I. RIGIDITY FOR THE CHSH GAME

Let us sketch the proof of the single-game CHSH rigidity theorem for the case that the devices' Hilbert spaces $\mathcal{H}_A$ and $\mathcal{H}_B$ are finite dimensional, and $\epsilon = 0$, i.e., the devices achieve the maximum correlation allowed by Tsirelson's inequality. Hilbert space completeness allows for truncating an infinite-dimensional space to finitely many dimensions, at an arbitrarily small cost.

A general strategy for the devices consists of measuring some shared state in $\mathcal{H}_A \otimes \mathcal{H}_B$. Since the success probability is extremal, i.e., $\epsilon = 0$, we may assume that the state is extremal, i.e., is pure. Each device measures its system using a reflection that depends on Eve's question, and returns the sign of the observed eigenvalue ± 1. The shared state $|\psi\rangle$ and Alice and Bob's four reflections R_0^A, R_1^A, R_0^B and R_1^B determine the strategy.

Jordan's Lemma[35] states that any two reflections acting on a finite-dimensional space can be simultaneously block-diagonalized into 1×1 and 2×2 blocks. (The same statement is false for infinite-dimensional spaces.) Apply the lemma to R_0^A and R_1^A, to obtain

$$R_a^A = \bigoplus_i R_a^A(i) \ , \tag{1}$$

where i labels the block index, and each $R_a^A(i)$ is a 1×1 or 2×2 reflection. By adding placeholder dimensions, we may assume without loss of generality that each block is 2×2. Eq. (1), which can equivalently be rewritten $R_a^A = \sum_i R_a^A(i) \otimes |i\rangle\langle i|$, gives a basis in which $\mathcal{H}_A = \mathbf{C}^2 \otimes \mathcal{H}'_A$, where $\mathcal{H}'_A$ is the Hilbert space with orthonormal basis $\{|i\rangle\}$. Thus Jordan's Lemma gives (an extension of) the a priori formless space $\mathcal{H}_A$ a tensor-product structure. It locates within $\mathcal{H}_A$ a qubit $\mathbf{C}^2$. However, Alice's operators need not act locally on this qubit; a controlled rotation is still needed to align her operators. After this rotation, we will show that Alice's strategy is close to the ideal CHSH game strategy that measures this qubit using the operators given in Fig. 1 in the main text.

Since the measurement of the block index commutes with both R_a^A, we may assume that Alice measures the block index first. Thus we reduce to the case that $\mathcal{H}_A = \mathbf{C}^2$, and, by a symmetrical argument, that $\mathcal{H}_B = \mathbf{C}^2$, still with $\epsilon = 0$.

If the R_α^D reflections act on $\mathbf{C}^2$ and are not equal to $\pm I$, then we can choose a basis such that $R_0^D = \sigma_z = \left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$, $R_1^A = \left(\begin{smallmatrix} \cos 2\theta & \sin 2\theta \\ \sin 2\theta & -\cos 2\theta \end{smallmatrix}\right)$ and $R_1^B = \left(\begin{smallmatrix} \cos 2\theta' & \sin 2\theta' \\ \sin 2\theta' & -\cos 2\theta' \end{smallmatrix}\right)$ for certain angles $\theta, \theta' \in [0, \frac{\pi}{2}]$. (As this basis depends on the block index, changing basis amounts to a controlled qubit rotation.) Letting $M_a = \frac{1}{2}(R_0^A + (-1)^a R_1^A) \otimes I - \frac{1}{\sqrt{2}} I \otimes R_a^B$ for $a \in \{0, 1\}$, the success probability satisfies

$$\begin{aligned} 2\sqrt{2} - 8\epsilon &\leq 8 \Pr[AB = X \oplus Y] - 4 \\ &= \langle\psi|\Big(\sum_{a,b \in \{0,1\}} (-1)^{ab} R_a^A \otimes R_b^B\Big)|\psi\rangle \\ &= 2\sqrt{2} - \sqrt{2}\langle\psi|(M_0^2 + M_1^2)|\psi\rangle \ . \end{aligned}$$

For $\epsilon = 0$, this means that $|\psi\rangle$ must lie in the intersection of the kernels of M_0 and M_1. The four eigenvalues of M_0 are $\pm \cos\theta \pm \frac{1}{\sqrt{2}}$. For the kernel to be nonempty, it must be that $\theta = \frac{\pi}{4}$. A symmetrical argument implies that $\theta' = \frac{\pi}{4}$. For small $\epsilon > 0$, $|\psi\rangle$ must lie close to small-eigenvalue subspaces of both M_0 and M_1, implying that θ and θ' are close to $\frac{\pi}{4}$. Thus the measurement operators are rigidly determined.

For $\theta = \theta' = \frac{\pi}{4}$, the kernel of $\sqrt{2}((HG) \otimes I)M_0((G^\dagger H) \otimes I) = \sigma_z \otimes I - I \otimes \sigma_z$ is spanned by the vectors $|00\rangle$ and $|11\rangle$. The kernel of $\sqrt{2}((HG) \otimes I)M_1((G^\dagger H) \otimes I) = \sigma_x \otimes I - I \otimes \sigma_x$ is spanned by the vectors $|+\rangle \otimes |+\rangle = \frac{1}{2}(|00\rangle + |01\rangle + |10\rangle + |11\rangle)$ and $|-\rangle \otimes |-\rangle = \frac{1}{2}(|00\rangle - |01\rangle - |10\rangle + |11\rangle)$. For the $|01\rangle$ and $|10\rangle$ terms to cancel out, a linear combination of these vectors must have equal coefficients. The intersection between the two kernels is therefore spanned by $|00\rangle + |11\rangle$. Thus the state $|\psi\rangle$ is rigidly determined.

The above argument conveys much of the intuition for the CHSH rigidity theorem. The case $\epsilon > 0$ can be handled by maintaining suitable approximations. However, we have not explained the derivation of the operators M_0 and M_1, chosen to satisfy $\sum_{a,b \in \{0,1\}} (-1)^{ab} R_a^A \otimes R_b^B = 2\sqrt{2} I \otimes I - \sqrt{2}(M_0^2 + M_1^2)$. In general, for a game in which Eve draws her questions from the distribution $p(a, b)$ and accepts if $x \oplus y = V(a, b)$, let $\Theta = \sum_{a,b} p(a, b)(-1)^{V(a,b)} |a\rangle\langle b|$ and $\hat{\Theta} = \left(\begin{smallmatrix} 0 & \Theta \\ \Theta^\dagger & 0 \end{smallmatrix}\right)$. Let ω^* be the optimal success probability. By the Tsirelson semi-definite program,[41] the optimal bias is $2\omega^* - 1 = \frac{1}{2} \max_{\Gamma \succeq 0, \Gamma \circ I = I} \langle\hat{\Theta}, \Gamma\rangle = \frac{1}{2} \min_{\Delta = \Delta \circ I \succeq \hat{\Theta}} \operatorname{Tr} \Delta$. Γ is the Gram matrix of the vectors $R_a^A|\psi\rangle$ and $R_b^B|\psi\rangle$. Letting Δ^* achieve the second optimum, we have $\frac{1}{2}\langle\hat{\Theta}, \Gamma\rangle = (2\omega^* - 1) - \frac{1}{2}\langle\Delta^* - \hat{\Theta}, \Gamma\rangle$. For the CHSH game, $\Delta^* = \frac{1}{2\sqrt{2}}\mathbf{1}$, and the matrices M_0, M_1 correspond to eigenvectors of $\Delta^* - \hat{\Theta}$.

II. RIGIDITY FOR SEQUENTIAL CHSH GAMES

For sequentially repeated CHSH games, our goal is to locate in $\mathcal{H}_A$ and $\mathcal{H}_B$ not one but many qubits, in tensor product, such that the devices' actual strategy is close to the ideal strategy that measures these qubits one at a time in sequence. In this section, we will introduce the notation and put together the theorems needed to prove rigidity for sequential CHSH games.

A. Notation

To make our claims precise, we begin with some notation for CHSH games played in sequence, one following the next, with no communication between games.

A strategy $\mathcal{S}$ for Alice and Bob to play n sequential CHSH games consists of the devices' Hilbert spaces, initial shared state and the reflections they use to play each game.

Initial state: Let $\mathcal{H}_A$ and $\mathcal{H}_B$ be Alice and Bob's respective Hilbert spaces, and $\mathcal{H}_C$ any external space. Let $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B \otimes \mathcal{H}_C$ be the devices' initial shared state.

Transcripts: Denote questions asked to Alice by $a_1, \ldots, a_n$, questions asked to Bob by $b_1, \ldots, b_n$, and possible answers by $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$, respectively. Write $h_j^A = (a_1, x_1, \ldots, a_j, x_j)$, $h_j^B = (b_1, y_1, \ldots, b_j, y_j)$ and $h_j = (h_j^A, h_j^B)$, a full transcript for games 1 through j. Write $h_{j,k}$ and $h_{j,k}^D$ for the full or partial transcripts for games j through k, inclusive.

Reflections: In game j, for questions a_j and b_j, let $R_{a_j}^A(h_{j-1}^A)$ and $R_{b_j}^B(h_{j-1}^B)$ be the reflections specifying Alice and Bob's respective strategies for game j, depending on the previous games' transcript. Define projections $P_j^A(h_j^A) = \frac{1}{2}(\mathbf{1}+(-1)^{x_j} R_{a_j}^A(h_{j-1}^A))$ and $P_j^B(h_j^B) = \frac{1}{2}(\mathbf{1} + (-1)^{y_j} R_{b_j}^B(h_{j-1}^B))$. For $D \in \{A, B\}$ and $j \leq k$, let $P_{j,k}^D(h_k^D) = P_k^D(h_k^D) \cdots\ P_{j+1}^D(h_{j+1}^D) P_j^D(h_j^D)$. Let $P_{j,k}^{AB}(h_k) = P_{j,k}^A(h_k^A) \otimes P_{j,k}^B(h_k^B)$.

Super-operators: For $j < k$ and partial transcript h_j, define super-operators $\mathcal{E}_k^{A|h_j^A}$ and $\mathcal{E}_k^{B|h_j^B}$ by

$$\begin{aligned}\mathcal{E}_k^{D|h_j^D}&(|h_{j+1,k-1}^D\rangle\langle h_{j+1,k-1}^D| \otimes \rho) \\ &= \frac{1}{2} \sum_{\alpha_k, \chi_k} |h_{j+1,k}^D\rangle\langle h_{j+1,k}^D| \otimes P_k^D(h_k^D) \rho P_k^D(h_k^D)\ ,\end{aligned} \tag{2}$$

where $h_k^D = (h_{k-1}^D, \alpha_k, \chi_k)$. These super-operators capture the effects of Alice and Bob playing game k, where games $j + 1$ to $k - 1$ of the transcript are stored in a separate register. For $\ell \geq k$, let $\mathcal{E}_{k,\ell}^{D|h_j^D} = \mathcal{E}_\ell^{D|h_j^D} \cdots \mathcal{E}_{k+1}^{D|h_j^D} \mathcal{E}_k^{D|h_j^D}$ and $\mathcal{E}_{k,\ell}^{AB|h_j} = \mathcal{E}_{k,\ell}^{A|h_j^A} \otimes \mathcal{E}_{k,\ell}^{B|h_j^B}$.

Let $\rho_1 = |\psi\rangle\langle\psi|$, $\rho_j(h_{j-1}) = P_{1,j-1}^{AB}(h_{j-1}) \rho_1 P_{1,j-1}^{AB}(h_{j-1})^\dagger \ / \operatorname{Tr}(P_{1,j-1}^{AB}(h_{j-1})\rho_1)$ be the state at the beginning of game j conditioned on h_{j-1}, and $\rho_j = \mathcal{E}_{1,j-1}^{AB}(\rho_1) = \frac{1}{4^{j-1}} \sum_{h_{j-1}} |h_{j-1}\rangle\langle h_{j-1}| \otimes P_{1,j-1}^{AB}(h_{j-1}) \rho_1 P_{1,j-1}^{AB}(h_{j-1})^\dagger$. Compare to Eq. (1) in the main text.

For fixed Hilbert spaces, a strategy $\mathcal{S}$ can be identified with the tuple $(\rho_1, \{\mathcal{E}_j^A\}, \{\mathcal{E}_j^B\})$. When considering multiple strategies, we will decorate this notation to indicate the corresponding strategy, e.g., $\tilde{\mathcal{S}} = (\tilde{\rho}_1, \{\tilde{\mathcal{E}}_j^A\}, \{\tilde{\mathcal{E}}_j^B\})$.

Call a strategy for a single CHSH game ϵ-structured if the probability of winning is at least $\omega^* - \epsilon/8$. In our theorem, we will assume that most games the devices play are ϵ-structured, in the following sense:

Definition II.1 (Structured strategy)**.** *A sequential CHSH game strategy $\mathcal{S}$ is* (δ, ϵ)-structured *if for every j, there is at least a $1 - \delta$ probability over transcripts h_{j-1} that game (j, h_{j-1}) is ϵ-structured. $\mathcal{S}$ is* ϵ-structured *if it is (ϵ, ϵ)-structured.*

We aim to show that the devices play nearly ideally:

Definition II.2 (Ideal strategy)**.** *A strategy $\mathcal{S}$ for n sequential CHSH games is an* ideal strategy *if there exist isometries $\mathcal{I}^D : \mathcal{H}_D \hookrightarrow (\mathbf{C}^2)^{\otimes n} \otimes \mathcal{H}'_D$ and a state $|\psi'\rangle \in \mathcal{H}'_A \otimes \mathcal{H}'_B \otimes \mathcal{H}_C$ such that for every j and h_{j-1},*

$$\begin{aligned}\mathcal{I}^A \otimes \mathcal{I}^B |\psi\rangle &= |\varphi\rangle^{\otimes n} \otimes |\psi'\rangle \\ R_\alpha^D(h_{j-1}^D) &= \mathcal{I}^{D\dagger} (R_\alpha^D)_j \mathcal{I}^D\ ,\end{aligned} \tag{3}$$

where $(R_\alpha^D)_j$ denotes the ideal reflection operator R_α^D from Fig. 1 in the main text acting on the jth qubit.

In order to compare strategies, define a notion of simulation:

Definition II.3 (Strategy simulation)**.** *Let $\mathcal{S}$ and $\tilde{\mathcal{S}}$ be two strategies for playing n sequential CHSH games. For $\epsilon \geq 0$, we say that strategy $\tilde{\mathcal{S}}$* ϵ-simulates *strategy $\mathcal{S}$ if they both use the same Hilbert spaces and for all j,*

$$\max_{D \in \{A,B\}} \|\mathcal{E}_{1,j}^D(\rho_1) - \tilde{\mathcal{E}}_{1,j}^D(\tilde{\rho}_1)\|_{\text{tr}} \leq \epsilon\ . \tag{4}$$

Say that $\tilde{\mathcal{S}}$ weakly ϵ-simulates *$\mathcal{S}$ if only the weaker inequality $\|\mathcal{E}_{1,j}^{AB}(\rho_1) - \tilde{\mathcal{E}}_{1,j}^{AB}(\tilde{\rho}_1)\|_{\text{tr}} \leq 2\epsilon$ holds.*

It is also convenient to allow a basis change by local unitaries or local isometries:

Definition II.4. *A strategy $\tilde{\mathcal{S}}$ is an* isometric extension *of $\mathcal{S}$ if there exist isometries $\mathcal{X}^D : \mathcal{H}_D \hookrightarrow \tilde{\mathcal{H}}_D$ such that $|\tilde{\psi}\rangle = \mathcal{X}^A \otimes \mathcal{X}^B |\psi\rangle$ and $\mathcal{X}^D R_\alpha^D(h_{j-1}^D) = \tilde{R}_\alpha^D(h_{j-1}^D)\mathcal{X}^D$ always. (Thus $\mathcal{X}^D \mathcal{E}_j^D = \tilde{\mathcal{E}}_j^D \mathcal{X}^D$.)*

B. Main rigidity theorem and proof outline

Our main theorem states that a structured strategy can be closely simulated by an ideal strategy:

Theorem II.5 (Rigidity theorem for sequential CHSH games)**.** *There exists a constant κ such that for any ϵ-structured strategy $\mathcal{S}$ for n sequential CHSH games, there exists an ideal strategy $\hat{\mathcal{S}}$ that $\kappa n^\kappa \epsilon^{1/\kappa}$-simulates an isometric extension of $\mathcal{S}$.*

The first step of the proof of Theorem II.5 is to replace the structured strategy $\mathcal{S}$ with one in which the devices play every game using the ideal CHSH game operators on some qubit, up to a local change in basis. See Fig. 1.

Definition II.6 (Single-qubit ideal strategy)**.** *A strategy $\mathcal{S}$ is a* single-qubit ideal strategy *if there exist unitaries $U_j^D(h_{j-1}^D) : \mathcal{H}_D \overset{\cong}{\rightarrow} \mathbf{C}^2 \otimes \mathcal{H}'_D$ such that always*

$$R_\alpha^D(h_{j-1}^D) = U_j^D(h_{j-1}^D)^\dagger (R_\alpha^D \otimes \mathbf{1}) U_j^D(h_{j-1}^D) \ . \tag{5}$$

That is, each device's reflections for game (j, h_{j-1}^D) are equivalent up to local unitaries to the ideal CHSH game reflections, although the qubits used need not be in tensor product.

Theorem II.7. *There exists a constant κ such that if $\mathcal{S}$ is an ϵ-structured strategy for n sequential CHSH games, then there is a single-qubit ideal strategy $\tilde{\mathcal{S}}$ that weakly $\kappa n^\kappa \epsilon^{1/\kappa}$-simulates an isometric extension of $\mathcal{S}$.*

Proof sketch. As explained in the main text, let $\tilde{\mathcal{E}}_j^D$ be the super-operator that replaces the device's measurement operators with the ideal operators promised by the CHSH rigidity theorem. Then $\tilde{\mathcal{S}} = (\rho_1, \{\tilde{\mathcal{E}}_j^A\}, \{\tilde{\mathcal{E}}_j^B\})$. If $\Pr[\text{game } j \text{ is } \epsilon\text{-structured}] \geq 1 - \delta$, then $\|\mathcal{E}_j^{AB}(\rho_j) - \tilde{\mathcal{E}}_j^{AB}(\rho_j)\|_{\mathrm{tr}} \leq 2\delta + O(\sqrt{\epsilon})$. (This expression combines bounds on the probability of the bad event and the $O(\sqrt{\epsilon})$ error from the good event.)

To show our goal, that $\mathcal{E}_{1,n}^{AB}(\rho_1) \approx \tilde{\mathcal{E}}_{1,n}^{AB}(\rho_1)$ in trace distance, use a hybrid argument that works backwards from game n to game 1 fixing each game's measurement operators one at a time. The error introduced from fixing a game j, by moving from $\mathcal{E}_j^{AB}(\rho_j)$ to $\tilde{\mathcal{E}}_j^{AB}(\rho_j)$, does not increase in later games because applying a super-operator cannot increase the trace distance. Mathematically, this hybrid argument is simply a triangle inequality using the expansion

$$\mathcal{E}_{1,n}^{AB}(\rho_1) - \tilde{\mathcal{E}}_{1,n}^{AB}(\rho_1) = \sum_{j \in [n]} \tilde{\mathcal{E}}_{j+1,n}^{AB}\big(\mathcal{E}_j^{AB}(\rho_j) - \tilde{\mathcal{E}}_j^{AB}(\rho_j)\big) \ .$$

□

Next, we find a nearby strategy in which the qubits for successive games are in tensor product.

Definition II.8 (Multi-qubit ideal strategy)**.** *A strategy $\mathcal{S}$ is a* multi-qubit ideal strategy *if there is a unitary isomorphism $\mathcal{Y}^D : \mathcal{H}_D \overset{\cong}{\rightarrow} (\mathbf{C}^2)^{\otimes n} \otimes \mathcal{H}'_D$ under which for unitaries $M_j^D(h_{j-1}^D) \in \mathcal{L}((\mathbf{C}^2)^{\otimes(n-j+1)} \otimes \mathcal{H}'_D)$ such that*

$$\begin{aligned} R_\alpha^D(h_{j-1}^D) = \quad & \mathcal{Y}^{D\dagger} M_1^{D\dagger} \ldots \\ & \big(\mathbf{1}_{\mathbf{C}^2}^{\otimes(j-1)} \otimes M_j^D(h_{j-1}^D)^\dagger\big)(R_\alpha^D)_j \big(\mathbf{1}_{\mathbf{C}^2}^{\otimes(j-1)} \otimes M_j^D(h_{j-1}^D)\big) \\ & \ldots M_1^D \mathcal{Y}^D \ . \end{aligned} \tag{6}$$

That is, $\mathcal{S}$ is a single-qubit ideal strategy in which the qubits used in each game must lie in tensor product with the qubits from previous games.

Theorem II.9. *There exists a constant κ such that if $\tilde{\mathcal{S}}$ is an ϵ-structured single-qubit ideal strategy for n sequential CHSH games, then there is a multi-qubit ideal strategy $\bar{\mathcal{S}}$ that weakly $\kappa n^\kappa \epsilon^{1/\kappa}$-simulates an isometric extension of $\tilde{\mathcal{S}}$.*

Proof sketch. The tensor-product structure is constructed beginning with a trivial transformation on $\tilde{\mathcal{S}}$: to each device, add n ancilla qubits each in state $|0\rangle$. Next, after a qubit has been measured, say as $|\alpha_j\rangle$ in game j, swap it with the jth ancilla qubit, then rotate this fresh qubit from $|0\rangle$ to $|\alpha_j\rangle$ and continue playing games $j+1, \ldots, n$. This defines a unitary change of basis that places the outcomes for games 1 to j in the first j ancilla qubits, and leaves the state in the original Hilbert space unchanged. Since qubits are set aside after being measured, the qubits for later games are automatically in tensor product with those for earlier games; the resulting strategy $\bar{\mathcal{S}}$ is multi-qubit ideal. At the end of the n games, swap back the ancilla qubits and undo their rotations, using the transcript.

The key to showing that $\bar{\mathcal{S}}$ is close to $\tilde{\mathcal{S}}$ is the fact that operations on one half of an EPR state can equivalently be performed on the other half, since for any 2×2 matrix M, $(M \otimes I)(|00\rangle + |11\rangle) = (I \otimes M^T)(|00\rangle + |11\rangle)$. This means that the outcome of an ϵ-structured CHSH game would be nearly unchanged if Bob were hypothetically to perform Alice's measurement before his own. By moving Alice's measurement operators for games $j+1$ to n over to Bob's side, we see that they cannot significantly affect the qubit $|\alpha_j\rangle$ from game j on her side. Therefore, undoing the original change of basis restores the ancilla qubits nearly to their initial state $|0^n\rangle$, and $\bar{\mathcal{S}} \approx \tilde{\mathcal{S}}$.

Formally, define a unitary super-operator $\mathcal{V}_j$ that rotates the jth ancilla qubit to $|\alpha_j\rangle$, depending on Alice's local transcript h_j^A. Define a unitary super-operator $\mathcal{T}_j$ to apply $\mathcal{V}_j$ and swap the jth ancilla qubit with the qubit Alice uses in game j (depending on h_{j-1}^A). Alice's multi-qubit ideal strategy is given by

$$\bar{\mathcal{E}}_j^A = \mathcal{T}_{1,j-1}^{-1}(\mathbf{1}_{\mathbf{C}^{2n}} \otimes \tilde{\mathcal{E}}_j^A)\mathcal{T}_{1,j-1} \ . \tag{7}$$

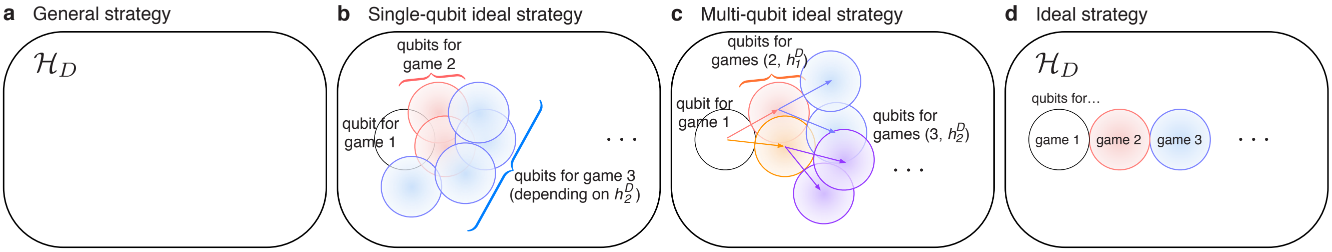

FIG. 1: **Proof outline for Theorem II.5.** **a**, Initially, each device $D \in \{A, B\}$ can play arbitrarily, measuring in game j one of two reflections $R^D_\alpha(h^D_{j-1})$, $\alpha \in \{0, 1\}$, that depend on the local transcript h^D_{j-1} for the previous games. No structure is given for the Hilbert space $\mathcal{H}_D$. **b**, We first show that D's strategy is close to a "single-qubit ideal strategy," in which for every game it measures some qubit using the ideal CHSH game strategy, but the qubit locations can be arbitrary. Here, the qubits are illustrated schematically as balls, and the overlaps indicate that they need not be in tensor product. **c**, We then construct a nearby "multi-qubit ideal strategy," in which the qubits used in each game must lie in tensor product with the qubits from previous games, but can overlap qubits used along other transcripts. **d**, Finally, we argue that the qubit locations cannot depend significantly on the transcript, and therefore that the original strategy is well-approximated by an ideal strategy that measures a fixed set of n qubits in sequence. (Note that these visualizations, representing qubits as balls, are inherently imprecise. A qubit's location in a Hilbert space is given not by a ball, but by the two anti-commuting reflection operators σ_x and σ_z.)

We aim to show that the strategy given by ρ_1, $\{\tilde{\mathcal{E}}^A_j\}$ and $\{\tilde{\mathcal{E}}^B_j\}$ is close to $\tilde{\mathcal{S}}$ up to the fixed isometry that prepends $|0^n\rangle\langle 0^n|$ to the state, i.e., that $|0^n\rangle\langle 0^n| \otimes \tilde{\mathcal{E}}^{AB}_{1,n}(\rho_1) \approx \bar{\mathcal{E}}^A_{1,n}\big(|0^n\rangle\langle 0^n| \otimes \tilde{\mathcal{E}}^B_{1,n}(\rho_1)\big)$. Define a super-operator $\tilde{\mathcal{F}}^{AB}_j$, in which Alice's measurements are made on *Bob's* Hilbert space $\mathcal{H}_B$, on the qubit determined by Bob's local transcript h^B_{j-1}. Since most games are ϵ-structured, by the CHSH rigidity theorem, $\tilde{\mathcal{F}}^{AB}_{j+1,k}(\tilde{\rho}_{j+1}) \approx \tilde{\mathcal{E}}^{AB}_{j+1,k}(\tilde{\rho}_{j+1}) = \tilde{\rho}_{k+1}$ for any $j \leq k$. Since $\tilde{\mathcal{F}}^{AB}_{j+1,k}$ acts on $\mathcal{H}_B$, it does not affect Alice's qubit $|\alpha_j\rangle$ from game j at all, and so this qubit must stay near $|\alpha_j\rangle$ in $\tilde{\rho}_{k+1}$ as well, i.e., the trace of the reduced density matrix against the projection $|\alpha_j\rangle\langle\alpha_j|$ stays close to one. As this holds for every j, $\mathcal{T}^{-1}_{1,n}$ indeed returns the ancillas almost to their initial state $|0^n\rangle$.

In more detail, let X_j be the operator that projects onto Alice's jth ancilla qubit and the qubit she uses in the jth game being $|0\rangle \otimes |\alpha_j\rangle$. By definition, $\operatorname{Tr}(X_j\, \tilde{\rho}_{j+1}) = 1$. By the Gentle Measurement Lemma,[42,43] it suffices to show that $\operatorname{Tr}(X_j\, \tilde{\rho}_{k+1}) = \operatorname{Tr} X_j \tilde{\mathcal{E}}^{AB}_{j+1,k}(\tilde{\rho}_{j+1}) \approx 1$. This is not obvious; since the operators for games $j+1$ to k do not act in tensor product, they can disturb the qubit measured in game j. However, since a super-operator on $\mathcal{H}_B$ cannot affect the expectation of an operator supported on $\mathcal{H}_A$, we find

$$\begin{aligned}\operatorname{Tr}(X_j\, \tilde{\rho}_{k+1}) &= \operatorname{Tr} X_j \tilde{\mathcal{E}}^{AB}_{j+1,k}(\tilde{\rho}_{j+1}) \\ &\approx \operatorname{Tr} X_j \tilde{\mathcal{F}}^{AB}_{j+1,k}(\tilde{\rho}_{j+1}) \\ &= \operatorname{Tr}(X_j\, \tilde{\rho}_{j+1}) \\ &= 1 \ .\end{aligned}$$

A symmetrical argument adjusts Bob's super-operators $\{\tilde{\mathcal{E}}^B_j\}$ to $\{\bar{\mathcal{E}}^B_j\}$, implying that $\bar{\mathcal{S}}$ weakly simulates $\tilde{\mathcal{S}}$. □

The last major step in the proof of Theorem II.5 is to argue that the qubit locations cannot depend significantly on the local transcripts, and therefore simulate a multi-qubit ideal strategy with an ideal strategy.

Theorem II.10. *There exists a constant κ such that if $\bar{\mathcal{S}}$ is an ϵ-structured multi-qubit ideal strategy for n sequential CHSH games, then there is an ideal strategy that weakly $\kappa n^\kappa \epsilon^{1/\kappa}$-simulates $\bar{\mathcal{S}}$.*

Proof sketch. Fix a transcript $\hat{h}_n$, chosen at random from the distribution of transcripts for $\bar{\mathcal{S}}$. Define a strategy $\hat{\mathcal{S}}$ to use the same initial state $\bar{\rho}_1$ as $\bar{\mathcal{S}}$ and the qubits specified by $\hat{h}_n$ in $\bar{\mathcal{S}}$, i.e., $\hat{R}^D_\alpha(h^D_{j-1}) = \bar{R}^D_\alpha(\hat{h}^D_{j-1})$ independent of h^D_{j-1}. Thus, as required in Definition II.2, $\hat{R}^D_\alpha(h^D_{j-1}) = \hat{\mathcal{I}}^{D\dagger}(R^D_\alpha)_j \hat{\mathcal{I}}^D$ for

$$\hat{\mathcal{I}}^D = \big(\mathbf{1}_{(\mathbf{C}^2)^{\otimes(n-1)}} \otimes \bar{M}^D_n(h^D_{n-1})\big) \ldots \bar{M}^D_1 \bar{\mathcal{Y}}^D \ .$$

We argue that $\hat{\mathcal{S}}$ closely approximates $\bar{\mathcal{S}}$, provided that $\hat{h}_n$ satisfies: for every j, conditioned on the partial transcript $\hat{h}_{j-1}$, (a) game j is ϵ-structured, and (b) there is a high probability that every subsequent game is ϵ-structured. By Markov inequalities, most transcripts satisfy these conditions.

We connect $\bar{\mathcal{S}}$ to $\hat{\mathcal{S}}$ by an argument that one game at a time switches play to locate qubits according to $\hat{h}_n$. The intermediate steps relate strategies in which the devices locate their qubits using a hybrid $(\hat{h}_j, h_{j+1,n})$ of $\hat{h}_n$ and the actual transcript h_n.

Consider a partial transcript h_j that differs from $\hat{h}_j$ only in the jth game, say on Alice's side. By (a) and the CHSH rigidity theorem, Alice's jth qubit is collapsed and nearly in tensor product with the rest of the state. Therefore, there exists a unitary V_j^A acting on this qubit such that

$$\bar{\rho}_{j+1}(h_j) \approx V_j^A \bar{\rho}_{j+1}(\hat{h}_j) V_j^{A\dagger} \ , \qquad (8)$$

up to error $O(\sqrt{\epsilon})$. Since applying a super-operator cannot increase trace distance and on Bob's side $h_j^B = \hat{h}_j^B$, therefore

$$\bar{\mathcal{F}}_{j+1,n}^{AB|h_j^B}\big(\bar{\rho}_{j+1}(h_j)\big) \approx V_j^A \bar{\mathcal{F}}_{j+1,n}^{AB|\hat{h}_j^B}\big(\bar{\rho}_{j+1}(\hat{h}_j)\big) V_j^{A\dagger} \ .$$

Here, $\bar{\mathcal{F}}_{j+1,n}^{AB|h_j^B}$ is the same super-operator used in the multi-qubit ideal strategy simulation step—that plays Alice's games on Bob's qubits—except conditioned on the local transcript h_j^B. By condition (b), these super-operators can be pulled back to Alice's side, to give

$$\bar{\mathcal{E}}_{j+1,n}^{AB|h_j}\big(\bar{\rho}_{j+1}(h_j)\big) \approx V_j^A \bar{\mathcal{E}}_{j+1,n}^{AB|\hat{h}_j}\big(\bar{\rho}_{j+1}(\hat{h}_j)\big) V_j^{A\dagger} \ .$$

Note that this approximation does not follow immediately from Eq. (8), because Alice's super-operators conditioned on h_j^A can be very different from her super-operators conditioned on $\hat{h}_j^A$.

By fixing the coordinates one at a time in this way, we find that for a typical transcript h_n, $\bar{\rho}_{n+1}(h_n) \approx V_{1,n}^{AB} \bar{\rho}_{n+1}(\hat{h}_n) V_{1,n}^{AB\dagger}$, and we conclude that $\bar{\mathcal{E}}_{1,n}^{AB}(\rho_1) \approx \hat{\mathcal{E}}_{1,n}^{AB}(\rho_1)$.

Since $\hat{\mathcal{E}}_{1,n}^{AB}$ measures qubits in tensor product with each other, by using the CHSH rigidity theorem one last time, it is not difficult to show that $\hat{\mathcal{E}}_{1,n}^{AB}(\rho_1) \approx \hat{\mathcal{E}}_{1,n}^{AB}(\hat{\rho}_1)$, where $\hat{\rho}_1$ has n EPR pairs in the qubit positions determined by $\hat{h}_n$. Thus $\bar{\mathcal{S}}$ is weakly simulated by the ideal strategy given by $(\hat{\rho}_1, \{\hat{\mathcal{E}}_j^A\}, \{\hat{\mathcal{E}}_j^A\})$. □

The last three theorems chain together via:

Lemma II.11. *Let $\mathcal{S}$ be a (δ, ϵ)-structured strategy for n sequential CHSH games. If $\tilde{\mathcal{S}}$ is a strategy that weakly η-simulates $\mathcal{S}$, then $\tilde{\mathcal{S}}$ is $(\delta + 2\sqrt{\eta}, \epsilon + 16\sqrt{\eta})$-structured.*

This lemma follows immediately from the definitions.

The conclusion from these theorems is that the devices' joint strategy is close to ideal: $\mathcal{E}_{1,n}^{AB}(\rho_1) \approx \hat{\mathcal{E}}_{1,n}^{AB}(\hat{\rho}_1)$. This weak simulation statement is not strong enough for our applications, in which sometimes Eve plays CHSH games with only one of the two devices. To prove Theorem II.5, we need to show a simulation statement, i.e., that the devices' strategies are *separately* close to ideal: $\mathcal{E}_{1,n}^{A}(\rho_1) \approx \hat{\mathcal{E}}_{1,n}^{A}(\hat{\rho}_1)$ and $\mathcal{E}_{1,n}^{B}(\rho_1) \approx \hat{\mathcal{E}}_{1,n}^{B}(\hat{\rho}_1)$. These estimates cannot be obtained directly because our main assumption, that every game j is usually ϵ-structured, is only of use if both devices have played games 1 through $j-1$—it gives information about $\mathcal{E}_j^D$ applied to $\mathcal{E}_{1,j-1}^{AB}(\rho_1)$, not about $\mathcal{E}_j^D$ applied to $\mathcal{E}_{1,j-1}^{D}(\rho_1)$. The key idea to obtain separate estimates is that applying both devices' super-operators is almost equivalent to applying Alice's super-operator, *guessing* Bob's measurement outcome from the ideal conditional distribution, and based on the guess applying a controlled unitary correction to his qubit. Since Alice's super-operator collapses both qubits of the EPR state, it is not actually necessary to measure Bob's qubit. Defining $\mathcal{G}_j^B$ to be this guess-and-correct super-operator, two hybrid arguments give $\mathcal{E}_{1,n}^{AB}(\rho_1) \approx \mathcal{G}_{1,n}^{B} \mathcal{E}_{1,n}^{A}(\rho_1)$ and $\tilde{\mathcal{E}}_{1,n}^{A} \mathcal{E}_{1,n}^{B}(\rho_1) \approx \mathcal{G}_{1,n}^{B} \tilde{\mathcal{E}}_{1,n}^{A}(\rho_1)$. Thus,

$$\mathcal{G}_{1,n}^{B} \mathcal{E}_{1,n}^{A}(\rho_1) \approx \mathcal{G}_{1,n}^{B} \tilde{\mathcal{E}}_{1,n}^{A}(\rho_1) \ .$$

The same super-operator $\mathcal{G}_{1,n}^B$ appears on both the left- and right-hand sides above. In general, applying a super-operator can reduce the trace distance. In this case, however, it does not; the correction part of $\mathcal{G}_{1,n}^B$ is unitary, and the guessing part is a stochastic map acting on a copy of Alice's classical transcript register. Therefore, indeed $\mathcal{E}_{1,n}^{A}(\rho_1) \approx \tilde{\mathcal{E}}_{1,n}^{A}(\rho_1)$. In general, this same argument implies:

Lemma II.12. *There exists a contant κ such that if $\mathcal{S} = (\rho_1, \{\mathcal{E}_j^A\}, \{\mathcal{E}_j^B\})$ is an ϵ-structured strategy that is weakly δ-simulated by $\tilde{\mathcal{S}} = (\rho_1, \{\tilde{\mathcal{E}}_j^A\}, \{\mathcal{E}_j^B\})$, a strategy differing only in Alice's reflection operators, then $\tilde{\mathcal{S}}$ also $\kappa n^\kappa (\delta + \epsilon)^{1/\kappa}$-simulates $\mathcal{S}$.*

This lemma suffices to strengthen Theorems II.7 and II.9 from weak simulation to simulation statements. The argument to strengthen Theorem II.10, and therefore conclude Theorem II.5, is similar, but more involved.

C. Rigidity theorem based on observed statistics

For $\zeta > 0$, call a strategy for n sequential CHSH games ζ-*ideal* if an isometric extension of the strategy is ζ-simulated by an ideal strategy. Theorem II.5 states that any ϵ-structured strategy is $\kappa n^\kappa \epsilon^{1/\kappa}$-ideal. Using a simple martingale argument, the structure assumption can be justified based on observed statistics:

Theorem II.13. *Let Alice and Bob play in sequence N sets each of n sequential CHSH games. Let $W \le Nn$ be the total number of games that Alice and Bob win. Fix $\epsilon > 0$, and let $G \le N$ be the number of sets of games for which the provers' joint strategy for that set, conditioned on the previous games' outcomes, is $\kappa n^\kappa \epsilon^{1/\kappa}$-ideal, where κ is the constant from Theorem II.5. Let $\eta > 0$. Then for any δ such that $t = \frac{1}{8}\epsilon^2 \eta N - \delta N n \ge 0$,*

$$\Pr\big[W \ge (\omega^* - \delta)Nn,\ G < (1-\eta)N\big] \le e^{-\frac{t^2}{2Nn}} \ . \qquad (9)$$

Informally, this theorem says that if the devices do not use a nearly ideal strategy for most subsequences of games, then they are unlikely to win too many games. It follows that if the devices have a high probability of winning many games—close to $\omega^* Nn$ games—then the strategy at the beginning of a random subsequence of n games is very likely to be $\kappa n^\kappa \epsilon^{1/\kappa}$-ideal.

III. VERIFIED QUANTUM DYNAMICS PROOF SKETCHES

A. *XZ*-determined states

In the CHSH game, each device has two measurement settings, that in the ideal strategy may be identified with Pauli σ_x and σ_z operators. For carrying out tomography, however, it is generally necessary to be able to measure in the σ_y basis as well. It is possible to extend the CHSH game to one in which the ideal strategy also uses σ_y operators on a shared EPR state. However, this extended game will not satisfy the rigidity property. The problem is that a device that consistently measures using $-\sigma_y$ will give indistinguishable statistics from one that uses $+\sigma_y$. (Switching the sign of σ_y is equivalent to taking an entry-wise complex conjugate in the computational basis.) It is impossible to fix the sign of the σ_y operator.

We therefore instead argue that for certain states, reliable tomography can be accomplished without needing to measure in the σ_y basis. This observation is due to McKague[38] and was suggested earlier by Magniez et al.[27] McKague shows that for $|\varphi\rangle = \frac{1}{\sqrt{2}}(|00\rangle + |11\rangle)$, an EPR pair, the states $(I \otimes U)|\varphi\rangle$, for any single-qubit real unitary U, and $\mathrm{CNOT}_{24}|\varphi\rangle_{12} \otimes |\varphi\rangle_{34}$, as well as finite tensor products of these states, are exactly determined by their traces against tensor products of I, σ_x and σ_z operators. That is, they are determined by the expectations of observables that can be estimated using measurements in the σ_x and σ_z bases. We call such states "*XZ*-determined." For our applications, we will need to show a larger class of states to be *XZ*-determined. However, characterizing the full set of *XZ*-determined states remains an open problem.

Definition III.1. *For a Hilbert space $\mathcal{H}$, a set of operators $S \subseteq \mathcal{L}(\mathcal{H})$ and $d > 0$, a state $\tau \in \mathcal{L}(\mathcal{H})$ is* determined by S with exponent d *if there exists $c > 0$ such that for all $\epsilon \geq 0$ and any state $\rho \in \mathcal{L}(\mathcal{H})$,*

$$\max_{P \in S} |\mathrm{Tr}\, P(\rho - \tau)| \leq \epsilon \quad \Longrightarrow \quad \|\rho - \tau\|_{\mathrm{tr}} \leq c\,\epsilon^d \ . \quad (10)$$

For $\mathcal{H} = (\mathbf{C}^2)^{\otimes n}$, a state τ is XZ-determined *if it is determined with some exponent $d > 0$ by the Pauli operators $\{I, \sigma_x, \sigma_z\}^{\otimes n}$.*

Robustness is important for applications, but previous work has considered only $\epsilon = 0$. By Łojasiewicz's inequality[44] (Prop. 2.3.11) in algebraic geometry, robustness follows from the $\epsilon = 0$ case:

Lemma III.2. *For a finite-dimensional Hilbert space $\mathcal{H}$, a state $\tau \in \mathcal{L}(\mathcal{H})$ is determined by a finite set $S \subset \mathcal{L}(\mathcal{H})$ if and only if for any state $\rho \in \mathcal{L}(\mathcal{H})$, the implication of Eq. (10) holds at $\epsilon = 0$.*

Recall that a *stabilizer state* is an n-qubit pure state $|\psi\rangle$ for which there exists a set of 2^n distinct and pairwise commuting operators $S \subset \{\pm P : P \in \{I, \sigma_x, \sigma_y, \sigma_z\}^{\otimes n}\}$, the *stabilizer group*, such that $P|\psi\rangle = |\psi\rangle$ for all $P \in S$.[45] Any set of n operators that generate the stabilizer group S are called stabilizer generators for $|\psi\rangle$. We prove:

Theorem III.3. *If $|\psi\rangle \in (\mathbf{C}^2)^{\otimes n}$ is a stabilizer state that has a set of stabilizer generators in $\{I, \sigma_x, \sigma_z\}^{\otimes n}$, and if U is the tensor product of any n single-qubit real unitaries, then $U|\psi\rangle$ is XZ-determined.*

In particular, the resource states needed in our verified quantum dynamics protocol, $|0\rangle \otimes (I \otimes H)|\varphi\rangle \otimes (I \otimes G)|\varphi\rangle \otimes \mathrm{CNOT}_{2,4}(|\varphi\rangle \otimes |\varphi\rangle)$, are *XZ*-determined, as are the same states with G applied transversally. This latter consideration is important because Alice's ideal measurement bases in the CHSH game are rotated by $\pi/4$ from σ_x and σ_z.

The proof of Theorem III.3 begins by showing easily that $|0\rangle$ is determined by $\{\sigma_z\}$. Then apply three closure properties. First, closure under tensor product for pure states implies that $|0\rangle^{\otimes n}$ is determined by the n operators $\sigma_z \otimes I^{\otimes(n-1)}, \ldots, I^{\otimes(n-1)} \otimes \sigma_z$. Second, observe by manipulating implication (10) that if τ is a state determined by S, then for any unitary U, $U\tau U^\dagger$ is determined with the same exponent by $\{UPU^\dagger : P \in S\}$. Since an arbitrary stabilizer state can be generated by applying Clifford operators to $|0\rangle^{\otimes n}$, this implies that a stabilizer state is determined by any of its sets of stabilizer generators. Furthermore, if τ is determined by $S = \{P_1, \ldots, P_s\}$, then for any invertible $s \times s$ matrix V, τ is determined with the same exponent by $\{\sum_{j=1}^s V_{ij} P_j : i = 1, \ldots, s\}$. This implies, e.g., that any *XZ*-determined state is also determined by the operators $\{I, \frac{1}{\sqrt{2}}(\sigma_z \pm \sigma_x)\}^{\otimes n}$, since the $\{I, \sigma_x, \sigma_z\}^{\otimes n}$ coefficients are functions of the $\{I, \frac{1}{\sqrt{2}}(\sigma_z \pm \sigma_x)\}^{\otimes n}$ coefficients.

B. State tomography

The state tomography protocol begins with $(K-1)m$ CHSH games, where K is chosen randomly. The multi-game rigidity theorem implies that at the beginning of the Kth block of m, with high probability Alice and Bob share a state that is close to m shared EPR states, possibly in tensor product with an additional state, and that their separate measurement strategies for the next

m games are close to the ideal strategy that uses one EPR state at a time. For the analysis of the state tomography protocol, we may therefore assume that Alice's strategy is exactly ideal and restrict consideration to these m EPR states.

Bob, on the other hand, does not play more CHSH games, but instead is given by Eve a random permutation of the m indices, and is asked to permute his qubits and prepare many copies of a particular resource state. (In the main text, this state is specified as $|\psi\rangle = |0\rangle \otimes (I \otimes H)|\varphi\rangle \otimes (I \otimes G)|\varphi\rangle \otimes \mathrm{CNOT}_{2,4}(|\varphi\rangle \otimes |\varphi\rangle)$. The $|0\rangle$ portion is for preparing the initial states in a computation and implementing the final measurements, and the other subsystems are for teleporting into each of the gates in a universal gate set, e.g., $(I \otimes G)|\varphi\rangle$ for teleporting into G. However, after teleporting into G, an H correction may or may not be required. To maintain the blindness property of the protocol, i.e., to avoid leaking any information about the computation to the separate devices, it is of technical use to have available the resource state $|\psi\rangle \otimes |\varphi\rangle$, where the extra EPR state is used for teleporting into the identity gate when an H correction is not needed.)

Note that Bob's reduced density matrix is maximally mixed, so the probability that he can measure the correct 11-qubit resource state is only $1/2^{11}$. However, since the states $\{(P \otimes I)|\varphi\rangle : P \in \{I, \sigma_x, \sigma_y, \sigma_z\}\}$ form an orthonormal basis, so too do the states

$$P^{(0)}|0\rangle \otimes (P^{(1)} \otimes I)|\varphi\rangle \otimes (P^{(2)} \otimes H)|\varphi\rangle \otimes (P^{(3)} \otimes G)|\varphi\rangle \otimes (P_1^{(4)} \otimes P_3^{(5)} \otimes \mathrm{CNOT}_{2,4})(|\varphi\rangle_{12} \otimes |\varphi\rangle_{34}) \ , \tag{11}$$

where $P^{(0)} \in \{I, \sigma_x\}$ and the other $P^{(j)}$ vary over $\{I, \sigma_x, \sigma_y, \sigma_z\}$. Any of the states in Eq. (11) are equally useful resources for computation by teleportation, as Eve can adjust for the $P^{(j)}$ operators in her classical Pauli frame.[46]

Therefore define an ideal state tomography protocol as one in which Alice and Bob's initial state consists of m shared EPR states, possibly in tensor product with an additional state; Alice plays honestly m CHSH games, directed by Eve; and Eve sends Bob a random m-item permutation and requests that he return the results of measuring consecutive 11-qubit blocks of permuted qubits in the basis of Eq. (11). Eve rejects if the tomography statistics returned by Alice are inconsistent with Bob's reported outcomes. More precisely, she rejects if the fraction of times Bob reports any particular state differs from $1/2^{11}$ by more than $\sqrt{(\log m)/m}$, i.e., about $\sqrt{\log m}$ standard deviations, or if for any state any of its $\{I, X, Z\}^{\otimes 11}$ Pauli coefficients differ from the corresponding observed estimates by more than $\sqrt{(\log m)/m}$. We show:

Theorem III.4. *In an ideal state tomography protocol, if Alice plays honestly, then:*

Completeness: *If Bob plays honestly, then Eve accepts with high probability.*

Soundness: *If Eve accepts with high probability, then there is a high probability that, after Bob and before Alice's play, for most of the consecutive 11-qubit subsystems, Alice's reduced state on the subsystem is close to the state in Eq. (11) that Bob reported to Eve.*

In these statements, "with high probability" means with probability inverse polynomially close to one, i.e., at least $1 - 1/m^c$, where the exact exponents are adjustable.

The completeness property of the protocol is a trivial application of Hoeffding's inequality; the probability of straying by more than $\sqrt{\log m}$ standard deviations is at most $\exp(-\Omega(\log m)) = m^{-\Omega(1)}$. The soundness property is also mostly a straightforward tomography argument, using that the states in Eq. (11) are all XZ-determined. The main technical complication is that the states of Alice's 11-qubit blocks need not be in tensor product. Therefore, her measurement results on different blocks need not be independent. They can be controlled with a suitable martingale.

First fix a permutation $\sigma \in S_m$ and a string $x \in (\{0,1\}^{11})^{m/11}$ such that, conditioned on Eve sending Bob σ and receiving back x, Eve accepts with high probability. The remaining randomness consists of Alice's measurement bases and results. Since Alice's measurements commute, we may assume that she measures her qubits in the permuted order, without changing the measurement statistics.

For $j = 1, \ldots, m/11$, let ρ_j be Alice's initial reduced state on her jth block of 11 qubits. Our goal is to control most of the states ρ_j. Let σ_j be the state of the same qubits just before she begins to measure them. The state σ_j is a random variable, but is a deterministic function of the transcript $h^A_{11(j-1)}$ of the earlier CHSH games with Alice. Conditioned on $h^A_{11(j-1)}$, Alice's measurement results for games $11(j-1)+1, \ldots, 11j$ are distributed according to the Pauli coordinates of σ_j.

For each $b \in \{0,1\}^{11}$, let π_b be the corresponding state of Eq. (11), and let τ_b be the average of those σ_j for which Bob reported $x_j = b$. Using a martingale argument, we can establish that with high probability, for all b, π_b and τ_b have similar $\{I, \sigma_x, \sigma_z\}^{\otimes 11}$ Pauli coordinates. Since π_b is XZ-determined, this implies that all Pauli coordinates of π_b and τ_b are close. Since π_b is a pure state, i.e., an extremal quantum state, a Markov inequality implies that for most j with $x_j = b$, π_b and σ_j have close Pauli coordinates. Finally, average back over the transcripts to get that for most j with $x_j = b$, π_b is close to $\rho_j = \sum_{h^A_{11(j-1)}} \Pr[h^A_{11(j-1)}] \sigma_j(h^A_{11(j-1)})$. This is the claim.

Theorem III.4 says that Bob's measurement usually prepares the correct state on Alice's side—but it says nothing about the distribution of his measurement results. Since Bob's half of the shared EPR states is maximally mixed, his measurement outcomes are in fact distributed nearly uniformly on most subsystems. Thus the effect of Bob's actual super-operator is close to that of the ideal super-operator, if we trace out everything except for a random subset of subsystems on Alice's side.

C. Process tomography

Computation by teleportation uses adaptively chosen two-qubit Bell measurements on prepared resource states. The state tomography protocol gives Eve a way of ensuring that Alice's initial m-qubit state consists of the desired resource states. The Bell states $\{(P \otimes I)|\varphi\rangle : P \in \{I, \sigma_x, \sigma_y, \sigma_z\}\}$, eigenstates of the Bell measurement, are themselves XZ-determined states. A symmetrical protocol, with the roles of Alice and Bob switched, could thus be used to prepare these states in Bob's initial m-qubit state, before his play begins. However, for verified quantum computation we need a stronger conclusion, that *after* Bob's measurements collapse Alice's initial state into copies of the resource states of Eq. (11), Alice's operations indeed implement Bell measurements. This is a stronger conclusion because, as explained earlier, it is possible that dishonest operations on Alice's side—far from Bell measurements—could still collapse Bob's initial state to many copies of Bell states.

The process tomography protocol, to ensure Alice's honesty, begins in the same way as state tomography, with $(K-1)m$ CHSH games refereed between both players. The multi-game rigidity theorem again allows us to assume that the devices then share m EPR states and would play the Kth block of games honestly, according to the ideal CHSH game strategy. However, Eve only continues playing CHSH games with Bob. To Alice, Eve repeatedly sends random pairs of indices, drawn uniformly without replacement from $\{1, \ldots, m\}$, requesting the results of Bell measurements on the specified qubits. Eve does not try to draw any statistical conclusions from the reported results, but instead rejects if Alice's reported result for any pair of qubits is inconsistent with Bob's reported outcomes. For example, if Alice reports measuring the Bell state $|\varphi\rangle$ and Bob is asked to measure the same two qubits both in the σ_z basis, then the parity of Bob's measurement results should be 0. This is because $\sigma_z \otimes \sigma_z$ is a stabilizer for $|\varphi\rangle$: $(\sigma_z \otimes \sigma_z)|\varphi\rangle = |\varphi\rangle$. The four Bell states are all stabilized by $\pm\sigma_z \otimes \sigma_z$ and $\pm\sigma_x \otimes \sigma_x$, with four different sign choices. If Bob is asked to measure one of the two qubits in the σ_x basis and one in the σ_z basis, then his results for that pair do not give a check on Alice.

From Alice's perspective, the interaction is indistinguishable from the questions she would be asked in an actual computation; she is entirely unaware that Bob is playing CHSH games instead of preparing resource states.

We show:

Theorem III.5. *In the process tomography protocol, if after the first $(K-1)m$ CHSH games, Bob plays honestly on his halves of m shared EPR states, then:*

Completeness: *If Alice plays honestly, Eve accepts with probability one.*

Soundness: *If Eve accepts with high probability, then the result of Alice's super-operator applied to the initial state is close to that of applying the ideal Bell measurements super-operator to the initial state.*

Again, the completeness statement is trivial. The soundness statement is not difficult. Fix a permutation of the qubits for which Eve accepts with high probability. We prove soundness for the protocol in which Alice is given the full permutation at the beginning instead of only two indices at a time; this can only give her more opportunities to cheat.

Let $\hat{\rho}_1$ be the initial state, consisting of m EPR states. Without loss of generality, Alice's strategy consists of measuring a complete set of 2^m orthogonal projections, and returning the outcome. For $j = 1, \ldots, m/2$, let $\mathcal{G}^A_j$ be Alice's super-operator that implements the jth alleged Bell measurement, by the appropriate marginal projective measurement. For $j \leq k$, let $\mathcal{G}^A_{j,k} = \mathcal{G}^A_k \cdots \mathcal{G}^A_{j+1}\mathcal{G}^A_j$. Alice's full strategy is implemented by $\mathcal{G}^A_{1,m/2}$. Let $\hat{\mathcal{G}}^A_j$ be the ideal super-operator that actually carries out a Bell measurement on the jth specified pair of qubits, and $\hat{\mathcal{G}}^A_{j,k} = \hat{\mathcal{G}}^A_k \cdots \hat{\mathcal{G}}^A_j$. Our goal is to show that in trace distance

$$\mathcal{G}^A_{1,m/2}(\hat{\rho}_1) \approx \hat{\mathcal{G}}^A_{1,m/2}(\hat{\rho}_1) \ . \tag{12}$$

Since Eve accepts all the tests with high probability, it must be that for every j, Eve's jth Bell measurement test passes with high probability. By the Gentle Measurement Lemma, this implies that $\mathcal{G}^A_j(\hat{\rho}_1) \approx \hat{\mathcal{G}}^A_j(\hat{\rho}_1)$. For $j = 1$, this gives the first step:

$$\mathcal{G}^A_{1,m/2}(\hat{\rho}_1) \approx \mathcal{G}^A_{2,m/2}\hat{\mathcal{G}}^A_1(\hat{\rho}_1) \ .$$

We cannot immediately apply $\mathcal{G}^A_2(\hat{\rho}_1) \approx \hat{\mathcal{G}}^A_2(\hat{\rho}_1)$ to continue, because although the $\mathcal{G}^A_j$ super-operators commute with each other, $\mathcal{G}^A_2$ might not commute with $\hat{\mathcal{G}}^A_1$.

An easy trick gets around the problem. Define $\hat{\mathcal{F}}^B_j$ to implement a Bell measurement on the jth specified pair of qubits on Bob's side. Let $\hat{\mathcal{F}}^B_{j,k} = \hat{\mathcal{F}}^B_k \cdots \hat{\mathcal{F}}^B_j$. Then $\hat{\mathcal{G}}^A_j(\hat{\rho}_1) = \hat{\mathcal{F}}^B_j(\hat{\rho}_1)$. Super-operators acting on Bob's Hilbert space automatically commute with those acting on Alice's space, so we can continue the above derivation:

$$\begin{aligned}\mathcal{G}^A_{1,m/2}(\hat\rho_1) &\approx \mathcal{G}^A_{2,m/2}\hat{\mathcal{G}}^A_1(\hat\rho_1) = \hat{\mathcal{F}}^B_1\mathcal{G}^A_{2,m/2}(\hat\rho_1)\\ &\approx \hat{\mathcal{F}}^B_1\mathcal{G}^A_{3,m/2}\hat{\mathcal{G}}^A_2(\hat\rho_1) = \hat{\mathcal{F}}^B_{1,2}\mathcal{G}^A_{3,m/2}(\hat\rho_1) \approx \cdots\\ &\approx \hat{\mathcal{F}}^B_{1,m/2}(\hat\rho_1) = \hat{\mathcal{G}}^A_{1,m/2}(\hat\rho_1)\ .\end{aligned}$$

The total approximation error is linear in m.

D. QMIP = MIP*

The way in which protocols for CHSH games, state and process tomography, and computation are combined to give verified quantum dynamics is sketched in the main text. We also describe there the main remaining technical obstacle, the issue of Eve choosing her questions adaptively. Formally, let ρ be the initial state, and let $\mathcal{B}$ be the super-operator describing Eve's interactions with Bob in state tomography. Roughly, state tomography implies that the states Bob prepares on Alice's side are correct up to a small error in trace distance, or

$$\mathrm{Tr}_B\,\mathcal{B}(\rho) \approx \mathrm{Tr}_B\,\hat{\mathcal{B}}(\hat\rho)\ , \tag{13}$$

where $\hat{\mathcal{B}}$ is the ideal super-operator and $\hat\rho$ is an ideal initial state consisting of perfect EPR states. Similarly, let $\mathcal{A}$ be the super-operator describing Eve's interactions with Alice in a process tomography protocol on Alice's operations; we have

$$\mathcal{A}(\rho) \approx \hat{\mathcal{A}}(\rho)\ . \tag{14}$$

Computation by teleportation can be implemented either by choosing Bob's state preparation questions nonadaptively and Alice's process questions adaptively, or vice versa. We show that these are exactly equivalent regardless of the devices' strategies, i.e.,

$$\mathcal{A}_{\mathrm{ad}}\mathcal{B} = \mathcal{B}_{\mathrm{ad}}\mathcal{A}\ , \tag{15}$$

where $\mathcal{A}_{\mathrm{ad}}$ and $\mathcal{B}_{\mathrm{ad}}$ are the same as $\mathcal{A}$ and $\mathcal{B}$, respectively, except with Eve choosing her questions adaptively based on the previous messages. Combining these steps, we therefore obtain

$$\begin{aligned}\mathrm{Tr}_B\,\mathcal{B}_{\mathrm{ad}}\mathcal{A}(\rho) &\approx \mathrm{Tr}_B\,\mathcal{B}_{\mathrm{ad}}\hat{\mathcal{A}}(\rho)\\ &= \hat{\mathcal{A}}_{\mathrm{ad}}\,\mathrm{Tr}_B\,\mathcal{B}(\rho)\\ &\approx \hat{\mathcal{A}}_{\mathrm{ad}}\,\mathrm{Tr}_B\,\hat{\mathcal{B}}_{\mathrm{ad}}(\hat\rho)\ ,\end{aligned}$$

and thus the actual computation by teleportation protocol leaves on Alice's side nearly the ideal output.

In this supplement, we would like to highlight two points of the proof that QMIP = MIP*: the conversion into a three-round protocol, and the addition of two new provers instead of reusing provers.

QMIP is the class of languages decidable by a polynomial-time quantum verifier exchanging polynomially many quantum messages with a polynomial number of quantum provers, who have unbounded computational power and share entanglement but cannot communicate among themselves. Kempe et al. have shown that any QMIP protocol can be converted into a three-turn protocol in which the provers send a quantum message to the verifier, the verifier broadcasts the result of a random coin flip, the provers each send a second quantum message, and then the verifier applies an efficient measurement to decide whether to accept or reject.[40] Beginning with this protocol transformation, our proof adds two new provers, Alice and Bob. The classical verifier, Eve, teleports both rounds of messages from the original k provers to Alice, and then directs Alice and Bob to run the original verifier's quantum circuit.

A natural question is whether it is necessary to add two new provers or if two of the provers already present can be used for implementing verified quantum dynamics. We conjecture that adding new provers is not necessary. Broadbent et al., for example, have suggested that a k-prover QMIP protocol can be converted to a k-prover MIP* protocol deciding the same language, and the scheme that they present indeed reuses the first two provers to simulate the verifier's quantum computations.[34] However, the analysis of this scheme does not consider all ways in which provers can play dishonestly.

In fact, any scheme with a structure along the lines we have presented will be unsound—if it does not first convert to a three-turn protocol or use more sophisticated tricks. Here is a general counterexample. Begin with an arbitrary QMIP protocol $\mathcal{P}$, deciding a language L. Modify the protocol by adding one new round at the end. In this last round, the provers can each send classical messages to the verifier, which the verifier simply broadcasts back to all of the provers. The effect of this final interaction is to allow the provers to communicate with each other. (Since they share entanglement, they can use quantum teleportation to communicate quantum information, if desired.) The modified protocol $\mathcal{P}'$ decides the same language L, with the exact same completeness and soundness parameters, though; communicating in the last step does not help the provers cheat.

Convert the modified protocol, somehow, into an MIP* protocol $\mathcal{P}''$ that uses the first two provers to simulate the verifier's quantum computations. In particular, they simulate the final acceptance predicate, by some procedure that has traps to detect cheating—in our scheme, CHSH games or state or process tomography. The problem is that the last round of messages in $\mathcal{P}'$ is useless, but in $\mathcal{P}''$ the intercommunication allows the provers to reveal to each other the traps set by the verifier. This allows them to avoid the traps and cheat freely. $\mathcal{P}''$ is unsound.

Converting to a three-turn protocol at the start deflects this general attack, as does using two fresh provers to run the verifier's quantum computation. Another way to avoid the attack might be to refresh the verifier's secrets—resetting any traps and re-hiding any quantum information—before revealing any message to a prover. Any message between provers potentially carries information that affects the security of the converted MIP* protocol differently from the original QMIP protocol.

[41] Cleve, R., Slofstra, W., Unger, F., and Upadhyay, S. Perfect parallel repetition theorem for quantum XOR proof systems. *Computational Complexity* **17**(2), 282–299 (2008).

[42] Winter, A. Coding theorem and strong converse for quantum channels. *IEEE Trans. Inf. Theory* **45**(7), 2481–2485 (1999).

[43] Ogawa, T. and Nagaoka, H. Making good codes for classical-quantum channel coding via quantum hypothesis testing. *IEEE Trans. Inf. Theory* **53**(6), 2261–2266 (2007).

[44] Benedetti, R. and Risler, J.-J. *Real algebraic and semi-algebraic sets*. Actualités Mathématiques. Hermann, (1990).

[45] Nielsen, M. A. and Chuang, I. L. *Quantum computation and quantum information*. Cambridge University Press, Cambridge, (2000).

[46] Knill, E. Quantum computing with realistically noisy devices. *Nature* **434**, 39–44 (2005).