The purpose of this note is to give an overview of my paper with Natarajan, "A Quantum Linearity Test for Robustly Verifying Entanglement". Since we first posted our paper on the quant-ph arXiv, Anand and I discovered that the test and its analysis could be reformulated in a more general framework of tests for group relations, and rounding of approximate group representations to exact group representations. This reformulation is stimulated by a beautiful paper by Gowers and Hatami on "Inverse and stability theorems for approximate representations of finite groups", which was first pointed to me by William Slofstra. The purpose of this post is to present the Gowers-Hatami result as a natural extension of the Blum-Luby-Rubinfeld linearity test to the non-abelian setting, with application to entanglement testing. (Of course Gowers and Hatami are well aware of this — though maybe not of the application to entanglement tests!) My hope in doing so is to make our result more accessible, and hopefully draw some of my readers from theoretical computer science into a beautiful area.

I will strive to make the post self-contained and accessible, with no quantum information background required — indeed, most of the content is purely — dare I say elegantly — mathematical. In the interest of being precise (and working out better parameters for our result than appear in our paper) I include essentially full proofs, though I may allow myself to skip a line or two in some of the calculations.

I am grateful to Anand, and Oded Regev and John Wright, for helpful comments on a preliminary version of this post.

1 Linearity testing

The Blum-Luby-Rubinfeld linearity test provides a means to certify that a function $f : \mathbb{Z}_2^n \to \{\pm 1\}$ is close to a linear function. The test can be formulated as a two-player game:

BLR linearity test:

(a) The referee selects $a, b \in \mathbb{Z}_2^n$ uniformly at random. He sends the pair (a, b) to one player, and either a, b, or $a + b$ (chosen uniformly at random) to the other.

(b) The first player replies with two bits, and the second player with a single bit. The referee accepts if and only if the player's answers satisfy the natural consistency constraint.

This test, as all others considered here, treats both players symmetrically. This allows us to restrict our attention to the case of players who both apply the same strategy, an assumption I will systematically make from now on.

Blum et al.'s result states that any strategy for the players in the linearity test must provide answers chosen according to a function that is close to linear. In this section I will provide a slight "matrix-valued" extension of the BLR result, that follows almost directly from the usual Fourier-analytic proof but will help clarify the extension to the non-abelian case.

1.1 Matrix-valued strategies

The "classical" analysis of the BLR test starts by modeling an arbitrary strategy for the players as a pair of functions $f : \mathbb{Z}_2^n \to \{\pm 1\}$ (for the second player, who receives a single string as query) and $f' : \mathbb{Z}_2^n \times \mathbb{Z}_2^n \to \{\pm 1\} \times \{\pm 1\}$ (for the first player, who receives a pair of strings as query). In doing so we are making an assumption: that the players are deterministic. More generally, we should allow "probabilistic strategies", which can be modeled via "probabilistic functions" $f : \mathbb{Z}_2^n \times \Omega \to \{\pm 1\}$ and $f' : \mathbb{Z}_2^n \times \mathbb{Z}_2^n \times \Omega \to \{\pm 1\} \times \{\pm 1\}$ respectively, where (Ω, μ) is an arbitrary probability space which plays the role of shared randomness between the players. Note that the usual claim that "probabilistic strategies are irrelevant because they can succeed no better than deterministic strategies" is somewhat moot here: the point is not to investigate success probabilities — it is easy to pass the BLR test with probability 1 — but rather derive structural consequences from the assumption that a certain strategy passes the test. In this respect, enlarging the kinds of strategies we consider valid can shed new light on the strengths, and weaknesses, of the test.

Thus, and with an eye towards the "quantum" analysis to come, let us consider an even broader set of strategies, which I'll refer to as "matrix-valued" strategies. A natural matrix-valued analogue of a function $f : \mathbb{Z}_2^n \to \{\pm 1\}$ is $F : \mathbb{Z}_2^n \to O_d(\mathbb{C})$, where $O_d(\mathbb{C})$ is the set of $d \times d$ Hermitian matrices that square to identity (equivalently, have all eigenvalues in $\{\pm 1\}$); these matrices are called "observables" in quantum mechanics. Similarly, we may generalize a function $f' : \mathbb{Z}_2^n \times \mathbb{Z}_2^n \to \{\pm 1\} \times \{\pm 1\}$ to a function $F' : \mathbb{Z}_2^n \times \mathbb{Z}_2^n \to O_d(\mathbb{C}) \times O_d(\mathbb{C})$. Here we'll impose an additional requirement: any pair (B, C) in the range of F' should be such that B and C commute. The latter condition is important so that we can make sense of the function as a strategy for the provers: we should be able to ascribe a probability distribution on outcomes $(a, (b, c))$ to any query $(x, (y, z))$ sent to the players. This is achieved by defining

$$\Pr\big((F(x), F'(y,z)) = (a, (b,c))\big) \;=\; \frac{1}{d}\,\mathrm{Tr}\big(F(x)^a F'(y,z)_1^b F'(y,z)_2^c\big), \qquad (1)$$

where for any observable O we denote O^{+1} and O^{-1} the projections on the $+1$ and -1 eigenspaces of O, respectively (so $O = O^{+1} - O^{-1}$ and $O^{+1} + O^{-1} = I$). The condition that $F'(y,z)_1$ and $F'(y,z)_2$ commute ensures that this expression is always non-negative; moreover it is easy to check that for all $(x, (y, z))$ it specifies a well-defined probability distribution on $\{\pm 1\} \times (\{\pm 1\} \times \{\pm 1\})$. Observe also that in case $d = 1$ we recover the classical deterministic case, for which with our notation $f(x)^a = 1_{f(x)=a}$. If all $F(x)$ and $F'(y,z)$ are simultaneously diagonal matrices we recover the probabilistic case, with the role of Ω (the shared randomness) played by the rows of the matrices (hence the normalization of $1/d$; we will see later how to incorporate the use of non-uniform weights).

With these notions in place we establish the following simple lemma, which states the only consequence of the BLR test we will need.

Lemma 1 *Let n be an integer, $\varepsilon \geq 0$, and $F : \mathbb{Z}_2^n \to O_d(\mathbb{C})$ and $F' : \mathbb{Z}_2^n \times \mathbb{Z}_2^n \to O_d(\mathbb{C}) \times O_d(\mathbb{C})$ a matrix strategy for the BLR test, such that players determining their answers according to this strategy (specifically, according to (1)) succeed in the test with probability at least $1 - \varepsilon$. Then*

$$\mathbb{E}_{x,y\in\mathbb{Z}_2^n} \frac{1}{d} \Re \operatorname{Tr}\big(F(x)F(y)F(x+y)\big) \;\geq\; 1 - O(\varepsilon).$$

Introducing a normalized inner product $\langle A, B\rangle_f = d^{-1}\mathrm{Tr}(AB^*)$ on the space of $d \times d$ matrices with complex entries (the * designates the conjugate-transpose), the conclusion of the lemma is that $\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\langle F(x)F(y),\, F(x+y)\rangle_f \;=\; 1 - O(\varepsilon)$.

Proof: Success with probability $1 - \varepsilon$ in the test implies the three conditions

$$\begin{aligned}
&\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\,\langle F'(x,y)_1, F(x)\rangle_f \;\geq\; 1 - 3\varepsilon,\\
&\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\,\langle F'(x,y)_2, F(y)\rangle_f \;\geq\; 1 - 3\varepsilon,\\
&\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\,\langle F'(x,y)_1F'(x,y)_2, F(x+y)\rangle_f \;\geq\; 1 - 3\varepsilon.
\end{aligned}$$

To conclude, use the triangle inequality as

$$\begin{aligned}
\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\quad &\big\|F(x)F(y) - F(x+y)\big\|_f^2\\
&\leq\; 3\Big(\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\,\big\|(F(x) - F'(x,y)_1)F(y)\big\|_f^2\\
&\qquad+\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\,\big\|(F(y) - F'(x,y)_2)F'(x,y)_1\big\|_f^2\\
&\qquad+\mathbb{E}_{x,y\in\mathbb{Z}_2^n}\,\big\|F'(x,y)_1F'(x,y)_2 - F(x+y)\big\|_f^2\Big),
\end{aligned}$$

where $\|A\|_f^2 = \langle A, A\rangle_f$ denotes the dimension-normalized Frobenius norm. Expanding each squared norm and using the preceding conditions and $F(x)^2 = 1$ for all x proves the lemma. □

1.2 The BLR theorem for matrix-valued strategies

Before stating a BLR theorem for matrix-valued strategies we need to define what it means for such a function $G : \mathbb{Z}_2^n \to O_d(\mathbb{C})$ to be *linear*. Consider first the case of probabilistic functions, i.e. G such that all $G(x)$ are diagonal, in the same basis. Any such G whose every diagonal entry is of the form $\chi_S(x) = (-1)^{S\cdot x}$ for some $S \in \{0,1\}^n$ *which may depend on the row/column number* will pass the BLR test. This shows that we cannot hope to force G to be a single linear function, we must allow "mixtures". Formally, call G linear if $G(x) = \sum_S \chi_S(x) P_S$ for some decomposition $\{P_S\}$ of the identity, i.e. the P_S are pairwsie orthogonal projections such that $\sum_S P_S = I$. Note that this indeed captures the probabilistic case; in fact, up to a basis change it is essentially equivalent to it. Thus the following may come as a surprise.

Theorem 2 *Let n be an integer, $\varepsilon \geq 0$, and $F : \mathbb{Z}_2^n \to O_d(\mathbb{C})$ such that*

$$\mathbb{E}_{x,y\in\mathbb{Z}_2^n} \frac{1}{d} \Re \langle F(x)F(y), F(x+y)\rangle_f \;\geq\; 1-\varepsilon. \tag{2}$$

Then there exists a $d' \geq d$, an isometry $V : \mathbb{C}^d \to \mathbb{C}^{d'}$, and a linear $G : \mathbb{Z}_2^n \to O_{d'}(\mathbb{C})$ such that

$$\mathbb{E}_{x\in\mathbb{Z}_2^n} \big\| F(x) - V^* G(x) V \big\|_f^2 \;\leq\; 2\,\varepsilon.$$

Note the role of V here, and the lack of control on d' (more on both aspects later). Even if F is a deterministic function f, i.e. $d = 1$, the function G returned by the theorem may be matrix-valued. In this case the isometry V is simply a unit vector $v \in \mathbb{C}^{d'}$, and expanding out the squared norm in the conclusion of the theorem yields the equivalent conclusion

$$\sum_S (v^* P_S v) \Big(\mathbb{E}_x f(x)\, \chi_S(x) \Big) \;\geq\; 1-\varepsilon,$$

where we expanded $G(x) = \sum_S \chi_S(x) P_S$ using our definition of a linear matrix-valued function. Note that $\{v^* P_S v\}$ defines a probability distribution on $\{0,1\}^n$. Thus by an averaging argument there must exist an S such that $f(x) = \chi_S(x)$ for a fraction at least $1-\varepsilon/2$ of all $x \in \mathbb{Z}_2^n$: the usual conclusion of the BLR theorem is recovered.

Proof: The proof of the theorem follows the classic Fourier-analytic proof of Bellare et al. Our first step is to define the isometry V. For a vector $u \in \mathbb{C}^d$, define

$$Vu = \sum_S \hat{F}(S) u \otimes e_S \in \mathbb{C}^d \otimes \mathbb{C}^{2^n},$$

where $\hat{F}(S) = \mathbb{E}_x \chi_S(x) F(x)$ is the matrix-valued Fourier coefficient of F at S and $\{e_S\}_{S\in\{0,1\}^n}$ an arbitrary orthonormal basis of $\mathbb{C}^{2^n}$. An easily verified extension of Parseval's formula shows $\sum_S \hat{F}(S)^2 = I$ (recall $F(x)^2 = I$ for all x), so that $V^*V = I$: V is indeed an isometry.

Next, define the linear probabilistic function G by $G(x) = \sum_S \chi_S(x) P_S$, where $P_S = I \otimes e_S e_S^*$ forms a partition of identity. We can evaluate

$$\begin{aligned} \mathbb{E}_x \tfrac{1}{d} \langle F(x), V^* G(x) V \rangle_f &= \mathbb{E}_x \sum_S \frac{1}{d} \langle F(x),\, \chi_S(x) \hat{F}(S)^2 \rangle_f \\ &= \mathbb{E}_{x,y} \frac{1}{d} \langle F(x+y),\, F(x)F(y) \rangle_f, \end{aligned}$$

where the last equality follows by expanding the Fourier coefficients and noticing the appropriate cancellation. Together with (2), this proves the theorem. □

At the risk of sounding yet more pedantic, it might be useful to comment on the relation between this proof and the usual argument. The main observation in Bellare et al.'s proof is that approximate linearity, expressed by (2), implies a lower bound on the sum of the *cubes* of the Fourier coefficients of f. Together with Parseval's formula, this bound implies the existence of a large Fourier coefficient, which identifies a close-by linear function.

The proof I gave decouples the argument. Its first step, the construction of the isometry V depends on F, but does not use anything regarding approximate linearity. It only uses Parseval's formula to argue that the isometry is well-defined. A noteworthy feature of this step is that the function G on the extended space is always well-defined as well: given a function F, it is always possible to consider the linear matrix-valued function which "samples S according to $\hat{F}(S)^2$" and then returns $\chi_S(x)$. The second step of the proof evaluates the correlation of F with the "pull-back" of G, and observes that this correlation is precisely our measure of "approximate linearity" of F, concluding the proof without having had to explicitly notice that there existed a large Fourier coefficient.

1.3 The group-theoretic perspective

Let's re-interpret the proof we just gave using group-theoretic language. A linear function $g : \mathbb{Z}_2^n \to \{\pm 1\}$ is, by definition, a mapping which respects the additive group structure on $\mathbb{Z}_2^n$, namely it is a representation. Since $G = (\mathbb{Z}_2^n, +)$ is an abelian group, it has $|G| = 2^n$ irreducible 1-dimensional representations, given by the characters χ_S. As such, the linear function defined in the proof of Theorem 2 is nothing but a list of all irreducible representations of G.

The condition (2) derived in the proof of the theorem can be interpreted as the condition that F is an "approximate representation" of G. Let's make this a general definition. For d-dimensional matrices A, B and σ such that σ is positive semidefinite, write

$$\langle A, B \rangle_\sigma = \mathrm{Tr}(AB^* \sigma),$$

where we use B^* to denote the conjugate-transpose. The following definition considers arbitrary finite groups (not necessarily abelian).

Definition 3 *Given a finite group G, an integer $d \geq 1$, $\varepsilon \geq 0$, and a d-dimensional positive semidefinite matrix σ with trace 1, an (ε, σ)-representation of G is a function $f : G \to U_d(\mathbb{C})$, the unitary group of $d \times d$ matrices, such that*

$$\mathbb{E}_{x,y \in G}\, \Re\big(\big\langle f(x)^* f(y), f(x^{-1}y)\big\rangle_\sigma\big) \;\geq\; 1 - \varepsilon, \tag{3}$$

where the expectation is taken under the uniform distribution over G.

The condition (3) in the definition is very closely related to Gowers's U^2 norm

$$\|f\|_{U^2}^4 \;=\; \mathbb{E}_{xy^{-1}=zw^{-1}}\, \big\langle f(x)f(y)^*, f(z)f(w)^*\big\rangle_\sigma.$$

While a large Gowers norm implies closeness to an affine function, we are interested in testing linear functions, and the condition (3) will arise naturally from our calculations in the next section.

If $G = (\mathbb{Z}_2^n, +)$, the product xy^{-1} should be written additively as $x - y = x + y$, so that the condition (2) is precisely that F is an (ε, σ)-representation of G, where $\sigma = d^{-1}I$. Theorem 2 can thus be reformulated as stating that for any (ε, σ)-approximate representation of the abelian group $G = (\mathbb{Z}_2^n, +)$ there exists an isometry $V : \mathbb{C}^d \to \mathbb{C}^d \otimes \mathbb{C}^{2^n}$ and an exact representation g of G on $\mathbb{C}^d \otimes \mathbb{C}^{2^n}$ such that f is well-approximated by the "pull-back" $V^* g V$ of g to $\mathbb{C}^d$. In the next section I will make the words in quotes precise and generalize the result to the case of arbitrary finite groups.

2 Approximate representations of non-abelian groups

2.1 The Gowers-Hatami theorem

In their paper Gowers and Hatami consider the problem of "rounding" approximate group representations to exact representations. I highly recommend the paper, which gives a thorough introduction to the topic, including multiple motivations. Here I will state and prove a slightly more general, but quantitatively weaker, variant of their result inspired by the somewhat convoluted analysis of the BLR test given in the previous section.

Theorem 4 (Gowers-Hatami) *Let G be a finite group, $\varepsilon \geq 0$, and $f : G \to U_d(\mathbb{C})$ an (ε, σ)-representation of G. Then there exists a $d' \geq d$, an isometry $V : \mathbb{C}^d \to \mathbb{C}^{d'}$, and a representation $g : G \to U_{d'}(\mathbb{C})$ such that*

$$\mathbb{E}_{x \in G}\, \big\| f(x) - V^* g(x) V \big\|_\sigma^2 \;\leq\; 2\,\varepsilon.$$

Gowers and Hatami limit themselves to the case of $\sigma = d^{-1}I_d$, which corresponds to the dimension-normalized Frobenius norm. In this scenario they in addition obtain a tight control of the dimension d', and show that one can always take $d' = (1+O(\varepsilon))d$ in the theorem. I will give a much shorter proof than theirs (the proof is implicit in their argument) that does not seem to allow to recover this estimate. (It is possible to adapt their proof to keep a control of d' even in the case of general σ, but I will not explain this here.) Essentially the same proof as the one sketched below has been extended to some classes of infinite groups by De Chiffre, Ozawa and Thom in a recent preprint.

Note that, contrary to the BLR theorem, where the "embedding" is not strictly necessary (if ε is small enough we can identify a single close-by linear function), as noted by Gowers and Hatami Theorem 4 does not in general hold with $d' = d$. The reason is that it is possible for G to have an approximate representation in some dimension d, but no exact representation of the same dimension: to obtain an example of this, take any group G that has all non-trivial irreducible representations of large enough dimension, and create an approximate representation in e.g. dimension one less by "cutting off" one row and column from an exact representation. The dimension normalization induced by the norm $\|\cdot\|_\sigma$ will barely notice this, but it will be impossible to "round" the approximate representation obtained to an exact one without modifying the dimension.

The necessity for the embedding helps distinguish the Gowers-Hatami result from other extensions of the linearity test to the non-abelian setting, such as the work by Ben-Or et al. on non-Abelian homomorphism testing (I thank Oded Regev for pointing me to the paper). In that paper the authors show that a function $f : G \to H$, where G and H are finite non-abelian groups, which satisfies $\Pr(f(x)f(y) = f(xy)) \geq 1-\varepsilon$, is $O(\varepsilon)$-close to a homomorphism $g : G \to H$. The main difference with the setting for the Gowers-Hatami result is that since H is finite, Ben-Or et al. use the Kronecker δ function as distance on H. This allows them to employ combinatorial arguments, and provide a rounding procedure that does not need to modify the range space (H). In contrast, here the unitary group is infinite.

The main ingredient needed to extend the analysis of the BLR test is an appropriate notion of Fourier transform over non-abelian groups. Given an irreducible representation $\rho : G \to U_{d_\rho}(\mathbb{C})$, define

$$\hat{f}(\rho) = \mathbb{E}_{x\in G}\, f(x) \otimes \overline{\rho(x)}. \tag{4}$$

In case G is abelian, we always have $d_\rho = 1$, the tensor product is a product, and (4) reduces to the usual definition of Fourier coefficient. The only properties we will need of irreducible representations is that they satisfy the relation

$$\sum_\rho d_\rho \,\mathrm{Tr}(\rho(x)) = |G|\delta_{xe}\ , \tag{5}$$

for any $x \in G$. Note that plugging in $x = e$ (the identity element in G) yields $\sum_\rho d_\rho^2 = |G|$.

Proof:[of Theorem 4] As in the proof of Theorem 2 our first step is to define an isometry $V : \mathbb{C}^d \to \mathbb{C}^d \otimes (\oplus_\rho \mathbb{C}^{d_\rho} \otimes \mathbb{C}^{d_\rho})$ by

$$V:\; u \in \mathbb{C}^d \;\mapsto \bigoplus_\rho d_\rho^{1/2} \sum_{i=1}^{d_\rho} \big(\hat{f}(\rho)(u \otimes e_i)\big) \otimes e_i,$$

where the direct sum ranges over all irreducible representations ρ of G and $\{e_i\}$ is the canonical basis. Note what V does: it "embeds" any vector $u \in \mathbb{C}^d$ into a direct sum, over irreducible representations ρ, of a d-dimensional vector of $d_\rho \times d_\rho$ matrices. Each (matrix) entry of this vector can be thought of as the Fourier coefficient of the corresponding entry of the vector $f(x)u$ associated with ρ. If $G = \mathbb{Z}_2^n$ and f ranges over $O_(\mathbb{C})$ this recovers the isometry defined in the proof of Theorem 2. And indeed, the fact that V is an isometry again follows from the appropriate extension of Parseval's formula:

$$\begin{aligned} V^*V &= \sum_\rho d_\rho \sum_i (I \otimes e_i^*) \hat{f}(\rho)^* \hat{f}(\rho) (I \otimes e_i) \\ &= \mathbb{E}_{x,y}\, f(x)^* f(y) \sum_\rho d_\rho \sum_i (e_i^* \rho(x)^T \overline{\rho(y)} e_i) \\ &= \sum_\rho \frac{d_\rho^2}{|G|} I = I, \end{aligned}$$

where for the second line we used the definition (4) of $\hat{f}(\rho)$ and for the third we used (5) and the fact that f takes values in the unitary group.

Following the same steps as in the proof of Theorem 2, we next define

$$g(x) = \bigoplus_\rho \big(I_d \otimes I_{d_\rho} \otimes \rho(x)\big),$$

a direct sum over all irreducible representations of G (hence itself a representation). Lets' first compute the "pull-back" of g by V: following a similar calculation as above, for any $x \in G$,

$$\begin{aligned}
V^* g(x) V &= \sum_\rho d_\rho \sum_{i,j} (I \otimes e_i^*) \hat{f}(\rho)^* \hat{f}(\rho)(I \otimes e_j) \otimes e_i^* \rho(x) e_j) \\
&= \mathbb{E}_{z,y}\, f(z)^* f(y) \sum_\rho d_\rho \sum_{i,j} (e_i^* \rho(z)^T \overline{\rho(y)} e_j)\big(e_i^* \rho(x) e_j\big) \\
&= \mathbb{E}_{z,y}\, f(z)^* f(y) \sum_\rho d_\rho \mathrm{Tr}\big(\rho(z)^T \overline{\rho(y)} \rho(x)^T\big) \\
&= \mathbb{E}_{z,y}\, f(z)^* f(y) \sum_\rho d_\rho \mathrm{Tr}\big(\rho(z^{-1} y x^{-1})\big) \\
&= \mathbb{E}_z\, f(z)^* f(zx),
\end{aligned}$$

where the last equality uses (5). It then follows that

$$\mathbb{E}_x \big\langle f(x), V^* g(x) V \big\rangle_\sigma \; = \mathbb{E}_{x,z} \mathrm{Tr}\big(f(x) f(zx)^* f(z) \sigma\big).$$

This relates correlation of f with $V^* g V$ to the quality of f as an approximate representation and proves the theorem. □

2.2 Application: the Weyl-Heisenberg group

In quantum information we care a lot about the Pauli group. For our purposes it will be be sufficient (and much more convenient, allowing us to avoid some trouble with complex conjugation) to consider the Weyl-Heisenberg group H, or "Pauli group modulo complex conjugation", which is the 8-element group $\{\pm\sigma_I, \pm\sigma_X, \pm\sigma_Z, \pm\sigma_W\}$ whose multiplication table matches that of the 2×2 matrices

$$\sigma_X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \qquad \sigma_Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \tag{6}$$

$\sigma_I = \sigma_X^2 = \sigma_Z^2$ and $\sigma_W = \sigma_X \sigma_Z = -\sigma_Z \sigma_X$. This group has four 1-dimensional representations, uniquely specified by the image of σ_X and σ_Z in $\{\pm 1\}$, and a single irreducible 2-dimensional representation, given by the matrices defined above. We can also consider the "n-qubit Weyl-Heisenberg group" $H^{(n)}$, the matrix group generated by n-fold tensor products of the 8 matrices identified above. The irreducible representations of $H^{(n)}$ are easily computed from those of H; for us the only thing that matters is that the only irreducible representation which satisfies $\rho(-I) = -\rho(I)$ has dimension 2^n and is given by the defining matrix representation (in fact, it is the only irreducible representation in dimension larger than 1).

With the upcoming application to entanglement testing in mind, I will state a version of Theorem 4 tailored to the group $H^{(n)}$ and a specific choice of presentation for the group relations. Towards this we first need to recall the notion of *Schmidt decomposition* of a bipartite state (i.e. unit vector) $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$. The Schmidt decomposition states that any such vector can be written as

$$\psi \,=\, \sum_i \sqrt{\lambda_i}\, u_i \otimes v_i, \tag{7}$$

for some orthonomal bases $\{u_i\}$ and $\{v_i\}$ of $\mathbb{C}^d$ (the "Schmidt vectors") and non-negative coefficients $\sqrt{\lambda_i}$ (the "Schmidt coefficients"). The decomposition can be obtained by "reshaping" $\psi = \sum_{i,j} \psi_{i,j} e_i \otimes e_j$ into a $d \times d$ matrix $K = (\psi_{i,j})_{1 \le i,j \le d}$ and performing the singular value decomposition. To ψ we associate the (uniquely defined) positive semidefinite matrix

$$\sigma \,=\, KK^* \,=\, \sum_i \lambda_i\, u_i u_i^* \;; \tag{8}$$

note that σ has trace 1. The matrix σ is called the *reduced density* of ψ (on the first system).

Corollary 5 *Let n, d be integer, $\varepsilon \ge 0$, $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$ a unit vector, σ the positive semidefinite matrix associated to ψ as in (8), and $f : \{X, Z\} \times \{0,1\}^n \to U(\mathbb{C}^d)$. For $a, b \in \{0,1\}^n$ let $X(a) = f(X, a)$, $Z(b) = f(Z, b)$, and assume $X(a)^2 = Z(b)^2 = I_d$ for all a, b (we call such operators, unitaries with eigenvalues in $\{\pm 1\}$, observables). Suppose that the following inequalities hold: consistency*

$$\mathbb{E}_a\, \psi^* \big(X(a) \otimes X(a)\big)\psi \,\ge\, 1 - \varepsilon, \qquad \mathbb{E}_b\, \psi^* \big(Z(b) \otimes Z(b)\big)\psi \,\ge\, 1 - \varepsilon, \tag{9}$$

linearity

$$\mathbb{E}_{a,a'} \big\|X(a)X(a') - X(a + a')\big\|_\sigma^2 \le \varepsilon, \qquad \mathbb{E}_{b,b'} \big\|Z(b)Z(b') - Z(b + b')\big\|_\sigma^2 \le \varepsilon, \tag{10}$$

and anti-commutation

$$\mathbb{E}_{a,b} \big\|X(a)Z(b) - (-1)^{a \cdot b} X(a)Z(b)\big\|_\sigma^2 \,\le\, \varepsilon. \tag{11}$$

Then there exists a $d' \ge d$, an isometry $V : \mathbb{C}^d \to \mathbb{C}^{d'}$, and a representation $g : H^{(n)} \to U_{d'}(\mathbb{C})$ such that $g(-I) = -I_{d'}$ and

$$\mathbb{E}_{a,b} \big\|X(a)Z(b) - V^* g(\sigma_X(a)\sigma_Z(b))V\big\|_\sigma^2 \,=\, O(\varepsilon).$$

Note that the conditions (10) and (11) in the corollary are very similar to the conditions required of an approximate representation of the group $H^{(n)}$; in fact it is easy to convince oneself that their exact analogue suffice to imply all the group relations. The reason for including only those relations is that they are the ones that it will be possible to test; see the next section for this. Condition (9) is necessary to derive the conditions of Theorem 4 from (10) and (11), and is also testable; see the proof.

Proof: To apply Theorem 4 we need to construct an (ε, σ)-representation f of the group $H^{(n)}$. Using that any element of $H^{(n)}$ has a unique representative of the form $\pm\sigma_X(a)\sigma_Z(b)$ for $a, b \in \{0,1\}^n$, we define $f(\pm\sigma_X(a)\sigma_Z(b)) = \pm X(a)Z(b)$. Next we need to verify (3). Let $x, y \in H^{(n)}$ be such that $x = \sigma_X(a_x)\sigma_Z(b_x)$ and $y = \sigma_X(a_y)\sigma_Z(b_y)$ for n-bit strings (a_x, b_x) and (a_y, b_y) respectively. Up to phase, we can exploit successive cancellations to decompose $(f(x)f(y)^* - f(xy^{-1})) \otimes I$ as

$$\begin{aligned}
&\big(X(a_x)Z(b_x)X(a_y)Z(b_y) - (-1)^{a_y\cdot b_x}X(a_x+a_y)Z(b_x+b_y)\big)\otimes I\\
&\quad = X(a_x)Z(b_x)X(a_y)\big(Z(b_y)\otimes I - I\otimes Z(b_y)\big)\\
&\qquad + X(a_x)\big(Z(b_x)X(a_y) - (-1)^{a_y\cdot b_x}X(a_y)Z(b_x)\big)\otimes Z(b_y)\\
&\qquad + (-1)^{a_y\cdot b_x}\big(X(a_x)X(a_y)\otimes Z(b_y)\big)\big(Z(b_x)\otimes I - I\otimes Z(b_x)\big)\\
&\qquad + (-1)^{a_y\cdot b_x}\big(X(a_x)X(a_y)\otimes Z(b_y)Z(b_x) - X(a_x+a_y)\otimes Z(b_x+b_y)\big)\\
&\qquad + (-1)^{a_y\cdot b_x}\big(X(a_x+a_y)\otimes I\big)\big(I\otimes Z(b_x+b_y) - Z(b_x+b_y)\otimes I\big).
\end{aligned}$$

(It is worth staring at this sequence of equations for a little bit. In particular, note the "player-switching" that takes place in the 2nd, 4th and 6th lines; this is used as a means to "commute" the appropriate unitaries, and is the reason for including (9) among the assumptions of the corollary.) Evaluating each term on the vector ψ, taking the squared Euclidean norm, and then the expectation over uniformly random a_x, a_y, b_x, b_y, the inequality $\|AB\psi\| \le \|A\|\|B\psi\|$ and the assumptions of the theorem let us bound the overlap of each term in the resulting summation by $O(\varepsilon)$. Using $\|(A\otimes I)\psi\| = \|A\|_\sigma$ by definition, we obtain the bound

$$\mathbb{E}_{x,y}\big\|f(x)f(y)^* - f(xy^{-1})\big\|_\sigma^2 = O(\varepsilon).$$

We are thus in a position to apply Theorem 4, which gives an isometry V and exact representation g such that

$$\mathbb{E}_{a,b}\Big\|X(a)Z(b) - \frac{1}{2}V^*\big(g(\sigma_X(a)\sigma_Z(b)) - g(-\sigma_X(a)\sigma_Z(b))\big)V\Big\|_\sigma^2 = O(\varepsilon). \qquad (12)$$

Using that g is a representation, $g(-\sigma_X(a)\sigma_Z(b)) = g(-I)g(\sigma_X(a)\sigma_Z(b))$. It follows from (12) that $\|g(-I)+I\|_\sigma^2 = O(\varepsilon)$, so we may restrict the range of V to the subspace where $g(-I) = -I$ without introducing much additional error. □

3 Entanglement testing

Our discussion so far has barely touched upon the notion of entanglement. Recall the Schmidt decopmosition (7) of a unit vector $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$, and the associated reduced density matrix σ defined in (8). The state ψ is called *entangled* if this matrix has rank larger than 1; equivalently, if there is more than one non-zero coefficient λ_i in (7). The *Schmidt rank* of ψ is the rank of σ, the number of non-zero terms in (7). It is a crude, but convenient, measure of entanglement; in particular it provides a lower bound on the local dimension d. A useful observation is that the Schmidt rank is invariant under local unitary operations: these may affect the Schmidt vectors $\{u_i\}$ and $\{v_i\}$, but not the number of non-zero terms.

3.1 A certificate for high-dimensional entanglement

Among all entangled states in dimension d, the *maximally entangled state* ϕ_d is the one which maximizes entanglement entropy, defined as the Shannon entropy of the distribution induced by the squares of the Schmidt coefficients:

$$\phi_d = \frac{1}{\sqrt{d}} \sum_{i=1}^{d} e_i \otimes e_i,$$

with entropy $\log d$. The following lemma gives a "robust" characterization of the maximally entangled state in dimension $d = 2^n$ as the unique common eigenvalue-1 eigenvector of all operators of the form $\sigma_P \otimes \sigma_P$, where σ_P ranges over the elements of the unique 2^n-dimensional irreducible representation of the Weyl-Heisenberg group $H^{(n)}$, i.e. the Pauli matrices (taken modulo $\sqrt{-1}$).

Lemma 6 *Let $\varepsilon \geq 0$, n an integer, $d = 2^n$, and $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$ a unit vector such that*

$$\mathbb{E}_{a,b}\, \psi^* \big(\sigma_X(a)\sigma_Z(b) \otimes \sigma_X(a)\sigma_Z(b)\big)\psi \geq 1 - \varepsilon. \tag{13}$$

Then $|\psi^* \phi_{2^n}|^2 \geq 1 - \varepsilon$. In particular, ψ has Schmidt rank at least $(1 - \varepsilon)2^n$.

Proof: Consider the case $n = 1$. The "swap" matrix

$$S = \frac{1}{4}\big(\sigma_I \otimes \sigma_I + \sigma_X \otimes \sigma_X + \sigma_Z \otimes \sigma_Z + \sigma_W \otimes \sigma_W\big)$$

squares to identity and has a unique eigenvalue-1 eigenvector, the vector $\phi_2 = (e_1 \otimes e_1 + e_2 \otimes e_2)/\sqrt{2}$ (a.k.a. "EPR pair"). Thus $\psi^* S \psi \geq 1 - \varepsilon$ implies $|\psi^* \phi|^2 \geq 1 - \varepsilon$. The same argument for general n shows $|\psi^* \phi_{2^n}|^2 \geq 1 - \varepsilon$. Any unit vector u of Schmidt rank at most r satisfies $|u^* \phi_{2^n}|^2 \leq r 2^{-n}$, concluding the proof. □

Lemma 6 provides an "experimental road-map" for establishing that a bipartite system is in a highly entangled state:

(i) Select a random $\sigma_P = \pm\sigma_X(a)\sigma_Z(b) \in H^{(n)}$;

(ii) Measure both halves of ψ using σ_P;

(iii) Check that the outcomes agree.

To explain the connection between the above "operational test" and the lemma I should review what a measurement in quantum mechanics is. For our purposes it is enough to talk about binary measurements (i.e. measurements with two outcomes, $+1$ and -1). Any such measurement is specified by a pair of orthogonal projections, M_+ and M_-, on $\mathbb{C}^d$ such that $M_+ + M_- = I_d$. The probability of obtaining outcome $\pm$ when measuring ψ is $\|M_\pm \psi\|^2$. We can represent a binary measurement succinctly through the *observable* $M = M_+ - M_-$. (In general, an observable is a Hermitian matrix which squares to identity.) It is then the case that if an observable M is applied on the first half of a state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$, and another observable N is applied on the second half, then the probability of agreement, minus the probability of disagreement, between the outcomes obtained is precisely $\psi^*(M \otimes N)\psi$, a number which lies in $[-1, 1]$. Thus the condition that the test described above accepts with probability $1 - \varepsilon$ when performed on a state ψ is precisely equivalent to the assumption (13) of Lemma 6.

Even though this provides a perfectly fine test for entanglement in principle, practitioners in the foundations of quantum mechanics know all too well that their opponents — e.g. "quantum-skeptics" — will not be satisfied with such an experiment. In particular, who is to guarantee that the measurement performed in step (ii) is really $\sigma_P \otimes \sigma_P$, as claimed? To the least, doesn't this already implicitly assume that the measured system has dimension 2^n?

This is where the notion of *device independence* comes in. Briefly, in this context the idea is to obtain the same conclusion (a certificate of high-dimensional entanglement) *without* any assumption on the measurement performed: the only information to be trusted is classical data (statistics generated by the experiment), but not the operational details of the experiment itself.

This is where Corollary 5 enters the picture. Reformulated in the present context, the corollary provides a means to *verify* that arbitrary measurements "all but behave" as Pauli measurements, provided they generate the right statistics. To explain how this can be done we need to provide additional "operational tests" that can be used to certify the assumptions of the corollary.

3.2 Testing the Weyl-Heisenberg group relations

Corollary 5 makes three assumptions about the observables $X(a)$ and $Z(b)$: that they satisfy approximate consistency (9), linearity (10), and anti-commutation (11). In this section I will describe two (somewhat well-known) tests that allow to certify these relations based only on the fact that the measurements generate statistics which pass the tests.

Linearity test:

(a) The referee selects $W \in \{X, Z\}$ and $a, a' \in \{0,1\}^n$ uniformly at random. He sends (W, a, a') to one player and (W, a), (W, a'), or $(W, a + a')$ to the other.

(b) The first player replies with two bits, and the second with a single bit. The referee accepts if and only if the player's answers are consistent.

As always in this note, the test treats both players simultaneously. As a result we can (and will) assume that the player's strategy is symmetric, and is specified by a permutation-invariant state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$ and a measurement for each question: an observable $W(a)$ associated to questions of the form (W, a), and a more complicated four-outcome measurement $\{W^{a,a'}\}$ associated with questions of the form (W, a, a') (It will not be necessary to go into the details of the formalism for such measurements).

The linearity test described above is exactly identical to the BLR linearity test described earlier, but for the use of the basis label $W \in \{X, Z\}$. The lemma below is a direct analogue of Lemma 1, which extends the analysis to the setting of players sharing entanglement. The lemma was first introduced in a joint paper with Ito, where we used an extension of the linearity test, Babai et al.'s multilinearity test, to show the inclusion of complexity classes NEXP⊆MIP*.

Lemma 7 *Suppose that a family of observables $\{W(a)\}$ for $W \in \{X, Z\}$ and $a \in \{0,1\}^n$, generates outcomes that succeed in the linearity test with probability $1 - \varepsilon$, when applied on a bipartite state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$. Then the following hold: approximate consistency*

$$\mathbb{E}_a\, \psi^* \big(X(a) \otimes X(a) \big) \psi \;=\; 1 - O(\varepsilon), \qquad \mathbb{E}_b\, \psi^* \big(Z(b) \otimes Z(b) \big) \psi \;\geq\; 1 - O(\varepsilon),$$

and linearity

$$\mathbb{E}_{a,a'} \big\| X(a)X(a') - X(a+a') \big\|_\sigma^2 = O(\varepsilon), \qquad \mathbb{E}_{b,b'} \big\| Z(b)Z(b') - Z(b+b') \big\|_\sigma^2 \;=\; O(\varepsilon).$$

Testing anti-commutation is slightly more involved. We will achieve this by using a two-player game called the Magic Square game. This is a fascinating game, but just as for the linearity test I will treat it superficially and only recall the part of the analysis that is useful for us (see e.g. the paper by Wu et al. for a description of the game and a proof of Lemma 8 below).

Lemma 8 (Magic Square) *The Magic Square game is a two-player game with nine possible questions (with binary answers) for one player and six possible questions (with two-bit answers) for the other player which has the following properties. The distribution on questions in the game is uniform. Two of the questions to the first player are labelled X and Z respectively. For any strategy for the players that succeeds in the game with probability at least $1-\varepsilon$ using a bipartite state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$ and observables X and Z for questions X and Z respectively, it holds that*

$$\big\|\big((XZ+ZX)\otimes I_d\big)\psi\big\|^2 = O\big(\sqrt{\varepsilon}\big). \tag{14}$$

Moreover, there exists a strategy which succeeds with probability 1 in the game, using $\psi = \phi_4$ and Pauli observables $\sigma_X \otimes I_2$ and $\sigma_Z \otimes I_2$ for questions X and Z respectively.

Based on the Magic Square game we devise the following "anti-commutation test".

Anti-commutation test:

(a) The referee selects $a, b \in \{0,1\}^n$ uniformly at random under the condition that $a \cdot b = 1$. He plays the Magic Square game with both players, with the following modifications: if the question to the first player is X or Z he sends (X, a) or (Z, b) instead; in all other cases he sends the original label of the question in the Magic Square game together with both strings a and b.

(b) Each player provides answers as in the Magic Square game. The referee accepts if and only if the player's answers would have been accepted in the game.

Using Lemma 8 it is straightforward to show the following.

Lemma 9 *Suppose a strategy for the players succeeds in the anti-commutation test with probability at least $1-\varepsilon$, when performed on a bipartite state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$. Then the observables $X(a)$ and $Z(b)$ applied by the player upon receipt of questions (X, a) and (Z, b) respectively satisfy*

$$\mathbb{E}_{a,b:\, a\cdot b=1} \big\|X(a)Z(b) - (-1)^{a\cdot b} Z(b)X(a)\big\|_\sigma^2 = O\big(\sqrt{\varepsilon}\big). \tag{15}$$

3.3 A robust test for high-dimensional entangled states

We are ready to state, and prove, our main theorem: a test for high-dimensional entanglement that is "robust", meaning that success probabilities that are a constant close to the optimal value suffice to certify that the underlying state is within a constant distance from the target state — in this case, a tensor product of n EPR pairs. Although arguably a direct "quantization" of the BLR result, this is the first test known which achieves constant robustness — all previous n-qubit tests required success that is inverse polynomially (in n) close to the optimum in order to provide any meaningful conclusion.

n-qubit Pauli braiding test: With probability 1/2 each,

(a) Execute the linearity test.

(b) Execute the anti-commutation test.

Theorem 10 *Suppose that a family of observables $W(a)$, for $W \in \{X, Z\}$ and $a \in \{0,1\}^n$, and a state $\psi \in \mathbb{C}^d \otimes \mathbb{C}^d$, generate outcomes that pass the n-qubit Pauli braiding test with probability at least $1-\varepsilon$. Then $d = (1 - O(\sqrt{\varepsilon}))2^n$.*

As should be apparent from the proof it is possible to state a stronger conclusion for the theorem, which includes a characterization of the observables $W(a)$ and the state ψ up to local isometries. For simplicity I only recorded the consequence on the dimension of ψ.

Proof: Using Lemma 7 and Lemma 9, success with probability $1-\varepsilon$ in the test implies that conditions (9), (10) and (11) in Corollary 5 are all satisfied, up to error $O(\sqrt{\varepsilon})$. (In fact, Lemma 9 only implies (11) for strings a, b such that $a \cdot b = 1$. The condition for string such that $a \cdot b = 0$ follows from the other conditions.) The conclusion of the corollary is that there exists an isometry V such that the observables $X(a)$ and $Z(b)$ satisfy

$$\mathbb{E}_{a,b} \big\| X(a)Z(b) - V^* g(\sigma_X(a)\sigma_Z(b))V \big\|_\sigma^2 = O(\sqrt{\varepsilon}).$$

Using again the consistency relations (9) that follow from part (a) of the test together with the above we get

$$\mathbb{E}_{a,b}\, \psi^*(V\otimes V)^*\big(\sigma_X(a)\sigma_Z(b) \otimes \sigma_X(a)\sigma_Z(b)\big)(V\otimes V)\psi = 1 - O(\sqrt{\varepsilon}).$$

Applying Lemma 6, $(V \otimes V)\psi$ has Schmidt rank at least $(1 - O(\sqrt{\varepsilon}))2^n$. But V is a local isometry, which cannot increase the Schmidt rank. □