delta} | W_S).$$

Observe that $\mathsf{P}(\neg W_i | W_S \wedge W_{>1-\delta})$ is the probability that, conditioned on winning all rounds in S, the randomly selected coordinate $i \in [n] - S$ happens to be one of the (at most) δn lost rounds. This is at most $\delta n/(n-t) \leq \varepsilon/4$. Now observe that

$$\mathop{\mathbb{E}}_{S} \mathsf{P}(\neg W_{>1-\delta} | W_S) \leq \mathop{\mathbb{E}}_{S} \frac{\mathsf{P}(W_S | \neg W_{>1-\delta})}{\mathsf{P}(W_S)} \leq \frac{1}{\mathsf{P}(W)}(1-\delta)^t \leq \varepsilon/4$$

where for the second inequality we used the fact that $\mathsf{P}(W_S) \geq \mathsf{P}(W)$. □

For the remainder of this proof we will fix a set S as given by Proposition 9. By renaming coordinates, we will assume without loss of generality that S is the last t coordinates of $[n]$. We will let $m = n - |S|$. We will refer to the games indexed by set S as the S-games.

2.3 Dependency-breaking variables

We define the k-player analogue of the dependency-breaking variable R that is used so crucially in information-theoretic proofs of parallel repetition [Raz98, Hol09, BG15]. R will consist of a variable Ω, which fixes the questions for the S-games, and at least $(k-1)$-of-k questions in every other coordinate, and a variable $Z = (A_S)$, which fixes the answers of S-games. More formally, $\Omega = (\Omega_1, \ldots, \Omega_m, X_S)$, where X_S are fixed questions for the S-games. Each $\Omega_i = (D_i, M_i)$, for $i \in \overline{S}$, where D_i is a uniformly random value in $[k]$, and

$$M_i = X_i^{-t} \text{ if } D_i = t$$

In other words, D_i specifies which player's question to omit; the other $(k-1)$ players are fixed.

For $i \notin S$, we let Ω_{-i} denote Ω with Ω_i omitted. We let $R_{-i} := (\Omega_{-i}, A_S)$. R_i will refer to Ω_i. We will use lowercase letters to denote instantiations of these random variables: e.g., r_{-i}, x_i^t refer to specific values of R_{-i}, X_i^t respectively.

Claim 10. *Conditioned on R, $\{X_{[n]}^t\}_{t \in [k]}$ are independent.*

In the following, P_I denotes the distribution of a uniformly random $i \in [m]$, and "$\mathsf{P} \approx_\delta \mathsf{Q}$" indicates that the probability distributions P and Q are δ-close in statistical distance. We will fix

$$\delta = \frac{1}{m}\left(\log \frac{1}{\mathsf{P}(W_S)} + |S| \log |\mathscr{A}|\right).$$

The next lemma states that for an average i, if we sample questions $x_i, \widehat{x}_i$ from the joint probability distribution $\rho(x_i, \widehat{x}_i)$, the distributions of the corresponding dependency-breaking variables will be close.

Lemma 11.

$$\frac{1}{m}\sum_i \sum_{x_i, \widehat{x}_i \in \mathscr{X}} \rho(x_i, \widehat{x}_i) \left\|\mathsf{P}_{R_{-i}|x_i, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}_i, W_S}\right\|_1 \leq O(\sqrt{\delta})$$

where $\rho(\cdot, \cdot)$ is the weight function of the $(k-1)$-connection graph H_G.

Proof. First, we establish the following: for all $t \in [k]$, we have

$$\frac{1}{m}\sum_i \sum_{x_i^{-t}} \mathsf{P}_{X_i^{-t}}(x_i^{-t}) \sum_{x_i^t, \widehat{x}_i^t} \mathsf{P}_{X_i^t|x_i^{-t}}(x_i^t) \cdot \mathsf{P}_{\widehat{X}_i^t|x_i^{-t}}(\widehat{x}_i^t) \left\|\mathsf{P}_{R_{-i}|x_i, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}_i, W_S}\right\|_1 \leq O(\sqrt{\delta}) \tag{2}$$

where we use the shorthand $x_i := x_i^{-t} x_i^t$ and $\widehat{x}_i := x_i^{-t} \widehat{x}_i^t$. This follows from the same arguments found in [Hol09, BG15]; for each player t, we can treat the other $(k-1)$ players as one "meta player", and apply the two-player analysis to obtain (2).

Observe that when $x_i^t \neq \widehat{x}_i^t$, we have

$$\mathsf{P}_{X_i^{-t}}(x_i^{-t}) \cdot \mathsf{P}_{X_i^t|x_i^{-t}}(x_i^t) \cdot \mathsf{P}_{\widehat{X}_i^t|x_i^{-t}}(\widehat{x}_i^t) = k\rho(x_i, \widehat{x}_i).$$

On the other hand, when $x_i^t = \widehat{x}_i^t$, $x_i = \widehat{x}_i$ so therefore $\left\|\mathsf{P}_{R_{-i}|x_i, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}_i, W_S}\right\|_1 = 0$. Furthermore, for x_i and $\widehat{x}_i$ that differ in more than 1 coordinate, we have $\rho(x_i, \widehat{x}_i) = 0$, and for every $x_i, \widehat{x}_i$ such that $\rho(x_i, \widehat{x}_i) \neq 0$, there exists a unique $t \in [k]$ such that $x_i^t \neq \widehat{x}_i^t$. Thus we can bound for every i:

$$\begin{aligned}
&\sum_{x_i, \widehat{x}_i \in \mathscr{X}} \rho(x_i, \widehat{x}_i) \left\| \mathsf{P}_{R_{-i}|x_i, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}_i, W_S} \right\|_1 \\
&= \sum_{t \in [k]} \left(\frac{1}{k} \sum_{x_i^{-t}, x_i^t, \widehat{x}_i^t} \mathsf{P}_{X_i^{-t}}(x_i^{-t}) \cdot \mathsf{P}_{X_i^t | x_i^{-t}}(x_i^t) \cdot \mathsf{P}_{\widehat{X}_i^t | x_i^{-t}}(\widehat{x}_i^t) \left\| \mathsf{P}_{R_{-i}|x_i, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}_i, W_S} \right\|_1 \right).
\end{aligned}$$

Averaging over i and using (2), we obtain the statement of the lemma. □

2.4 Avoiding correlated sampling using expansion

At this point, ideally, every player would like to sample from $R_{-i}|x_i, W_S$. Lemma 11 establishes that $R_{-i}|x_i, W_S$ is close to $R_{-i}|x_i^{-t}, W_S$ for each $t \in [k]$. None of the players alone has knowledge of x_i^{-t}, however. We will show now that nevertheless, there is a *global* distribution known to all the players, from which the players can approximately sample $R_{-i}|x_i, W_S$.

Lemma 12. *For all $i \in [m]$ there exists a distribution $\widetilde{\mathsf{P}}_{R_{-i}}$ over R_{-i} such that*

$$\frac{1}{m} \sum_i \sum_x \rho(x) \| \mathsf{P}_{R_{-i}|x, W_S} - \widetilde{\mathsf{P}}_{R_{-i}} \|_1 \leq O\left(\frac{\delta^{1/4}}{\sqrt{\lambda}} \right).$$

Proof. For each i, define the vector-valued function $g_i : \mathscr{X} \to \mathbb{R}^{R_{-i}}$ as follows: for all $x \in \mathscr{X}$,[4]

$$g_i(x) = \sqrt{\mathsf{P}_{R_{-i}|x, W_S}}$$

where $\sqrt{\mathsf{P}_{R_{-i}|x, W_S}}$ denotes the entry-wise square root of the probability distribution $\mathsf{P}_{R_{-i}|x, W_S}$, viewed as a vector. In other words, the entries of $g_i(x)$ are indexed by different values r_{-i} of the random variable R_{-i}. Thus, g_i is a unit vector in the ℓ_2 norm.

For any i and any $x, \widehat{x} \in \mathscr{X}$, the quantity $\|g_i(x) - g_i(\widehat{x})\|^2$ is simply the square of the *Hellinger distance* between $\mathsf{P}_{R_{-i}|x, W_S}$ and $\mathsf{P}_{R_{-i}|\widehat{x}, W_S}$, which can be related to their statistical distance by

$$\|g_i(x) - g_i(\widehat{x})\|^2 \leq \| \mathsf{P}_{R_{-i}|x, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}, W_S} \|_1.$$

By Lemma 11, we can average the above inequality over all i and choosing $x, \widehat{x}$ according to the probability distribution $\rho(x, x')$, we get

$$\frac{1}{m} \sum_i \sum_{x, \widehat{x}} \rho(x, \widehat{x}) \|g_i(x) - g_i(\widehat{x})\|^2 \leq \mathop{\mathbb{E}}_i \sum_{x, \widehat{x}} \rho(x, \widehat{x}) \| \mathsf{P}_{R_{-i}|x, W_S} - \mathsf{P}_{R_{-i}|\widehat{x}, W_S} \|_1 \leq O(\sqrt{\delta})$$

But now we can leverage Equation (1). For every i, define the vector $\overline{g}_i = \sum_x \mathsf{P}_X(x) g_i(x)$. This is not necessarily a unit vector, but we have the relation

$$\frac{1}{m} \sum_i \sum_x \rho(x) \|g_i(x) - \overline{g}_i\|^2 \leq \frac{1}{\lambda m} \sum_i \sum_{x, \widehat{x}} \rho(x, \widehat{x}) \; \|g_i(x) - g_i(\widehat{x})\|^2 \leq O\left(\frac{\sqrt{\delta}}{\lambda} \right).$$

[4] Here, when we write x, we are implicitly mean x_i; we drop the subscript i for notational convenience.

If $O(\sqrt{\delta}/\lambda)$ is small, then this implies that on average, the vectors $g_i(x)$ are all close to a fixed state $\overline{g}_i$. Since $g_i(x)$ are all unit vectors, this implies that $\overline{g}_i$ is close to a unit vector. By increasing the error by a constant factor, we can assume that $\overline{g}_i$ is in fact a unit vector. Thus we can construct the probability distribution

$$\widetilde{\mathsf{P}}_{R_{-i}}(r_{-i}) = \overline{g}_i(r_{-i})^2.$$

Using that the statistical distance is at most (up to constant factors) the square root of the Hellinger distance, we get that

$$\frac{1}{m}\sum_i \sum_x \rho(x) \| \mathsf{P}_{R_{-i}|x,W_S} - \widetilde{\mathsf{P}}_{R_{-i}} \|_1 \leq O(\delta^{1/4}\lambda^{-1/2}).$$

□

2.5 Finishing the proof

Let $\{f^t\}$ be an optimal strategy for the game $G^{\otimes n}$. If $\mathsf{P}(W) \leq \frac{4}{\varepsilon}2^{-\varepsilon n/16}$, then we are done. Otherwise, suppose $\log 1/\mathsf{P}(W) \leq \varepsilon n/16 - \log 4/\varepsilon$. Let the subset S be as given by Proposition 9, and assume the coordinates are numbered so that S is the last $|S|$ coordinates of $[n]$. For all $i \in [m]$, let $\widetilde{\mathsf{P}}_{R_{-i}}$ be as given by Lemma 12. Consider the following single-shot strategy by the players, where x is drawn from μ and x^t is given to player t:

1. Using shared randomness, the players sample an $i \in [m]$ uniformly at random, and sample r_{-i} from $\widetilde{\mathsf{P}}_{R_{-i}}$. Each player t then sets x_i^t to be their "true" question x^t they received from the referee.

2. Using private randomness, each player t samples x_{-i}^t from $\mathsf{P}_{X_{-i}^t|x_i^t,r_{-i}}$. That is, each player samples questions for the n coordinates that come from the repeated game, conditioned on their own true input x_i^t and the dependency-breaking variable r_{-i}.

3. Player t outputs the i'th component of the answer vector $f^t(x_{[n]}^t)$.

Lemma 12 implies that after the first step, the sample r_{-i} each player possesses will be, up to statistical error $O(\delta^{1/4}/\sqrt{\lambda})$, distributed according to $\mathsf{P}_{R_{-i}|x,W_S}$ (on average over i and x). Then, by Claim 10, the joint distribution of the random variables $\{X_{[n]}^t\}$ that the players have sampled is

$$\mathsf{P}_{X_{[n]}^1|x_i^t,r_{-i}} \times \cdots \times \mathsf{P}_{X_{[n]}^k|x_i^t,r_{-i}} = \mathsf{P}_{X_{[n]}|x_i,r_{-i}}.$$

Thus, conditioned on r_{-i} and x_i, the distribution of their answers a_i will be distributed according to $\mathsf{P}_{A_i|x_i,r_{-i}}$. Averaging over i, x_i, and r_{-i}, we get that their answers are $O(\delta^{1/4}/\sqrt{\lambda})$-close to being distributed according to

$$\mathsf{P}_I \cdot \mathsf{P}_{X_i} \cdot \mathsf{P}_{R_{-i}|X_i,W_S} \cdot \mathsf{P}_{A_i|X_i,R_{-i}}$$

where P_I stands for the uniform distribution over $i \in [m]$. We also have that, on average i, $\mathsf{P}_{X_i|W_S}$ is $O(\sqrt{\delta})$-close in statistical distance to P_{X_i}. Thus their answers are $O(\delta^{1/4}/\sqrt{\lambda}) + O(\sqrt{\delta})$ close to being distributed as

$$\mathsf{P}_I \cdot \mathsf{P}_{A_i|W_S}.$$

Thus by Proposition 9, the probability that the players win G is at least

$$1 - \varepsilon/2 - \left(O(\delta^{1/4}/\sqrt{\lambda}) + O(\sqrt{\delta}) \right).$$

If $O(\delta^{1/4}/\sqrt{\lambda}) + O(\sqrt{\delta}) < \varepsilon/2$, then we would contradict the fact that $\mathrm{val}(G) = 1 - \varepsilon$. This implies that we must have $\delta = \Omega(\varepsilon^4 \lambda^2)$. If we let $\mathsf{P}(W) = 2^{-\gamma n}$, then we can write

$$\delta \leq \frac{16}{\varepsilon} \left[\frac{1}{n} \log \frac{4}{\varepsilon} + 2 \log |\mathscr{A}| \gamma \right]$$

where we plugged in the bound on $|S| \leq n/2$ from Proposition 9. This implies the lower bound

$$\gamma \geq \Omega \left(\frac{\varepsilon^5 \lambda^2}{\log |\mathscr{A}|} \right) \tag{3}$$

when $n \geq \frac{\log 4/\varepsilon}{\varepsilon^5 \lambda^2}$, proving the theorem.

References

[BG15] Mark Braverman and Ankit Garg. *Small value parallel repetition for general games*. In *Proc. 47th ACM Symp. on Theory of Computing (STOC)*, pages 335–340. 2015. eccc:TR14-095, doi:10.1145/2746539.2746565.

[BVY15] Mohammad Bavarian, Thomas Vidick, and Henry Yuen. *Anchoring games for parallel repetition*, 2015. (manuscript). arXiv:1509.07466.

[BVY16] ———. *Parallel repetition via fortification: analytic view and the quantum case*, 2016. (manuscript). arXiv:1603.05349, eccc:TR16-047.

[CWY15] Kai-Min Chung, Xiaodi Wu, and Henry Yuen. *Parallel repetition for entangled k-player games via fast quantum search*. In *Proc. 30th Comput. Complexity Conf.*, pages 512–536. 2015. arXiv:1501.00033, doi:10.4230/LIPIcs.CCC.2015.512.

[DS91] Persi Diaconis and Daniel Stroock. *Geometric bounds for eigenvalues of Markov chains*. Ann. Appl. Probab., 1(1):36–61, 1991. doi:10.1214/aoap/1177005980.

[GHSZ90] Daniel M. Greenberger, Michael A. Horne, Abner Shimony, and Anton Zeilinger. *Bells theorem without inequalities*. Am. J. Phys., 58(12):1131–1143, 1990. doi:10.1119/1.16243.

[Gur16] Venkatesan Guruswami. *Rapidly mixing Markov chains: a comparison of techniques*, 2016. (survey). arXiv:1603.01512.

[Hol09] Thomas Holenstein. *Parallel repetition: Simplification and the no-signaling case*. Theory Comput., 5(1):141–172, 2009. (Preliminary version in 39*th STOC*, 2007). arXiv:cs/0607139, doi:10.4086/toc.2009.v005a008.

[Mos14] Dana Moshkovitz. *Parallel repetition from fortification.* In *Proc. 55th IEEE Symp. on Foundations of Comp. Science (FOCS)*, pages 414–423. 2014. eccc:TR14-054, doi:10.1109/FOCS.2014.51.

[Pol12] D. H. J. Polymath. *A new proof of the density Hales-Jewett theorem.* Ann. of Math., 175:1283–1327, 2012. arXiv:0910.3926, doi:10.4007/annals.2012.175.3.6.

[Raz98] Ran Raz. *A parallel repetition theorem.* SIAM J. Comput., 27(3):763–803, June 1998. (Preliminary version in *27th STOC*, 1995). doi:10.1137/S0097539795280895.

[Sin92] Alistair Sinclair. *Improved bounds for mixing rates of Markov chains and multicommodity flow.* Combin. Probab. Comput., 1(4):351–370, 1992. (Preliminary version in *1st LATIN*, 1992). doi:10.1017/S0963548300000390.

[Ver96] Oleg Verbitsky. *Towards the parallel repetition conjecture.* Theoret. Comput. Sci., 157(2):277–282, 1996. (Preliminary version in *9th Structure in Complexity Theory Conference*, 1994). doi:10.1016/0304-3975(95)00165-4.

A Proof of Corollary 7

For each type of game, we compute a lower bound on the second-smallest eigenvalue of the corresponding $(k-1)$-connection graph. Applying Theorem 6 then yields the statements of the corollary.

A.1 Free games

For simplicity, assume that $\mu(x)$ is the uniform distribution over $[d]^k$, where $d = |\mathscr{X}^1| = \cdots = |\mathscr{X}^k|$.[5] Then the $(k-1)$-connection graph is a weighted version of the d-ary, k-dimensional hypercube (with self loops). Indeed, the corresponding weight function ρ behaves as follows: for $x, x' \in [d]^k$, we have $\rho(x,x) = d^{-(k+1)}$, and $\rho(x,x') = d^{-(k+1)}/k$ when x and x' differ in exactly one coordinate, and is 0 otherwise. If we compute the normalized Laplacian L_H, we get that

$$(L_H)_{u,v} = \begin{cases} 1 - \frac{1}{d} & \text{if } u = v \text{ and } \rho(v) \neq 0 \\ -\frac{1}{kd} & \text{if } \rho(u), \rho(v) \neq 0 \\ 0 & \text{otherwise} \end{cases}$$

This is the normalized Laplacian corresponding to the Cayley graph over the Abelian group $(\mathbb{Z}/d\mathbb{Z})^k$ with (weighted) generators $\{g \in (\mathbb{Z}/d\mathbb{Z})^k : |g| \leq 1\}$ where $|g|$ is the number of non-zero components of g. If $g = (0,0,\ldots,0)$, then its weight is d^{-1}, and if $|g| = 1$, then its weight is $(kd)^{-1}$. The spectrum of Cayley graphs is well understood; we have that the smallest non-zero eigenvalue of L_H is therefore $\lambda(H) = \frac{1}{k}$. Thus H_G is a $1/k$-expander.

[5] Indeed, by letting d be large enough, we can approximate μ arbitrarily well through discretization and identifying $[d]$ with $\mathscr{X}^t$ for $t = 1, \ldots, k$ in a many-to-one-fashion. Our bounds will not depend on d, so d can be taken to be arbitrarily large.

A.2 Anchored games

Here, we prove a lower bound on the second eigenvalue of the $(k-1)$-connection graph of an *anchored* game, and show that it is at least $8k/\alpha^k$. Plugging in this bound into Theorem 6 gives us

$$\text{val}(G^{\otimes n}) \leq \exp\left(-\frac{c\,\alpha^{2k}\,\varepsilon^5\,n}{64\,k^2\log|\mathscr{A}|}\right).$$

This asymptotically matches the bounds obtained in [BVY15] in terms of the dependence on α and k.

Let us first recall the definition of an anchored game.

Definition 13 (α-anchored games [BVY15])**.** *Given a k-prover game G, and a parameter $\alpha < 1$ we define the α* anchored *game $G_\perp$ as follows: the referee chooses a question tuple $(x^1,\dots,x^k)$, according to G, and independently, for every $t \in [k]$, replaces x^t by the anchoring symbol $\perp$ with probability α to get the tuple $(x'^1,\dots x'^k)$. The new domain is thus $\mathscr{X}'^1 \times \mathscr{X}'^2 \dots \mathscr{X}'^k$, where $X'^i = \mathscr{X} \cup \{\perp\}$. If any of the x''s are $\perp$, the verifier accepts trivially, otherwise the verifier accepts according to the predicate of the game G.*

For convenience, we will denote the α-anchored game itself by G in this section, and its $(k-1)$-connection graph by H_G. We will show the following lemma.

Lemma 14. [6] *$\lambda(H_G) \geq \alpha^k/8k$, when $\alpha < 1/2$.*

In order to prove Lemma 14, we need to make a couple of observations. First, note that the 1-connection graph H_G's vertices can be partitioned into disjoint sets $V_0, V_1, \dots, V_k$, where V_i has vertices of all question-tuples with exactly i bottom symbols. Thus, V_0 has vertices corresponding to the original question tuples, and $V_k = \{(\perp,\perp,\dots,\perp)\}$. While V_0 has edges between its own vertices (corresponding to edges in the 1-connection graph of the un-anchored game), all other edges in H_G go between V_i and V_{i+1}.

We will lower bound $\lambda(H_G)$ using the notion of *congestion* in the graph. This technique was first introduced by Diaconis and Strook [DS91], and improved by Sinclair [Sin92]. The below form can be found in the survey [Gur16, Section 4].

Let us view H_G as an undirected graph[7], with weight function ρ on the edges. Since $\rho(x,y) = \rho(y,x)$ by our definition, this is well-defined. A set of canonical paths in H_G is a set $\mathcal{P}$ of simple paths, one between every ordered pair (x,y) in H_G. The *path congestion parameter* of this set of canonical paths is defined as follows:

$$\zeta(\mathcal{P}) \triangleq \max_{e\in E(H_G)} \frac{1}{\rho(e)} \sum_{p_{xy}\ni e} \rho(x)\rho(y)|p_{xy}|$$

[6] Although the proof of the lemma can be easily seen to show a bound dependent only on α and k for all $\alpha < 1$, the anchored game definition in [BVY15] sets α to be a constant $< 1/2$. We only state this case, for clarity of exposition and comparison to their result.

[7] On the other hand, if viewed as a directed Markov chain, the transition probability $\Pr[y \mid x]$ for moving from x to y is exactly $\rho(x,y)/\rho(x)$. The stationary distribution on every vertex is $\rho(x)$.

Above, p_{xy} denotes the path from x to y in $\mathcal{P}$, and $|p_{xy}|$ is its length. Intuitively, the numerator in the above equation defines the 'load' on the edge (x, y), while $\rho(x, y)$ can be interpreted as its capacity. Thus, one would naturally expect that if we could find a set of canonical paths with low congestion parameter, the graph must be expanding in some sense. This is formalized in the following theorem:

Theorem 15 ([Sin92], see also [Gur16, Theorem 4.3])**.** *For any set of canonical paths $\mathcal{P}$,*

$$\lambda(H_G) \geq \frac{1}{\zeta(\mathcal{P})}$$

We will prove Lemma 14, by choosing a good set of canonical paths in H_G.

Proof of Lemma 14. Consider two vertices x, y in H_G. Let $\Delta(x, y) = \{i_1, \ldots, i_s\} \subseteq [k]$ be the set of (player) indices where the tuples differ, with $i_1 \leq i_2 \leq \ldots i_s$. We will define the canonical path from x to y to be the one obtained by flipping each of $x^{i_1}, \ldots x^{i_s}$ to $\perp$ in order, and then flip these to $y^{i_1}, \ldots y^{i_s}$, but in the reverse order $i_s \to \ldots \to i_1$. Each flip corresponds to moving along an edge in H_G. Call the set of these canonical paths $\mathcal{P}$. The path from x to y in $\mathcal{P}$ is exactly the reverse of the path from y to x.

We will upper bound the congestion through any edge $e = \{u, v\}$ caused by $\mathcal{P}$. If $u, v \in V_0$, then no path in $\mathcal{P}$ passes through this edge, and hence the congestion on e is 0. Suppose that $u \in V_l$, and $v \in V_{l+1}$ for some $l < k$.

We need to find which vertices x would use a canonical path that passes from u to v to reach another vertex. To identify this set, define $B_v \triangleq \{i \in [k] \ : \ v_i = \perp\}$, and similarly $B_u \triangleq \{i \in [k] \ : \ u_i = \perp\}$. Clearly $|B_v| = l + 1, |B_u| = l$, and $B_u \subseteq B_v$. Let us write u as $u = (\perp^l, z_u)$, where the indices are appropriately ordered (with z_u in $\overline{B}_u$).

For $0 \leq r \leq l$, a vertex $w \in V_r$ will be said to be in the r-th *shadow* of u (denoted by $S_r(u)$), if:

(a) $w|_{\overline{B}_u} = z_u$, and

(b) If $B_u = \{j_1, \ldots, j_l\}$, with $j_1 \leq \ldots \leq j_l$, then $w_{j_q} \neq \perp$ for every $q > l - r$.

The following Claim is easy to verify:

Claim 16. $\rho(S_r(u)) = \Pr_{x \sim \rho}[x^{\overline{B}_u} = z_u] \times \alpha^r (1 - \alpha)^{l-r}$

Proof. Any vertex in $S_r(u)$ can be seen to be generated by the verifier in the following way: pick a random question in the original (un-anchored) game conditioned on $x^{\overline{B}_u} = z_u$, then flip $j_1, \ldots j_r$ to $\perp$ (happens with probability α^r), and leave the others unflipped (happens with probability $(1 - \alpha)^{k-r}$). The probability of not flipping $\overline{B}_u$ (i.e. $(1 - \alpha)^{k-l}$) is accounted for in the distribution ρ of the anchored game. This yields the measure of the set $S_r(u)$ as being the expression given above. □

Any path in $\mathcal{P}$ that passes through u will necessarily either originate or end in one of its shadows. The length of any canonical path as defined above is at most $2k$. Hence, the load through the edge (u, v) can be upper bounded as follows (denoting $\Pr_{x \sim \rho}[x^{\overline{B}_u} = z_u]$ by $\Pr[z_u]$ for clarity):

$$\begin{aligned}
\sum_{p_{xy} \ni e} \rho(x)\rho(y)|p_{xy}| &\leq 2k \sum_{p_{xy} \ni e} \rho(x)\rho(y) \\
&\leq 4k \sum_{r=0}^{l} \rho(S_r(x)) \\
&= 4k \sum_{r=0}^{l} \Pr_{x \sim \rho}[x^{B_u} = z_u] \times \alpha^r (1-\alpha)^{l-r} \\
&= 4k(1-\alpha)^l \ \Pr[z_u] \sum_{r=0}^{l} \left(\frac{\alpha}{1-\alpha} \right)^r \\
&\leq 4k(l+1)(1-\alpha)^l \ \Pr[z_u] && \ldots \text{ since } \alpha < 1/2 \\
&\leq 8k \ \Pr[z_u]
\end{aligned}$$

The capacity of edge (u,v) is $\rho(u,v) = \Pr[z_u] \times \alpha^l$. Thus, the congestion along the edge is bounded by

$$\zeta(e) \leq \frac{8k \ \Pr[z_u]}{\Pr[z_u] \times \alpha^l} = \frac{8k}{\alpha^l}$$

Hence, the maximum congestion is bounded by $\zeta(\mathcal{P}) \leq \frac{8k}{\alpha^k}$, which yields the lower bound $\lambda(H_G) > \frac{\alpha^k}{8k}$, by invoking Theorem 15. □

A.3 Connected games

This follows from the observation that $\lambda(H) \geq \rho_{min}$ when the graph H is connected. The "in particular" statement follows from the definition of the weight function ρ of the $(k-1)$-connection graph: $\mathsf{P}_X(x)$ is simply $1/|S|$, and $\mathsf{P}_{X^t|x^{-t}}(x'^t)$ is also at least $1/|S|$.