

Minimal Measurement Dependence Required to Simulate GHZ Correlations

Qihang Wang,¹ Zhiyuan Yao,² and Kun Chen^{3,*}

¹*School of Mathematical Sciences, Peking University, Beijing 100871, China*

²*Lanzhou Center for Theoretical Physics, Key Laboratory of Theoretical Physics of Gansu Province,
Key Laboratory of Quantum Theory and Applications of MoE,*

Gansu Provincial Research Center for Basic Disciplines of Quantum Physics, Lanzhou University, Lanzhou 730000, China

³*Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, China*

Local models can simulate Bell nonlocality if settings depend on hidden variables. How much dependence is needed to reproduce all outcome probabilities? We find the minimum total-variation cost for Pauli X and Y measurements on n -qubit Greenberger-Horne-Zeilinger states at any visibility and coherence phase, for every $n \geq 3$. For odd n , the stronger of the real and imaginary coherence components sets the cost; for even n , the weaker one can raise it. Mermin settings contain an even number of Y measurements, so their data miss the imaginary part and do not fix the cost in general.

Introduction.—Bell’s theorem assumes that the measurement settings are statistically independent of the variables that describe the source [1, 2]. Bell tests with settings drawn from cosmic sources or human choices go to great lengths to justify this assumption [3, 4], because local models that relax it, even partially, can reproduce quantum correlations [5–7]. This raises a quantitative question: how strongly must the hidden-variable distribution depend on the settings to reproduce a given experiment? The size of a Bell violation alone need not answer it.

Measurement dependence can be quantified by the largest total-variation (TV) distance between the hidden-variable distributions for two joint settings [5]. This distance vanishes when one distribution serves every setting and grows as a model gives up more measurement independence. For a given experiment, the required dependence is its smallest value among local models that reproduce every joint outcome probability. Matching a Bell expectation alone imposes a weaker constraint: it does not fix the full outcome statistics.

Greenberger-Horne-Zeilinger (GHZ) correlations make this distinction sharp [8]. Their Mermin violations grow exponentially with the number of parties n [9], and Mermin-type observables are used to certify GHZ entanglement from three photons [10] to large quantum processors [11]. With Pauli X and Y measurements, settings with an even number of Y measurements probe the real part of the GHZ coherence, and settings with an odd number probe its imaginary part. Changing the imaginary part therefore leaves the whole even-setting Mermin table unchanged, even when that table violates a Bell inequality. A local model of all settings, however, must reproduce both parts with the same response functions.

Previous work has solved parts of this problem. Hall constructed a local model of the perfect tripartite GHZ correlations [7]. Other studies let measurement independence fail separately for each observer [12] or quantified it by information costs for given input distributions [13]. Using the same TV measure as here, Alai showed that reproducing all outcome probabilities, rather than only the

correlators, costs nothing extra. For the even settings at perfect visibility, he proved a general lower bound, found local models that attain it up to $n = 13$, and conjectured that the resulting values, equal for $n = 2k + 1$ and $n = 2k + 2$, hold for all n [14]. These settings, however, do not probe the imaginary part of the coherence.

What additional dependence is needed to reproduce both parts at once? The Fourier criterion of Werner and Wolf and of Żukowski and Brukner determines when a single setting-independent distribution suffices [15, 16]. Outside this local region, answering the question requires both an explicit model and a proof that no model does better. For three parties, local deterministic models of arbitrary equatorial measurements are known, but their optimality was not established [17]. Here each party has one orthogonal measurement pair, and the number of parties is arbitrary.

A test on finite data must also distinguish measurement dependence from bias in the choice of inputs. Finite-data bounds on measurement dependence already exist [18], and sequential Bell tests can handle memory and imperfect inputs [19]. With memory, the relevant target is the dependence averaged over trials, each conditioned on its past, rather than a cost inferred from one pooled outcome table.

In this Letter, we determine the minimum measurement dependence required by the full X/Y experiment for every $n \geq 3$, for GHZ-subspace states of arbitrary visibility and coherence phase. We construct optimal local models and prove matching lower bounds; in particular, we prove Alai’s conjecture for all n and extend it to dephased states. For odd n , an optimal model for the stronger coherence component can reproduce the weaker one at no extra cost. For even n , the joint cost can exceed both separate costs. States with identical, already nonlocal Mermin tables can therefore have different costs. For even n this joint constraint also bounds the cost directly from measured correlations, without fitting a GHZ state or a noise model. We turn this bound into a finite-sample certificate for trials with memory and give an input-calibration protocol under stated assumptions on the setting generator.

Setting and cost measure.—Consider

$$|\text{GHZ}_n\rangle = \frac{|0\rangle^{\otimes n} + |1\rangle^{\otimes n}}{\sqrt{2}} \quad (1)$$

and let party i choose $x_i \in \{0, 1\}$, corresponding to the Pauli observables $A_i^{(0)} = X_i$ and $A_i^{(1)} = Y_i$, with outcome $a_i \in \{\pm 1\}$. For the even-parity Mermin settings $\mathcal{X}_n = \{\mathbf{x} \in \mathbb{F}_2^n : \sum_i x_i = 0\}$, define

$$q(\mathbf{x}) = \sum_{i < j} x_i x_j \pmod{2}. \quad (2)$$

Write $\chi(\mathbf{x}) = \sum_i x_i \pmod{2}$ for the setting parity. On $\mathcal{X}_n$, the dephased GHZ state has full correlator $E_v(\mathbf{x}) = v(-1)^{q(\mathbf{x})}$, $0 \leq v \leq 1$, and zero nonempty proper correlators. Its full outcome probabilities are $P_v(\mathbf{a}|\mathbf{x}) = 2^{-n}[1 + E_v(\mathbf{x}) \prod_i a_i]$ [Eq. (A2)]. Dephasing scales every full correlator by the same visibility v ; it is not a model of arbitrary laboratory noise. The full experiment below also includes odd-parity settings.

In a local deterministic model, the hidden variable λ specifies the response to either measurement at each site. Measurement dependence allows its probability distribution to change with the joint setting:

$$P(\mathbf{a}|\mathbf{x}) = \sum_{\lambda} p(\lambda|\mathbf{x}) \prod_{i=1}^n \delta_{a_i, A_i(x_i, \lambda)}, \quad (3)$$

where $A_i(x_i, \lambda) \in \{\pm 1\}$. Its departure from measurement independence is measured by

$$F = \max_{\mathbf{x}, \mathbf{y} \in \mathcal{X}_n} D_{\text{TV}}[p(\lambda|\mathbf{x}), p(\lambda|\mathbf{y})],$$

$$D_{\text{TV}}[r, s] = \frac{1}{2} \sum_{\lambda} |r(\lambda) - s(\lambda)|. \quad (4)$$

An observer who knows λ can distinguish two equally likely settings attaining F with success probability $(1 + F)/2$. A model is *faithful* if it reproduces every probability in Eq. (A2), not merely the Mermin average [14]. In Hall's notation, the measurement-dependence parameter is $M_{\text{Hall}} = 2F$ and the retained fraction is $F_{\text{Hall}} = 1 - M_{\text{Hall}}/2 = 1 - F$ [5, 7].

Other measures address different questions: bounds on setting probabilities [20], dependence for each observer separately [21], mutual information [22], or causal-network constraints [13]. Pal *et al.* consider both distances to the input-weighted average and the pairwise distance used here, and find tight bounds for a perfectly won tripartite GHZ game [18]. The objective here is to minimize the pairwise distance subject to reproducing the full outcome table.

Minimal dependence on the Mermin settings.—For every integer $n \geq 3$ and $0 \leq v \leq 1$, put $R = 2^{\lfloor (n-1)/2 \rfloor}$. The minimum of Eq. (4) over all faithful models of Eq. (A2) is

$$F_n(v) = \frac{R^2}{2(R^2 - 1)} \left(v - \frac{1}{R} \right)_+. \quad (5)$$

Here $(z)_+ = \max\{z, 0\}$. Up to $v = 1/R$, one common hidden-variable distribution suffices. Above this threshold, the minimum dependence increases linearly with visibility. For the Mermin expression and its local bound,

$$\mathcal{B}_n = \sum_{\mathbf{x} \in \mathcal{X}_n} (-1)^{q(\mathbf{x})} E(\mathbf{x}), \quad \mathcal{B}_n^L = \frac{|\mathcal{X}_n|}{R}, \quad (6)$$

the Bell ratio of the dephased table is $\beta = \mathcal{B}_n/\mathcal{B}_n^L = Rv$, so that

$$F_n(\beta) = \frac{R}{2(R^2 - 1)} (\beta - 1)_+, \quad 0 \leq \beta \leq R. \quad (7)$$

Equation (7) holds for the dephased table; for general noise, a single Bell value does not determine the outcome table.

The complete X/Y experiment.—We now retain both even and odd settings. The state is supported on the GHZ subspace spanned by $|0^n\rangle$ and $|1^n\rangle$:

$$\varrho_{z; c_0, c_1} = \begin{pmatrix} z & (c_0 - ic_1)/2 \\ (c_0 + ic_1)/2 & 1 - z \end{pmatrix}. \quad (8)$$

Positivity requires $0 \leq z \leq 1$ and $c_0^2 + c_1^2 \leq 4z(1 - z)$. Here z is a basis-state population, and c_0, c_1 are the two real components of the coherence. Rotating the orthogonal measurement pair at site i to azimuths φ_i and $\varphi_i + \pi/2$ rotates the coherence seen by the measurements. With $\Phi = \sum_i \varphi_i$, the measured components are

$$\begin{pmatrix} \bar{c}_0 \\ \bar{c}_1 \end{pmatrix} = \begin{pmatrix} \cos \Phi & \sin \Phi \\ -\sin \Phi & \cos \Phi \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \end{pmatrix}. \quad (9)$$

An oppositely oriented pair differs only by a local outcome sign. All proper equatorial correlators vanish, regardless of z , and the full correlator is

$$E_{\bar{c}_0, \bar{c}_1}(\mathbf{x}) = (-1)^{q(\mathbf{x})} \begin{cases} \bar{c}_0, & \chi(\mathbf{x}) = 0, \\ \bar{c}_1, & \chi(\mathbf{x}) = 1. \end{cases} \quad (10)$$

Every outcome probability is therefore $2^{-n}[1 + E(\mathbf{x}) \prod_i a_i]$. The experiment includes *all* 2^n settings, not only those with an even number of Y measurements. These measurements do not detect z . Writing the initial coherence as $c_0 + ic_1 = ve^{i\theta}$, they detect its phase relative to the axes, $\theta_{\text{eff}} = \theta - \Phi$.

Let $F_n^{XY}(c_0, c_1)$ be the minimum measurement dependence needed to reproduce this table. We call it the cost and use c_0, c_1 for the measured components; stronger means larger in magnitude. The maximization in Eq. (4) is replaced by $\mathbf{x}, \mathbf{y} \in \mathbb{F}_2^n$. As before, the settings may be correlated with λ in an arbitrary way. We write the baseline cost in Eq. (5) as $C_R(t) = (R^2 t - R)_+/[2(R^2 - 1)]$, with $R > 1$. For every odd $n \geq 3$, put $R = 2^{(n-1)/2}$. Then

$$F_n^{XY}(c_0, c_1) = C_R(\max\{|c_0|, |c_1|\}). \quad (11)$$

For physical visibility v and effective phase θ_{eff} , the argument is $v \max\{|\cos \theta_{\text{eff}}|, |\sin \theta_{\text{eff}}|\}$. To obtain this cost, start with a model whose two coherence components are equal. Auxiliary random signs, chosen independently of the settings, then reduce either component without increasing the distance between distributions. Applying Eq. (5) to the even settings and, after a relabeling, to the odd settings gives the matching lower bound (Appendix B).

For every $n \geq 3$, the full real axis has the baseline cost $F_n^{XY}(v, 0) = C_R(v)$, with $R = 2^{\lfloor (n-1)/2 \rfloor}$. For even $n \geq 4$, equal magnitudes instead give

$$F_n^{XY}(t, t) = C_{2R}(t), \quad R = 2^{(n-2)/2}. \quad (12)$$

Outcome sign flips that depend only on the setting parity give the same cost for all four sign choices $(\pm t, \pm t)$. One common hidden-variable distribution suffices in the odd-party square $\max\{|c_0|, |c_1|\} \leq 1/R$ and the even-party diamond $|c_0| + |c_1| \leq 1/R$, intersected with the physical disk. These local regions follow from the Fourier criterion [15, 16]; the question is the minimum dependence outside them.

Cost of both components for even parties.—Write $c_> = \max\{|c_0|, |c_1|\}$ and $c_< = \min\{|c_0|, |c_1|\}$. For every even $n \geq 4$, put $R = 2^{(n-2)/2}$, $K_R = R^2 + 2R - 1$, and $d_R = (2R + 1)(R + 3)(R - 1)$. The complete answer is

$$F_n^{XY} = \max\left\{0, C_R(c_>), \frac{R^2(c_> + c_<) - R}{4R^2 - 1}, \frac{RK_R c_> + (R^2 - 1)c_< - K_R}{d_R}\right\}. \quad (13)$$

The four-party case is

$$F_4^{XY} = \max\left\{0, \frac{2c_> - 1}{3}, \frac{4(c_> + c_<) - 2}{15}, \frac{14c_> + 3c_< - 7}{25}\right\}. \quad (14)$$

The last term is essential even for physical states: at $(c_0, c_1) = (4/5, 2/5)$ it gives 27/125, whereas the other positive bounds give only 1/5.

The four terms describe different constraints on a local model. Put $v_* = (R^2 + R + 2)/[2(R + 1)^2]$ and $t_* = (Rc_> - 1)/(R - 1)$. Table I gives the controlling regimes. At fixed $c_> \geq 1/R$, adding a small second component need not increase the cost. Once $c_<$ exceeds $v_* t_*$, reproducing both components requires extra dependence, governed by the last term. Above t_* , the term involving their sum is largest. Rotating the coherence phase at fixed visibility, $(c_>, c_<) = v(\cos \vartheta, \sin \vartheta)$ with $0 \leq \vartheta \leq \pi/4$ after relabeling, changes both components; each site's two measurements remain orthogonal.

The extra cost comes from using the same local response functions for both setting parities, not from a

TABLE I. Which constraint determines the even-party cost in Eq. (13). Here $c_>$ and $c_<$ are the larger and smaller coherence magnitudes, $v_* = (R^2 + R + 2)/[2(R + 1)^2]$, and $t_* = (Rc_> - 1)/(R - 1)$. The active bound is respectively zero, the single-component bound, the joint bound, or the sum bound. Regimes overlap where bounds tie. Physical states also require $c_>^2 + c_<^2 \leq 4z(1 - z)$.

Regime	Condition
Local	$c_> + c_< \leq 1/R$
Single component	$c_> \geq 1/R, \quad c_< \leq v_* t_*$
Joint constraint	$c_> \geq 1/R, \quad v_* t_* \leq c_< \leq t_*$
Sum constraint	$c_> + c_< \geq 1/R, \quad c_< \geq t_*$

larger Mermin violation. For even n , exchanging X and Y at every site maps each setting to a complementary setting of the same parity. In any local deterministic model, the full products at complementary settings differ by the same sign in both parity classes, whereas the GHZ signs require opposite signs, so a model serving both can need more dependence than either class alone.

An explicit model at $(c_>, c_<) = (1, v_*)$ attains the crossover cost. A weighted comparison of setting pairs gives the matching lower bound. Mixtures of six models then attain the cost in every regime (Appendix C). Some of these building blocks are valid classical outcome tables outside the quantum disk; we do not claim that they are quantum states.

Reduction to full products.—The even-setting construction provides the baseline for the full experiment. Its main steps follow; the extensions to both setting parities are proved in Appendixes B and C. On $\mathcal{X}_n$, the product of the local deterministic responses in Eq. (3) is a sign times the parity of a subset of setting bits,

$$\prod_i A_i(x_i, \lambda) = (-1)^{\phi + \gamma \cdot \mathbf{x}}, \quad (\phi, \gamma) \in \mathbb{F}_2 \times \mathbb{F}_2^n. \quad (15)$$

Grouping assignments by the label (ϕ, γ) cannot increase any TV distance. Spreading the weight of each group evenly over its 2^{n-1} assignments preserves every distance and sets all proper correlators to zero [14], so only full products matter. On even settings, γ and $\gamma + \mathbf{1}$ give the same product but remain separate groups (Appendix A). The GHZ signs $(-1)^{q(\mathbf{x})}$ are quadratic in the setting bits, with bilinear form

$$B(\mathbf{x}, \mathbf{y}) = q(\mathbf{x} + \mathbf{y}) + q(\mathbf{x}) + q(\mathbf{y}) \\ = \mathbf{x} \cdot \mathbf{y} + \chi(\mathbf{x})\chi(\mathbf{y}), \quad \chi(\mathbf{x}) = \sum_i x_i. \quad (16)$$

Construction and optimality.—All nonzero Fourier coefficients of the sign pattern $(-1)^{q(\mathbf{x})}$ have the same magnitude. A setting-independent mixture μ of the corresponding assignments therefore reproduces the table at visibility $1/R$. Mixing with uniform outcomes gives every $v \leq 1/R$ at zero cost. At perfect visibility, $\nu_{\mathbf{x}}$ instead

gives equal weight to those assignments that match the most GHZ signs and agree with the sign at setting $\mathbf{x}$, as in the ansatz of Ref. [14]. Counting their overlaps for every n gives cost $R/[2(R+1)]$. For even n , complementary settings select identical assignments; this is why $n = 2k + 1$ and $n = 2k + 2$ have the same cost (Appendix A). For $v \geq 1/R$, use

$$p(\lambda|\mathbf{x}) = (1 - \eta)\mu(\lambda) + \eta\nu_{\mathbf{x}}(\lambda), \quad \eta = \frac{Rv - 1}{R - 1}. \quad (17)$$

The common component contributes nothing to the distance, so

$$F_n(v) \leq \eta F_n(1) = \frac{R(Rv - 1)}{2(R^2 - 1)}. \quad (18)$$

For the lower bound, choose R^2 settings on which the correlators of any deterministic assignment, weighted by the GHZ signs, sum to at most R . Replacing the conditional distributions on these settings by their average changes this sum by at most $2(R^2 - 1)F$. Thus $R^2v \leq R + 2(R^2 - 1)F$, matching Eq. (18); Appendix A gives the proof. At $v = 1$ this reduces to Alai's overlap bound [14].

Physical consequences.—For every even $n \geq 4$, take GHZ-subspace states with $z = 1/2$ and coherence components $(c_0, c_1) = (t, 0)$ and (t, t) , where $t = 1/R$; both are valid states since $2/R^2 \leq 1$. Their Mermin tables coincide and sit at the local bound, but the costs of the full experiment differ:

$$F_n^{XY}(t, 0) = 0, \quad F_n^{XY}(t, t) = \frac{R}{4R^2 - 1} > 0. \quad (19)$$

For four particles these values are 0 and $2/15$. Above threshold, four-party states with $(4/5, 0)$ and $(4/5, 2/5)$ have identical even-setting tables and Mermin ratio $8/5 > 1$, but their full costs are $1/5$ and $27/125$, respectively. The last term in Eq. (14) detects a difference absent from their Mermin data. The cost therefore depends on the state, the measured axes and which settings are included, not on a single Bell expectation.

This joint constraint can also be tested without assuming a GHZ state. For arbitrary even-party correlators, define the sign-weighted averages over even ($e = 0$) and odd ($e = 1$) settings,

$$S_e = 2^{-(n-1)} \sum_{\chi(\mathbf{x})=e} (-1)^{q(\mathbf{x})} E(\mathbf{x}), \quad e = 0, 1. \quad (20)$$

Every setting enters S_e with the same weight, however often it is chosen. Any common hidden-variable distribution obeys $RK_R S_0 + (R^2 - 1)S_1 \leq K_R$; exceeding this limit requires measurement dependence of at least

$$F \geq \frac{RK_R S_0 + (R^2 - 1)S_1 - K_R}{d_R}. \quad (21)$$

Here F compares all pairs of settings, and the proper correlators need not vanish (Appendix C); at $n = 4$ the bound is $(14S_0 + 3S_1 - 7)/25$. For general data it is only a lower bound; for GHZ-subspace states it is tight in the joint-constraint regime of Table I, up to sign flips and an exchange of the two parity classes.

Finite-sample certification.—Fix the bound and its sign convention before recording N_s trials. Write $a = RK_R/d_R$, $b = (R^2 - 1)/d_R$, $k = K_R/d_R$, and $L = a + b$. Choose an even setting with total probability a/L and an odd setting with probability b/L , uniformly within each class. Call this setting distribution $q_*(\mathbf{x})$; λ may still depend on the settings. Each trial contributes the signed product $Z_t = (-1)^{q(\mathbf{x}_t)} \prod_i a_{i,t} \in \{\pm 1\}$. Let F_t be the largest TV distance between the hidden-variable distributions of trial t for different settings, given its observed past. Write $\bar{F}_{N_s} = N_s^{-1} \sum_t F_t$ and $\bar{Z} = N_s^{-1} \sum_t Z_t$. If, given any past record, the settings are drawn from q_* , then with probability at least $1 - \delta_s$,

$$\bar{F}_{N_s} \geq \left[L\bar{Z} - k - L\sqrt{\frac{2\ln(1/\delta_s)}{N_s}} \right]_+. \quad (22)$$

The hidden variables and responses may depend on the past and may drift, but not on future settings; this per-trial model differs from the block-wise measurement dependence of Ref. [23]. The bound concerns the average of F_t , not each trial or a pooled table. Appendix D gives the proof and a version valid for data-dependent stopping. Only one signed product per trial is needed, not estimates of exponentially many correlators.

Imperfect inputs require calibration [19, 24]. If every setting has nonzero probability and, given the past, the actual setting distribution is within TV distance ε_t of q_* , then Eq. (22) holds with $2L\bar{\varepsilon}$ subtracted, where $\bar{\varepsilon}$ is the mean of ε_t . The factor $2L$ cannot be reduced, even when the even and odd frequencies are correct. A distribution known before each trial can instead be reweighted, at the cost of larger fluctuations (Appendix E). For a generator that draws the parity with bounded drift and then a uniform setting of that parity, earlier input records supply the error bounds (Appendix F).

Conclusion.—We have determined the minimum measurement dependence needed to reproduce the full X/Y experiment on GHZ-subspace states of arbitrary coherence for every $n \geq 3$. Although the Mermin violation grows exponentially with n , the cost stays bounded. For real coherence $(c_0, c_1) = (v, 0)$, it approaches $v/2$ at fixed visibility $v > 0$; at $v = 1$, it approaches Alai's universal ceiling $1/2$ from below [14]. The parity of n decides how the two coherence components affect the cost. For odd n , adding the weaker component costs nothing extra. For even n , local responses fix the same relative sign between complementary settings in both parity classes, whereas the GHZ signs require opposite ones, so reproducing both

parts can require additional dependence. Measuring the imaginary part reveals differences that identical, already nonlocal Mermin tables miss.

For even n the same constraint gives a lower bound on the cost directly from measured correlations, which a single signed score turns into a finite-sample certificate for trials with memory and calibrated input errors. Extending these results to nonorthogonal measurement axes, detection loss, and other stabilizer states remains open.

Z.Y. and K.C. acknowledge support from the Strategic Priority Research Program of the Chinese Academy of Sciences under Grant No. XDB1680102. Z.Y. also acknowledges support from the National Natural Science Foundation of China under Grant No. 12247101. We acknowledge Gewu Intelligence Lab for providing the collaborative research environment in which this project was developed.

Data availability.—No experimental data were generated or analyzed. The results are analytical, and every construction and proof is given in the Appendixes. Linear programs for small n guided the constructions but enter no proof. The preceding even-setting manuscript and its source are archived at Zenodo.

Disclosure of automated assistance.—OpenAI GPT-6.1 Sol, OpenAI GPT-6 Astra, OpenAI GPT-5.6 Sol, Anthropic Claude Opus 5.5, Anthropic Claude Fable 5, and Grok 4.6 contributed to a collective workflow of analytical derivation, local-model construction, literature comparison, and manuscript preparation. Across this workflow, the systems developed and checked derivations, local models, optimality proofs, and sequential confidence and calibration bounds, checked references, and revised the physical exposition and wording.

* chenkun@itp.ac.cn

- [1] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Physics Physique Fizika* **1**, 195 (1964).
- [2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014).
- [3] J. Handsteiner, A. S. Friedman, D. Rauch, J. Gallicchio, B. Liu, H. Hosp, J. Kofler, D. Bricher, M. Fink, C. Leung, A. Mark, H. T. Nguyen, I. Sanders, F. Steinlechner, R. Ursin, S. Wengerowsky, A. H. Guth, D. I. Kaiser, T. Scheidl, and A. Zeilinger, Cosmic Bell test: Measurement settings from Milky Way stars, *Phys. Rev. Lett.* **118**, 060401 (2017).
- [4] The BIG Bell Test Collaboration, Challenging local realism with human choices, *Nature (London)* **557**, 212 (2018).
- [5] M. J. W. Hall, Local deterministic model of singlet state correlations based on relaxing measurement independence, *Phys. Rev. Lett.* **105**, 250404 (2010); Erratum, *Phys. Rev. Lett.* **116**, 219902 (2016).
- [6] J. Barrett and N. Gisin, How much measurement independence is needed to demonstrate nonlocality?, *Phys. Rev. Lett.* **106**, 100406 (2011).
- [7] M. J. W. Hall, Relaxed Bell inequalities and Kochen–Specker theorems, *Phys. Rev. A* **84**, 022102 (2011).
- [8] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, Bell’s theorem without inequalities, *Am. J. Phys.* **58**, 1131 (1990).
- [9] N. D. Mermin, Extreme quantum entanglement in a superposition of macroscopically distinct states, *Phys. Rev. Lett.* **65**, 1838 (1990).
- [10] J.-W. Pan, D. Bouwmeester, M. Daniell, H. Weinfurter, and A. Zeilinger, Experimental test of quantum nonlocality in three-photon Greenberger–Horne–Zeilinger entanglement, *Nature (London)* **403**, 515 (2000).
- [11] J. Cai, J. Huang, F. Otto, Y. Li, C. de Gois, T. Jiang, S. Cao, F. Chen, H. Fu, J. Lin, W. Xie, N. Zhou, S. Tang, X.-Y. Li, C.-Z. Peng, X. Yuan, O. Gühne, and M. Gong, Generalized Mermin inequalities for benchmarking large-scale GHZ states, arXiv:2607.23574 (2026).
- [12] S. S. Hossain, M. K. Molla, A. Kundu, B. Paul, I. Chattopadhyay, and D. Sarkar, Measurement dependence in tripartite non-locality, *Quantum Inf. Process.* **23**, 297 (2024); arXiv:2107.12495.
- [13] R. Chaves, G. Moreno, E. Polino, D. Poderini, I. Agresti, A. Suprano, M. R. Barros, G. Carvacho, E. Wolfe, A. Canabarro, R. W. Spekkens, and F. Sciarrino, Causal networks and freedom of choice in Bell’s theorem, *PRX Quantum* **2**, 040323 (2021).
- [14] A. Alai, Exact minimum measurement dependence for faithful local deterministic models of multipartite GHZ–Mermin correlations, arXiv:2608.00124 (2026).
- [15] R. F. Werner and M. M. Wolf, All-multipartite Bell-correlation inequalities for two dichotomic observables per site, *Phys. Rev. A* **64**, 032112 (2001).
- [16] M. Żukowski and Č. Brukner, Bell’s theorem for general N -qubit states, *Phys. Rev. Lett.* **88**, 210401 (2002).
- [17] A. Roy, A. Mukherjee, S. S. Bhattacharya, M. Banik, and S. Das, Local deterministic simulation of equatorial von Neumann measurements on tripartite GHZ state, *Quantum Inf. Process.* **14**, 217 (2015).
- [18] P. Pal, G. Sharma, and R. Podila, Loophole-Robust Certification of Quantum Advantage, arXiv:2607.13090 (2026).
- [19] D. Elkouss and S. Wehner, (Nearly) optimal P values for all Bell inequalities, *npj Quantum Inf.* **2**, 16026 (2016).
- [20] G. Pütz, D. Rosset, T. J. Barnea, Y.-C. Liang, and N. Gisin, Arbitrarily small amount of measurement independence is sufficient to manifest quantum nonlocality, *Phys. Rev. Lett.* **113**, 190402 (2014).
- [21] A. S. Friedman, A. H. Guth, M. J. W. Hall, D. I. Kaiser, and J. Gallicchio, Relaxed Bell inequalities with arbitrary measurement dependence for each observer, *Phys. Rev. A* **99**, 012121 (2019).
- [22] M. J. W. Hall and C. Branciard, Measurement-dependence cost for Bell nonlocality: Causal versus retro-causal models, *Phys. Rev. A* **102**, 052228 (2020).
- [23] E. Y.-Z. Tan, Y. Cai, and V. Scarani, Measurement-dependent locality beyond independent and identically distributed runs, *Phys. Rev. A* **94**, 032117 (2016).
- [24] J. Kofler, M. Giustina, J.-Å. Larsson, and M. W. Mitchell, Requirements for a loophole-free photonic Bell test using imperfect setting generators, *Phys. Rev. A* **93**, 032115 (2016).
- [25] J. Komiyama, E. Fouché, and J. Honda, Finite-time Analysis of Globally Nonstationary Multi-Armed Bandits, *J. Mach. Learn. Res.* **25**, 112 (2024).

Appendix A: Faithful lift and support counts

The dephased GHZ outcome probabilities. The dephased state

$$\varrho_v = v|\text{GHZ}_n\rangle\langle\text{GHZ}_n| + \frac{1-v}{2}(|0^n\rangle\langle 0^n| + |1^n\rangle\langle 1^n|), \quad 0 \leq v \leq 1, \quad (\text{A1})$$

has full correlator $v(-1)^{q(\mathbf{x})}$ on even-parity settings and zero nonempty proper equatorial correlators. Binary moment inversion gives

$$P_v(\mathbf{a}|\mathbf{x}) = 2^{-n} \left[1 + v(-1)^{q(\mathbf{x})} \prod_{i=1}^n a_i \right], \quad \mathbf{x} \in \mathcal{X}_n. \quad (\text{A2})$$

The two basis populations are equal here; Appendix B treats the general populations and the full setting domain.

Recovering individual outcomes from their product. Following Ref. [14], group all assignments of individual responses with the same deterministic full-product prediction. These groups, called fibers below, are disjoint. Uniform probabilities within a group cancel every proper correlator while keeping the full product and the TV distances. Write the deterministic responses as $A_i(0, \lambda) = (-1)^{\alpha_i}$ and $A_i(1, \lambda) = (-1)^{\alpha_i + \gamma_i}$. Here (ϕ, γ) is an unreduced label: γ and $\gamma + \mathbf{1}$ are identified only when counting Walsh characters, not when defining strategy fibers. For a fixed label $c = (\phi, \gamma)$, its fiber consists of the 2^{n-1} choices of α with $\sum_i \alpha_i = \phi$. If $p_{\mathbf{x}}(c)$ is a class distribution, the uniform lift is $\tilde{p}_{\mathbf{x}}(c, \alpha) = p_{\mathbf{x}}(c)/2^{n-1}$. Hence

$$\frac{1}{2} \sum_c \sum_{\alpha: \sum_i \alpha_i = \phi} |\tilde{p}_{\mathbf{x}}(c, \alpha) - \tilde{p}_{\mathbf{y}}(c, \alpha)| = \frac{1}{2} \sum_c |p_{\mathbf{x}}(c) - p_{\mathbf{y}}(c)|. \quad (\text{A3})$$

Flipping one output sign inside a nonempty proper subset and one outside preserves the fiber but reverses the subset product, so its mean vanishes. Coarse graining contracts total variation by the triangle inequality; together with Eq. (A3), this proves the reduction [14].

Fourier coefficients of the target signs. The Walsh transform is the discrete Fourier transform on binary setting strings. It identifies the deterministic product predictions needed for a common model and counts their agreement with the GHZ signs. All arithmetic in exponents is over $\mathbb{F}_2$. Let $V = \mathcal{X}_n$ and let $V_0 = \{\mathbf{r} \in V : B(\mathbf{r}, \mathbf{x}) = 0 \text{ for all } \mathbf{x} \in V\}$ be the radical of the polar form. For a character represented by γ ,

$$W_{\gamma} = \sum_{\mathbf{x} \in V} (-1)^{q(\mathbf{x}) + \gamma \cdot \mathbf{x}}, \quad W_{\gamma}^2 = |V| \sum_{\mathbf{r} \in V_0} (-1)^{q(\mathbf{r}) + \gamma \cdot \mathbf{r}}. \quad (\text{A4})$$

The square identity follows by changing summation variables to their difference $\mathbf{r}$; the inner character sum vanishes outside V_0 . On V , the polar form is the dot product. Its radical is trivial for odd n , and $V_0 = \{\mathbf{0}, \mathbf{1}\}$ for even n , where $\mathbf{1}$ is the all-ones vector. Thus the odd case has R^2 coefficients of magnitude R . In the even case, every nonzero coefficient has magnitude $2R$, and Parseval's identity gives exactly R^2 such coefficients. Characters are counted once per coset of $\mathbb{F}_2^n / \langle \mathbf{1} \rangle$. Fourier inversion then gives Eq. (A5), including its normalization.

Explicitly, let $f(\mathbf{x}) = (-1)^{q(\mathbf{x})}$ and count each character once in $\widehat{\mathcal{X}}_n \simeq \mathbb{F}_2^n / \langle \mathbf{1} \rangle$. The signed Fourier ensemble satisfies

$$\frac{1}{R^2} \sum_{\gamma: W_{\gamma} \neq 0} \text{sgn}(W_{\gamma}) (-1)^{\gamma \cdot \mathbf{x}} = \frac{f(\mathbf{x})}{R}. \quad (\text{A5})$$

Equal weight on these deterministic classes is setting independent and realizes the faithful table at $v = 1/R$.

Perfect-visibility models. Both parity branches below have $N = R^2$ effective labels and

$$G = \frac{N + R}{2} \quad (\text{A6})$$

supporting labels per setting. Two supports either coincide or intersect in $h = (N + 2R)/4$ labels, so their largest distance is

$$F_n(1) = 1 - \frac{h}{G} = \frac{R}{2(R + 1)}. \quad (\text{A7})$$

Each label is locally realizable, and the disjoint-fiber lift restores all proper marginals. These supports consist of best-satisfying classes, as in the ansatz evaluated through $n = 13$ in Ref. [14]. The following counts establish the normalization and all intersections for every n rather than assuming them.

For odd n , put $\mathcal{E} = V$ and $N = |\mathcal{E}| = R^2$. Write $W_0 = \sigma R$ and choose $(-1)^\delta = \sigma$. For $\mathbf{b} \in \mathcal{E}$ and $\epsilon \in \mathbb{F}_2$, take

$$\phi_{\mathbf{b}} = q(\mathbf{b}) + \delta, \quad \gamma = \mathbf{b} + \epsilon \mathbf{1}. \quad (\text{A8})$$

Agreement means $U_{\mathbf{x}}(\mathbf{b}) = (-1)^{q(\mathbf{x}+\mathbf{b})+\delta} = +1$. Translation gives $\sum_{\mathbf{b}} U_{\mathbf{x}}(\mathbf{b}) = R$. For distinct $\mathbf{x}, \mathbf{y}$, the product $U_{\mathbf{x}}U_{\mathbf{y}}$ is a constant sign times the nontrivial character $(-1)^{B(\mathbf{x}+\mathbf{y}, \mathbf{b})}$, whose sum is zero. Consequently the support and intersection sizes are

$$G = \sum_{\mathbf{b}} \frac{1 + U_{\mathbf{x}}(\mathbf{b})}{2} = \frac{N + R}{2}, \quad h = \sum_{\mathbf{b}} \frac{[1 + U_{\mathbf{x}}(\mathbf{b})][1 + U_{\mathbf{y}}(\mathbf{b})]}{4} = \frac{N + 2R}{4}. \quad (\text{A9})$$

Weight $1/(N + R)$ on each of the $2G = N + R$ supported $(\mathbf{b}, \epsilon)$ normalizes the model and gives overlap h/G .

For even n , put $d = n - 1$ and write a setting as $\mathbf{s}(\mathbf{x}) = (\mathbf{x}, \chi(\mathbf{x}))$ with $\mathbf{x} \in \mathbb{F}_2^d$. Its target bit is $q(\mathbf{x}) + \chi(\mathbf{x})$. Let $H_j = \{\mathbf{y} : \chi(\mathbf{y}) = j\}$, $\mathcal{E} = H_0$, and $N = |\mathcal{E}| = R^2$. Here q, χ, B use their defining formulas on $\mathbb{F}_2^d$, and $\mathbf{1} \in \mathbb{F}_2^d$. The parity-restricted Gauss sums obey

$$C_j = \sum_{\mathbf{y} \in H_j} (-1)^{q(\mathbf{y})}, \quad C_0 + iC_1 = (1 + i)^d, \quad |C_0| = |C_1| = R. \quad (\text{A10})$$

Indeed, a vector of weight k contributes i^k to the complex sum. Choose $(-1)^{\delta_j} = C_j/R$, put $c = \delta_0$, $u = 1 + \delta_0 + \delta_1$, and use

$$\phi_{\mathbf{b}} = q(\mathbf{b}) + c, \quad \gamma_n = \epsilon, \quad \gamma_i = b_i + u + \epsilon \quad (1 \leq i \leq d), \quad \mathbf{b} \in \mathcal{E}, \quad \epsilon \in \mathbb{F}_2. \quad (\text{A11})$$

Agreement now means $q(\mathbf{x} + \mathbf{b}) = \delta_{\chi(\mathbf{x})}$. Set $U_{\mathbf{x}}(\mathbf{b}) = (-1)^{q(\mathbf{x}+\mathbf{b})+\delta_{\chi(\mathbf{x})}}$. Since $\mathbf{x} + \mathcal{E} = H_{\chi(\mathbf{x})}$, $\sum_{\mathbf{b}} U_{\mathbf{x}}(\mathbf{b}) = R$, giving the same G . For displacement $\mathbf{a} = \mathbf{x} + \mathbf{y}$, the product $U_{\mathbf{x}}U_{\mathbf{y}}$ is a constant sign times $(-1)^{\mathbf{a} \cdot \mathbf{b}}$. The orthogonal complement of $\mathcal{E}$ is $\{\mathbf{0}, \mathbf{1}\}$. Thus the product sum vanishes for $\mathbf{a} \notin \{\mathbf{0}, \mathbf{1}\}$, and Eq. (A9) gives the same h .

For the exceptional displacement $\mathbf{a} = \mathbf{1}$, the odd dimension d implies $B(\mathbf{z}, \mathbf{1}) = 0$, so $q(\mathbf{z} + \mathbf{1}) = q(\mathbf{z}) + q(\mathbf{1})$. Translation exchanges the parity hyperplanes, yielding $C_{j+1} = (-1)^{q(\mathbf{1})}C_j$ and $\delta_{j+1} = \delta_j + q(\mathbf{1})$, with $j + 1$ understood modulo two. It follows that $U_{\mathbf{x}+\mathbf{1}} = U_{\mathbf{x}}$: the supports coincide. The same weight $1/(N + R)$ on supported $(\mathbf{b}, \epsilon)$ therefore gives distance zero for this pair and $1 - h/G$ for every generic pair. This proves the perfect-visibility upper bound for both parities.

Why a smaller cost is impossible. We compare each setting's hidden-variable distribution with the average over selected settings. The deterministic correlation bound applies to that common average. Any excess target correlation must then be supplied by changes between the conditional distributions. For odd n , select the entire even-parity setting space $\mathcal{X}_n$. For even n , select

$$\mathcal{S} = \{(\mathbf{x}, 0) : \mathbf{x} \in \mathbb{F}_2^{n-1}, \chi(\mathbf{x}) = 0\}.$$

In either case this subsystem has $N = R^2$ settings and a nondegenerate quadratic polar form. For every deterministic class c , let $g_{\mathbf{x}}(c) \in \{\pm 1\}$ be its full prediction multiplied by the target sign. The quadratic Walsh spectrum gives

$$\sum_{\mathbf{x}} g_{\mathbf{x}}(c) \leq R. \quad (\text{A12})$$

For an arbitrary visibility- v model define $\bar{p} = N^{-1} \sum_{\mathbf{x}} p_{\mathbf{x}}$. Since $\mathbb{E}_{p_{\mathbf{x}}} g_{\mathbf{x}} = v$ and $|\mathbb{E}_r g - \mathbb{E}_s g| \leq 2D_{\text{TV}}[r, s]$,

$$\begin{aligned} Nv &\leq R + 2 \sum_{\mathbf{x}} D_{\text{TV}}[p_{\mathbf{x}}, \bar{p}] \\ &\leq R + 2(N - 1)F. \end{aligned} \quad (\text{A13})$$

The last inequality uses convexity, $D_{\text{TV}}[p_{\mathbf{x}}, \bar{p}] \leq N^{-1} \sum_{\mathbf{y}} D_{\text{TV}}[p_{\mathbf{x}}, p_{\mathbf{y}}]$. Thus $F \geq (Nv - R)/[2(N - 1)]$. At $v = 1$ this is the overlap bound of Ref. [14] for this subsystem, whose maximum satisfiable count is $(N + R)/2$. Together with $F \geq 0$ and the mixture upper bound, this proves Eq. (5) for every $n \geq 3$. At perfect visibility the calibrated even-setting law gives

$$\begin{aligned} F_3 = F_4 &= \frac{1}{3}, & F_5 = F_6 &= \frac{2}{5}, \\ F_7 = F_8 &= \frac{4}{9}, & F_{13} = F_{14} &= \frac{32}{65}. \end{aligned} \quad (\text{A14})$$

Appendix B: Full-setting quadratic mechanism and phase dependence

The construction in Appendix A also works when all setting strings are included. Each signed character represents a deterministic full-product response. Randomizing the individual responses within its fiber makes every proper moment vanish without changing that product or the distances between distributions. It therefore suffices to find conditional probabilities for the signed characters. We allow abstract outcome tables with $|c_0|, |c_1| \leq 1$ in the proof; quantum states must also satisfy the positivity condition in Eq. (8).

Cost of a quadratic pattern of correlation signs. The first step is a construction for correlations whose signs are set by a binary quadratic function. The rank of its bilinear form determines the number of distinct agreement patterns and their overlap, hence the optimal TV distance. Let Q be a binary quadratic function on $\mathbb{F}_2^m$, and let its polar form have rank $2r > 0$. For the complete probability table with full correlator $t(-1)^{Q(\mathbf{x})}$, zero proper moments, and $0 \leq t \leq 1$, the minimum cost is $C_{2r}(t)$. This is a statement about probability tables and local simulation, not a quantum realization of every Q .

An affine addition to Q only relabels signed characters. Remove its constant part, and write K for the polar radical. The restriction $Q|_K$ is linear. Extend it to a linear function ℓ on the entire space. Then $Q' = Q + \ell$ vanishes on K and is constant on its cosets. A linear complement $\mathcal{E}$ of K has dimension $2r$ and nondegenerate polar form B_Q . Put $R_Q = 2^r$ and $N = |\mathcal{E}| = R_Q^2$. Squaring the quadratic character sum gives

$$W(\mathbf{b}) = \sum_{\mathbf{x} \in \mathcal{E}} (-1)^{Q'(\mathbf{x}) + B_Q(\mathbf{b}, \mathbf{x})}, \quad W(\mathbf{b})^2 = N. \quad (\text{B1})$$

Indeed, with difference variable $\mathbf{d}$, the inner sum is a nontrivial character unless $\mathbf{d}$ is in the radical, leaving only $\mathbf{d} = \mathbf{0}$. Thus $W(\mathbf{0}) = \sigma R_Q$ for $\sigma = \pm 1$. Choose $(-1)^\delta = \sigma$, and let $\mathbf{x}_\mathcal{E}$ be the projection along K . For every $\mathbf{b} \in \mathcal{E}$ use the affine prediction

$$h_{\mathbf{b}}(\mathbf{x}) = (-1)^{Q'(\mathbf{b}) + \delta + B_Q(\mathbf{x}_\mathcal{E}, \mathbf{b}) + \ell(\mathbf{x})}. \quad (\text{B2})$$

Its agreement with the target sign is $U_{\mathbf{x}}(\mathbf{b}) = (-1)^{Q'(\mathbf{x}_\mathcal{E} + \mathbf{b}) + \delta}$. The sum over labels is R_Q . For distinct projections the product of two agreement signs is a nontrivial character, whose sum is zero. Therefore their agreeing supports have sizes $G = (N + R_Q)/2$ and intersection $h = (N + 2R_Q)/4$. The uniform agreeing-support ensemble realizes $t = 1$ with worst-pair distance $1 - h/G = R_Q/[2(R_Q + 1)]$; equal projections have distance zero. The affine predictions are products of local deterministic responses.

On the full space, the Fourier transform of $(-1)^Q$ has exactly N nonzero coefficients, each of normalized magnitude $1/R_Q$. Equal weights $1/N$ on the corresponding signed characters therefore realize the target at $t = 1/R_Q$ with a common ensemble. A common zero-correlation ensemble covers smaller t by mixing. Above this threshold, mix the common and perfect ensembles on disjoint hidden-variable sectors with weight $(R_Q t - 1)/(R_Q - 1)$ on the latter. This gives $C_{R_Q}(t)$.

For the converse, restrict any model to $\mathcal{E}$. Each deterministic signed character has quadratic signed sum at most R_Q . The mean-ensemble argument in Eq. (A13) therefore gives

$$Nt \leq R_Q + 2(N - 1)F, \quad (\text{B3})$$

matching the construction. A rank-zero quadratic is affine; a common mixture of its character and its negative realizes every t with zero cost. The apparent singularity of C_1 is thus not used.

Adjusting the two components for odd parties. For the GHZ quadratic, Eq. (16) has radical $\langle \mathbf{1} \rangle$ when n is odd, hence rank $n - 1$. The rank law realizes the diagonal table (t, t) at cost $C_R(t)$, where $R = 2^{(n-1)/2}$. Put $t = \max\{|c_0|, |c_1|\}$. For $t > 0$, choose independent signs u, w with means $c_0/t, c_1/t$. These auxiliary random signs are the flags used in the construction. Each fixed flag pair multiplies the full output product by u on even settings and w on odd settings: multiply one site's outputs by u and every site's setting-1 output by w/u . The multiplier is $u(w/u)^{x(\mathbf{x})}$ and is locally implementable. Appending a common flag distribution on disjoint sectors leaves every pairwise TV distance unchanged. Averaging gives the desired components and keeps all proper moments zero. The zero case uses uniform outputs.

The even-sector restriction costs at least $C_R(|c_0|)$ by Eq. (5). Translation by an odd vector maps the odd sector to the even sector and changes q only by an affine function. Its restriction thus costs at least $C_R(|c_1|)$. Their maximum matches the flag construction, proving Eq. (11).

One nonzero component and two equal components. Adding the missing odd settings to a real-coherence model does not increase its cost. Equal components for even parties instead use the quadratic pattern on the entire setting space. To extend an optimal even-setting model on the real axis, average each class distribution with its image under

$\gamma \mapsto \gamma + 1$. Even-setting predictions are unchanged, and distances cannot increase. Every resulting ensemble has zero full prediction at every odd setting. Use their common even-setting mean for each odd setting. Its distance to each even ensemble is at most the original worst-pair distance by convexity, while odd–odd distances vanish. The faithful lift completes the outcome table. Restriction to even settings gives the matching lower bound. Swapping one site’s setting labels translates $\mathbf{x} \mapsto \mathbf{x} + \mathbf{e}_i$; the induced affine sign from $q(\mathbf{x} + \mathbf{e}_i) = q(\mathbf{x}) + B(\mathbf{x}, \mathbf{e}_i)$ exchanges the axes without changing the cost.

For even n , the full polar form is nondegenerate of rank n . The rank law on (t, t) therefore uses $2R = 2^{n/2}$, proving Eq. (12). Parity-dependent output flags remove either component sign. The same construction also gives, for general even n ,

$$\begin{aligned} F_n^{XY}(c_0, c_1) &\leq C_{2R}(\max\{|c_0|, |c_1|\}), \\ F_n^{XY}(c_0, c_1) &\geq \max\left\{C_R(|c_0|), C_R(|c_1|), \frac{R^2(|c_0| + |c_1|) - R}{4R^2 - 1}\right\}. \end{aligned} \quad (\text{B4})$$

The last lower bound averages all $4R^2$ settings: the deterministic quadratic sum is at most $2R$, whereas the target sum is $2R^2(|c_0| + |c_1|)$. These bounds agree on the axes and diagonal. The joint bound and construction in Appendix C complete the cost away from the diagonal.

From the quantum state to the measured correlations. An equatorial observable at azimuth α has upper-right entry $e^{-i\alpha}$. Taking the trace with Eq. (8) gives

$$E(\mathbf{x}) = \text{Re} \left[(c_0 - ic_1) e^{i(\Phi + |\mathbf{x}| \pi/2)} \right]. \quad (\text{B5})$$

Every nonempty proper product flips a proper subset of computational basis bits, so it has zero matrix elements within the GHZ subspace. This proves the complete table and population independence. Mixtures of preparation phases simply average the complex coherence, provided preparation is not conditioned on the later settings.

For the effective axes $\Phi = 0$, the normalized Fourier coefficients are

$$\widehat{E}(\gamma) = 2^{-n/2} \text{Re} \left[(c_0 - ic_1) e^{i\pi(n-2|\gamma|)/4} \right]. \quad (\text{B6})$$

The necessary and sufficient full-correlation locality criterion of Werner–Wolf and Żukowski–Brukner is $\sum_{\gamma} |\widehat{E}| \leq 1$ [15, 16]. The sum here is $R \max\{|c_0|, |c_1|\}$ for odd n , and $R(|c_0| + |c_1|)$ for even n . This evaluates an established criterion; it does not give the positive dependence cost by itself. Its sufficiency for our entire table follows by lifting the common signed-character ensemble. Any unused probability weight is split equally between a character and its negative, contributing zero full correlation. The resulting lift cancels all proper moments.

Appendix C: Uniform even-party optimum

Let $n = 2m \geq 4$, $R = 2^{m-1}$, $K_R = R^2 + 2R - 1$, and $d_R = (2R + 1)(R + 3)(R - 1)$. Local parity-dependent output signs and one input swap reduce the components to $0 \leq c_{<} \leq c_{>} \leq 1$. The argument covers this entire abstract-table triangle; physical GHZ-subspace states also satisfy $c_0^2 + c_1^2 \leq 4z(1 - z)$. The proof has three steps. We construct a model at the crossover point $(1, v_*)$, derive the joint lower bound by comparing distributions for pairs of settings, and mix six models to cover every parameter value. All three steps hold for arbitrary even n , without extrapolating a finite optimization.

A model at the crossover between constraints. We assign probabilities to signed deterministic classes according to the Hamming weight of $\gamma + \mathbf{x}$ modulo four and whether the predicted product agrees with the target sign. This rule is the same for every even particle number. For a signed class $c = (\phi, \gamma)$, define

$$\mathbf{r} = \gamma + \mathbf{x}, \quad k = |\mathbf{r}|, \quad s_{\mathbf{x}}(c) = (-1)^{q(\mathbf{x}) + \phi + \gamma \cdot \mathbf{x}}. \quad (\text{C1})$$

At fixed $\mathbf{x}$, the map $c \mapsto (\mathbf{r}, s_{\mathbf{x}})$ is a bijection. Put $\sigma = m \bmod 4$, $G_* = R(R + 1)$, and

$$v_* = \frac{R^2 + R + 2}{2(R + 1)^2}, \quad f_* = \frac{R}{2(R + 1)}. \quad (\text{C2})$$

For even $\mathbf{x}$, assign probability $1/G_*$ to classes with $k \equiv \sigma \pmod{4}$ and $s_{\mathbf{x}} = +1$, and zero otherwise. For odd $\mathbf{x}$, assign the per-class probabilities

$$p_{\mathbf{x}}(c) = \begin{cases} u = \frac{1}{R(R+1)^3}, & k \equiv \sigma, s_{\mathbf{x}} = +1, \\ v = \frac{R+3}{4R(R+1)^2}, & k \equiv \sigma-1, s_{\mathbf{x}} = -1, \\ w = \frac{3R+5}{4R(R+1)^2}, & k \equiv \sigma+1, s_{\mathbf{x}} = +1, \\ 0, & \text{otherwise,} \end{cases} \quad (\text{C3})$$

where residue congruences are modulo four. Summing powers of i selects precisely the weights in each residue class:

$$C_a = \#\{\mathbf{r} : |\mathbf{r}| \equiv a \pmod{4}\} = \frac{1}{4} \sum_{\ell=0}^3 i^{-a\ell} (1+i^\ell)^n = R^2 + R \cos \frac{\pi(m-a)}{2}. \quad (\text{C4})$$

Hence $C_\sigma = G_*$ and $C_{\sigma \pm 1} = R^2$. The even ensemble is normalized with signed mean one, while $G_*u + R^2(v+w) = 1$ and $G_*u + R^2(w-v) = v_*$. The faithful lift completes the outcome table, so this model realizes $(c_>, c_<) = (1, v_*)$.

To evaluate every pair, put $\mathbf{z} = \mathbf{x} + \mathbf{y}$ and $d = |\mathbf{z}|$. If $j = |\mathbf{r} \cap \mathbf{z}|$, then $k' = k + d - 2j$ and

$$\frac{s_{\mathbf{y}}}{s_{\mathbf{x}}} = (-1)^{\binom{d}{2} + j + \chi(\mathbf{x})d}. \quad (\text{C5})$$

The last term follows from the polar form and is essential across parity. The joint residue count is exactly

$$N_{ab}(d) = \frac{1}{16} \sum_{\ell, t=0}^3 i^{-a\ell - bt} (1+i^{\ell+t})^{n-d} (i^\ell + i^t)^d. \quad (\text{C6})$$

For interior even d , only $(\ell, t) = (0, 0), (2, 2)$ and the eight mixed even/odd index pairs survive. Their sum is $2^{n+1} + 16R$ when $a = b = \sigma$, and 2^{n+1} for $a, b \in \{\sigma-1, \sigma+1\}$. Thus $N_{\sigma\sigma} = R^2/2 + R$ and each adjacent-residue intersection has size $R^2/2$. For odd d , the character sum

$$\sum_{|\mathbf{r}| \equiv \sigma} (-1)^{\mathbf{r} \cdot \mathbf{z}} = \frac{1}{4} \sum_{\ell=0}^3 i^{-\sigma\ell} (1+i^\ell)^{n-d} (1-i^\ell)^d = 0$$

splits the sigma support equally between its two adjacent images. Equation (C5) preserves the selected sign on same-residue same-parity transitions and reverses it between the adjacent residues. Across parity it maps the even support to the negative sigma-minus-one and positive sigma-plus-one supports. Complementing all bits fixes the sigma residue and exchanges its neighbors.

Because $0 < u < v < w < 1/G_*$ for $R \geq 2$, these counts give Table II. The interior odd-odd value is $R/[4(R+1)] + R/[2(R+1)^3] \leq f_*$. The model therefore has cost f_* , matching the axis lower bound at $c_> = 1$.

TABLE II. TV distances between the conditional hidden-variable distributions for every setting-pair type of the model at $(1, v_*)$. The faithful lift preserves all distances. Identical settings have TV zero.

Setting parities	Hamming distance d	TV
even/even	even $0 < d < n$	$R^2/(2G_*) = f_*$
even/even	$d = n$	0
even/odd	any odd d	$1 - G_*(v+w)/2 = f_*$
odd/odd	even $0 < d < n$	$R^2(u+w-v)/2$
odd/odd	$d = n$	$R^2(w-v) = f_*$

A matching joint lower bound. Regard each setting as a graph vertex. Assign a nonnegative weight to each unordered pair, called an edge. These weights sum to one, so the weighted average of pairwise TV distances cannot exceed F . Three edge types have nonzero weights, or capacities:

$$t_E = \frac{R+1}{R^2 d_R}, \quad t_C = \frac{R-1}{2R^3 d_R}, \quad t_O = \frac{R^2-1}{R^2 d_R}. \quad (\text{C7})$$

Assign t_E to even-even edges except complement pairs, t_C to every even-odd edge, and t_O to odd-odd complement pairs. All other edges have zero capacity. Their respective numbers are $2R^2(R^2 - 1)$, $4R^4$, and R^2 , giving total capacity one.

For each signed class, call it type I when $\chi(\gamma) = m \bmod 2$, and type II otherwise. The full quadratic Walsh sum has magnitude $2R$, and shifting γ by $\mathbf{1}$ multiplies it by $(-1)^{m+\chi(\gamma)}$. Consequently, for some $\tau \in \{\pm 1\}$, the even/odd sums of $s_{\mathbf{x}}$ are $(2\tau R, 0)$ for type I and $(0, 2\tau R)$ for type II. This also follows directly by translating the full sum by $\mathbf{1}$, using $q(\mathbf{1}) = m \bmod 2$ and $B(\mathbf{x}, \mathbf{1}) = \chi(\mathbf{x})$. Moreover,

$$s_{\mathbf{x}+\mathbf{1}}/s_{\mathbf{x}} = (-1)^{m+\chi(\gamma)+\chi(\mathbf{x})}.$$

Type I thus has even sign counts $R^2 \pm \tau R$ and same-sign even complements, whereas its odd complements have opposite signs. Type II has balanced even signs and opposite-sign even complements; its odd sign counts are $R^2 \pm \tau R$.

For each deterministic class we assign a signed quantity to oriented edges. Reversing orientation reverses its sign; this is the flow below. Its divergence is the sum of outgoing values at a setting. Bounding each value by half the edge weight will turn the divergence inequalities into a TV lower bound. Define the flow as follows. On an even-even edge with opposite signs, put $f_{\mathbf{xy}} = (t_E/2)s_{\mathbf{x}}$; for equal signs put zero. For an even-to-odd edge, put $f_{\mathbf{xy}} = (t_C/2)s_{\mathbf{x}}$ for type I, and $f_{\mathbf{xy}} = -(t_C/2)s_{\mathbf{y}}$ for type II. Reverse these edges by antisymmetry. On odd-odd complement edges put $f_{\mathbf{xy}} = (t_O/2)s_{\mathbf{x}}$ for type I, and zero for type II. The type-I signs are opposite there. Every edge obeys $|f_{\mathbf{xy}}| \leq t_{\mathbf{xy}}/2$.

Direct sign counting gives the divergence at a vertex of sign s :

	even vertex	odd vertex
I	$sR^2(t_E/2 + t_C) - \tau Rt_E/2$	$st_O/2 - \tau Rt_C$
II	$s(R^2 - 1)t_E/2 - \tau Rt_C$	$sR^2 t_C$

(C8)

For all classes these dominate the demands

$$d_E(s) = sR^2(t_E/2 + t_C) - Rt_E/2, \quad d_O(s) = st_O/2 - Rt_C. \quad (C9)$$

Indeed type-I excesses are $(1-\tau)Rt_E/2$ and $(1-\tau)Rt_C$. The identities $(R-1)t_E/2 = R(R+1)t_C$ and $t_O/2 = R(R+1)t_C$ show that type-II positive signs have even excess $(1-\tau)Rt_C$ and odd excess zero. For negative signs the excesses are $(R+1)t_E/2 + (R^2 - \tau R)t_C$ and $2Rt_C$, respectively. All are nonnegative.

Multiply the demand inequalities by arbitrary normalized $p_{\mathbf{x}}(c) \geq 0$, sum, and pair oriented edges. Their expectation for the target table is

$$\begin{aligned}
& 2R^2[R^2(t_E/2 + t_C)c_{>} - Rt_E/2 + (t_O/2)c_{<} - Rt_C] \\
&= \frac{RK_R c_{>} + (R^2 - 1)c_{<} - K_R}{d_R} \\
&\leq \sum_{\{\mathbf{x}, \mathbf{y}\}, c} f_{\mathbf{xy}}(c)[p_{\mathbf{x}}(c) - p_{\mathbf{y}}(c)] \\
&\leq \sum_{\{\mathbf{x}, \mathbf{y}\}} t_{\mathbf{xy}} D_{\text{TV}}[p_{\mathbf{x}}, p_{\mathbf{y}}] \leq F.
\end{aligned} \quad (C10)$$

The inequality was proved for every deterministic class before averaging. It therefore applies to arbitrary conditional class distributions, without assuming that the model has the target's symmetries. Restriction to one sector and averaging over all settings also give $C_R(c_{>})$ and $[R^2(c_{>} + c_{<}) - R]/(4R^2 - 1)$.

For arbitrary measured outcome tables, coarse graining onto signed classes contracts TV and preserves $\sum_c p_{\mathbf{x}}(c)s_{\mathbf{x}}(c) = (-1)^{q(\mathbf{x})}E(\mathbf{x})$. The same demand expectation uses the signed parity averages S_0, S_1 of Eq. (20), without assuming they are constant within a sector. This proves Eq. (21) even for nonzero proper marginals. These are equal-weight averages of conditional correlators, not averages over setting-source frequencies. No sign/order convention for S_e is needed for this raw lower inequality.

Models attaining the bound throughout the triangle. We attain the joint constraint and the other three terms by mixing explicit models at six points:

$$O = (0, 0), \quad L = (1/R, 0), \quad T = (1/(2R), 1/(2R)), \quad U = (1, 0), \quad V_* = (1, v_*), \quad D_* = (1, 1). \quad (C11)$$

Their costs are $0, 0, 0, f_*, f_*, R/(2R+1)$. Apart from V_* , they follow from the axis and diagonal constructions in Appendix B. A common setting-independent flag on disjoint classical sectors mixes these models with cost at most the

weighted sum. Put $t_* = (Rc_{>} - 1)/(R - 1)$. The following weights are nonnegative, sum to one, and reproduce the desired components in their regions:

$$\begin{aligned}
c_{>} + c_{<} \leq 1/R &: (w_O, w_L, w_T) = (1 - R(c_{>} + c_{<}), R(c_{>} - c_{<}), 2Rc_{<}), \\
c_{>} \geq 1/R, c_{<} \leq v_* t_* &: (w_L, w_U, w_{V_*}) = (1 - t_*, t_* - c_{<}/v_*, c_{<}/v_*), \\
c_{>} \geq 1/R, v_* t_* \leq c_{<} \leq t_* &: (w_L, w_{V_*}, w_{D_*}) = \left(1 - t_*, \frac{t_* - c_{<}}{1 - v_*}, \frac{c_{<} - v_* t_*}{1 - v_*}\right), \\
c_{>} + c_{<} \geq 1/R, c_{<} \geq t_* &: (w_L, w_T, w_{D_*}) = \left(R(c_{>} - c_{<}), \frac{2R[1 - Rc_{>} + (R - 1)c_{<}]}{2R - 1}, \frac{R(c_{>} + c_{<} - 1)}{2R - 1}\right). \quad (\text{C12})
\end{aligned}$$

These regions exhaust $0 \leq c_{<} \leq c_{>} \leq 1$. Their respective costs are zero, the axis plane, the mixed plane, and the full quadratic-mean plane. Every upper cost is a valid lower plane, so it equals the maximum in Eq. (13). Sign flips and the component swap complete the square. At $R = 2$, the model and envelope recover Eq. (14).

Appendix D: A finite-sample certificate for the mixed witness

Equation (21) concerns the true conditional correlations. For finite data, the input distribution is chosen so that the mean of one measured product equals the weighted correlation sum in that bound. Conditional concentration then controls the statistical error even when responses depend on previous trials. Finite-data dependence bounds and Bell-score tests with memory are established methods [18, 19]; their application here bounds the average resource F for the even-party joint constraint. The following model specifies the temporal assumptions and the quantity estimated.

Sequential model and input design. Fix even $n = 2m \geq 4$, the sign/sector orientation of the witness, a sample size $N_s \geq 1$, and $0 < \delta_s < 1$ before the test data. Let $\mathcal{H}_{t-1}$ denote the complete observed past settings, outcomes, and recorded public protocol choices. At each history the tested model has conditional current-input local responses,

$$P_t(\mathbf{a}|\mathbf{x}, \mathcal{H}_{t-1}) = \sum_{\lambda} p_t(\lambda|\mathbf{x}, \mathcal{H}_{t-1}) \prod_i \delta_{a_i, A_{i,t}(x_i, \lambda, \mathcal{H}_{t-1})}. \quad (\text{D1})$$

Private memory and shared randomness may be included in λ . Responses may drift and use arbitrary past history, but not another party's current input except through λ , or future inputs. This causal per-trial model is not unrestricted block-local measurement dependence [23]. For each trial, compare the hidden-variable distributions for all possible current settings at the same observed past. Their maximum TV distance is

$$F_t = \max_{\mathbf{x}, \mathbf{y}} D_{\text{TV}}[p_t(\lambda|\mathbf{x}, \mathcal{H}_{t-1}), p_t(\lambda|\mathbf{y}, \mathcal{H}_{t-1})], \quad \bar{F}_{N_s} = \frac{1}{N_s} \sum_{t=1}^{N_s} F_t. \quad (\text{D2})$$

This average depends on the history actually observed and is generally random. It is not a single static optimum or the optimum of a pooled outcome table.

With R, K_R, d_R as in Eq. (13), put

$$a = \frac{RK_R}{d_R}, \quad b = \frac{R^2 - 1}{d_R}, \quad k = \frac{K_R}{d_R}, \quad L = a + b, \quad (\pi_0, \pi_1) = \left(\frac{a}{L}, \frac{b}{L}\right). \quad (\text{D3})$$

The actual known setting marginal, conditional on every observed past, must obey

$$\Pr(\mathbf{X}_t = \mathbf{x}|\mathcal{H}_{t-1}) = \frac{\pi_{\mathbf{x}}(\mathbf{x})}{2^{n-1}}. \quad (\text{D4})$$

Both sectors have positive support. This joint correlated-setting design is not a product of independent local generators. Equation (D4) specifies the conditional marginal of settings, not independence between settings and the current hidden variable. A nominal programmed law or observed input frequencies alone do not establish this sampling assumption.

Every registered trial must supply a raw ± 1 outcome at every site. There is no data-dependent exclusion or replacement by fractional, readout-mitigated counts. Detection, heralding, or postselection protocols require their own trial definition and sampling argument. Selecting the witness orientation on the same test data requires a separate multiple-testing or predictable-adaptation proof. Such procedures are not included in Eq. (22).

From the measured score to a confidence bound. Define $Z_t = (-1)^{q(\mathbf{X}_t)} \prod_i a_{i,t}$ and $Y_t = LZ_t$. Let $S_{e,t}$ be Eq. (20) evaluated on the conditional table at history $\mathcal{H}_{t-1}$. Applying Eq. (21) at each history and using the sampling law gives

$$\mu_t = \mathbb{E}[Y_t | \mathcal{H}_{t-1}] = L(\pi_0 S_{0,t} + \pi_1 S_{1,t}) = aS_{0,t} + bS_{1,t} \leq k + F_t. \quad (\text{D5})$$

Since $Y_t \in [-L, L]$, the centered increment $Y_t - \mu_t$ has conditional mean zero and lies in the history-measurable interval $[-L - \mu_t, L - \mu_t]$ of length $2L$. Its absolute value need not be at most L . The conditional form of Hoeffding's lemma gives

$$\mathbb{E}[e^{s(Y_t - \mu_t)} | \mathcal{H}_{t-1}] \leq e^{s^2 L^2 / 2}, \quad s > 0. \quad (\text{D6})$$

For completeness, if a centered variable U lies in an interval of length h , set $\psi(s) = \ln \mathbb{E} e^{sU}$. The exponentially tilted distribution has variance at most $h^2/4$, so $\psi''(s) \leq h^2/4$. Integrating twice from $\psi(0) = \psi'(0) = 0$ yields $\psi(s) \leq s^2 h^2 / 8$. The variance bound follows from $\text{Var}(U) \leq (\mathbb{E}U - u_-)(u_+ - \mathbb{E}U) \leq (u_+ - u_-)^2 / 4$ on any interval $[u_-, u_+]$. Apply this argument conditionally with $h = 2L$.

Iterating conditional expectation in Eq. (D6) and applying Markov's inequality gives

$$\Pr \left\{ \sum_{t=1}^{N_s} (Y_t - \mu_t) \geq u \right\} \leq \inf_{s>0} e^{-su + N_s s^2 L^2 / 2} = e^{-u^2 / (2N_s L^2)}, \quad u > 0. \quad (\text{D7})$$

Choose $u = L\sqrt{2N_s \ln(1/\delta_s)}$. Except on an event of probability at most δ_s ,

$$\frac{1}{N_s} \sum_t \mu_t \geq L\bar{Z} - L\sqrt{\frac{2\ln(1/\delta_s)}{N_s}}. \quad (\text{D8})$$

Combining this with Eq. (D5) and $\bar{F}_{N_s} \geq 0$ proves Eq. (22). No stationary Bernoulli parameter or independence across trials was used. The conclusion bounds the realized average conditional diameter; it does not lower-bound the resource of each trial. In particular, binomial inversion under a uniform conditional winning cap is a different inference problem from an average-only resource budget [19].

Why use these even and odd sampling probabilities? For any fixed parity-zero probability $p \in (0, 1)$, with uniform strings within sectors, the unbiased importance score for $aS_0 + bS_1$ is

$$Y_t(p) = \begin{cases} aZ_t/p, & \chi(\mathbf{X}_t) = 0, \\ bZ_t/(1-p), & \chi(\mathbf{X}_t) = 1. \end{cases} \quad (\text{D9})$$

Its worst absolute value is $M(p) = \max\{a/p, b/(1-p)\}$. The inequalities $a \leq Mp$ and $b \leq M(1-p)$ imply $M \geq a+b$, with equality exactly at $p = a/L$. Our allocation therefore minimizes the worst-case score range within these fixed two-sector importance scores. This is not global optimization of variance, power, p-values, sample complexity, or all witnesses. Inverse-probability Bell scoring is an established method [19].

Also $0 < L < 1$ for every $R \geq 2$, since

$$d_R - (RK_R + R^2 - 1) = R^3 + 2R^2 - 3R - 2 = (R-2)(R^2 + 4R + 5) + 8 > 0. \quad (\text{D10})$$

The confidence penalty for this fixed score has no 2^n simultaneous-estimation prefactor. A chosen absolute error allowance $\epsilon > 0$ is reached once $N_s \geq 2L^2 \ln(1/\delta_s)/\epsilon^2$. This only bounds the penalty: a physical gap may shrink with n , and positive certification or n -independent resolution of that gap does not follow. The raw inequality is valid for general noise, but a single score does not determine its minimal simulation cost.

An explicit stopping-safe variant. The fixed- N_s confidence statement cannot silently be used at a data-dependent stopping time. For a sequential process defined for all $j \geq 1$, retain the fixed witness and conditional sampling law and allocate $\delta_{s,j} = \delta_s/[j(j+1)]$. Because $\sum_{j \geq 1} 1/[j(j+1)] = 1$, applying Eq. (D7) at every j and taking a countable union gives, with probability at least $1 - \delta_s$, simultaneously for all j ,

$$\bar{F}_j \geq \left[L\bar{Z}_j - k - L\sqrt{\frac{2\ln[j(j+1)/\delta_s]}{j}} \right]_+. \quad (\text{D11})$$

This permits evaluation at any almost-surely finite integer-valued stopping time. It is a conservative simultaneous corollary, not an optimal confidence sequence; the larger penalty must not be replaced by the fixed-sample one.

Appendix E: Calibrated input deviations and their sharp systematic cost

Imperfect setting generators and conservative calibration already enter memory-robust Bell analysis [19, 24]. Our target is the average measurement dependence defined in Appendix D, not the generator's bias itself. There are two cases. If the actual input probabilities are unknown, a calibrated bound on their deviation gives a systematic correction. If they are exactly known before each trial, reweighting removes the bias. We retain the same conditional model, fixed witness and complete raw trials.

An unknown input law with a bounded deviation. Denote the designed input distribution by $q_*(\mathbf{x}) = \pi_{\chi(\mathbf{x})}/2^{n-1}$; the star distinguishes this probability from the quadratic $q(\mathbf{x})$. Let $r_t(\mathbf{x}|\mathcal{H}_{t-1})$ be the actual conditional setting law, with positive support on every setting. Suppose a calibrated bound

$$D_{\text{TV}}[r_t, q_*] \leq \varepsilon_t, \quad \bar{\varepsilon} = \frac{1}{N_s} \sum_{t=1}^{N_s} \varepsilon_t \quad (\text{E1})$$

holds at the relevant histories. Predictable envelopes suffice. Then, with probability at least $1 - \delta_s$,

$$\bar{F}_{N_s} \geq \left[L\bar{Z} - k - 2L\bar{\varepsilon} - L\sqrt{\frac{2\ln(1/\delta_s)}{N_s}} \right]_+ \quad (\text{E2})$$

The envelope controls the entire joint setting law, not merely parity frequencies. If only the parity mass changes and strings remain uniform within each sector, its TV is the absolute parity-mass deviation. Full support is a separate assumption: a small TV tolerance alone does not ensure that every setting has positive probability.

To prove the result, set $f_t(\mathbf{x}) = (-1)^{q(\mathbf{x})} E_t(\mathbf{x}|\mathcal{H}_{t-1}) \in [-1, 1]$. On this same conditional table, the nominal score mean is $\mu_t^* = L \sum_{\mathbf{x}} q_*(\mathbf{x}) f_t(\mathbf{x}) = aS_{0,t} + bS_{1,t} \leq k + F_t$. For the actual mean $\mu_t = \mathbb{E}[LZ_t|\mathcal{H}_{t-1}]$,

$$\mu_t - \mu_t^* = L \sum_{\mathbf{x}} [r_t(\mathbf{x}) - q_*(\mathbf{x})] f_t(\mathbf{x}) \leq 2LD_{\text{TV}}[r_t, q_*] \leq 2L\varepsilon_t. \quad (\text{E3})$$

The positive and negative parts of $r_t - q_*$ each have mass $D_{\text{TV}}[r_t, q_*]$, which proves the factor two for any bounded f_t . The raw score LZ_t still lies in $[-L, L]$. Apply the conditional Hoeffding argument to its *actual* conditional mean, then combine its lower confidence bound with $\mu_t \leq k + F_t + 2L\varepsilon_t$. This proves Eq. (E2) without stationarity or independence across trials. A fixed tolerance leaves a systematic floor $2L\bar{\varepsilon}$ even as statistical fluctuation decreases.

Sharpness within deterministic local models. Choose a type-I signed class with positive even Walsh sum in Appendix C. Its agreement signs $g_c(\mathbf{x})$ have even and odd sums $2R$ and zero. Since each sector has $2R^2$ settings, the nominal mean is $L\pi_0/R = a/R = k$. The odd-sector signs are balanced, so select odd settings $\mathbf{x}_+, \mathbf{x}_-$ with signs $+1, -1$. For $0 < \varepsilon < q_*(\mathbf{x}_-)$, define

$$r(\mathbf{x}) = q_*(\mathbf{x}) + \varepsilon\delta_{\mathbf{x}, \mathbf{x}_+} - \varepsilon\delta_{\mathbf{x}, \mathbf{x}_-}. \quad (\text{E4})$$

The law is normalized and positive, its TV deviation is ε , and both parity masses are unchanged. Fixing the same deterministic hidden class for every input gives $F = 0$, but actual mean score $k + 2L\varepsilon$. Thus no smaller uniform coefficient than $2L$ in the affine budget $\mu \leq k + F + C\varepsilon$ is valid, even for local models. This is sharp systematic sensitivity of this witness, not an optimal finite-sample tail or the minimal cost for arbitrary noise. The transferable mass can be small at large n ; saturation alone does not promise a large experimentally resolvable margin.

Known law and the statistical cost of correction. Suppose instead that the actual law is known exactly before the current input and outcome, may depend on recorded past, and satisfies $r_t(\mathbf{x}|\mathcal{H}_{t-1}) \geq \zeta q_*(\mathbf{x})$ for a fixed deterministic $0 < \zeta \leq 1$ at every history and setting. Use the inverse-probability score [19]

$$W_t = L \frac{q_*(\mathbf{X}_t)}{r_t(\mathbf{X}_t|\mathcal{H}_{t-1})} Z_t. \quad (\text{E5})$$

Then $\mathbb{E}[W_t|\mathcal{H}_{t-1}] = \mu_t^* \leq k + F_t$ exactly, while $|W_t| \leq L/\zeta$. Conditional Hoeffding with this common deterministic range gives, with probability at least $1 - \delta_s$,

$$\bar{F}_{N_s} \geq \left[\bar{W} - k - \frac{L}{\zeta} \sqrt{\frac{2\ln(1/\delta_s)}{N_s}} \right]_+, \quad \bar{W} = \frac{1}{N_s} \sum_t W_t. \quad (\text{E6})$$

Known drift is removed at the price of a larger statistical range. A relative lower-probability tolerance $r_t \geq (1-\eta)q_*$ uses $\zeta = 1-\eta$. A TV tolerance alone does not provide an n -independent ratio floor: almost all mass at one small-probability setting can disappear while TV remains small. Estimated or pooled input probabilities cannot replace the exact conditional law in Eq. (E5). Nor does a predictable variable range justify optimizing a Chernoff parameter using a future realized variance sum; that requires a separate self-normalized argument.

Combining calibration and sampling errors. Calibration is a separate inference task [24]. Suppose a valid external procedure supplies an event $\mathcal{C}$ of probability at least $1 - \delta_{\text{cal}}$ on which all claimed conditional envelopes hold along the tested path. The bounded-score martingale event $\mathcal{M}$ is proved unconditionally. On their intersection Eq. (E2) holds, and $\Pr(\mathcal{C} \cap \mathcal{M}) \geq 1 - \delta_{\text{cal}} - \delta_s$. No independence is required; conditioning the martingale proof on calibration success would be a different argument. A confidence bound on pooled input frequencies does not by itself supply $\mathcal{C}$. A calibration procedure under explicit generator assumptions follows in Appendix F. It does not make estimated probabilities exact for Eq. (E5).

For a process defined for all prefixes, allocating $\delta_{s,j} = \delta_s/[j(j+1)]$ gives, simultaneously for all $j \geq 1$,

$$\bar{F}_j \geq \left[L\bar{Z}_j - k - 2L\bar{\epsilon}_j - L\sqrt{\frac{2\ln[j(j+1)/\delta_s]}{j}} \right]_+. \quad (\text{E7})$$

Its probability is at least $1 - \delta_s$, or $1 - \delta_s - \delta_{\text{cal}}$ when calibration covers all prefixes only on $\mathcal{C}$. The same union argument applies to Eq. (E6) with $\bar{W}_j, L/\zeta$. These simultaneous statements permit any almost-surely finite integer-valued stopping time with their larger penalties.

Why pooled input frequencies are insufficient. The following example has no measurement dependence at any trial, yet its average input probabilities equal the design and its uncorrected score exceeds the local limit. Input bias that changes with the past can therefore mislead a test based only on pooled frequencies. Cycle equally often through all 2^{n+1} signed deterministic classes $c = (\phi, \gamma)$, recording the cycle in the pretrial protocol. Put $g_c(\mathbf{x}) = (-1)^{q(\mathbf{x})+\phi+\gamma \cdot \mathbf{x}}$ and $m_c = \sum_{\mathbf{x}} q_*(\mathbf{x})g_c(\mathbf{x})$. At a trial assigned class c , fix $\lambda = c$ independently of the current input, giving $F_t = 0$, and choose

$$r_c(\mathbf{x}) = q_*(\mathbf{x}) \left[1 + \frac{g_c(\mathbf{x}) - m_c}{2} \right]. \quad (\text{E8})$$

It is normalized, and positive because $|m_c| < 1$. Sign partners ϕ and $\phi + 1$ average their two input laws to q_* exactly. Thus the cycle-averaged marginal equals the design. Yet the expected signed score in a trial is $m_c + (1 - m_c^2)/2$. The even-parity spectrum splits the classes equally between $|m_c| = \pi_0/R$ and π_1/R , so the cycle mean is

$$\langle Z \rangle_{\text{cycle}} = \frac{1}{2} \left[1 - \frac{\pi_0^2 + \pi_1^2}{2R^2} \right]. \quad (\text{E9})$$

For four parties this is $2107/4624 > 7/17 = k/L$, although every conditional diameter vanishes. Independent input draws across repeated cycles make pooled frequencies approach q_* while the uncorrected nominal certificate becomes positive. The calibrated bound remains valid. The relevant diagnostic is the history-conditional input law, not only its aggregate histogram or parity frequencies.

Appendix F: Causal calibration under an explicit generator model

The input-error correction in Appendix E requires a bound on the setting distribution before each trial. For the bounded-drift generator specified here, past inputs supply such a bound. The construction applies established multi-window concentration and drift-estimation tools [25] to the parity probability of the GHZ sampling protocol. It assumes the generator properties stated next; it does not verify them.

Generator and drift assumptions. Retain the fixed even-parity witness and the complete observed-history filtration of Appendix D. Let $J_t = 1$ select even parity and $J_t = 0$ odd parity, with unknown predictable probability, determined by the recorded past, $\omega_t = \Pr(J_t = 1 | \mathcal{H}_{t-1})$. Conditional on both $(\mathcal{H}_{t-1}, J_t)$, draw a uniform mask $\mathbf{U}_t \in \mathbb{F}_2^{n-1}$ and assign

$$X_{i,t} = U_{i,t} \ (i < n), \quad X_{n,t} = 1 - J_t + \sum_{i < n} U_{i,t} \pmod{2}. \quad (\text{F1})$$

Thus each sector is uniform, and

$$r_t(\mathbf{x}|\mathcal{H}_{t-1}) = \begin{cases} \omega_t/2^{n-1}, & \chi(\mathbf{x}) = 0, \\ (1 - \omega_t)/2^{n-1}, & \chi(\mathbf{x}) = 1, \end{cases} \quad D_{\text{TV}}[r_t, q_*] = |\omega_t - \pi_0|. \quad (\text{F2})$$

The map within each sector is a bijection; the two sector contributions to the ℓ_1 distance are each $|\omega_t - \pi_0|$. No setting–hidden-variable independence is imposed.

Specify deterministic acquisition constants $0 < \omega_{\min} < \min\{\pi_0, \pi_1\}$ and $\Delta_{\text{in}} \geq 0$ such that, almost surely at all relevant histories,

$$\omega_{\min} \leq \omega_t \leq 1 - \omega_{\min}, \quad |\omega_{t+1} - \omega_t| \leq \Delta_{\text{in}}. \quad (\text{F3})$$

The probability can depend on recorded past outcomes and inputs. The rate bound, full support, and mask uniformity conditional on both history and parity are external model assumptions. Pooled frequency checks, unconditional fair bits, or feedback stability do not establish them. This correlated joint generator is not a product of independent local setting sources and is not a demonstrated spacelike protocol.

Calibration before each trial at a fixed horizon. Before data collection fix N_s , $\delta_{\text{cal}} \in (0, 1)$ and a finite nonempty grid $\mathcal{W}$ of positive integer window lengths. Record $\ell_{\max} = \max \mathcal{W}$ complete warmup trials, excluding their outcomes from the reported score by this predetermined chronology. Score all of the next N_s complete raw trials, $t = \ell_{\max} + 1, \dots, \ell_{\max} + N_s$. Before each scored trial t , compute from past parity bits only

$$\begin{aligned} \hat{\omega}_{t,\ell} &= \frac{1}{\ell} \sum_{s=t-\ell}^{t-1} J_s, & h_\ell &= \sqrt{\frac{\ln(2N_s|\mathcal{W}|/\delta_{\text{cal}})}{2\ell}}, \\ u_{t,\ell} &= |\hat{\omega}_{t,\ell} - \pi_0| + h_\ell + \frac{\Delta_{\text{in}}(\ell + 1)}{2}. \end{aligned} \quad (\text{F4})$$

The prior interval supplies $\varepsilon_{\text{prior}} = \max\{|\omega_{\min} - \pi_0|, |1 - \omega_{\min} - \pi_0|\}$. Store the pretrial envelope

$$\varepsilon_t = \min \left\{ \varepsilon_{\text{prior}}, \min_{\ell \in \mathcal{W}} u_{t,\ell} \right\}. \quad (\text{F5})$$

Generate the input, record J_t and the complete outcome string, and add its signed product Z_t to the score. The window can be chosen from this grid using the observed past; no unbudgeted window or current J_t is used to calibrate that same trial.

Here is the coverage proof. For each deterministic (t, ℓ) , $J_s - \omega_s$ is a martingale difference with conditional interval length one. Conditional Hoeffding, as proved in Appendix D, gives

$$\Pr \left\{ \left| \hat{\omega}_{t,\ell} - \frac{1}{\ell} \sum_{s=t-\ell}^{t-1} \omega_s \right| > h \right\} \leq 2e^{-2\ell h^2}. \quad (\text{F6})$$

The means may be random and history-adaptive. A union bound over the $N_s|\mathcal{W}|$ prescribed windows controls all deviations by h_ℓ with probability at least $1 - \delta_{\text{cal}}$. Overlapping windows need not be independent. Pathwise telescoping of Eq. (F3) yields

$$\left| \omega_t - \frac{1}{\ell} \sum_{s=t-\ell}^{t-1} \omega_s \right| \leq \frac{\Delta_{\text{in}}}{\ell} \sum_{j=1}^{\ell} j = \frac{\Delta_{\text{in}}(\ell + 1)}{2}. \quad (\text{F7})$$

On this simultaneous event, every $u_{t,\ell}$ and the prior cap bound $|\omega_t - \pi_0|$, so Eq. (F5) is valid for all scored trials. More past inputs reduce the statistical radius but increase the drift allowance.

Intersect this event with the unconditional raw-score sampling event. With probability at least $1 - \delta_{\text{cal}} - \delta_s$,

$$\bar{F}_{N_s}^{\text{test}} \geq \left[L\bar{Z}^{\text{test}} - k - 2L\bar{\varepsilon}^{\text{test}} - L\sqrt{\frac{2\ln(1/\delta_s)}{N_s}} \right]_+. \quad (\text{F8})$$

All bars here average only the prescribed scored trials; their conditional diameters use the same full history as before. Calibration and score data can overlap. No independence between the two confidence events or conditioning on calibration success is used. Choose their total failure budget below one for a nonvacuous confidence level. Estimated probabilities enter only as covered TV envelopes, not exact probabilities in the importance score of Eq. (E5).

Calibration when the stopping time is not fixed. For a process defined at all times, a separate schedule covers every $t \geq 2$ and $\ell = 1, \dots, t-1$. Allocate

$$\alpha_{t,\ell} = \frac{2\delta_{\text{cal}}}{t(t+1)(t-1)}, \quad h_{t,\ell} = \sqrt{\frac{\ln[t(t+1)(t-1)/\delta_{\text{cal}}]}{2\ell}}. \quad (\text{F9})$$

The sum of $\alpha_{t,\ell}$ over all windows is δ_{cal} . Use Eq. (F5) with all preceding windows and these radii. The resulting envelope is simultaneously valid at all times. Score trials $2, \dots, j+1$ and apply the sampling allocation $\delta_{s,j} = \delta_s/[j(j+1)]$. With probability at least $1 - \delta_{\text{cal}} - \delta_s$, Eq. (F8) holds for every scored prefix j , with all horizon-indexed averages and denominators taken over these j scored trials and its sampling logarithm replaced by $\ln[j(j+1)/\delta_s]$. It can therefore be evaluated at any almost-surely finite integer-valued stopping time in scored-trial count. The fixed-horizon calibration and sampling radii cannot replace these simultaneous penalties at an arbitrary stopping time.

A separately bounded imperfect mask. Let $\kappa_t \in [0, 1]$ be deterministic or $\mathcal{H}_{t-1}$ -measurable and known before the current parity, mask and outcomes are generated. If, for both parity branches at each relevant history, the conditional mask law is within TV κ_t of uniform, the sector-uniform intermediate law in Eq. (F2) differs from the actual law by at most $\omega_t \kappa_t + (1 - \omega_t) \kappa_t = \kappa_t$. Thus $\min\{1, \varepsilon_t + \kappa_t\}$ is a joint-law envelope. If the mask contract holds only on an external event of failure probability at most δ_{mask} , that event must cover both branches and all scored histories simultaneously, or all times in the stopping variant. Intersect it unconditionally and add δ_{mask} once to the total risk. Full support remains a separate requirement; TV closeness alone does not ensure it. Parity logs do not calibrate κ_t or verify Eq. (F3). The procedure supplies error bounds within the stated generator model; it neither verifies randomness without that model nor demonstrates an implemented experimental test.