Noise-Robust Self-Testing: Detecting Non-Locality in Noisy Non-Local Inputs

by

Romi Lifshitz

Submitted to the Department of Computer Science and Mathematics in partial fulfillment of the requirements for the degree of

MASTER'S OF COMPUTER SCIENCE AND MATHEMATICS

at the

WEIZMANN INSTITUTE OF SCIENCE

April 2025

© 2025 Romi Lifshitz. All rights reserved.

Authored by: Romi Lifshitz

Department of Computer Science and Mathematics

April 2, 2025

Supervised by: Thomas Vidick

Professor of Computer Science and Applied Mathematics, Thesis Supervisor

Accepted on: April 30, 2025

Contents

I.INTRODUCTION	. 4
A.Motivation	. 4
B.Contribution and Main Ideas	. 5
II.PRELIMINARIES	. 6
A.Quantum Information	. 6
B.Probability & Statistics	. 7
C.Non-Local Games	. 7
D.Self-Testing & Common Self-Tests	. 9
D.1.CHSH: An initial non-local game	. 9
D.2.2-CHSH Game	. 10
D.3.The Magic Square Game	. 10
III.NOISE-ROBUSTNESS FRAMEWORK FOR NON-LOCAL GAMES	. 12
A.Noise-Tolerance	. 12
B.Convincingness	. 14
B.1.Motivation	. 14
B.2.Construction	. 14
B.3.Simplified Convincingness	. 15
B.4. Noise-Robustness Framework for Comparative Scores	. 16
IV.COMPUTATIONAL METHODS	. 17
A.Analysis Across Noise Levels	. 18
B.Optimization for Specific Noise Regimes	. 18
B.1.Optimized 2-CHSH Construction	. 18
B.2.Linear Program	. 20
C.Noise-Tolerance Analysis	. 20
V.ANALYTIC METHODS	. 21
A.Gapped Expressions	. 21
B.Gap-Based Scoring	. 22
C.Polynomial Comparisons	. 24
VI.RESULTS AND DISCUSSION	. 25
A.Convincingness Analysis	. 25
B.Noise-Tolerance Analysis	. 30
VII.DEFINING NOISE-ROBUSTNESS OF NON-LOCAL GAMES	. 32
TIII.CONCLUSION	. 33
IX.APPLICATIONS	. 33
X.STRENGTHS & WEAKNESSES	. 34
XI.FUTURE RESEARCH AVENUES & OPEN QUESTIONS	. 34
A.Non-Local Reductions	
B.Further Investigations of Noise-Robustness	

XII.ACKNOWLEDGMENTS	
XIII.(APPENDIX) 2-CHSH-OPT Score Derivation	
XIV.(APPENDIX) Convincingness Computations f	or Optimized 2-CHSH Games 41
XV.(APPENDIX) 2-CHSH-OPT Local Bound Cor	nputation
XVI.(APPENDIX) Proof of Noise-Tolerance Relati	on to Convincingness
XVII.(APPENDIX) Impossibility of DI-LOSR Conv	version from the Magic-Square to Tsirelson Box \cdot 44
List of Figures	
 Optimization Axes for Non-Local Games. Axes Visualizations for Evaluating Noise-Robustness Convincingness Behaviour Analysis for Finite Resonance Convincingness Behaviour Analysis for Near-Infinite Convincingness Behaviour Analysis for Near-Infinite Configurations. Order of Crossings (1 – η†) for Games at Different I Comparing Candidate Predictors of Crossing Order Noise-Tolerance Analysis. List of Tables Gapped Expressions for Non-Local Games Comparison of Non-Local Games and Their Propert 	12 16 17 18 19 10 11 12 13 14 15 16 17 17 18 19 10 10 10 10 11 12 12 12 12 12 12 12 12 12 13 14
List of Abl	oreviations
CHSH Game – Clauser, Horne, Shimony, & Holt Game	• CPTP Map – Completely Positive & Trace Preserving
• MSG – Magic Square Game	Map • XOR – Exclusively-OR
• W – Werner State	• 2-CHSH Game – 2-Clauser, Horne, Shimony, & Holt
DI – Device-Independent CHZ – Charach annual Harma Zailianna	Game
 GHZ – Greenberger-Horne-Zeilinger EPR – Einstein-Podolsky-Rosen 	• Def(s). – Definition(s)
• \mathfrak{C}_G – Convincingness of Game G	• Fig(s). – Figure(s)
• κ_G – Gapped Score of Game G	• Sec(s). – Section(s)
 <i>H</i> - Hilbert Space 	• 2-CHSH-OPT – Optimized 2-CHSH Game
• I – Identity operator	• LP – Linear Program(ming)
POVM – Positive Operator-Valued Measure	• Const. – Constant
• LHV (or LHVM) – Local Hidden Variable (Model)	 Coeff Coefficient IFF - if & only if

ABSTRACT

Non-local games test for non-locality and entanglement in quantum systems and are used in self-tests for certifying quantum states in untrusted devices. However, these protocols are tailored to ideal states, so realistic noise prevents maximal violations and leaves many partially non-local states undetected. Selecting self-tests based on their 'robustness' to noise can tailor protocols to specific applications, but current literature lacks a standardized measure of noise-robustness. Creating such a measure is challenging as there is no operational measure for comparing tests of different dimensionalities and input-output settings. We propose and study three comparative measures: noise-tolerance, convincingness, and an analytic approximation of convincingness called the gapped score. Our computational experiments and analytic framework demonstrate that convincingness provides the most nuanced measure for noise-robustness. We then show that the CHSH game has the highest noise-robustness compared to more complex games (2-CHSH variants and the Magic Square Game) when given equal resources, while with unequal resources, some 2-CHSH variants can outperform CHSH at a high resource cost. This work provides the first systematic and operational framework for comparing noise-robustness in self-testing protocols, laying a foundation for theoretical advances in understanding noise-robustness of self-tests and practical improvements in quantum resource utilization.

I. INTRODUCTION

A. Motivation

Bell tests, fundamental to quantum computing and communication, demonstrate non-locality (and hence, entanglement) in physical systems by showing violations of inequalities that cannot be explained by classical, local hidden-variable theories [1]. These tests have evolved into protocols called non-local games, which are crucial for detecting entanglement in quantum resources [2, 3, 4]. Self-tests, a specific class of non-local games, enable device-independent (DI) certification of non-locality in quantum states and measurements [5]. In these tests, a referee interacts with spatially isolated players who have prepared some physical system in a device. Using a self-test, the referee can determine the presence of non-locality in the device based solely on its input-output statistics and the assumption that quantum mechanics is true. Self-tests have proven valuable in DI Quantum Key Distribution [6, 7, 8], randomness certification [9], and experimental demonstrations of entanglement [10, 11, 12]. Moreover, they certify select quantum properties more efficiently than tomography [13].

However, the DI property of self-tests comes at a cost: self-tests are optimized for specific pure entangled states. For instance, the Clauser, Horne, Shimony, Holt (CHSH) self-test [14] achieves maximal quantum violation only with a maximally entangled state, the EPR pair up to local isometries. This presents practical challenges, as physical imperfections and noise in real experiments make achieving maximal violations impossible with current technology. Consequently, mixed states with partial non-locality may go undetected, leading to the waste of potentially useful quantum resources. Moreover, no established methodology exists to rank different self-tests based on their ability to detect non-locality in noisy states. This prevents practitioners from selecting tests with appropriate sensitivity levels for specific applications—whether they need robust detection of noisy entanglement (as in entanglement distillation) or stringent verification of more secure states (as in quantum key distribution).

While previous works have explored methods for detecting non-locality in noisy systems, they either rely on device-dependent approaches—such as weak measurements [15], local noise-reversal channels [16], or optimized entanglement witnesses [17]—or require extensive resources, like the DI parallel-repetition-based approach proposed in [18]. Other studies have characterized lower bounds on the violations achieved for specific input states (i.e., GHZ, W, graph states) affected by noisy quantum channels [19, 20, 21, 22, 23, 24]. However, these works interchangeably use the terms noise-robustness, -tolerance, or -resistance of a game to describe its ability to detect non-locality in noisy systems, with unclear and un-unified definitions. Each approach is uniquely tied to the specific noise models, games, and experimental setups in a work. This fragmentation reveals a critical gap. The field lacks: (1) a unified definition and measure of the noise-robustness of self-tests, and (2) a systematic framework for comparing different tests' abilities to detect non-locality in noisy systems. This thesis addresses these limitations by developing rigorous methods for defining, computing, and comparing the noise-robustness of self-tests, particularly towards noisy states rather than noisy measurements. In doing so, we provide tools

for experimentalists and theorists to select a self-test which is more suitable to the level of robustness needed in their application, noise model, and resource constraints. We are, hence, motivated to ask the following research questions to help define the noise-robustness of self-tests:

Question 1. Given a set of self-tests, which test is able to detect non-locality in the largest regime of noisy states, under the assumption of a particular noise model?

Question 2. Given a fixed set of noisy resources (under a noise model) and a set of candidate self-tests, which self-test should one use to produce the largest deviation from local deterministic statistics, knowing that the statistics will not be ideal? In other words, how should we use a noisy entanglement resource to demonstrate non-locality most effectively?

Question 3. Is it possible to expand the set of noisy states whose non-locality is detected by a self-test?

Questions 1 and 2 capture different perspectives on noise-robustness—either focusing on the regime of noisy states a game can certify or how to best leverage a limited, imperfect resource supply—while 3 targets the ultimate aim of devising more noise-robust games. The challenge in answering these questions is that, due to the distinct setup and norms of games, there is no straightforward way to compare their: (1) sets of violating states, and (2) violations and scores. For example, consider comparing the CHSH game, whose ideal state is a 2-qubit EPR pair, versus the Magic Square Game (MSG), whose ideal state is two EPR pairs (4 qubits). The games operate optimally on different dimensionalities, their scores exist on different scales, and a violation of the local bound (ω_c) in one game cannot be directly mapped to a meaningful violation in the other. While a ratio of the score attained from an experiment to the local bound has been used as a comparative measure [25, 26], it is inadequate as it lacks operational interpretation and normalization between games.

From a geometric perspective, consider how different games relate to the quantum and local sets [27, 28, 29]. Each non-local game defines a distinct axis through these sets, creating different directional cuts through the space of correlations. This geometric view highlights fundamental questions: How are points compared along different axes? How do we characterize and compare the sets of states violating different games? These questions remain poorly understood, highlighting the lack of theoretical tools for systematically comparing non-local games and understanding their relative capabilities.

We note that throughout the thesis, we use the terms non-local games and self-tests interchangeably (unless stated otherwise), as we are only concerned with non-local games which are self-tests.

B. Contribution and Main Ideas

We answer the above questions by proposing novel methods for comparing the abilities of non-local games to detect non-locality in noisy states (Sec. III.). Using these methods, we construct a rigorous definition for noise-robustness (Secs. III., V.), and a framework for testing and optimizing the noise-robustness of non-local games (Sec. B.4.). Further, we apply the new methods computationally and analytically (Sec. IV., V., VI.) to compare the noise-robustness abilities of existing, and new non-local games in finite and infinite (noisy) resource regimes. Specifically, we construct the following measures of noise-robustness:

- 1. Noise-tolerance of G: noise-tolerance characterizes a non-local game's ability to detect non-locality in noisy versions of its winning state, where a game is considered (η, ε) -tolerant if it can detect non-locality in its winning state under visibility (probability of no noise) η in a noise model ε . To compare noise-tolerance between different games, we examine whether they can detect non-locality in their respective winning states under the same visibility η , with appropriate extensions of the noise channel for games of different dimensionalities (i.e., comparing CHSH's tolerance of a noisy EPR pair versus MSG's tolerance of two noisy EPR pairs).
- 2. Convincingness of $G(\mathfrak{C}_G)$: Convincingness is a statistical measure that quantifies how likely a game's observed statistics could have occurred due to locality, using a p-value calculation¹. A higher convincingness (lower p-value)

¹During the course of this thesis, we observed that [30] employs an identical *p*-value formulation, focused on optimizing Bell-inequality violations for single-shot rejection of local-hidden-variable models. In contrast, we derive the *p*-value independently to define noise-robustness and develop a comprehensive framework for comparing and ranking the noise-robustness of non-local games with differing structures, physical noise channels, and resource constraints.

indicates stronger evidence of non-locality in the input state, allowing comparison of different games' effectiveness at detecting non-locality in noisy states on a normalized scale regardless of their setup or bounds.

3. Gapped Score (κ_G): The gapped score is an efficient approximation of the convincingness measure for ranking games' effectiveness at detecting non-locality in noisy states, without evaluating \mathfrak{C}_G on input states. While convincingness offers a complete analysis, our gapped score provides a more analytic, simpler calculation based on a game's sensitivity to a noise model and its quantum-classical separation. The score reliably ranks games when moderate to high numbers of noisy resources are available for proving non-locality. See Def. V.41 for the formal definition.

We refer the reader to Sec. III. for more detail. Equipped with these new definitions, we utilize them to guide designing games with improved noise-robustness. Lastly, after investigating the behaviour of these noise-robustness measures, we arrive at the proposed definition of noise-robustness in Def. VII.42: A non-local game G_1 is more noise-robust to a noisy channel than G_2 if G_1 is significantly convincing for higher noise levels on the input state than G_2 . This definition can be further simplified to Def. VII.43 if a moderate to high resource supply is available for games to prove non-locality: A non-local game G_1 is more noise-robust to a noisy channel than G_2 if $\kappa_{G_1} > \kappa_{G_2}$. These definitions can provide practical guidance, based in theory, to experimentalists who wish to identify entanglement in imperfect hardware, while preserving their resources. Further, as the latter definition depends on an analytically computable expression (κ_G), it is suitable for theoretical studies of noise-robustness in asymptotic resource regimes. To our knowledge, we are the first to build a framework for evaluating and studying the noise-robustness of non-local games. Moreover, we are the first to use the p-value for comparing between different non-local games' robustness to noisy inputs.

II. PRELIMINARIES

A. Quantum Information

Let us define relevant quantum information concepts which we use throughout the work. For more information, see [31].

Definition II.1 (Hilbert Space). A d-dimensional Hilbert space $\mathcal{H} = \mathbb{C}^d$ is a complete inner product space over \mathbb{C} .

Definition II.2 (Bra-ket). For a vector $|\psi\rangle \in \mathcal{H}$, its dual vector $\langle \psi| = (|\psi\rangle)^{\dagger} \in \mathcal{H}^*$, where \dagger is the conjugate transpose.

Definition II.3 (Qubit and n-Qubit System). A qubit is a quantum state in $\mathcal{H} = \mathbb{C}^2$. The computational basis states are denoted $\{|0\rangle, |1\rangle\}$. An n qubit system is a quantum state in $\mathcal{H} = (\mathbb{C}^2)^{\otimes n} \cong \mathbb{C}^{2^n}$.

Definition II.4 (Quantum State). A quantum state of a system with Hilbert space \mathcal{H} is described by a density operator ρ acting on \mathcal{H} . The state is:

- Pure if and only if $\rho^2 = \rho$ (equivalently, $Tr(\rho^2) = 1$)
- Mixed if and only if $\rho^2 \neq \rho$ (equivalently, $Tr(\rho^2) < 1$)

A pure state can alternatively be represented by a unit vector $|\psi\rangle \in \mathcal{H}$, with corresponding density operator $\rho = |\psi\rangle\langle\psi|$.

Definition II.5 (Separable & Entangled States). The state ρ_{AB} is separable (not entangled) if and only if it can be written as $\rho_{AB} = \sum_i p_i \rho_i^A \otimes \rho_i^B$, where p_i are probabilities ($\sum_i p_i = 1$) and ρ_i^A , ρ_i^B are density matrices for systems A and B. ρ_{AB} is entangled if and only if it is not separable.

Definition II.6 (Bipartite System). A bipartite quantum system is a composite system with $\mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B$, where \mathcal{H}_A and \mathcal{H}_B are the Hilbert spaces of subsystems A and B, respectively.

Definition II.7 (Maximally Mixed State). The maximally mixed state for a quantum system $\rho \in \mathbb{C}^{d \times d}$ is $\rho = \frac{\mathbb{I}_d}{d}$.

Definition II.8 (Maximally Entangled State). A bipartite system, ρ_{AB} , with subsystems A, B in dimension d, is maximally entangled if and only if ρ_{AB} is pure and the reduced densitiy matrices, $\rho_A = Tr_B(\rho_{AB}), \rho_B = Tr_A(\rho_{AB})$, are maximally mixed. The EPR pair is the maximally entangled state, $|\psi\rangle = \frac{|00\rangle + |11\rangle}{\sqrt{2}}$, up to local isometries.

Definition II.9 (POVM). A Positive Operator-Valued Measure (POVM) is a set $\{E_i\}$ where $E_i \succeq 0$ (positive semidefinite), $\sum_i E_i = I$ (completeness). For input state ρ , the probability of measuring outcome i is $Tr(E_i\rho)$.

Definition II.10 (Projective Measurement). A projective measurement is a POVM $\{P_i\}$ where $P_i^2 = P_i$ and $P_iP_j = \delta_{ij}P_i$.

Definition II.11 (Quantum Channel). A quantum channel is a completely positive and trace preserving (CPTP) map T that maps density operators on \mathcal{H}_A to density operators on \mathcal{H}_B .

Lemma 1 (Tensor Product of Quantum Channels). If T_1 and T_2 are CPTP maps, then $T_1 \otimes T_2$ is also a CPTP map.

Definition II.12 (2-Qubit Depolarizing Channel). The two-qubit depolarizing channel with visibility $\eta \in [0,1]$ is the CPTP map $\varepsilon_{\eta}^{2-qubit}(\rho) = \eta^2 \rho + (1-\eta^2) \frac{\mathbb{I}_4}{4}$, where ρ is a two-qubit state (i.e., $\rho \in \mathbb{C}^{4\times 4}$).

Definition II.13 (4-Qubit Depolarizing Channel). The four-qubit depolarizing channel with visibility $\eta \in [0,1]$ is the CPTP map $\varepsilon_{\eta}^{4-qubit}(\rho) = \varepsilon_{\eta}^{2-qubit}(\rho) \otimes \varepsilon_{\eta}^{2-qubit}(\rho)$ where each factor $\varepsilon_{\eta}^{2-qubit}$ acts independently on a two-qubit subsystem.

Note that when the input state is known, we drop the ρ specification in the channels.

B. Probability & Statistics

We define some probability and statistics terms, which we later use to define noise-robustness. For more details, see [32].

Definition II.14 (Probability Distribution). A probability distribution is a function P mapping a sample space $\Omega \to [0,1]$ such that: (1) For a discrete random variable, $X: \sum_{\omega \in \Omega} P(\omega) = 1$ and P(X = x) gives probability of specific values, and (2) For a continuous random variable, $X: \int_{\Omega} f(x)dx = 1$ and $P(a \le X \le b) = \int_a^b f(x)dx$ gives probability over intervals.

Definition II.15 (Hypothesis Test). A hypothesis test evaluates competing probability distributions for observed data. Let θ represent the true, unknown parameter value that determines the probability distribution of our data. The test compares: (1) The null hypothesis $H_0: \theta = \theta_0$, which specifies a single probability distribution for the data, and (2) The alternative hypothesis $H_1: \theta = \theta_1$ (or $\theta > \theta_0$, or $\theta < \theta_0$), which specifies a single (or set of) distribution(s). The test uses a test statistic T computed from the data to determine which distribution better explains the observations.

Definition II.16 (p-value (upper-tailed)). Given a hypothesis test comparing $H_0: \theta = \theta_0$ with $H_1: \theta > \theta_0$, the p-value (upper-tailed) is the probability, computed under the distribution specified by θ_0 , of obtaining a test statistic value at least as extreme as the observed value t. Formally, $p = P(T \ge t \mid H_0) = \sum_{j=t}^n P(T = j \mid H_0)$, where the summation is over all outcomes that are as extreme or more extreme than t in the direction specified by H_1 .

C. Non-Local Games

Non-local games are interactive protocols (games) between multiple parties, used to detect non-locality in quantum devices through violations of Bell inequalities. In order to understand them, we require the following information [33, 34]:

Definition II.17 (Local Hidden Variable Model). A local hidden variable model (LHVM) for a bipartite system describes correlations between measurement outcomes a, b for settings x, y as probability distributions of the form: $P(a, b|x, y) = \sum_{\lambda} p_{\lambda} P(a|x, \lambda) P(b|y, \lambda)$ where λ represents hidden variables with probability distribution p_{λ} , and the probabilities $P(a|x, \lambda)$, $P(b|y, \lambda)$ depend only on local measurements and the hidden variable.

Definition II.18 (Bell Inequality). A Bell inequality is a linear constraint on probability distributions P(a, b|x, y) of measurement outcomes a, b for settings x, y that must be satisfied by any local hidden variable model.

Theorem II.1 (Non-locality implies Entanglement). If a quantum state ρ_{AB} exhibits non-locality by violating a Bell inequality, with local measurements, then ρ_{AB} must be entangled (violation is sufficient, but not necessary for entanglement).

With this in mind, let us now define the setup, strategies, and scoring of 2-player games, which may be extended to k-player games (for a more detailed explanation, see [34, 35]). Intuitively, in a 2-player non-local game, a referee, Charlie,

sends questions $x, y \in \mathcal{X}, \mathcal{Y}$ to each player, Alice and Bob. Each player then returns an output $a, b \in \mathcal{A}, \mathcal{B}$, respectively, to Charlie, who then checks a relationship (i.e., an inequality) defined by a predicate, V. The players win if the inequality is satisfied, and lose otherwise. Games with more players can be defined analogously, with an additional set of questions and answers per player, and an updated predicate. Formally:

Definition II.19 (2-Player Non-Local Game). A 2-player non-local game can be defined generally, as a 6-tuple $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$ where:

- The finite non-empty sets of $\mathcal{X} = (x_1, x_2, \dots, x_k), \mathcal{Y} = (y_1, y_2, \dots, y_k)$ are questions and the finite non-empty sets of $\mathcal{A} = (a_1, a_2, \dots, a_k), \mathcal{B} = (b_1, b_2, \dots, b_k)$, are answers, for arbitrary k
- The sample space $\Omega = \mathcal{X} \times \mathcal{Y} \times \mathcal{A} \times \mathcal{B}$ consists of all possible combinations of questions and answers
- $\pi \in \mathcal{P}(\mathcal{X} \times \mathcal{Y})$ is a probability vector denoting the probability of the referee selecting $(x, y) \in \mathcal{X} \times \mathcal{Y}$
- V: Ω → {0,1} is a measurable function (predicate) determining winning outcomes, where V(a, b|x, y) denotes if (a, b) win (1) or lose (0) given (x, y).

Importantly, players can coordinate a *strategy* for playing the game. For example, a strategy might be that they each return 1 upon receiving a 0. If the resources used by the players are classical, the strategy is expressed by two functions f_1, f_2 such that $f_1 : \mathcal{X} \to \mathcal{A}$ and $f_2 : \mathcal{Y} \to \mathcal{B}$. However, the general definition of either a classical or quantum strategy is presented in Def. II.20. Moreover, by playing multiple *rounds* of the game with a strategy, a statistical *score* can be computed for it, by the means of II.21. Each game also has a unique maximal score using a classical or quantum strategy (up to local isometries) known as its *local* and *quantum bounds* (Defs. II.22 and II.23). Finally, the local bound is used to determine whether a strategy's state (input state) has non-locality, using the concept of a *violation* (Def. II.24).

Definition II.20 (Strategy). A strategy of a 2-player non-local game can be generally defined as the 3-tuple: $S = (\rho, \mathcal{M}_1, \mathcal{M}_2)$, where ρ is some (input) state in the Hilbert space \mathcal{H} and $\mathcal{M}_{i \in \{1,2\}}$ is a set of measurement operators (observables of projectors) used by players $P_{i \in \{1,2\}}$ on their share of the subspace, \mathcal{H}_1 or \mathcal{H}_2 , respectively. During the strategy, each player selects a measurement, based on their question, x or y, and applies it on their subspace. A classical or quantum strategy is one in which the players use only classical or quantum resources, respectively.

Definition II.21 (Score). For each trial $t \in \Omega$, define the win/loss indicator $W : \Omega \to \{0,1\}$ as W(x,y,a,b) = V(a,b|x,y). The score, or value, of game G is the expectation $\mathbb{E}[W]$, which, for a given strategy $S = (\rho, \mathcal{M}_1, \mathcal{M}_2)$ can be computed as $\langle \mathbf{S}, \rho \rangle = Tr(\rho \mathbf{S}) := \omega(G)$, where \mathbf{S} , is a Hermitian operator combining \mathcal{M}_1 and \mathcal{M}_2 . \mathbf{S} should account for correlations corresponding to all input and output combinations contributing nonzero probabilities to the score.

Definition II.22 (Local Bound). The maximal score of a game, G, achieved with a classical strategy is the best score achieved under the local hidden variables (LHV) model and is called the local bound (or, winning classical value) of G:

$$\omega_{c_G} = \max_{f,g} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \pi(x,y)V(f(x),g(y)|x,y). \tag{1}$$

Definition II.23 (Quantum Bound). The maximal score of a game, G, achieved with a quantum strategy is the best score achieved under the assumption of quantum mechanics and is called the quantum bound (or, winning quantum value) of G. To compute it, let us define each operator in the sets $\mathcal{M}_1, \mathcal{M}_2$ according to POVM elements corresponding to each player's possible questions and answers. Let $W_{a|x}$ be a POVM element for an outcome a, given x, acting on \mathcal{H}_1 , and similarly let $Q_{b|y}$ be a POVM element acting on \mathcal{H}_2 . Then the quantum bound is:

$$\omega_{q_G} = \sup_{\mathcal{S}} \sum_{(x,y)\in\mathcal{X}\times\mathcal{Y}} \pi(x,y) \sum_{(a,b)\in\mathcal{A}\times\mathcal{B}} V(a,b|x,y) \langle W_{a|x} \otimes Q_{b|y}, \rho \rangle. \tag{2}$$

Definition II.24 (Violation). A violation of a non-local game results when some strategy achieves $\omega(G) > \omega_c(G)$, implying that non-locality and, therefore, entanglement has been used in the strategy; that LHV has been violated. If a game has been violated, this is a 'win', and otherwise, it is a 'loss'.

Let us also make a note about the comparability of scores and violations in games. If **S** is a projector, the score is a probability and is called the success probability, whereas if **S** is an observable, the score is on a shifted scale. If the game is a *XOR game*, the two scales may be easily interconverted. Otherwise, the interconversion is more challenging. Moreover, due to the different input-output settings of games, measurements, and normalizations π , the scores and violations are incomparable between different games. Also note that we simplify $\omega_{c_G} = \omega_c, \omega_{q_G} = \omega_q$ when the reference game is clear.

D. Self-Testing & Common Self-Tests

In this thesis, we are concerned with non-local games which enforce a *local strategy* and are *device-independent*. Such non-local games are called self-tests, or self-testing protocols, and are used to certify the non-locality of either a state or measurements used by the device during the process of a game [5]. In particular:

Definition II.25 (Self-Test). A self-test is a non-local game defined by the following rules:

- 1. The players are restricted to local strategies during the protocol; however, they may share a resource such as randomness or entanglement, established before the game starts.
- 2. The physical systems of the two players comprise a device that is considered to be a black box (device-independence). That is, the referee sends questions to and receives answers from the players, making no assumptions about the device's inner workings. The only assumption is that the laws of quantum mechanics hold. Hence, only the statistics of the inputs and outputs of the device are used to discern a property (i.e., non-locality) of the system inside the box.

D.1. CHSH: An initial non-local game

Let us consider various commonly played 2-player non-local games. The first is the *Clauser, Horne, Shimony, and Holt* (CHSH) game (Fig. 1a) [14], one of the most commonly used components in self-testing protocols:

Definition II.26 (CHSH Game). The CHSH Game is the 2-player, 2-input, 2-output non-local game where the question and answer sets are binary values, $\mathcal{X} = \mathcal{Y} = \mathcal{A} = \mathcal{B} = \{0,1\}$, and the probability distribution across all possible questions, π , is uniform such that $\pi(0,0) = \pi(0,1) = \pi(1,0) = \pi(1,1) = \frac{1}{4}$. Further, the predicate for the CHSH game V is:

$$V(a, b|x, y) = \begin{cases} 1 & \text{if } a \oplus b = x \land y \\ 0 & \text{if } a \oplus b \neq x \land y. \end{cases}$$
 (3)

More intuitively, the CHSH game entails two non-signalling players, Alice (A) and Bob (B). A classical referee, Charlie (C), selects bits $x,y \in \{0,1\}$ uniformly at random, sending x to A and y to B. The two players then carry out a preestablished strategy for satisfying the CHSH inequality: $a \oplus b = x \wedge y$. The score of the game, the quantum and local bounds, and optimal quantum strategy are as follows:

Definition II.27 (General CHSH Strategy & Score). Using projective measurements, where the correlations across settings are denoted by $P(a,b|x,y) = Tr(\rho(M_{a|x} \otimes M_{b|y}))$, the CHSH score (as a success probability) for any state $\rho \in \mathcal{H}^{AB}$ is:

$$\omega_{chsh} = \frac{1}{4} \sum_{x,y,a,b} V_{\rho}(a,b|x,y) P(a,b|x,y) = \frac{1}{4} \sum_{\substack{x,y,a,b\\ a \oplus b = x \land y}} Tr(\rho(M_{a|x} \otimes M_{b|y})), \tag{4}$$

where $M_{a|x}$ and $M_{b|y}$ are Alice and Bob's measurements, respectively. Alternatively, the same score is commonly computed on a shifted scale as $\frac{Tr(\rho \mathbf{S})}{8} + \frac{1}{2}$ using a Hermitian operator, $\mathbf{S}_{chsh} = A_0B_0 + A_0B_1 + A_1B_0 - A_1B_1$, with:

$$A_0 = M_{a=0|x=0} - M_{a=1|x=0}, \ A_1 = M_{a=0|x=1} - M_{a=1|x=1}$$

$$B_0 = M_{b=0|y=0} - M_{b=1|y=0}, \ B_1 = M_{b=0|y=1} - M_{b=1|y=1}.$$

Theorem II.2 (CHSH Optimal Scores). The local and quantum bounds of CHSH are 75% and 85.36%, respectively, in terms of success probabilities, and 2 and $2\sqrt{2}$, respectively, on the shifted scale attained with the observable operator [36].

Theorem II.3 (CHSH Optimal Quantum Strategy). The optimal CHSH strategy can be defined as $S = (\rho_{EPR}, \mathbf{S})$ where ρ_{EPR} is the EPR pair, up to local isometries and \mathbf{S}_{chsh} with $A_0 = \sigma_z, A_1 = \sigma_x, B_0 = \frac{\sigma_z + \sigma_x}{\sqrt{2}}$, and $B_1 = \frac{\sigma_z - \sigma_x}{\sqrt{2}}$ for Pauli matrices σ_x, σ_z [33].

D.2. 2-CHSH Game

The 2-CHSH game (Fig. 1b) is an extension of the CHSH, wherein Alice and Bob play two CHSH games simultaneously with one referee. The referee sends and receives two inputs (x_1, x_2) and outputs (a_1, a_2) from Alice and two inputs (y_1, y_2) and outputs (b_1, b_2) from Bob. The players win IFF both CHSH games win. More formally, the game is defined as follows:

Definition II.28 (2-CHSH Game). The 2-CHSH is a non-local game with two players, each receiving two questions and giving two answers to the referee, corresponding to the binary values, $\mathcal{X}_1 = \mathcal{Y}_1 = \mathcal{X}_2 = \mathcal{Y}_2 = \mathcal{A}_1 = \mathcal{B}_1 = \mathcal{A}_2 = \mathcal{B}_2 = \{0,1\},$ with uniform probability across questions, π , such that $\pi(x_1, y_1, x_2, y_2) = \frac{1}{16}$. The predicate V_2 is defined as:

$$V_2(a_1, b_1, a_2, b_2 | x_1, y_1, x_2, y_2) = V(a_1, b_1 | x_1, y_1) \wedge V(a_2, b_2 | x_2, y_2)$$
(5)

$$= \begin{cases} 1 & if (a_1 \oplus b_1 = x_1 \wedge y_1) \wedge (a_2 \oplus b_2 = x_2 \wedge y_2) \\ 0 & otherwise, \end{cases}$$
 (6)

where the game is won if and only if both constituent CHSH games are won simultaneously.

Definition II.29 (General 2-CHSH Strategy & Score). The score for the 2-CHSH can be computed using projective measurements for convenience, where $P(a_1, b_1, a_2, b_2 | x_1, y_1, x_2, y_2) = Tr(\rho(M_{a_1|x_1} \otimes M_{a_2|x_2} \otimes M_{b_1|y_1} \otimes M_{b_2|y_2}))$:

$$\omega_{2chsh} = \frac{1}{16} \sum_{x_1, y_1, x_2, y_2, a_1, b_1, a_2, b_2} V_2(a_1, b_1, a_2, b_2 | x_1, y_1, x_2, y_2) P(a_1, b_1, a_2, b_2 | x_1, y_1, x_2, y_2)$$

$$(7)$$

$$\omega_{2chsh} = \frac{1}{16} \sum_{x_1, y_1, x_2, y_2, a_1, b_1, a_2, b_2} V_2(a_1, b_1, a_2, b_2 | x_1, y_1, x_2, y_2) P(a_1, b_1, a_2, b_2 | x_1, y_1, x_2, y_2)
= \frac{1}{16} Tr \left(\rho \cdot \sum_{\substack{x_1, y_1, x_2, y_2, a_1, b_1, a_2, b_2 \\ (a_1 \oplus b_1 = x_1 \land y_1) \land (a_2 \oplus b_2 = x_2 \land y_2)}} (M_{a_1 | x_1} \otimes M_{a_2 | x_2} \otimes M_{b_1 | y_1} \otimes M_{b_2 | y_2}) \right).$$
(8)

The measurements are defined for Alice as $M_{0|0} = \frac{\mathbb{I} + A_0}{2}, M_{1|0} = \frac{\mathbb{I} - A_0}{2}, M_{0|1} = \frac{\mathbb{I} + A_1}{2}, M_{1|1} = \frac{\mathbb{I} - A_1}{2}$, and for Bob as $M_{0|0} = \frac{\mathbb{I} + B_0}{2}, M_{1|0} = \frac{\mathbb{I} - B_0}{2}, M_{0|1} = \frac{\mathbb{I} + B_1}{2}, M_{1|1} = \frac{\mathbb{I} - B_1}{2}.$

Note that in the code, we do $M_{a_1|x_1} \otimes M_{b_1|y_1} \otimes M_{a_2|x_2} \otimes M_{b_2|y_2}$ due to how the order of the qubits is stored in our code quantum objects.

Definition II.30 (2-CHSH Optimal Strategy, & Bounds). The optimal strategy for the 2-CHSH is to play the optimal CHSH strategy on each of the two CHSH instances and output a win only if both CHSH games are won (i.e., take the AND of the two outcomes). The quantum bound is ≈ 0.729 , achieved with $\rho_{EPR} \otimes \rho_{EPR}$, and the local bound is $\frac{10}{16}$ [37].

D.3. The Magic Square Game

We introduce another common game, the Magic Square Game (MSG). It is a 2-player game (Alice and Bob) with one referee (Charlie). As an overview, in the MSG, Charlie first samples 2 bits $x, y \in \mathcal{X}, \mathcal{Y}$ uniformly at random, and sends them to Alice and Bob, respectively. Next, Alice returns 3 bits, $a_{x_i} \in \{\pm 1\}$, to Charlie, and Bob returns one bit, $b_{x_n} \in \{\pm 1\}$. Lastly, the referee checks that the bits satisfy the parity and consistency checks in a predicate, V. If both conditions are satisfied, accept (win). Otherwise, reject (loss). The game and general strategy for the MSG are formally defined below.

Definition II.31 (Magic Square Game). The Magic Square Game is the 2-player, (6,3)-input, (3,1)-output non-local game where the question and answer sets are $\mathcal{X} = \{1, \dots, 6\}$, $\mathcal{Y} = \{1, 2, 3\}$, $\mathcal{A} = \{\pm 1\}^3$, $\mathcal{B} = \{\pm 1\}$, and the probability distribution across all possible questions, π , is uniform such that $\pi(x,y) = \frac{1}{18} \quad \forall x \in \mathcal{X}, y \in \mathcal{Y}$. Further, the predicate is:

$$V(a,b|x,y) = \begin{cases} 1 \ (accept) & if \ (a_{x_1}a_{x_2}a_{x_3} = c_x \ (parity \ check)) \land (a_{x_y} = b_{x_y} \ (consistency \ check)) \\ 0 \ (reject) & otherwise \end{cases}$$

where $a = (a_{x_1}, a_{x_2}, a_{x_3}) \in \mathcal{A}$ is Alice's output triple, $b \in \mathcal{B}$ is Bob's output bit, and c_x is defined as:

$$c_x = \begin{cases} -1 & if \ x = 6 \ (third \ column) \\ +1 & otherwise \end{cases}$$

Definition II.32 (General MSG Strategy & Score). The players possess some state, ρ . They define a grid of mutually commuting observables to choose from, where the products of observables in each row and column are as follows:

To obtain her three bits, Alice measures the three observables, which we denote by $O_{x_1}, O_{x_2}, O_{x_3}$, in the x'th block (row or column) on her share of ρ . To obtain his bit, Bob measures the y'th observable in block x of the grid, $O_{x_y}^T$, on his share of ρ . It is also noteworthy to mention that when considering this, the parity check can be written as $a_{x_1}a_{x_2}a_{x_3}=c_x$, where $c_xI=O_{x_1}O_{x_2}O_{x_3}$. More broadly, a scoring operator can be derived by combining Alice and Bob's measurements and all possible output combinations:

$$\mathbf{S}_{MSG} = \frac{1}{18} \sum_{x \in \{1, \dots, 6\}, a = a_{x_1}, a_{x_2}, a_{x_3}, b = b_{x_1}, b_{x_2}, b_{x_3}} A^{j}_{a_{x_1}, a_{x_2}, a_{x_3}} \otimes (B^{y=1}_{b_{x_1}} + B^{y=2}_{b_{x_2}} + B^{y=3}_{b_{x_3}}) \text{ where,}$$
(9)

$$A_{a_{x_1},a_{x_2},a_{x_3}}^x = \frac{\left(\mathbb{I} + a_{x_1}O_{x_1}\right)}{2} \frac{\left(\mathbb{I} + a_{x_2}O_{x_2}\right)}{2} \frac{\left(\mathbb{I} + a_{x_3}O_{x_3}\right)}{2}, B_{b_{x_y}}^k = \frac{\left(\mathbb{I} + b_{x_y}O_{x_y}\right)}{2}. \tag{10}$$

The expectation value of the operator with an input state ρ , $\langle S_{MSG}, \rho \rangle$, can be computed to determine the MSG score.

We note that to our knowledge, we are the first to score the MSG for noisy input states using this construction. That is, we derived the score operator as we were unable to find the use of any such scoring operator in the MSG literature. Note that in Eqn. 9, the summation loops over all possible combinations of outcomes for a, and likewise over the outcomes for b. The 1/18 comes from the uniform distribution across input settings. Also note that, by the definitions above, $O_s = B_{+1}^s - B_{-1}^s$ for $s \in \{1, \ldots, 9\}$.

Definition II.33 (MSG Local and Quantum Bounds). The optimal quantum strategy is played with two shared EPR pairs (up to local isometries): $|\psi\rangle = \frac{(|00\rangle_{12} + |11\rangle_{12})}{\sqrt{2}} \otimes \frac{(|00\rangle_{34} + |11\rangle_{34})}{\sqrt{2}} = \frac{1}{2}(|0000\rangle + |0011\rangle + |1100\rangle + |1111\rangle)$ where Alice possesses qubits 1 and 3, and Bob qubits 2 and 4. The grid for the optimal quantum strategy is defined with the observables:

where each entry (i.e., IZ) acts on one of the party's set of qubits. The first of two observables (i.e., I) acts on the measuring party's first qubit and the second (i.e., Z) acts on their second qubit. The quantum bound, attained with the above strategy, is 1 (i.e., the players win 100% of the time). Under LHV, the maximal score is 17/18 (local bound).

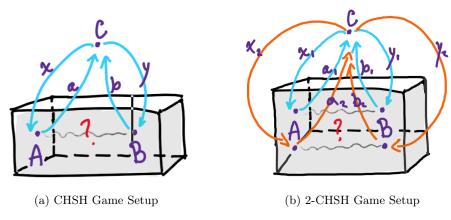


Figure 1: The CHSH game has 2 inputs and 2 outputs while the 2-CHSH game has 4 inputs and 4 outputs. What is inside the box (quantum or classical) is unknown to the referee, as these are self-tests.

III. NOISE-ROBUSTNESS FRAMEWORK FOR NON-LOCAL GAMES

We want to be able to formalize a reliable method for assessing the noise-robustness of a non-local game to noisy inputs, which, intuitively, is the ability of a game to detect non-locality in noisy input states. Ideally, we want to be able to compare the noise-robustness of games with different dimensionalities and input-output settings. In this section, we propose various methods for doing so, and introduce new definitions that prove useful for this objective along the way.

A. Noise-Tolerance

An intuitive, natural way to think about noise-robustness is to simply test whether a game is violated for a particular noisy form of the game's winning state, ρ_{win} (up to local isometries). If the noisy state violates the game, then we can say that the game is tolerant of that state. For example, in the CHSH game, the winning state, up to local isometries, is the EPR pair. Hence, we can test whether certain noisy versions of the EPR pair violate the CHSH game to conclude whether the game is tolerant of these states. Further, we can expand this intuition to study a set of noisy states defined by a noise model, which has an explicit noise level applied to the winning state. Doing so will allow for an analysis of tolerated noise levels. This leads to two key questions:

Question 4. Let us define a state, the winning state, ρ_{win} (up to local isometries), of a non-local game, G, and a noisy version of it, ρ'_{win} (up to local isometries). Does G detect the non-locality remaining in ρ'_{win} ? That is, does ρ'_{win} violate G's winning condition?

Question 5. Define an assumed noise model by a quantum channel yielding the noisy state, $\rho'_{win} = \varepsilon(\eta, \rho_{win})$, where the probability of no noise (visibility) applied to the winning state of a non-local game G is η . What are the noise levels which still permit the noisy winning state to violate G?

In the first question, we ask whether a game is tolerant of a specific noisy state, and in the second, we ask if a game is tolerant of a noise interval. Importantly, we specifically consider noisy forms of the winning state of a game as this will allow comparison of noise-tolerance across games. That is, it is not possible to compare the tolerance of explicit states between games with different input dimensionalities (i.e., the CHSH and MSG), but all games have winning states up to local isometries. Hence, we can compare the levels of deviation from the winning state that still violate games. A quantum channel model for noise helps in establishing a noise parameter to quantify the deviations from the winning state. Thus, the choice of comparing channel noise levels applied to winning states helps to standardize a way for comparing the noise-tolerance between different games. We formally define the notion of noise-tolerance as follows:

Definition III.34 (Noise-tolerance of a non-local game). Define a non-local game G with a local bound ω_{c_G} , and let \mathbf{S}_G denote some fixed observable for the player's strategy in G. Further, let us define $\rho_{ref} \in \mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ to be an n-qubit state, and $\Phi(\rho_{ref}, \eta)$ to be a noisy quantum channel, where η is the visibility affecting ρ_{ref} . Then G is (η', Φ) -tolerant if $Tr(\mathbf{S}_G \Phi(\eta', \rho_{ref})) > \omega_{c_G}$, for some η' , noise model ε .

That is, G is (η', Φ) -tolerant if the strategy $S_G = (\rho' = \Phi(\eta', \rho_{ref}), \mathbf{S}_G)$ violates the maximal classical value (ω_{c_G}) , for some η' and noise model Φ . Note that we leave ρ_{ref} to be a general state in the definition so that it can be flexibly chosen for the desired question or analysis. However, for our purposes we are concerned with $\rho_{ref} = \rho_{win}$ for the reasons stated above. Let us take an example of applying the definition. We can fix a game, G_1 , to be the CHSH game. Its optimal state, ρ_{win_1} , is the EPR pair. Then, we could analyze the violations of $\omega_{c_{G_1}}$ yielded by strategies $S_{G_1}^i = (\rho_i = \Phi(\eta_i, \rho_{win_1}), \mathbf{S}_{G_1})$, for varying visibilities $\eta_i \in \{\eta_1, \eta_2, \dots, \eta_n\}$ in a given noise model Φ . Consequently, we would attain information about whether G_1 is (η_i, Φ) -tolerant. Taking $\eta_{min} = (min\{\eta_1, \eta_2, \dots, \eta_l\} \text{ s.t. } G_1 \text{ is } (\eta_i, \Phi)$ -tolerant), we can claim that G_1 is at least (η_{min}, Φ) -tolerant. Let us now define the protocol for comparing the noise-tolerance of two different non-local games:

Definition III.35 (Protocol for comparing noise-tolerance of two games). Given two non-local games, G_1 and G_2 , each with local bound $\omega_{c_{G_1}}$ and $\omega_{c_{G_2}}$, respectively, let $\mathbf{S}_{G_1}, \mathbf{S}_{G_2}$ denote some fixed Hermitian operator for the players' strategies in G_1, G_2 , respectively. Let us also define $\rho_{win_1} \in \mathcal{H} = (\mathbb{C}^2)^{\otimes n}$ to be the n-qubit winning state of G_1 , and $\rho_{win_2} \in \mathcal{H} = (\mathbb{C}^2)^{\otimes m}$ to be the m-qubit winning state of G_2 .

Further, let $\rho'_1 = \Phi(\eta, \rho_{win_1})$ and $\rho'_2 = \Phi(\eta, \rho_{win_2})$ be the noisy input states with visibility level η to G_1 and G_2 , respectively. If the games have different dimensionalities for the input Hilbert space (i.e., m > n), design an extension from the n-dimensional channel to the m-dimensional one, which preserves a clear relationship between the two channel's noise parameters and is CPTP. To compare the noise-tolerance of G_1 and G_2 , for some η^* :

- 1. Decide $Tr(\mathbf{S}_{G_1} \Phi(\eta^*, \rho_{win_1})) > \omega_{c_{G_1}}$.
- 2. Decide $Tr(\mathbf{S}_{G_2} \Phi(\eta^*, \rho_{win_2})) > \omega_{c_{G_2}}$.

If one game is tolerant to its noisy state and the other is not, the former is more η^* -tolerant of its winning state than the other, under the noisy channels defined. Otherwise, they are both η^* -tolerant of their winning states under the noisy channels defined.

Put more simply: if both conditions 1 and 2 are satisfied, then the games are tolerant to η^* noise on their winning state under the noisy channel. Otherwise, one is more tolerant to η^* noise. Note that the noise level applied when checking conditions 1 and 2 should be the same to both channels, regardless of whether one is an extension of the other or not. This allows us to observe the effect of noise on larger dimensions or cases where games have parallel repetition.

As can be seen in Def. III.35, comparing the noise-tolerance of two games whose winning states have different dimensionalities can be complex. The situation becomes simpler if the winning state of the higher dimensional game is a tensored repetition of the winning state of the lower dimensional game. Let us illustrate this with an example. Say we wish to compare between the noise-tolerance of the CHSH and MSG games towards some quantum channel Φ . The CHSH has a 2-qubit winning state (a single EPR pair), and the MSG a 4-qubit winning state (two EPR pairs), so we choose: $\rho_{ref}^{msg} = \rho_{EPR} \otimes \rho_{EPR}$ and $\rho_{ref}^{chsh} = \rho_{EPR}$. In this case, it is natural to define the noise-tolerance comparison as:

$$\operatorname{Tr}(\mathbf{S}_{CHSH} \Phi(\eta, \rho_{EPR})) > \omega_{c_{chsh}}$$
 (11)

$$\operatorname{Tr}(\mathbf{S}_{MSG} \Phi(\eta, \rho_{EPR}) \otimes \Phi(\eta, \rho_{EPR})) > \omega_{c_{msg}}$$
 (12)

The channel for the MSG is simply the tensor of two noisy CHSH inputs. The channels studied in this thesis, the 2-and 4-qubit depolarizing channels (Defs. II.12 and II.13, respectively, with $\rho = \rho_{EPR}$), are based on this tensor product extension. By this construction, if both Eqns. 11 and 12 are satisfied $\forall \eta$, then both games are tolerant to Φ -type noise on the EPR state. Otherwise, one game is more tolerant; that is, it can detect the non-locality of a Φ -noisy EPR state better than the other. Note also that in this simple EPR extended case, in considering CHSH and MSG, we can see if increasing the number of copies of a noisy resource available in a round may improve the ability to detect non-locality in a noisy EPR pair. Later, we also consider a parallel repetition of the CHSH game, the 2-CHSH game. Further, note that when the noise-model is clearly apparent, the Φ specification in the noise-tolerance description is dropped.

B. Convincingness

B.1. Motivation

The second measure of noise-robustness is more specifically designed to capture how efficient non-local games are for detecting non-locality in noisy inputs. There were three main objectives when designing this measure:

- 1. The measure should be a general, continuous score that is **comparative** among games which have different numbers of inputs, outputs and input state dimensions. For example, it should provide a continuous scale for comparing and ranking the CHSH, 2-CHSH, and MSG's robustness to noisy inputs. This is highly non-trivial and does not yet exist. The challenge is that the scores of different games are on completely different scales which are normalized differently, and the games are in different Hilbert Space dimensions.
- 2. The measure should have an operational meaning tied to noise-robustness of each non-local game.
- 3. The measure should be possible to assess for **finite and infinite resources** of noisy states. This comes from both practical motivation, to understand which game is most efficient for experimentalists to utilize for non-locality detection in noisy resources, and from a theoretical motivation to pave way for understanding the asymptotic strengths or limitations of noise-robustness in different games.

B.2. Construction

We call the new measure the *convincingness* of a non-local game. Intuitively, the convincingness characterizes the likelihood that statistics observed from n trials of a self-test have occurred due to locality (i.e., the likelihood that the statistics are due to maximal or high noise levels in the input state). This measure processes two metrics, (1) the local bound of a game and (2) the success probability of a given input state to the game, into a statistical measure that is comparable amongst different games. One can think of this new score as normalizing the violations of each game to the same scale, allowing a more fair comparison of games' abilities to detect non-locality than simply comparing raw violations. We aim for the convincingness to be a common-ground score which can be computed in a normalized manner amongst all games.

Let us now describe how the convincingness score is constructed. Its definition is based on the p-value, which is a statistical tool that has been used to present very strong evidence against local realism in loophole-free Bell Inequality experiments [4]. In particular, it has been used to characterize the strength (or significance) of experimental Bell violations results. The formal definition of the p-value can be found in Def. II.16. Intuitively, the p-value represents the probability that observed results occurred due to a pre-defined null hypothesis. A low (upper-tailed) p-value indicates that it is highly improbable for our results, or more extreme ones, to occur under the assumption that the null hypothesis is true. If the likelihood is below some significance threshold, α , this suggests that the results may be due to an effect unaccounted for by the null hypothesis (there is significant evidence against the null hypothesis). So, practically, the p-value computes the likelihood that an observed output statistical distribution can be explained by an assumed statistical distribution.

It is natural to extend this to self-testing, where we deal with statistical distributions of physical systems. The construction is as follows. Say that we run n rounds of a non-local game, G, noting down the violation measured for each of the n rounds. Then, for each trial $t \in \Omega$, define the win/loss indicator $W: \Omega \to \{0,1\}$ as W(x,y,a,b) = V(a,b|x,y). Under the null hypothesis of local realism, $W \sim \text{Bernoulli}(\omega_{c_G})$, while for a state with visibility η and success probability ω_{v_G} , $W \sim \text{Bernoulli}(\omega_{v_G})$. Hence, we may compare the likelihood that the outcome distribution was explained due to a local distribution. Upon performing a binomial test to compare the observed number of successes, k, from Binomial (n, ω_{v_G}) to what we would expect under Binomial (n, ω_{c_G}) , we can take an upper-tailed test: sum over the binomial test statistic for each possible number of successes $j \in [k, n]; j \in \mathbb{Z}$, yielding the convincingness of a non-local game in Eqn. 14:

$$P(X \ge k|H_0 = \omega_{c_G}) = \sum_{j=k}^{n} P(X = j|H_0 = \omega_{c_G})$$
(13)

$$\iff P(X \ge \text{round}(n \cdot \omega_{v_G}) | H_0 = \omega_{c_G}) = \sum_{\text{round}(n \cdot \omega_{v_G})}^n \binom{n}{j} \omega_{c_G}^j (1 - \omega_{c_G})^{n-j}$$
(14)

```
Algorithm 1: Convincingness Calculation for Non-Local Games
```

```
Input: Number of trials n, classical bound \omega_c, observed success probability \omega_v
Output: Convincingness \mathfrak{C}_G (p-value for non-local game, G)
Let X be a random variable where:
X=1 represents a game violation (success)
X = 0 represents no violation (failure)
// Define probabilistic model under null hypothesis
Set \omega_c as bias under H_0 (classical bound of G)
X \sim \text{Bernoulli}(\omega_c) \text{ where } P(X=1) = \omega_c
// Define probabilistic model under alternative hypothesis
Set \omega_v as bias under H_1 (observed success probability for some input to G)
X \sim \text{Bernoulli}(\omega_v) \text{ where } P(X=1) = \omega_v
// Sample from binomial distribution
Draw n independent samples X_1, X_2, \ldots, X_n with probability \omega_v
Let k be the number of successes observed in these n trials
Then, k is a realization of the random variable K \sim \text{Binomial}(n, \omega_v)
Expected successes: k \sim \mathbb{E}[X] = n \cdot \omega_v
// Compute upper-tailed p-value
\mathfrak{C}_G = P(X \ge \text{round}(n \cdot \omega_v)|H_0 = \omega_c)
  \sum_{j=\text{round}(n\cdot\omega_v)}^n \binom{n}{j} \omega_c^j (1-\omega_c)^{n-j}
if \mathfrak{C}_G \leq 0.05 then
| Reject null hypothesis of locality
end
```

Concretely, this sum gives the probability of observing k or more game violations under a local model. This convincingness test also provides an easy check: if the likelihood (Eqn. 14) is below a certain significance level ($\leq \alpha$), this indicates that the observed data is sufficiently inconsistent with the null hypothesis of locality for it to be rejected in favour of non-locality. We define $\alpha = 0.05$ as this is the standard significance level for the p-value across different disciplines. This idea is formally defined in Def. III.36 and the full computation is defined in Algorithm 1. By our definition and setup, if $\mathfrak{C}_G \leq 0.05$, the results of the non-local game are significantly unlikely to be due to locality, and hence, are highly likely to be due to non-locality (entanglement). In this case, we would therefore say that G is a convincing game for detecting entanglement. Hence, convincingness is a simple yet powerful extension of the p-value definition to non-local games.

Definition III.36 (Convincingness). Define G to be a non-local game with local bound ω_{c_G} , and let a measured violation in one round of G with a particular strategy (trial) be denoted by ω_{v_G} . The convincingness, \mathfrak{C}_G , of G is the upper-tailed g-value with the Null Hypothesis, g-value with the Null Hypothesis, g-value g-val

Can one game be more or less convincing than the other? When looking at the bare p-value, the lower the p-value, the greater the statistical significance of the outcome. Hence, for games G_1, G_2 , if $\mathfrak{C}_{G_1} < \mathfrak{C}_{G_2}$, then G_1 is more 'convincing' of non-locality (and hence, entanglement) than G_2 .

By extension, we later focus on applying this to mixed states—that is, we will use the convincingness score to study which games provide more evidence of non-locality in noisy states.

B.3. Simplified Convincingness

We can further simplify the equation for convincingness by borrowing knowledge about the behaviour of the p-value as in Theorem III.4. The proof follows.

²Note that the Alternative Hypothesis can come from any input state, so this is a general construction and not specific to knowing a noise model with visibility η .

Theorem III.4 (Convincingness Bound). For a non-local game G with $\omega_{c_G}, \omega_{v_G}, n$ as in Def. III.36, and $\omega_{c_G} \neq \omega_{v_G}$, we can bound $\mathfrak{C}_G \sim \exp\left(-n\left(\omega_{v_G} - \omega_{c_G}\right)^2\right)$, where we define $G(n) \sim H(n)$ to denote that the function G(n) scales proportionally to some function H(n) for asymptotically large n.

Proof. It is known that the lower-tailed p-value is upper bounded as follows:

$$F(k; n, p) = \Pr(X \le k) \le \exp\left(-2n\left(p - \frac{k}{n}\right)^2\right),\tag{15}$$

which extends to the following for the upper-tailed p-value:

$$F(n-k; n, 1-p) = \Pr(X \ge k) \le \exp\left(-2n\left(\frac{k}{n} - p\right)^2\right). \tag{16}$$

Plugging in the null and alternative hypotheses for our construction of the convincingness metric, we attain:

$$F\left(n - \text{round}(n \cdot \omega_{v_G}); n, 1 - \omega_{c_G}\right) \le \exp\left(-2n\left(\frac{\text{round}(n \cdot \omega_{v_G})}{n} - \omega_{c_G}\right)^2\right)$$
(17)

$$= \exp\left(-2n\left(\omega_{v_G} - \omega_{c_G}\right)^2\right). \tag{18}$$

Let $G(n) \sim H(n)$ denote that the function G(n) scales proportionally to some function H(n) for asymptotically large n. Then assuming $\omega_{c_G} \neq \omega_{v_G}$, we can accurately approximate the convincingness for $n \to \infty$:

$$\mathfrak{C}_G \sim \exp\left(-n\left(\omega_{v_G} - \omega_{c_G}\right)^2\right),\tag{19}$$

up to a universal constant factor in the exponent, where the rounding error in line (17) is negligible. We ignore the factor of 2 in line (18) as it is constant for all games and hence does not affect their comparative scaling.

B.4. Noise-Robustness Framework for Comparative Scores

Given that we now have a continuous, comparative score across different games (\mathfrak{C}_G) , we now use it to develop a framework for studying the noise-robustness of games to noisy input states, and particularly, to noisy channels on the input. Let us first understand a nuance which we glossed over in the prior section. While we have been describing \mathfrak{C}_G as the convincingness of a 'game', both the 'game' and 'strategy' are hiding within $\omega_{v_G} = \text{Tr}(\mathbf{S}_G \rho_{v_G})$ and $\omega_{c_G} = \text{Tr}(\mathbf{S}_G \rho_{c_G})$ of \mathfrak{C}_G . That is, the state ρ and the measurement operators of S come purely from the strategy $S = (\rho, \mathcal{M}_1, \mathcal{M}_2)$. In contrast, the structure of S-the weights $\pi(x,y)$ and the coefficients (on the measurements) derived from V(a, b|x, y)-comes purely from the game definition $G = (\mathcal{X}, \mathcal{Y}, \mathcal{A}, \mathcal{B}, \pi, V)$. This means that there are four axes, assuming that the input-output settings and number of players defining the game are fixed, as in Fig. 2. We call a fixed choice of the four axes a configuration

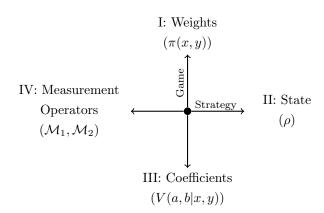


Figure 2: Optimization axes for non-local games, illustrating the four adjustable components of a non-local game configuration. That is, we can optimize or evaluate over the following–I: The weights, $\pi(x,y)$, of the measurements in \mathbf{S} , II: Quantum state ρ , III: The coefficients derived from V(a,b|x,y) for the operators in \mathbf{S} , and IV: The measurement operators $\mathcal{M}_1, \mathcal{M}_2$ used to construct \mathbf{S} .

of a non-local game. This framework provides a natural decomposition for understanding noise-robustness in non-local games: we can conceptualize it both through the analysis of game performance across varying noise levels and through the systematic optimization of game parameters for a single or set of noisy states. Specifically, we can do the following.

- 1. Analysis across noise levels (Fig. 3a): We can vary the state axis (II) while holding all other axes constant. That is, first establish a configuration for which you wish to test noise-robustness. Second, evaluate the comparative score of the configuration at different input state noise levels. In this work, we choose to vary the states according to a quantum noise channel (depolarizing channel), and choose \mathfrak{C}_G for the score as it is comparable among different games and considers the quantity of resources physically available. However, other noisy channels or more complex noise models can be chosen to define the state axis, and other scores can be used if they are comparable among games. By evaluating configurations across different noise levels, this analysis can elucidate differences in robustness: some configurations may only detect entanglement in highly entangled states while failing with weakly entangled states. In contrast, particularly robust configurations can identify entanglement even in weakly entangled states.
- 2. Optimization for specific noise regimes (Fig. 3b): We can select a specific noisy state or regime (set of states) to define the state axis on (II), and optimize one or more of the axes for that noisy regime. For example, by optimizing the operator coefficient axis (III) for a fixed η' -noisy state (see Fig. 3b), we can identify configurations particularly well-suited to specific noise levels, potentially revealing new configurations with enhance noise-robustness. These optimized configurations can then be evaluated for their robustness across noise levels using the first approach.

The convincingness measure also provides an analytical framework to understand this behaviour. For weakly entangled states, ω_{v_G} approaches ω_{c_G} , resulting in $\mathfrak{C}_G \to 1$, which mathematically captures their consistency with local behaviour. Conversely, highly entangled states drive ω_{v_G} toward its maximal quantum value, resulting in $\mathfrak{C}_G \to 0$ and indicating strong non-local behaviour. For an arbitrary input state ρ_{in} , this relationship can be expressed as: $\mathfrak{C}_G \sim \exp\left(-n\left(\text{Tr}(\mathbf{S}_G\rho_{in}) - \omega_{c_G}\right)^2\right)$. Note that if we wish to specify a particular noise level η for some noise model, we can substitute an η -noisy input ρ_{η} , which achieves the value $\omega_{\eta_G} = \text{Tr}(\mathbf{S}_G\rho_{\eta})$. Moreover, in later sections we derive a single numeric, continuous score based on the simplified \mathfrak{C}_G , in order to rank the convincingness behaviour, and hence, noise-robustness, of different games.

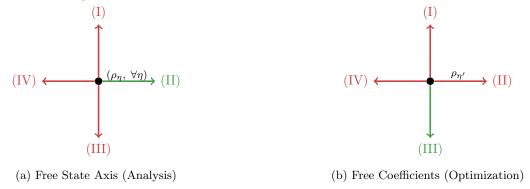


Figure 3: Experiments for evaluating noise-robustness with a comparative score. Illustrating two ways to look at how a game can be robust to noise: robust across a region, or robust for a particular noise type. Green axes are varied and Red axes are fixed. 3a (left): Fixing known (i.e., optimal) weights, coefficients, and measurements while varying the visibility η of the input state ρ_{η} . 3b (right): Optimizing the coefficients for a pre-selected state with visibility η' , fixing the rest of the axes. The resulting configuration from 3b can be analyzed using the method in 3a.

IV. COMPUTATIONAL METHODS

We demonstrate how to compare the noise-robustness of 2-player games using our noise-robustness framework for comparative scores, and the noise-tolerance. The games we focus on as an example are the CHSH, 2-CHSH, MSG, and optimized 2-CHSH variants (defined in Sec. B.1.). Comparing this set is natural as we can see if parallel repetition of the CHSH (i.e., the 2-CHSH) helps to improve its noise-robustness, and whether the 2-CHSH or its optimized versions are comparable to the MSG which uses the same amount of quantum resources (qubits) in its optimal quantum strategy. We expect these experiments to also provide a good benchmark of the quality of the convincingness for comparing the noise-robustness of games which have different number of inputs, outputs, and input state dimensions. Using the convincingness we can

use this analysis to discern whether the non-locality in noisy, partially depolarized states are best detected in the CHSH, 2-CHSH, MSG, or optimized 2-CHSH games. Overall, computationally we do the following:

- 1. Evaluate the convincingness for states with different noise levels, for different games, and compare which game permits more noisy states to be convincing. Plots of such experiments will be shown in Section VI..
- 2. Search for 2-CHSH configurations which are optimal for particular noise levels, included as well in the plots of (1).
- 3. Analyze the convincingness of non-local games at fixed and variable, finite and near-infinite resource regimes.
- 4. Evaluate and plot the noise-tolerance of non-local games, relating these results to convincingness results.

A. Analysis Across Noise Levels

Let $NLG = \{\text{CHSH, MSG, 2-CHSH, 2-CHSH-OPT}\}$, where the first three games are as defined in Sec. C., and the 2-CHSH-OPT will be momentarily defined. Also let ρ_{η} be some η -noisy input to $G \in NLG$ for a noise model with visibility η (probability of noise $1 - \eta$), yielding score $\omega_{\eta_G} = \text{Tr}(\mathbf{S}_G \rho_{\eta})$. \mathbf{S}_G is the operator for G composed of the measurement coefficients, and weights which achieve optimal score for a maximally entangled state $\rho_{\eta=1}$ (up to local isometries).

In this experiment, we fix the configuration of measurements, coefficients, and weights (axes IV, III, I) to yield the maximal score at $\rho_{\eta=1}$ for each game. Further, we choose the noise model on the input state to be the depolarizing channel, defining ρ_{η} for the 2-qubit and 4-qubit Hilbert subspaces, respectively, as follows:

$$\varepsilon_{\eta}^{2-qubit}(\rho_{EPR}) = \eta^{2}\rho_{EPR} + (1-\eta^{2})\frac{\mathbb{I}_{4}}{4}, \quad \varepsilon_{\eta}^{4-qubit}(\rho_{EPR}) = \eta^{4}(\rho_{EPR}\otimes\rho_{EPR}) + \eta^{2}(1-\eta^{2})(\rho_{EPR}\otimes\frac{\mathbb{I}_{4}}{4}) + \eta^{2}(1-\eta^{2})(\frac{\mathbb{I}_{4}}{4}\otimes\rho_{EPR}) + (1-\eta^{2})^{2}(\frac{\mathbb{I}_{4}}{4}\otimes\frac{\mathbb{I}_{4}}{4}).$$

We also now define \mathfrak{C}_G specifically for hypotheses $H_0 = \omega_{c_G}$ (local bound of G) and $H_1 = \omega_{\eta_G}$ (score with input ρ_{η}). As this is the convincingness we'll be dealing with from now on, we often refer interchangeably to \mathfrak{C}_G and $\mathfrak{C}_G(\eta)$.

With this setup, we take the operator \mathbf{S}_G and compute \mathfrak{C}_G for all $\eta \in [0,1]$. For each game, this produces a curve $\{(\eta,\mathfrak{C}_G) \mid \forall \eta \in [0,1]\}$, and all curves are plotted for comparison. Importantly, the order in which the curves cross the significance threshold α can be determined from this plot, revealing the maximal noise levels on ρ_{EPR} and $\rho_{EPR} \otimes \rho_{EPR}$ at which each game is convincing of non-locality.

Hence, the above experiment aims to answer: How does \mathfrak{C}_G behave as the noise level of a state ρ is varied? Which game is convincing (crosses $\mathfrak{C}_G = \alpha$), for a higher amount of noise (lower visibility, η)? Answering the latter reveals which game is most noise-robust by the convincingness measure. Beyond these questions, the convincingness enables our study of how the amount of resources available affects each game's likelihood of detecting non-locality. That is, given a supply of identical noisy resources (N_{res}) to play each game, which game is most convincing for the highest noise? We study this question both at finite and near-infinite N_{res} regimes. Note that in our experiments, N_{res} defines the number of noisy EPR pairs available. Hence, games with one or two EPR pairs as their winning state can play N_{res} or $N_{res}/2$ rounds of the game, respectively, if we restrict them to use the same number of resources.

B. Optimization for Specific Noise Regimes

We demonstrate how one can optimize a non-local game for noisy states, using the 2-CHSH game as an example. We choose to optimize the 2-CHSH game since, as will be seen in Sec. VI., the 2-CHSH shows worse noise-robustness than the CHSH but better noise-robustness than the MSG. Hence, the objective of the optimization is to show whether optimizing an axis of the 2-CHSH game's configuration can help improve upon and potentially surpass the CHSH's noise-robustness.

B.1. Optimized 2-CHSH Construction

The optimization is based on the work from István et. al. [38], extending their linear programming method to find the optimal Bell coefficients (predicate, axis III) that will maximize the achievable quantum violation of the 2-CHSH when

subject to a specific level of noise (η') in the input state (axis II) (see Sec. B.2. for details on the LP). In theory, thinking about noise-robustness, sequential optimizations along multiple axes can be done while fixing the state axis in order to find a truly optimal configuration. In addition, this method is extendable to optimizing over a noisy regime defined by a noise interval on η . However, for simplicity and illustrative purposes, we only optimize over the coefficient axis for a specified noisy state (Fig. 3b). This is also a good opportunity to illustrate the convenience of the noise-robustness framework. Looking to the game axes, we can concisely say that [38] optimizes over the coefficients (axis III) for specified measurements on axis IV, whereas we now optimize the coefficients (axis III) for specified noisy states on axis II holding all other axes constant. Note that we will refer to configurations resulting from this process as Optimized 2-CHSH configurations.

Let us now define the Optimized 2-CHSH Game (2-CHSH-OPT). As before, the game has two players and a referee, with $\mathcal{X}_1 = \mathcal{Y}_1 = \mathcal{X}_2 = \mathcal{Y}_2 = \mathcal{A}_1 = \mathcal{B}_1 = \mathcal{A}_2 = \mathcal{B}_2 = \{0,1\}$ and uniform $\pi(x_1, y_1, x_2, y_2) = \frac{1}{16}$. The measurements for the strategy are as defined for the 2-CHSH game. The measurements, are thus, constructed as follows.

Definition IV.37 (2-CHSH-OPT Strategy). Let us define the single-qubit measurement settings of Alice and Bob to be $A_{a_k}^{x_k}$ and $B_{b_k}^{y_k}$, respectively. Alice's single-qubit basis is defined by σ_x and σ_z , and Bob's by $\sigma_{x+z} = \frac{\sigma_x + \sigma_z}{\sqrt{2}}$ and $\sigma_{x-z} = \frac{\sigma_x - \sigma_z}{\sqrt{2}}$. From these, we can construct a set of projectors:

$$A_0^+ := \frac{\mathbb{I}_2 + \sigma_x}{2}, \quad A_0^- := \frac{\mathbb{I}_2 - \sigma_x}{2}, \quad A_1^+ := \frac{\mathbb{I}_2 + \sigma_z}{2}, \quad A_1^- := \frac{\mathbb{I}_2 - \sigma_z}{2}$$

$$B_0^+ := \frac{\mathbb{I}_2 + \sigma_{x+z}}{2}, B_0^- := \frac{\mathbb{I}_2 - \sigma_{x+z}}{2}, B_1^+ := \frac{\mathbb{I}_2 + \sigma_{x-z}}{2}, B_1^- := \frac{\mathbb{I}_2 - \sigma_{x-z}}{2}.$$

Combining these 1-qubit projectors, Alice and Bob's 2-qubit projective measurements are defined, respectively, as:

$$\begin{split} \mathbb{A} &= [(A_0^+ \otimes A_0^+), (A_0^+ \otimes A_0^-), (A_0^- \otimes A_0^+), (A_0^- \otimes A_0^-), (A_0^+ \otimes A_1^+), (A_0^+ \otimes A_1^-), (A_0^- \otimes A_1^+), (A_0^- \otimes A_1^-), \\ & (A_1^+ \otimes A_0^+), (A_1^+ \otimes A_0^-), (A_1^- \otimes A_0^+), (A_1^- \otimes A_0^-), (A_1^+ \otimes A_1^+), (A_1^+ \otimes A_1^-), (A_1^- \otimes A_1^+), (A_1^- \otimes A_1^-)], \\ \mathbb{B} &= [(B_0^+ \otimes B_0^+), (B_0^+ \otimes B_0^-), (B_0^- \otimes B_0^+), (B_0^- \otimes B_0^-), (B_0^+ \otimes B_1^+), (B_0^+ \otimes B_1^-), (B_0^- \otimes B_1^+), (B_0^- \otimes B_1^-), (B_0^+ \otimes B_1^-), (B_0^- \otimes B_1^-), (B_0^- \otimes B_1^-), (B_0^- \otimes B_1^-), (B_0^+ \otimes B_0^-), (B_0^+ \otimes B$$

Moreover, the general input state for the game is $\rho \in \mathbb{C}^{16 \times 16}$.

We have now fixed the measurement configurations. While the measurements would typically be used to construct the score operator ($\mathbf{S}_{2chsh,\eta'}$) using a set of weights and coefficients, the score is computed using the output of the linear program, which is in a different format. The linear program outputs a Bell matrix, $\mathbf{B}_{\eta'}$, whose elements are the coefficients which have been optimized for a fixed η' -noisy input; see Sec. B.2. for its construction. The score is computed by multiplying $\mathbf{B}_{\eta'}$ as in Def. IV.38 with a probability matrix, \mathbf{P}_{η} , which encodes the strategy (state and measurement) and weights axes. In this way, we consider both the predicate, the weights, and the strategy.

Definition IV.38 (2-CHSH-OPT Score). Given a 2-CHSH game whose coefficients have been optimized for an η' -noisy input, its score is defined as:

$$\omega_{\eta_{2chsh,\eta'}} = Tr(\mathbf{P}_{\eta}\bar{\mathbf{B}}_{\eta'}^T) \text{ for } \bar{\mathbf{B}}_{\eta'} \coloneqq \frac{\mathbf{B}_{\eta'} - \mathbf{B}_{\eta^{min}}^{min}}{(\mathbf{B}_{\eta'} - \mathbf{B}_{\eta'}^{min})^{max}}.$$
 (20)

where \mathbf{X}^{max} and \mathbf{X}^{min} denote the max and min elements of a matrix, \mathbf{X} , respectively. $\mathbf{P}_{\eta} \in \mathbb{R}^{16 \times 16}$ is a probability matrix for which each element is the probability of an outcome, given an η -noisy input, $\rho_{\eta} \in \mathbb{C}^{16 \times 16}$, and measurement settings $A_i = A_{a_1}^{x_1} \otimes A_{a_2}^{x_2}, B_j = B_{b_1}^{y_1} \otimes B_{b_2}^{y_2}$: $\mathbf{P}_{\eta}^{element} = P(a_1, a_2, b_1, b_2 | x_1, x_2, y_1, y_2) = Tr(\rho_{\eta}(A_{a_1}^{x_1} \otimes A_{a_2}^{x_2} \otimes B_{b_1}^{y_1} \otimes B_{b_2}^{y_2})) = Tr(\rho_{\eta}(A_i \otimes B_j))$. The whole matrix is defined as follows with the uniform probability of input questions $\pi = 1/16$:

$$\mathbf{P}_{\eta} = \pi \cdot \sum_{i,j=1}^{16} Tr(\rho_{\eta}(A_i \otimes B_j)) |e_i\rangle\langle e_j|. \tag{21}$$

The matrix $\mathbf{B}_{\eta'} \in \mathbb{R}^{16 \times 16}$ is the Bell matrix, which encodes the coefficients used to combine the expectations in \mathbf{P}_{η} .

We make a few comments about the score construction. Note that η represents the noise level of the strategy being tested by the game, and η' represents the visibility that the game's coefficients have been optimized for. In our experiment, we once again define the noise model on the state to be the depolarizing channel on the input space: $\rho_{\eta} = \varepsilon_{\eta}^{4-qubit}(\rho_{EPR})$. Note also that the affine transform on $\mathbf{B}_{\eta'}$ is added after the linear program for our purposes, to ensure that the final score in Eqn. 20 is in [0,1] and, hence, can be used in the convincingness computation. Further, note that the probability matrix is constructed by summing over all possible measurement setting combinations (16×16) . This is important for the normalization of the score as well. It is also worth mentioning that the indexing of the projector lists in Def. IV.37 matters for the code and how $\mathbf{P}_{\eta}^{element}$ is constructed. Measurement A_i corresponds to the i'th element of $\mathbb A$ and measurement B_j corresponds to the j'th element of $\mathbb B$. For example, $A_0 \otimes B_1 = (A_0^+ \otimes A_0^+) \otimes (B_0^+ \otimes B_0^+)$.

Finally, by Def. IV.39, the local bound of a 2-CHSH-OPT game can be simply computed given the game's Bell matrix and the quantum bound can be simply computed given both the game's Bell matrix and $\mathbf{P}_{\eta=1}$.

Definition IV.39 (2-CHSH-OPT Local and Quantum Bounds). For a 2-CHSH-OPT configuration, the quantum bound is $\omega_{\eta_{2chsh,\eta'}}$ for $\eta=1$. The local bound is the maximum score achieved by deterministic strategies using the game's $\bar{\mathbf{B}}_{\eta'}$ (see App. XV.). By convexity, this maximum over deterministic strategies completely characterizes the local bound.

B.2. Linear Program

Now that we have defined the construction for an optimized 2-CHSH game, we may understand how the linear program (LP) is set up to optimize it for a particular noise level in our noise model. We modify the LP and code from [38]. In [38], the LP maximizes the achievable quantum violation of the 2-CHSH when subject to noise in the measurements. Given a probability matrix $\mathbf{P}(\gamma)$ defined for γ noise on measurements (detection efficiency), the LP outputs the optimized Bell coefficients. As $\mathbf{P}(\gamma)$ is defined both by the input state and measurements, we set perfect detection efficiency and encode our depolarizing noise model, $\varepsilon_{\eta}^{4-qubit}(\rho_{EPR})$, in the probability matrix for a given visibility η' as in Eqn. 21. So, $\mathbf{P}_{\eta'} := \mathbf{P}(\gamma = 1, \eta')$. The objective function and constraints are then the same as [38], with a modified probability matrix:

minimize
$$-\sum_{i,j=1}^{16} \mathbf{P}_{\eta'}^{ij} c_{ij}$$
subject to
$$\sum_{i,j=1}^{16} L_k^{ij} c_{ij} \ge 1 \quad \forall k \in \{1,\dots,4^8\}$$

$$-1 \le c_{ij} \le 1 \quad \forall i,j \in \{1,\dots,16\}$$

where the optimization variables c_{ij} represent $\mathbf{B}_{\eta'}^{ij}$, the coefficients for a measurement combination, where i, j range from 1 to 16 to cover all of Alice and Bob's possible measurements, respectively. The constraints ensure that no local strategy violates the Bell inequality, using the vector of local vertices L to encode all deterministic LHV strategies. The code is written in MATLAB [39], which we translated and extended into Python [40].

C. Noise-Tolerance Analysis

We compute a simple noise-tolerance analysis for all the games tested. For each game $G \in \{CHSH, MSG, 2\text{-}CHSH, 2\text{-}CHSH\text{-}OPT Games}\}$, the setup is as follows:

- 1. For CHSH, set the reference state for the noise-tolerance to be its the winning state, ρ_{EPR} , and hence the depolarized input is $\rho_{\eta} = \varepsilon_{\eta}^{2-qubit}(\rho_{EPR})$. For the other 4-dimensional games, use the tensor extended depolarized input $\rho_{\eta} = \varepsilon_{\eta}^{4-qubit}(\rho_{EPR})$. The strategy is then: $S_{G,\eta} = (\rho_{\eta}, \mathbf{S}_G)$ for the input state corresponding to G.
- 2. Compute a set of scores, $\omega_{\eta_G} = \text{Tr}(\rho_{\eta} \mathbf{S}_{\mathbf{G}})$, over the interval $\eta \in [0, 1]$.
- 3. Plot the η values versus the scores for the game, producing a curve.
- 4. Find the η at which the game crosses its local bound, ω_{c_G} . Let us rename this η to η^* . Then, the game is η^* -tolerant.

We can plot all curves on the same axis, made possible by the fact that we evaluate different games for the same η values in corresponding noise models. The tolerances of the games can then be compared, allowing us to determine which game is more tolerant to depolarizing noise. That is, we may determine which game detects non-locality for greater depolarization.

V. ANALYTIC METHODS

We also develop a more theoretical analysis. Specifically, we analytically derive a single numeric score which we coin, the gapped score, which can be mapped to each game to indicate the order in which games pass the significance threshold. The score's structure not only helps to dissect the reasons for the observed results, but also allows to have a more convenient measure for noise-robustness of a non-local game. It allows us to have a single number to rank the convincingness behaviour of a non-local game across all $\eta \in [0, 1]$. Moreover, we offer a method for performing a more nuanced analysis for games which have very similar convincingness behaviour and hence similar gapped scores.

A. Gapped Expressions

We start with the simplified convincingness expression for an η -noisy state, which reflects the behaviour of the convincingness for high n: $\mathfrak{C}_G \sim \exp\left(-n\left(\omega_{\eta_G}-\omega_{c_G}\right)^2\right)$. In particular, we note that lower \mathfrak{C}_G corresponds to a higher gap: $\Delta_G = \omega_{\eta_G} - \omega_{c_G}$. Therefore, the gap can be an indicator of \mathfrak{C}_G for high n. Expanding Δ_G and Δ_G^2 for the depolarizing channel noise model, one will find that the expression is at most an eighth degree polynomial, with even degrees:

$$\Delta_G = d + c_1 \eta^2 + c_2 \eta^4 - \omega_{c_G}, \quad \Delta_G^2 = z + r_1 \eta^2 + r_2 \eta^4 + r_3 \eta^6 + r_4 \eta^8.$$

Hence, there are multiple candidates for a numeric score which can model convincingness across noisy regimes. We consider Δ_G , Δ_G^2 , the most significant coefficients of both (constants, η^2 , η^4), or some unique transform of the coefficients, constants and local bound. Let us obtain the candidate scores for each game by deriving their coefficients and constants in Δ_G . We start with a convenient way to obtain these for 4-qubit games.

Theorem V.5. For $G \in \{2\text{-}CHSH, MSG, 2\text{-}CHSH\text{-}OPT\})$ receiving input $\varepsilon_{\eta}^{4-qubit}(\rho_{EPR})$ for η , the gap is,

$$\Delta_G = \eta^4 k_{ideal} + \eta^2 (1 - \eta^2) k_{part-mixed} + (1 - \eta^2)^2 k_{mixed} - \omega_{c_G}, \tag{22}$$

where k_{ideal} , $k_{part-mixed}$, and k_{mixed} are constants specific to each G.

Proof. Define **O** to be some score operator of a game G and π to be some normalization. Then the value for G played with an η -depolarized input is:

$$\omega_{\eta_{G}} = \pi \operatorname{Tr} \left[\varepsilon_{\eta}^{4-qubit}(\rho_{EPR}) \mathbf{O} \right] = \pi \operatorname{Tr} \left[\left(\left(\eta^{2} \rho_{EPR} + (1 - \eta^{2}) \frac{\mathbb{I}_{4}}{4} \right) \otimes \left(\eta^{2} \rho_{EPR} + (1 - \eta^{2}) \frac{\mathbb{I}_{4}}{4} \right) \right) \mathbf{O} \right]$$

$$= \eta^{4} \operatorname{Tr} \left[\pi (\rho_{EPR} \otimes \rho_{EPR}) \mathbf{O} \right] + \eta^{2} (1 - \eta^{2}) \operatorname{Tr} \left[\pi (\rho_{EPR} \otimes \frac{\mathbb{I}_{4}}{4}) \mathbf{O} + \pi (\frac{\mathbb{I}_{4}}{4} \otimes \rho_{EPR}) \mathbf{O} \right] + (1 - \eta^{2})^{2} \operatorname{Tr} \left[\pi (\frac{\mathbb{I}_{4}}{4} \otimes \frac{\mathbb{I}_{4}}{4}) \mathbf{O} \right]$$

$$:= \eta^{4} k_{ideal} + \eta^{2} (1 - \eta^{2}) k_{part-mixed} + (1 - \eta^{2})^{2} k_{mixed},$$

for constants $k_{\text{ideal}}, k_{\text{part-mixed}}, k_{\text{mixed}}$. Note that the score for the 2-CHSH and MSG is computed this way. That is, as defined in II.29, for the 2-CHSH $\pi = 1/16$ and $\mathbf{O} = \sum_{\substack{(a_1 \oplus b_1 = x_1 \land y_1) \land (a_2 \oplus b_2 = x_2 \land y_2)}} ((M_{a_1|x_1} \otimes M_{a_2|x_2}) \otimes (M_{b_1|y_1} \otimes M_{b_2|y_2}))$ where the sum is over all measurement combinations:

$$(a_i, b_i, x_i, y_i) \in \{(0, 0, 0, 0), (1, 1, 0, 0), (0, 0, 0, 1), (1, 1, 0, 1), (0, 0, 1, 0), (1, 1, 1, 0), (0, 1, 1, 1), (1, 0, 1, 1)\}.$$

Further, the operator for the MSG is Eqn. 9 with $\pi = 1/18$. Since 2-CHSH-OPT games deal with separate objects, it

isn't clear if the above applies. Hence, we derive the value of an Optimized 2-CHSH game:

$$\omega_{\eta_{2chsh,\eta'}} = \operatorname{Tr}(\mathbf{P}_{\eta}\bar{\mathbf{B}}_{\eta'}^{T}) = \operatorname{Tr}\left[\frac{1}{16}\left(\sum_{a,b=1}^{16}\operatorname{Tr}(\rho_{in}(A_{a}\otimes B_{b}))|e_{a}\rangle\langle e_{b}|\right)\left(\sum_{i,j=1}^{16}\bar{\mathbf{B}}_{\eta'}^{i,j}|e_{j}\rangle\langle e_{i}|\right)\right]$$

$$= \frac{1}{16}\operatorname{Tr}\left(\rho_{in}\sum_{a,b=1}^{16}(A_{a}\otimes B_{b})\bar{\mathbf{B}}_{\eta'}^{a,b}\right)$$

$$= \frac{1}{16}\left(\eta^{4}\operatorname{Tr}[(\rho_{\text{EPR}}^{\otimes 2})\sum_{a,b=1}^{16}(A_{a}\otimes B_{b})\bar{\mathbf{B}}_{\eta'}^{a,b}] + (1-\eta^{2})^{2}\operatorname{Tr}[(\frac{\mathbb{I}_{4}^{\otimes 2}}{16})\sum_{a,b=1}^{16}(A_{a}\otimes B_{b})\bar{\mathbf{B}}_{\eta'}^{a,b}]\right)$$

$$+\eta^{2}(1-\eta^{2})\operatorname{Tr}[(\rho_{\text{EPR}}\otimes\frac{\mathbb{I}_{4}}{4}+\frac{\mathbb{I}_{4}}{4}\otimes\rho_{\text{EPR}})\sum_{a,b=1}^{16}(A_{a}\otimes B_{b})\bar{\mathbf{B}}_{\eta'}^{a,b}]$$

$$:= \eta^{4}k_{\text{ideal}} + \eta^{2}(1-\eta^{2})k_{\text{part-mixed}} + (1-\eta^{2})^{2}k_{\text{mixed}},$$
(by linearity & cyclicity of trace)
$$= \frac{1}{16}\left(A_{a}\otimes B_{b}\right)\bar{\mathbf{B}}_{\eta'}^{a,b}$$

where intermediate steps are shown in App. XIII.. Therefore, for $G \in \{2\text{-CHSH}, \text{MSG}, 2\text{-CHSH-OPT}\}$, $\Delta_G = \eta^4 k_{\text{ideal}} + \eta^2 (1 - \eta^2) k_{\text{part-mixed}} + (1 - \eta^2)^2 k_{\text{mixed}} - \omega_{c_G}$ for constants k_{ideal} , $k_{\text{part-mixed}}$, and k_{mixed} specific to each G. As desired. \square

Note that: (1) k_{ideal} is the probability that both EPR pairs are noise-free (i.e., the quantum bound of G with the maximally entangled state); (2) $k_{\text{part-mixed}}$ is the probability that exactly one pair is noisy; and (3) k_{mixed} is the probability that both pairs are noisy (the score with the maximally mixed state). Following the method in the proof, we obtain:

$$\omega_{\eta_{2chsh}} = 0.72855\eta^4 + 0.85355\eta^2(1 - \eta^2) + 0.25000(1 - \eta^2)^2 \tag{23}$$

$$\omega_{\eta_{2chsh},\eta'=1} = 0.10937\eta^4 + 0.30936\eta^2(1-\eta^2) + 0.21875(1-\eta^2)^2 \tag{24}$$

$$\omega_{\eta_{msg}} = \eta^4 + 1.22222\eta^2 (1 - \eta^2) + 0.5(1 - \eta^2)^2 \tag{25}$$

where we show the 2-CHSH optimized for $\eta' = 1$ as an example. Recall that η' is the visibility used to optimize and fix a 2-CHSH-OPT's Bell matrix, and η is the visibility of the input used to play the now fixed configuration. For CHSH, we compute its value independently for input $\varepsilon_{\eta}^{2-qubit}(\rho_{EPR}) = \eta^2 \rho_{EPR} + \frac{\mathbb{I}_4}{4}(1-\eta^2)$:

$$\omega_{\eta_{chsh}} = \frac{1}{8} \text{Tr}(\mathbf{S}_{G_{chsh}} \varepsilon_{\eta}^{2-qubit}(\rho_{EPR})) + \frac{1}{2} = \eta^2 \frac{\text{Tr}(\mathbf{S}_{G_{chsh}} \rho_{EPR})}{8} + 0.25(1 - \eta^2) \frac{\text{Tr}(\mathbf{S}_{G_{chsh}})}{8} + \frac{1}{2} = \frac{\sqrt{2}}{4} \eta^2 + \frac{1}{2}, \quad (26)$$

where the last equality follows by definition of $\mathbf{S}_{G_{chsh}}$ (the expectation value that $\mathbf{S}_{G_{chsh}}$ yields given either the EPR pair, $2\sqrt{2}$, or maximally mixed state as an input state, 0). Hence, we can observe from the final equation that the quantum value of the CHSH is 0.85355, as expected. Note that the CHSH value also has coefficients corresponding to the probabilities of maximally entangled and maximally mixed input states. The full Δ_G expressions for all games considered are shown in Table 1. Further, App. XIV. contains the raw ω_{η_G} unsimplified expressions showing the explicit k constants.

We note that for 2-CHSH-OPT configurations, the η' values were selected as the results were plotted, having a more fine-grained analysis between 0.83-0.89 as this region had some of the most convincing configurations, and eventually became unconvincing for $\eta' = 0.83, 0.84$. Also note that throughout the rest of the work, the Δ_G expressions are re-written with explicit η -dependence, as in Table 1, and are referred to as both gapped expressions and noise-dependent gaps.

B. Gap-Based Scoring

Using the gapped expressions in Table 1, we explored multiple approaches for constructing a single numerical score for a non-local game, such that the score is a proxy measure for its convincingness in comparison to other games. In particular, we searched for a score which reflects the following intuition: if game G_1 scores higher than game G_2 , then G_1 should become statistically convincing at a lower η value than G_2 . This aligns with our goal of identifying games that are more robust to experimental imperfections while still providing strong evidence of non-locality, as a higher score for G_1 would

Table 1: Gapped Expressions for Non-Local Games

G	$\Delta_G = \omega_{\eta_G} - \omega_{c_G}$
CHSH	$0.35355\eta^2 + 0.5 - 0.75$
2-CHSH	$0.35355\eta^2 + 0.12500\eta^4 + 0.25000 - 0.625$
2-CHSH-OPT $\eta' = 1$	$0.30936\eta^2 + 0.10937\eta^4 + 0.21875 - 0.546875$
2-CHSH-OPT $\eta' = 0.95$	$0.25411\eta^2 - 0.00097\eta^4 + 0.36133 - 0.5625$
2-CHSH-OPT $\eta' = 0.90$	$0.25411\eta^2 - 0.00097\eta^4 + 0.36133 - 0.546875$
2-CHSH-OPT $\eta' = 0.89$	$0.24307\eta^2 + 0.34375 - 0.515625$
2-CHSH-OPT $\eta' = 0.88$	$0.24307\eta^2 + 0.34375 - 0.515625$
2-CHSH-OPT $\eta' = 0.87$	$0.24307\eta^2 + 0.34375 - 0.515625$
2-CHSH-OPT $\eta' = 0.86$	$0.24307\eta^2 + 0.34375 - 0.515625$
2-CHSH-OPT $\eta' = 0.85$	$0.24307\eta^2 + 0.34375 - 0.515625$
2-CHSH-OPT $\eta' = 0.84$	0.53125 - 0.53125
2-CHSH-OPT $\eta' = 0.83$	0.53125 - 0.53125
MSG	$0.22222\eta^2 + 0.27778\eta^4 + 0.5 - 0.94444$

imply that G_1 remains convincing at higher noise levels. This can be formalized as follows:

Definition V.40 (Significance Crossing). For a game G, let $\eta_G^{\dagger} = \inf \eta \in [0,1]$: $\mathfrak{C}_G(\eta) \leq \alpha$ be the visibility at which the game first becomes statistically significant at level α , and refer to η_G^{\dagger} as G's 'significance crossing'.

Question 6. Can we define a metric κ_G to rank games such that for any two games G_1 and G_2 : $\kappa_{G_1} > \kappa_{G_2} \implies \eta_{G_1}^{\dagger} < \eta_{G_2}^{\dagger} \pm \epsilon$? ϵ is a small tolerance parameter accounting for numerical precision and cases where the convincingness curves $\mathfrak{C}_{G_1}(\eta)$ and $\mathfrak{C}_{G_2}(\eta)$ are nearly identical near their crossing points.

Ideally, we want to use κ_G to correctly rank the order of the significance crossings for different games, which would, importantly, rank the games' convincing abilities under our noise model. We construct a new score which we coin the gapped score (Def. V.41), inspired by convincingness, for modeling the above behaviour. As will be shown in Sec. VI., out of the existing ratio and gap-inspired scores explored, Def. V.41 best captures the desired behaviour of κ_G . The optimal candidate for κ_G , is presented here, which is based on noise-dependent gaps.

Definition V.41. [Quadratic Noisy Gapped Ratio Scoring: Depolarizing Channel] For a non-local game G whose input undergoes the depolarizing channel with probability of no noise η , G has the **gapped expression**:

$$\omega_{\eta_G} - \omega_{c_G} = c_1 \eta^2 + c_2 \eta^4 + d - \omega_{c_G}.$$

Then we define the gapped score of G, played with N_{res} η -noisy inputs as:

$$\kappa_G := \left(\frac{c_1 + c_2}{|d - \omega_{c_G}|}\right)^2 \cdot \frac{N_{res}}{-\ln(\alpha)} \tag{27}$$

where the parameter α defines the significance threshold.

Note that in contrast to the convincingness measure, which must be evaluated separately for each noise level, the gapped score provides a single number that characterizes a game's overall convincingness behaviour for all possible noise levels.

We now explain the intuition behind the design of the gapped score. The score in Def. V.41 captures essential features that determine when a game becomes statistically significant for non-locality. The first factor, $\left(\frac{c_1+c_2}{|d-\omega_{c_G}|}\right)^2$, captures the relative strength of the noise-dependent terms versus the constant separation in the gap expression:

- The numerator, $c_1 + c_2$, represents the "strength" of the η -dependent terms. We use $c_1 + c_2$ rather than a weighting of the terms to capture the full contribution of the noise-dependent terms when $\eta = 1$ (i.e., when there is no noise), ensuring that the score reflects the maximum possible rate of gap growth under ideal conditions.
- The denominator $|d \omega_{c_G}|$ represents the "baseline separation" that needs to be overcome.

The ratio between these terms indicates how quickly the gap expression can cross the significance threshold as η increases. It enforces that games with higher scores have lower η^{\dagger} , indicating greater robustness to noise. Moreover, the quadratic relationship amplifies the differences between gap expressions, providing a more informative score. We do find that enforcing a linear relationship instead is also characteristic of the games' relative η^{\dagger} 's, albeit, yielding less distinct scores.

Further, the scaling factor, $\frac{N_{res}}{-\ln(\alpha)}$, relates our score to the experimental conditions needed for statistical significance. To understand where this comes from, consider Δ_G becomes significant when its p-value, which we model by $p \sim \exp(-N_{res}(\omega_{\eta_G} - \omega_{c_G})^2)$, falls below α . Taking the negative logarithm of both sides of $\exp(-N_{res}(\omega_{\eta_G} - \omega_{c_G})^2) < \alpha$, we find that significance occurs when $(\omega_{\eta_G} - \omega_{c_G})^2 > \frac{-\ln(\alpha)}{N_{res}}$. The scaling factor in our score is the inverse of this threshold, which aligns with intuition: more experimental resources (N_{res}) make it easier to achieve significance and thus increase the score, while stricter significance requirements (smaller α) make it harder and thus decrease the score.

Putting it all together, higher scores indicate earlier crossing of the significance threshold. If \mathfrak{C}_G crosses the threshold earlier, we expect a larger coefficient strength and smaller baseline separation yielding a higher κ_G . We also expect increased experimental resources to increase κ_G . Of course, a greater α will increase κ_G as well.

Note that the definition of κ_G can be extended to a different noise model by deriving a gapped expression for that model. Only the degree of the polynomial, and the corresponding coefficients, c_i , and constant, d, will differ. Simply summing all of the new model's coefficients as is for the numerator may work, or a weighting may need to be derived to reliably capture the significance crossing behaviour of that noise model, by a similar process to that presented in Sec. VI.

C. Polynomial Comparisons

As will be seen in Section VI., the gapped score achieves the desired behaviour, but has a limitation that games with extremely similar convincingness behaviour have the same or similar scores. A more fine-grained analysis for games that are similar in convincingness behaviour can be done via subtracting the gapped expressions and analyzing the behaviour of the resulting curve. Let us demonstrate this approach by comparing the standard 2-CHSH and the $(\eta' = 1)$ -optimized 2-CHSH games. Define the difference function $D(\eta) := \Delta_{2chsh}(\eta) - \Delta_{2chsh,\eta'=1}(\eta)$, where,

$$\Delta_{2chsh}(\eta) = 0.35355\eta^2 + 0.12500\eta^4 - 0.375,$$

$$\Delta_{2chsh,\eta'=1}(\eta) = 0.30936\eta^2 + 0.10937\eta^4 - 0.32813.$$

This yields $D(\eta) = 0.04419\eta^2 + 0.01563\eta^4 - 0.04688$. Now, the behaviour of $D(\eta)$ reveals several key insights. At maximal noise $(\eta = 0)$, we have D(0) = -0.04688 < 0, indicating that the $(\eta' = 1)$ -optimized 2-CHSH has a larger gap. Conversely, in the noiseless limit $(\eta = 1)$, we find D(1) = 0.04419 + 0.01563 - 0.04688 = +0.01294 > 0, showing the standard 2-CHSH achieves a larger gap. To characterize the complete behaviour of $D(\eta)$, note that its derivative $D'(\eta) = 0.08838\eta + 0.06252\eta^3$ is strictly positive for all $\eta > 0$. Therefore, $D(\eta)$ is strictly increasing on [0, 1], and by the Intermediate Value Theorem, must cross zero exactly once on this domain. Numerically, this crossing occurs at $\eta \approx 0.91$, leading to the conclusion:

$$D(\eta) < 0$$
 for $\eta < 0.91$, $D(\eta) > 0$ for $\eta > 0.91$,

These results reveal a fundamental trade-off: the $\eta'=1$ optimized 2-CHSH performs better (has a larger gap) in highnoise regimes ($\eta < 0.91$), while the standard 2-CHSH achieves superior performance in low-noise regimes ($\eta > 0.91$). This suggests that optimization of the two-copy measurement bases enhanced the detection of non-locality under moderate noise, at the cost of reduced violation magnitude at the noiseless limit.

We do note that D(0) and D(1) have relatively small values, which indicates the two games have very similar convincingness behaviour—as expected, since the 2-CHSH-OPT game is optimized for no noise on the input state. As we will

see, the above conclusions match exactly the behaviour observed in the simulation of convincingness in Sec. VI. (Fig. 4, left). Despite the fact that both curves for the exact convincingness score are nearly overlapping, the gap accurately describes the behaviour. Therefore, this polynomial comparison method complements the gapped score approach: while the gapped score identifies which games first cross the significance threshold, direct polynomial comparison reveals the detailed comparative behaviour of games and configurations which have highly overlapping convincingness curves.

VI. RESULTS AND DISCUSSION

We now discuss the results of the noise-robustness analysis, first summarizing the overall behaviours. The results of the convincingness analyses indicate that the gapped score is a reliable measure of the noise-robustness of a non-local game for stable resource regions, to be defined below, and more accurately reflects the order of games' significance crossings (η_G^{\dagger} 's) compared to the ratio ($\omega_{\eta=1_G}/\omega_{c_G}$), the gap between the quantum bound and local bound ($\omega_{\eta=1_G}-\omega_{c_G}$), and all other new measures proposed in this work. Further, the noise-tolerance behaviour observed exactly matches the trends observed for the convincingness under near-infinite resources, suggesting that the convincingness is a more nuanced measure of noise-robustness than noise-tolerance. Note that simulations used Python [40], MATLAB [39] and QuTiP [41].

A. Convincingness Analysis

Figs. 4 and 5 exhibit the results for the convincingness analysis across depolarized noise levels of the EPR pair(s). Let us recall the setup (from Sec. A.). Each non-local game is played at different visibilities, η , with input $\rho_{\eta} = \varepsilon_{\eta}^{2-qubit}(\rho_{EPR})$ for CHSH and $\rho_{\eta} = \varepsilon_{\eta}^{4-qubit}(\rho_{EPR} \otimes \rho_{EPR})$ for the MSG and all 2-CHSH configurations, yielding a ω_{η_G} for each round. Algorithm 1 is then used to evaluate each $\mathfrak{C}_G(\eta) \ \forall \eta \in [0,1]$, using N_{res} noisy EPR pairs for each game. This translates to N_{res} and $\frac{N_{res}}{2}$ rounds played for CHSH and the rest of the games, respectively. The curves $\{(\eta, \mathfrak{C}_G) \ \forall \eta \in [0,1]\}$ are shown in Figs. 4 and 5 for different N_{res} values. The 2-CHSH-OPT configurations were optimized for particular visibilities $(\eta' \in \{0.85, 0.86, \dots, 1\})$, using the method in Sec. B. and the curves for these games are also plotted. Note that $\Delta_G = 0$ for 0.83- and 0.84-optimized games, so they are completely unconvincing configurations and hence excluded from figures.

Let us now discuss the results from the convincingness simulations. One of the most important properties on the figures is the visibility levels at which $\mathfrak{C}_G = \alpha$, the significance crossings η_G^{\dagger} (V.40). Note that our $\alpha = 0.05$. The η_G^{\dagger} crossings are the visibility levels for which a game (or configuration), G, becomes convincing of non-locality, or in other words, will be highly likely to detect non-locality, for the given input state $\rho_{\eta_G^{\dagger}}$. Based on analyzing the order of the crossings of different games, we can deduce a number of major observations:

Observation 1. Let $\{2\text{-}CHSH, MSG, 2\text{-}CHSH\text{-}OPT\ configurations}\}$ be the set of candidate games and configurations being tested for noise-robustness, where lower η_G^{\dagger} implies greater 'noise-robustness'. Then, CHSH is the most noise-robust candidate to the depolarizing channel, assuming an equal number of noisy resources is available to all candidates.

Let us discuss Observation 1. If we look at Figs. 4 and 5, where the N_{res} used by each game to prove non-locality is the same (i.e., $N_{res} = 1000, 10000, 100000, 1000000$, or 1000000), then the CHSH has the lowest η^{\dagger} out of all candidates. That is,

$$\eta_{chsh}^{\dagger} < \eta_F^{\dagger} \; \forall F \in \{\text{2-CHSH}, \text{MSG}, \text{2-CHSH-OPT configurations}\}.$$

Moreover, consider the 2-CHSH games optimized for $\eta' \in \mathcal{Q} = \{0.85, 0.86, \dots, 0.89\}$, which have approximately the same η^{\dagger} as each other. A pattern is seen across all experiments. As N_{res} grows, the \mathfrak{C} curves of these particular 2-CHSH optimized games grow closest to η_{chsh}^{\dagger} out of all other games, but are unable to beat it. That is, as $N_{res} \to \infty$, we see:

$$\eta_{2chsh.0.85}^{\dagger} \approx \eta_{2chsh.0.86}^{\dagger} \approx \eta_{2chsh.0.87}^{\dagger} \approx \eta_{2chsh.0.88}^{\dagger} \approx \eta_{2chsh.0.89}^{\dagger} \rightarrow \eta_{chsh}^{\dagger}.$$
 (28)

Note that we tried optimizing the 2-CHSH game for smaller visibilities ($\eta = 0.83, 0.84$), which yielded unconvincing behaviour ($\Delta_{2chsh,\eta} = 0$). Hence, for the purpose of this analysis we can assume that 2-CHSH optimized for Q are the best configurations of the game for the depolarizing channel. Based on this and Eqn. 28, we question whether this trend is revealing of a fundamental limitation in the axes fixed for the 2-CHSH optimized non-local games. One possible limitation

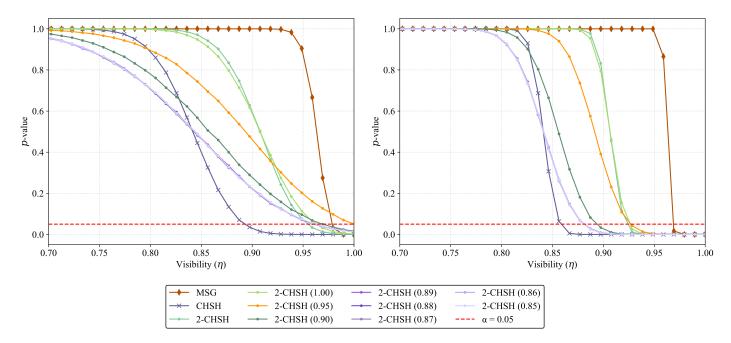


Figure 4: Convincingness Behaviour Analysis for Finite Resources (low N_{res}). The convincingness curves were computed for the CHSH, MSG, 2-CHSH, and 2-CHSH-OPT games using the methods proposed in Sec. IV.. Each point on a curve was computed assuming N_{res} η -noisy EPR pairs are available to that game. The η -level for which 2-CHSH-OPT games were optimized is noted in brackets in the legend. Games which have similar κ_G , computed according to the methods in Sec. B., are coloured similarly. The dotted red line denotes the significance threshold, $\alpha = 0.05$. The left figure shows the $\mathfrak{C}_G(\eta)$ scores (p-values), assuming $N_{res} = 1000$ are available for all games. The right figure shows the same, but assumes $N_{res} = 10,000$. All plotted values were averaged over 10 runs, each with a random seed, where the p-value computed for each random seed is an average of three runs with this seed.

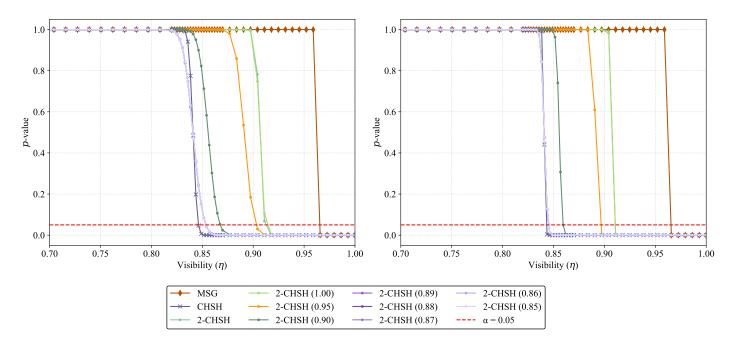


Figure 5: Convincingness Behaviour Analysis for Near-Infinite Resources (high N_{res}). The convincingness was computed the same way as in Figure 4, with the exception that left figure shows the convincingness scores for when $N_{res} = 100,000$ for all games, and the right shows the same for when $N_{res} = 1,000,000$.

is the restriction on both games to use the same amount of resources in order to prove the presence of non-locality; perhaps the 2-CHSH-OPT games require more rounds, and hence, more resources than the CHSH game in order to do so. We, thus, test whether allowing the 2-CHSH-OPT configurations corresponding to \mathcal{Q} to use more resources than the CHSH permits their $\eta_{2chsh,opt}^{\dagger}$ values to move below η_{chsh}^{\dagger} , and by Fig. 6, observe the following positive result:

Observation 2. Let us define noise-robustness as before, where a game G is more 'noise-robust' than game G' if $\eta_G^{\dagger} < \eta_{G'}^{\dagger}$. Let F' denote the set of 2-CHSH games optimized for $\eta' \in \mathcal{Q}$, and denote $N_{res}^{G \in F'}$ and N_{res}^{chsh} the resources each game can use to prove non-locality. $\forall G \in F'$, if $N_{res}^{G \in F'} >> N_{res}^{chsh}$, then the configurations in F' are more noise-robust than CHSH.

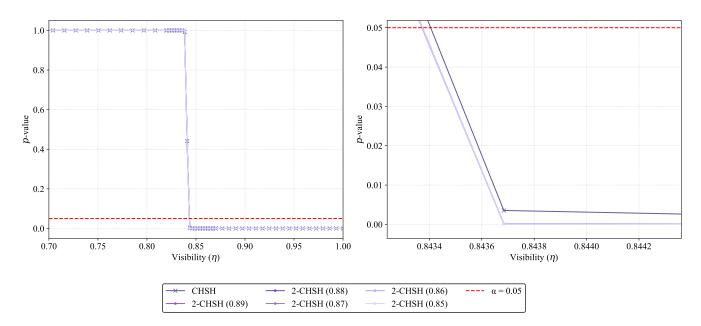


Figure 6: Convincingness Behaviour Analysis for Near-Infinite, Unequal Resources Between CHSH and 2-CHSH-OPT Configurations. The right figure is a zoom-in of the left figure. We took the most convincing 2-CHSH-OPT games, and permitted them to have much more noisy resources (many more rounds) than the CHSH in order to see if the 2-CHSH-OPT's convincingness was fundamentally bounded by the CHSH's convincingness. When $N_{res} = 1,000,000$ for the CHSH and $N_{res} = 10 \times 1,000,000$ for the 2-CHSH-OPT configurations, the 2-CHSH-OPT configurations outperform the CHSH. That is, $\eta^{\dagger}_{2chsh,\eta'} < \eta^{\dagger}_{chsh}$ for $\eta' \in \mathcal{Q}$. All simulations were conducted as before and once again, all plotted values were averaged over 10 runs, each with a random seed, where the *p*-value computed for each random seed is an average of three runs with this seed. The dotted red line denotes the significance threshold, $\alpha = 0.05$.

Through varying the relative resources of the two in the simulations, and testing the asymptotic resource regime with resources in the millions of noisy EPR pairs, we find that when the CHSH game has $N_{res} = 1M$ resources to use and the 2-CHSH-OPT configurations optimized for $\eta' \in \mathcal{Q}$ have 10 times that amount to use, $\eta_{2chsh,\eta'}^{\dagger} < \eta_{chsh}^{\dagger}$ (Fig. 6). This observation is a counter-example to the hypothesis that Eqn. 28 holds across all noisy resource amounts, and suggests that 2-CHSH-OPT requires more resources to be robust to η_{chsh}^{\dagger} visibility than the CHSH. It is valid to question whether a possible reason for these results is some error bound in the simulation, potentially stemming from the Monte Carlo seeds used to randomly select the number of successes in the p-value computation. However, since our figures plot the average over many random seeds, this is unlikely. Moreover, the finding aligns with our expectations for the 2-CHSH-OPT game: we expect that the 2-CHSH-OPT would perform better or the same as the CHSH either due to parallel repetition through the use of simultaneous measurements on a greater Hilbert space (to increase the probability of detecting non-locality), or due to the fact that an optimized version of the 2-CHSH game can choose to use only one pair of qubits and reach, at the very least, performance of the CHSH. Our results also align with the work of Arnon-Friedman and Yuen [18], whose parallel-repetition DI protocol needs on the order of 10^6 parallel CHSH rounds before it can certify that the high-dimensional input state—built from many noisy two-qubit copies—contains a large amount of entanglement. Our finding, therefore, motivates future study into exactly why substantially higher noisy resources are needed in parallel repetition

non-local game schemes in order to certify non-locality. A possible explanation is that using multiple noisy states per round fundamentally limits a game's performance, but a rigorous theoretical analysis remains open for future work.

We also note another interesting finding. In the finite resource regime there are specific choices of significance thresholds, α (target p-values), and number of EPR pairs (per game round) where the CHSH game is not the most noise-robust. For example, in the $N_{res} = 1000$ experiment (Fig. 4, left), we observe that $\eta^{\dagger}_{2chsh,i\in\mathcal{Q}\cup\{0.90,0.95\}} < \eta^{\dagger}_{chsh}$ for lower visibilities when $\alpha \approx 0.5$. While 0.5 is a very lenient threshold and leaves room for potential false positive detections of non-locality (likely ill-advised), these observations show that the convincingness of games is much more nuanced for low, more practical resource levels. It also demonstrates the importance of the experimenter's choice of certainty α and prompts future study into the impact of choosing higher α 's on comparing the noise-robustness of games. It may be the case that the $\alpha = 0.05$ we use is a more stringent condition than practically necessary. More generally:

Observation 3. For a low (finite) η -noisy N_{res} supply available to each game, a more lenient α choice can enable the noise-robustness of some games to improve beyond that of CHSH; where we define noise-robustness as before by η_G^{\dagger} .

By all of the above observations, we note that our original objective has been attained. We have shown a clear, computational method, for determining which non-local game, in a set of candidate games with different input-output settings and winning state dimensionalities, is most likely to detect non-locality in noisy (depolarized) states, under different resource or certainty conditions (N_{res} available, or certainty of non-locality, α , decided by the experimenter). While the above methods heavily rely on simulations, we now discuss a more analytic method to arrive at similar conclusions. Specifically, since the η_G^{\dagger} values can be used to compare the noise-robustness of games, in the convincingness-sense of noise-robustness, our goal is to find a closed form analytic expression for a score which best matches the order of η_G^{\dagger} 's across games. In particular, the candidate predictors which we consider are: the ratio ($\omega_{\eta=1_G}/\omega_{c_G}$), the raw gap ($\omega_{\eta=1_G}-\omega_{c_G}$), the coefficients and constant in Δ_G and Δ_G^2 , and the gapped score (κ_G) derived in Sec. V.. All of these predictors were derived from the analytic expression of the gap, Δ_G , since, as we know from Eqn. 19, the gap heavily influences the convincingness and hence the order of significance crossings. The values of each predictor can be seen in Table 2.

CHSH Property 2-CHSH 2-CHSH OPT MSG $\eta = 0.89$ $\eta = 0.88$ $\eta = 0.87$ $\eta = 0.85$ $\eta = 0.90$ $\eta = 0.95$ = 0.86 η Local Bound (ω_{c_G}) 0.75 0.625 0.516 0.516 0.516 0.516 0.516 0.547 0.547 0.562 0.944 Quantum Bound $(\omega_{\eta=1_G})$ 0.85 0.7290.5870.5870.5870.5870.5870.6140.6370.614Score $(\kappa_G, N_{res} = 1K)$ 667.60 43.26 543.61 667.59 667.59 667.59 667.59 667.59 621.33 543.59528.566675.89 6675.89432.62 Score $(\kappa_G, N_{res} = 10K)$ 6676.04 5436.12 6675.89 6675.89 6675.89 6308.87 5435.925285.57 Ratio $(\omega_{\eta=1_G}/\omega_{c_G})$ 1.1331.138 1.138 1.1381.1381.1231.1651.0921.059 1.166 1.138 Raw Gap $(\omega_{\eta=1_G} - \omega_{c_G})$ 0.104 0.0710.071 0.071 0.0710.0710.067 0.09 0.0510.056 0.1Constant in Δ_G^2 0.0620.1410.03 0.03 0.030.030.030.0340.108 0.040.198 η^2 Coefficient in Δ_G^2 -0.084-0.084-0.094 -0.177-0.265-0.084-0.084-0.084-0.203-0.102-0.198 η^4 Coefficient in Δ_G^2 0.059 0.059 0.059 0.1250.031 0.059 0.059 0.0650.024 0.065 -0.198Constant in Δ_G -0.25-0.375-0.172-0.172-0.172-0.172-0.172-0.186-0.328-0.201-1.389 η^2 Coefficient in Δ_G 0.3540.3540.2430.2430.2430.2430.2430.2540.3090.2540.222 η^4 Coefficient in Δ_G -0.001 0 0.1250 0 0 0 0 0.109 -0.0010.278Sig. Crossing $(\eta_G^{\dagger}, N_{res} = 1K)$ 0.107 0.04 0.038 0.038 0.038 0.04 0.03 0.037 0.046 N/A 0.022 Sig. Crossing $(\eta_G^{\dagger}, N_{res} = 10K)$ 0.1420.0780.120.120.120.1190.120.107 0.0750.0730.032

Table 2: Comparison of Non-Local Games and Their Properties.

Note: This table compares various candidate measures for approximating the order of η_G^{\dagger} across games (at $\alpha = 0.05$).

Prior to getting into the analysis of these values, let us summarize our main observations:

Observation 4. For an unstable N_{res} region, the order of the η_G^{\dagger} crossings is not predicted reliably by the candidate scores and should be computed using the exact convincingness simulation.

Observation 5. For a stable N_{res} region, the order of the η_G^{\dagger} crossings is reliably captured by the gapped score, supplemented by the polynomial expression analysis for games or configurations which have the same scores. Moreover, the gapped score is the best crossing predictor in this resource region out of all candidates.

To arrive at these observations, it is important to understand what is meant by a stable and unstable N_{res} region. In Fig. 7, we plot the order of the crossings that is attained when all games and configurations are given access to N_{res}

 η -noisy resources. Each curve shows an experiment where all games had access to N_{res} resources, and we particularly plot $1-\eta^{\dagger}$ in order to see the trend more easily. For low N_{res} , it is observed that the order of the crossings is varied, or unstable. At $N_{res} = 500$, only four out of the 11 games cross the significance threshold, and at $N_{res} = 1000$ ten out of the 11 games cross the threshold. However, for mid to high N_{res} , starting at $N_{res} = 2000$, the ordering of the crossings starts to take on a consistent, stable, pattern. Lastly, we also make the observation that as N_{res} increases, the values of the crossings themselves (the y-axis) increases as well, plateauing eventually for very high N_{res} (O(1M)). Hence, we conclude that the order of the crossings, which is the pattern we wish to predict with a candidate measure, has unstable and stable regions, and that each crossing $(1-\eta^{\dagger})$ for a game appears to have a maximum value. Note that these observations make sense, since by Eqn. 19, for large N_{res} , any $\Delta_G > 0$ will eventually produce significance (as captured by the c_i coefficients), whereas for small N_{res} the absolute magnitude of Δ_G critically determines the convincingness, making its prediction more challenging.

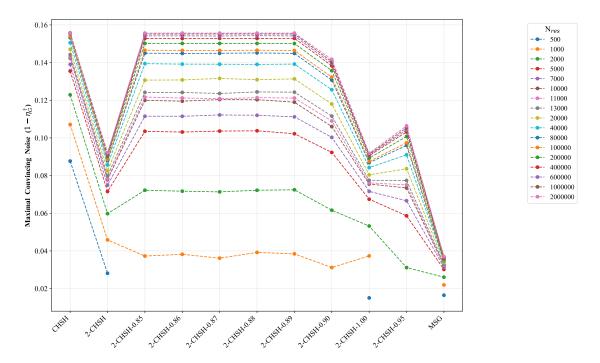


Figure 7: Order of Crossings $(1 - \eta_G^{\dagger})$ for Games at Different N_{res} .

With these observations in mind, we compare the crossing orders of experiments conducted for unstable ($N_{res} = 500, 1000$) and stable ($N_{res} = 2000$) resource regions to the values of all candidate predictors in Fig. 8. Doing so will enable us to ascertain which candidate is the most reliable estimator of the crossing orders in both regions.

Overall, we observe that no candidate predictor is a reliable predictor for the unstable region, whereas the gapped score is clearly the most reliable predictor of the crossing orders in the stable region (Fig. 8). In particular, Fig. 8 (right) we observe that for a stable crossing order, the gapped score excels at capturing: (1) the increases and decreases between games which have distinct convincingness behaviours (i.e., CHSH vs 2-CHSH vs MSG), and, (2) the nuances between similarly scoring games and configurations. For an unstable crossing order, the gapped score is less reliable. In contrast, in Fig. 8 (left) we observe that the candidates do not reliably predict the unstable and stable regions. Firstly, many candidates do not capture the main increases or decreases. For example, from CHSH to 2-CHSH, the crossings all decrease. However, many candidates (the Const. in Δ_G^2 , Raw Gap, Ratio, η^4 Coeff. in Δ_G , η^2 Coeff. in Δ_G) increase or stay the same. The most similar candidate curves to the stable crossing order are η^2 Coeff. in Δ_G^2 , η^2 Coeff. in Δ_G , Const in Δ_G , and η^4 Coeff. in Δ_G^2 . These capture major increases or decreases between completely different games (i.e., CHSH, 2-CHSH, MSG), but fail to capture the differences between configurations (i.e., 2-CHSH-1.00 to 2-CHSH-0.95).

The gapped score is not perfect, however. We observe that the CHSH and 2-CHSH-OPT games optimized for $\eta' \in \mathcal{Q}$ are predicted to have the same κ_G . On one hand, this aligns with the fact that these 2-CHSH-OPT configurations consistently have the closest η^{\dagger} to η_{chsh}^{\dagger} . On the other, if we truly want to know the difference in the crossings for

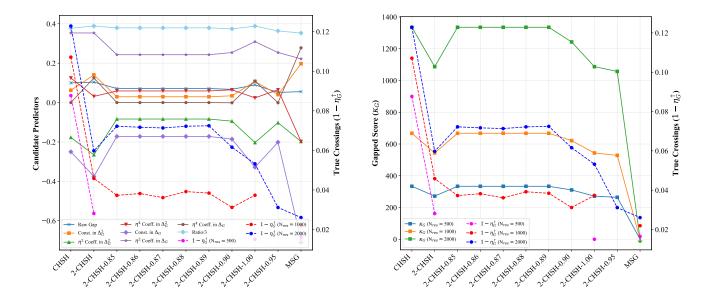


Figure 8: Comparing Candidate Predictors of Crossing Order. The figures show the true crossings attained at $N_{res} = 500,1000$, or 2000 across all experiments versus the gapped score (right) and all other candidate metrics (left). Some scaling was applied to metrics in order to make them fit on a plot; however, no trends were affected. In particular, the MSG points of some curves were scaled down by 0.5, indicated by downward arrows, and all ratios were divided by 3.

these games, this is unhelpful. Hence, we suggest complementing the gapped score with an even more nuanced theoretic polynomial analysis detailed in Sec. C.. This method uses only the gapped expressions, Δ_G , to detect precise behaviours in the exact convincingness curves of games which have highly similar crossings and/or overlapping convincingness curves. By these observations and analysis, we therefore conclude Observations 4 and 5. Note that the latter observation holds for finite resources, as the crossings stabilize as early as 2000 noisy EPR pairs. We have, thus, overall devised a single, simple closed form expression for a score which can reliably rank the crossing order of games with different input-output numbers, supplemented by an additional simple polynomial analysis for games or configurations which have similar scores.

Lastly, let us discuss more theoretically why κ_G is observed to be the best predictor of the crossings. First, it is not surprising that the raw gap and the ratio are unsuccessful predictors as they are fixed constants for a game which do not consider the noise model. The coefficients and constants, in comparison, do come from an expression which depends on our noise model: $\Delta_G = \omega_{\eta_G} - \omega_{c_G} = c_1 \eta + c_2 \eta^2 + c_3 \eta^3 + c_4 \eta^4 + \cdots + d - \omega_{c_G}$. In our noise model, only the even degree terms have nonzero coefficients. Our initial hypothesis was that c_2 , being the coefficient of the most significant term (η^2) , would be a strong predictor for the crossings. However, c_2 fails to be a good predictor. By a simple plot of $c_2 \eta^2 + d - \omega_{c_G}$, one can see that c_2 is not a good approximator of Δ_G especially at the η regimes important to us (≈ 0.8 onward). Since c_2 is the most significant contributor, this explains why the other coefficients are unreliable predictors as well. A similar argument can be made for the Δ_G^2 related predictors. In comparison, κ_G takes all of the coefficients in Δ_G into account. Further, κ_G considers the significance threshold and the number of resources used to play the game. Hence, it is the candidate that captures the most information about both the noise model and the convincingness experiment.

B. Noise-Tolerance Analysis

Fig. 9 shows the result of the noise-tolerance analysis (Sec. C.). The curves have a quadratic form due to the quadratic and quartic dependence of the depolarizing channel on the noise parameter. We note that two main properties of this graph influence the noise-tolerance: (1) the gap between local and quantum bounds, and (2) the slope of the curves. In fact, the experiment designed in Sec. C. exactly plots the ω_{η_G} values derived later on in Sec. V.. Hence, the slope is attributed to the coefficients on the η factors in the score expressions. The gap itself is defined as usual by the particular configuration of the game.

Interestingly, we see that the order of the noise-tolerances of the games (see Fig. 9 (right)) is the same as the order

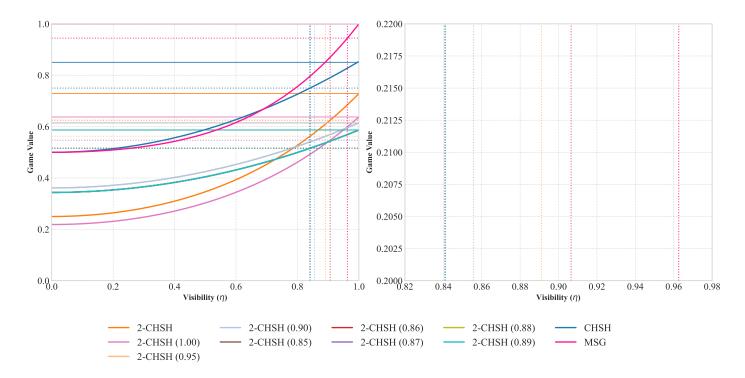


Figure 9: **Noise-Tolerance Analysis.** Left: The horizontal dotted and filled lines represent the local and quantum bounds, respectively, for the curve of that colour. The vertical dotted lines indicate the intersection of the local bound with the curves, indicating the noise-tolerance for that game. Right: This is a zoom in on the vertical lines from the left figure. It is observed that the order of the game's noise-tolerances matches the order of the convincingness significance crossings for N=1,000,000 in Fig. 5.

of the significance crossings in the near-asymptotic convincingness experiment (Fig. 5). Intuitively, this is because the noise-tolerance of a game, G, is defined for η_G^* where $\omega_{\eta_G^*} = \omega_{c_G} \iff \Delta_G(\eta_G^*) = 0$. By the definition of non-local games, this is saying that the strategy's score is the local bound, and hence, G does not detect any non-locality in that strategy. Relating this to convincingness, a gap of 0 implies that the game achieves completely unconvincing behaviour at η_G^* under asymptotic resources ($\mathfrak{C}_G = 1$ at η_G^*). These are important realizations, as they show that the noise-tolerance and convincingness are two sides of the same coin, in the asymptotic resource limit. We can also see this directly on the convincingness figure: the noise-tolerance corresponds to the point just to the left of a significance crossing for a game, which is at $\mathfrak{C}_G = 1$. This shows that for our examples, the noise-tolerance results are captured by convincingness simulations. Note that we prove the equivalence between noise-tolerance and unconvincingness at asymptotic resources in Lemma 3 (App. XVI.). Further, we prove that in the asymptotic resource limit, the noise-tolerance level $\eta_G^* < \eta_G^{\dagger}$ for arbitrary $\eta_G^*, \eta_G^{\dagger}$ (Thm. VI.6). Whether the order of the noise-tolerances and the order of the asymptotic convincingness crossings is the same for all possible games and, moreover, all noise models remains an open question. Our results show, however, that for the games tested, the convincingness can provide more nuanced information about their robustness to noise; likely stemming from the additional parameters in the convincingness model (i.e., N_{res} , α). Moreover, the fact that the noise-tolerance results align with the asymptotic \mathfrak{C}_G results suggests that the noise-tolerance, as a noise-robustness measure, is less practical for modern experimental design applications. Whether or not noise-tolerance has implications for the noise-robustness of games in the finite resource regime remains an open question. A future direction is to further develop the noise-tolerance model by adding some resource-dependent scaling factor. Although, as the convincingness naturally encodes this parameter in its statistical base, this motivates its practical use for the study of noise-robustness.

Theorem VI.6. Let η_G^* represent the visibility level at which a convincingness curve for a game, G, crosses the significance threshold, α . Let η_G^* represent the minimum visibility level (maximal noise) that G is tolerant of. Then $\eta_G^* < \eta_G^{\dagger}$ holds for any $N_{res} \in \mathbb{N}$.

Proof. Let us determine η_G^{\dagger} by analyzing the convincingness function. By Equation 19, we have:

$$\mathfrak{C}_G(\eta_G^{\dagger}) = \alpha \iff \mathfrak{C}_G(\eta_G^{\dagger}) \sim \exp\left(-N_{res}(\Delta_G(\eta_G^{\dagger}))^2\right) = \alpha,$$

where \sim denotes asymptotic equivalence as $N_{res} \to \infty$. Taking the natural logarithm of both sides yields: $\Delta_G(\eta_G^{\dagger}) = \sqrt{-\frac{\ln(\alpha)}{N_{res}}}$. Note that this value is strictly positive for any $N_{res} \in \mathbb{N}$ since $\alpha < 1$. Now, we can establish that $\eta_G^{\dagger} > \eta_G^*$ through the following steps. By Lemma 2, $\Delta_G(\eta)$ is a strictly increasing function. Further, by definition, the noise-tolerance η_G^* is the visibility level where the local bound and ω_{η_G} become equal. Hence, $\omega_{c_G} = \omega_{\eta_G} \iff \Delta_G(\eta_G^*) = \omega_{\eta_G} - \omega_{c_G} = 0$. Therefore, since $\Delta_G(\eta)$ is strictly increasing, $\Delta_G(\eta_G^*) = 0$, and $\Delta_G(\eta_G^{\dagger}) = \sqrt{-\ln(\alpha)/N_{res}} > 0$ for any $N_{res} \in \mathbb{N}$, we conclude that $\eta_G^* < \eta_G^{\dagger}$ for any $N_{res} \in \mathbb{N}$.

Lemma 2. Given some depolarizing channel visibility $\eta > 0$ and game G, $\Delta_G(\eta)$ is strictly increasing.

Proof. The value of a game played with input state $\varepsilon(\eta, \rho_{win})$ is defined as: $\omega_{\eta_G} = \text{Tr}(\mathbf{S}_G \cdot \varepsilon(\eta, \rho_{win}))$. Taking the 2-qubit depolarizing channel as an example, $\varepsilon(\eta, \rho_{win}) = \eta^2 \rho_{win} + (1 - \eta^2)^{\frac{\pi}{4}}$. Differentiating with respect to η :

$$\frac{\partial \omega_{\eta_G}}{\partial \eta} = \text{Tr}(\mathbf{S}_G(2\eta \rho_{win} - 2\eta \frac{\mathbb{I}}{4})) \iff \frac{\partial \omega_{\eta_G}}{\partial \eta} = 2\eta \text{Tr}(\mathbf{S}_G \cdot (\rho_{win} - \frac{\mathbb{I}}{4})).$$

Since $\text{Tr}(\mathbf{S}_G \cdot \rho_{win}) > \text{Tr}(\mathbf{S}_G \cdot \frac{\mathbb{I}}{4})$ (the winning state always achieves a higher score than the maximally mixed state), we have: $\frac{\partial \omega_{\eta_G}}{\partial \eta} > 0$. Thus, ω_{η_G} is a strictly increasing function of η . This argument extends by a simple computation of the derivative to the 4-qubit depolarizing channel. Since $\Delta_G = \omega_{\eta_G} - \omega_{c_G}$, and ω_{c_G} is constant, the strictly increasing nature of ω_{η_G} implies: $\frac{\partial \Delta_G}{\partial \eta} = \frac{\partial \omega_{\eta_G}}{\partial \eta} > 0$. Thus, Δ_G is a strictly increasing function of η .

VII. DEFINING NOISE-ROBUSTNESS OF NON-LOCAL GAMES

We observed in Sec. VI. that the convincingness threshold crossings, and the more analytic gapped score, can be used to rank the probability of different non-local games to detect non-locality in noisy states. This naturally suggests the following, intuitive, definition of noise-robustness:

The noise-robustness of a non-local game is the maximal noise (on its ideal state) for which it can detect non-locality with high probability.

Mathematically, the convincingness is an excellent tool for quantifying the maximal noise on the input state that is supported by a non-local game. Moreover, as we have seen, the convincingness is more practically applicable to finite resource regimes than the noise-tolerance measure. Hence, we propose the following formal definition of noise-robustness:

Definition VII.42 (Noise-Robustness using Convincingness). A game G_1 is more noise-robust to a noisy channel, $\Phi(\eta)$, than G_2 if G_1 is significantly convincing for lower visibilities η than G_2 . That is, G_1 is more $\Phi(\eta)$ -noise-robust than G_2 if $\eta_{G_1}^{\dagger} < \eta_{G_2}^{\dagger}$, where η is the channel's visibility η (i.e., noise $= 1 - \eta$) on the ideal state of the game and $\eta_{G_i}^{\dagger}$ is the visibility level at which the convincingness curve of game G_i crosses the chosen significance threshold, α , using a supply of N_{res} $\eta_{G_i}^{\dagger}$ -noisy resources. Note that α is the target likelihood of a game detecting locality.

Further, for resource regimes which yield a stable ordering of η_G^{\dagger} crossings, we know that the gapped score can reliably predict the trends in the η_G^{\dagger} of different games. Hence, for these stable resource regimes we define noise-robustness more simply, presented in Def. VII.43 below. This second definition offers a key advantage: it can be computed analytically. This makes it both more practical for experimental work and more suitable for theoretical studies of noise-robustness. Note that Def. VII.43 is very convenient for theoretical questions about the noise-robustness in the asymptotic resource limit, since the asymptotic resource regime naturally tends towards stability.

Definition VII.43 (Noise-Robustness using the Gapped Score). A game G_1 is more noise-robust to a noisy channel, $\Phi(\eta)$, than G_2 if $\kappa_{G_1} > \kappa_{G_2}$, where the games are played in a stable N_{res} region.

VIII. CONCLUSION

This thesis presents a rigorous framework for analyzing and comparing the noise-robustness of non-local games. In doing so, it addresses a critical gap in self-testing theory and practice. Through the development of novel comparison methods—particularly the convincingness measure, its analytical approximation (the gapped score), and noise-tolerance—we have established quantitative tools for evaluating how effectively different games can detect non-locality in noisy quantum states. In analyzing the capabilities of each method, we arrived at a precise and reliable definition of noise-robustness which uses the significance crossings of the convincingness. Formally, a game G_1 is more noise-robust to a noisy channel than G_2 if G_1 has a lower significance crossing than G_2 (i.e., G_1 is significantly convincing for higher noise levels than G_2) (Def. VII.42). For stable resource regimes, this simplifies to analytically comparing the games' gapped scores (Def. VII.43).

We note that the proposed noise-robustness framework successfully answers the questions posed in Sec. I.. First, it identifies the most noise-robust game out of a set of candidates, for a given noise model. Second, it pinpoints the test that maximizes deviation from local statistics under specified resource budgets. Third, by allowing resource allocation to vary or Bell coefficients to be optimized, it systematically enlarges the set of noisy states whose non-locality can be certified. Even more, our analysis revealed that games with lower noise-robustness under equal resource constraints can transition to being more robust when unequal resource allocations are allowed, highlighting that CHSH is not the most noise-robust game if unequal resources are permitted.

Importantly, we note that because the convincingness metric embeds resource constraints directly into its definition, it supports meaningful comparisons in both finite and asymptotic regimes—even when candidate games differ in input—output settings or state dimension. In practice, convincingness yields finer experimental guidance than the coarse noise-tolerance threshold while converging to the same ranking in the infinite-resource limit. For analytical work, the gapped score offers a closed-form proxy for convincingness, enabling asymptotic studies without the computational overhead of computing significance crossings. Thus, the framework gives experimentalists a direct noise-robustness score and theorists a closed-form asymptotic proxy. Together, the convincingness and gapped scores let one choose non-local games that match a given noise model and resource budget—whether the aim is maximal robustness for detecting noisy entanglement and preserving these resources, or low noise-robustness for stringent verification, security-critical tasks.

IX. APPLICATIONS

To illustrate the applicability of this framework, we note that non-local games serve as fundamental subroutines in multiple quantum information processing tasks, including quantum algorithms, communication protocols, and cryptographic applications. For instance, in applications such as DI quantum key distribution (DIQKD) [6, 42] and randomness generation [9, 43]—where security relies on the use of near-maximally entangled states—it may be advantageous to employ less noise-robust games than CHSH (i.e., certain 2-CHSH variants) that yield stronger guarantees under ideal conditions. Conversely, in applications such as entanglement distillation [44, 45, 46], which can operate effectively on lower-fidelity entangled states, more noise-robust games (i.e., CHSH) may be preferable, as they enable the certification and utilization of a larger set of imperfect quantum resources, thereby improving the overall efficiency of the protocol. Hence, our methods can be used to make these schemes more efficient and hopefully preserve quantum resources in the process.

Another prospective application is the use of more noise-robust tests in DIQKD scenarios where the secret-key rate is zero, or nearly-zero. We present two such scenarios. The first is theoretical rate studies that model full quantum-repeater chains, exemplified in [47]. The second is single-segment, memory-assisted entanglement-swapping prototypes—such as the trapped-ion experiment in [48] and the rubidium-atom experiment in [49]—which realize the core repeater building block without yet chaining multiple segments. In both settings every entanglement-swapping stage reduces the visibility of the shared pair; lifting the visibility back above the violating threshold of the self-tests used in the DIQKD protocols often demands an additional distillation round that halves the surviving-pair rate and can exceed the memories' decoherence window, driving the key rate to zero. A self-test with a slightly lower threshold visibility (i.e., lower significance crossing) could potentially certify those same swapped pairs without invoking that costly distillation layer, providing a small—but

strictly positive—secret-key where current tests yield none. It is important to note, however, that the certified randomness of the outcome will likely be lower, at the cost of ultimately achieving a positive key rate. Hence, the trade-off between certified randomness and the secret-key rate should be considered when increasing the noise-robustness of self-tests in DIQKD methods. A systematic investigation of this trade-off could inform the optimal choice of non-local games in different applications.

Looking to entanglement theoretic research, the noise-robustness framework also opens a possibility for researching partially entangled mixed states. Specifically, as the advantages of partially entangled mixed states in quantum protocols remain largely unexplored–possibly due to detection difficulties–our framework opens new avenues for investigating their application and fundamental properties. In contrast, as partially entangled *pure* states have known advantages in applications, such as random bit generation [50], future work could extend the analysis in this work to noise models producing partially entangled pure states. In this way, more of these states can be preserved for random bit generation schemes.

Lastly, the noise-robustness framework enables evaluating the noise-robustness of additional non-local games, under different noise models and resource constraints. The convincingness score can also be used as a general comparative score between non-local games, outside of the noise-robustness context. That is, \mathfrak{C} can be used as a comparable, operational score which normalizes the scores of games with different input-output settings and state dimensionalities.

X. STRENGTHS & WEAKNESSES

Our framework offers several strengths. It provides a non-experimental method for experimentalists to determine the most suitable game based on their noise model and resources. Its flexibility accommodates various noise models, enabling the inclusion of multiple imperfections, and the convincingness measure facilitates comparison across games with different dimensionalities, input-output settings, and resource constraints. Additionally, while focused on noise-robustness to input states, the framework could extend to robustness against measurement noise by comparing games optimized for specific detection inefficiencies. Finally, the gapped score provides an analytic alternative to classical simulations, allowing noise-robustness analyses in higher-dimensional Hilbert spaces.

However, there are some limitations. The framework requires knowledge of the system's noise model, which can be challenging to characterize, though advancements in quantum noise characterization make this increasingly feasible. Experimentalists can also adopt more conservative noise models by, for example, considering a combination of multiple noisy channels that are known to be affecting a device (i.e., depolarizing and dephasing noise). Another requirement is knowledge of each game's local bound, which is a one-time calculation. Practical implementation of noise-robust games on quantum hardware also presents challenges.

XI. FUTURE RESEARCH AVENUES & OPEN QUESTIONS

Below we outline some of the several avenues for future investigation into the noise-robustness of non-local games.

A. Non-Local Reductions

Throughout this work we repeatedly returned to the idea of comparing games by how far into the noisy regime they still certify non-locality. Motivated by this, we propose a systematic *non-local reduction* hierarchy for non-local games and point to its formal development as a promising avenue for future research. For non-local games A and B, we write:

 $B \leq_{NR} A \iff$ every state that violates the classical bound of B also violates that of A (equivalently, $\mathcal{V}(B) \subseteq \mathcal{V}(A)$).

The relation \leq_{NR} is reflexive and transitive, so it is a preorder. Games higher in this hierarchy are strictly more noise-robust: they certify non-locality for all states witnessed by any lower game and for additional, noisier states. Thus the "harder" (more noise-robust) games are precisely the maximal elements, because they witness entanglement for the largest possible class of noisy states.

This definition is natural in the resource theory of non-locality, whose free operations are local operations and shared randomness (LOSR). Because LOSR cannot create Bell violations, it orders resources by convertibility:

$$X \xrightarrow{\text{LOSR}} Y \iff \text{there exists an LOSR protocol mapping } X \text{ to } Y.$$
 (29)

Extensive work on abstract resource theories has introduced LOSR monotones and closed sets that quantify convertibility under *state-level* LOSR operations [51, 52]. In our setting this implies that, whenever $\mathcal{V}(B) \subseteq \mathcal{V}(A)$, any state that violates B, using LOSR alone (and fresh measurements), can be transformed into an A-violating one; hence, the resource detected by B is therefore always convertible into one detected by A.

A related—but stricter—framework orders behaviour boxes by device-independent LOSR wirings [53, 54]. Here the free operations act only on the observed input—output probability distributions P(ab|xy), leaving the underlying quantum state and measurements untouched. Such box-level convertibility can fail even when state-level convertibility holds: for example, it can be shown that the perfect Magic-Square box cannot be wired into the Tsirelson CHSH box under DI-LOSR (see App. XVII.), although the same two-singlet state violates CHSH once new measurements are chosen. These limitations motivate the state-based, game-level preorder proposed here.

Relating resource-theoretic monotones to the state sets that violate given non-local games would lay the quantitative groundwork for defining *non-local reductions* between games, which by extension, can enable a ranking of games' robustness to noisy (partially non-local) states. Specifically, developing this preorder—finding complete monotones, conversion rates, and maximal elements—would add an order-theoretic method for ranking games' noise-robustness. To our knowledge, no rigorous study of such inter-game reductions yet exists, making this an open and promising research direction.

B. Further Investigations of Noise-Robustness

The noise-robustness framework proposed in this work and our results naturally prompt the following research questions:

- 1. What limits the noise-robustness of parallel-repetition schemes like 2-CHSH compared to simpler tests like CHSH?
- 2. In finite resource regimes, is CHSH still the most robust when considering noise models other than depolarizing noise, such as more complex models used in laboratory settings?
- 3. Could optimizing games along multiple axes enhance their noise-robustness?
- 4. How can the stability of games in lower resource regimes be improved?
- 5. Is there a precise relationship between the significance threshold α and other measures of entanglement of the tested state?
- 6. How might this framework be applied to more complex non-local games?
- 7. What effects do entanglement distillation schemes have on convincingness and gapped scores?

Additional future directions include using the proposed framework to study the noise-robustness of more complex games with greater input-output settings or input state dimension, and applying the framework to the applications discussed in Sec. IX..

XII. ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my thesis supervisor, Prof. Thomas Vidick. Your guidance and insight throughout this thesis journey have been invaluable, and I am truly thankful for all the time and care you have devoted to mentoring, supporting, and developing my academic rigour. You have taught me so much, both through your mentorship and by example, and I will always be grateful for the opportunity to learn from you. I would also like to thank Dr. Elie Wolfe, Dr. Anand Natarajan, Prof. Debbie Leung, Prof. Dr. Jens Eisert, Prof. Nathan Wiebe, Noam Avidan, Ilya

Merkulov, Itammar Steinberg, and Thomas Hahn for helpful discussions throughout. Lastly, I would like to thank my family—my mother, father, and brother—for supporting me throughout anything and everything, and pushing me to be the best version of myself, always.

References

- [1] John S Bell. On the einstein podolsky rosen paradox. Physics Physique Fizika, 1(3):195, 1964.
- [2] Bas Hensen, Hannes Bernien, Anaïs E Dréau, Andreas Reiserer, Norbert Kalb, Machiel S Blok, Just Ruitenberg, Raymond FL Vermeulen, Raymond N Schouten, Carlos Abellán, et al. Loophole-free bell inequality violation using electron spins separated by 1.3 kilometres. *Nature*, 526(7575):682–686, 2015.
- [3] Lynden K Shalm, Evan Meyer-Scott, Bradley G Christensen, Peter Bierhorst, Michael A Wayne, Martin J Stevens, Thomas Gerrits, Scott Glancy, Deny R Hamel, Michael S Allman, et al. Strong loophole-free test of local realism. *Physical review letters*, 115(25):250402, 2015.
- [4] Marissa Giustina, Marijn AM Versteegh, Sören Wengerowsky, Johannes Handsteiner, Armin Hochrainer, Kevin Phelan, Fabian Steinlechner, Johannes Kofler, Jan-Åke Larsson, Carlos Abellán, et al. Significant-loophole-free test of bell's theorem with entangled photons. *Physical review letters*, 115(25):250401, 2015.
- [5] Ivan Supić and Joseph Bowles. Self-testing of quantum systems: a review. Quantum, 4:337, 2020.
- [6] Umesh Vazirani and Thomas Vidick. Fully device independent quantum key distribution. Communications of the ACM, 62(4):133-133, 2019.
- [7] Rotem Arnon-Friedman, Renato Renner, and Thomas Vidick. Simple and tight device-independent security proofs. SIAM Journal on Computing, 48(1):181–225, 2019.
- [8] Juan Yin, Yuan Cao, Yu-Huai Li, Sheng-Kai Liao, Liang Zhang, Ji-Gang Ren, Wen-Qi Cai, Wei-Yue Liu, Bo Li, Hui Dai, et al. Satellite-based entanglement distribution over 1200 kilometers. *Science*, 356(6343):1140–1144, 2017.
- [9] Stefano Pironio, Antonio Acín, Serge Massar, A Boyer de La Giroday, Dzmitry N Matsukevich, Peter Maunz, Steven Olmschenk, David Hayes, Le Luo, T Andrew Manning, et al. Random numbers certified by bell's theorem. *Nature*, 464(7291):1021–1024, 2010.
- [10] Wen-Hao Zhang, Geng Chen, Xing-Xiang Peng, Xiao-Min Hu, Zhi-Bo Hou, Shang Yu, Xiang-Jun Ye, Zong-Quan Zou, Xiao-Ye Xu, Jian-Shun Tang, et al. Experimental self-testing of entangled states. arXiv preprint arXiv:1803.10961, 2018.
- [11] Dian Wu, Qi Zhao, Xue-Mei Gu, Han-Sen Zhong, You Zhou, Li-Chao Peng, Jian Qin, Yi-Han Luo, Kai Chen, Li Li, et al. Robust self-testing of multiparticle entanglement. *Physical Review Letters*, 127(23):230503, 2021.
- [12] Xiao-Min Hu, Yi Xie, Atul Singh Arora, Ming-Zhong Ai, Kishor Bharti, Jie Zhang, Wei Wu, Ping-Xing Chen, Jin-Ming Cui, Bi-Heng Liu, et al. Self-testing of a single quantum system from theory to experiment. npj Quantum Information, 9(1):103, 2023.
- [13] Koon Tong Goh, Chithrabhanu Perumangatt, Zhi Xian Lee, Alexander Ling, and Valerio Scarani. Experimental comparison of tomography and self-testing in certifying entanglement. *Physical Review A*, 100(2):022305, 2019.
- [14] John F Clauser, Michael A Horne, Abner Shimony, and Richard A Holt. Proposed experiment to test local hiddenvariable theories. *Physical review letters*, 23(15):880, 1969.
- [15] Parvinder Singh and Atul Kumar. Analysing nonlocality robustness in multiqubit systems under noisy conditions and weak measurements. *Quantum Information Processing*, 17:1–33, 2018.
- [16] Łukasz Pawela, Piotr Gawron, Zbigniew Puchała, and Jan Sładkowski. Enhancing pseudo-telepathy in the magic square game. *PloS one*, 8(6):e64694, 2013.
- [17] Leandro Aolita, Fernando De Melo, and Luiz Davidovich. Open-system dynamics of entanglement: a key issues review. Reports on Progress in Physics, 78(4):042001, 2015.

- [18] Rotem Arnon-Friedman and Henry Yuen. Noise-tolerant testing of high entanglement of formation. arXiv preprint arXiv:1712.09368, 2017.
- [19] Rafael Chaves, Antonio Acín, Leandro Aolita, and Daniel Cavalcanti. Detecting nonlocality of noisy multipartite states with the chsh inequality. arXiv preprint arXiv:1311.4678, 2013.
- [20] Péter Diviánszky, Réka Trencsényi, Erika Bene, and Tamás Vértesi. Bounding the persistency of the nonlocality of w states. Physical Review A, 93(4):042113, 2016.
- [21] Marcin Wieśniak. Symmetrized persistency of bell correlations for dicke states and ghz-based mixtures: Studying the limits of monogamy. arXiv preprint arXiv:2102.08141, 2021.
- [22] Wieslaw Laskowski, Tamas Vertesi, and Marcin Wiesniak. Highly noise resistant multipartite quantum correlations. arXiv preprint arXiv:1412.8745, 2014.
- [23] Adel Sohbi, Isabelle Zaquine, Eleni Diamanti, and Damian Markham. Decoherence effects on the nonlocality of symmetric states. *Physical Review A*, 91(2):022101, 2015.
- [24] Piotr Gawron, Jarosław Miszczak, and Jan Sładkowski. Noise effects in quantum magic squares game. *International Journal of Quantum Information*, 6(supp01):667–673, 2008.
- [25] David Pérez-García, Michael M Wolf, Carlos Palazuelos, Ignacio Villanueva, and Marius Junge. Unbounded violation of tripartite bell inequalities. Communications in Mathematical Physics, 279:455–486, 2008.
- [26] Carlos Palazuelos and Thomas Vidick. Survey on nonlocal games and operator space theory. *Journal of Mathematical Physics*, 57(1), 2016.
- [27] Boris S Cirel'son. Quantum generalizations of bell's inequality. Letters in Mathematical Physics, 4:93–100, 1980.
- [28] Sandu Popescu and Daniel Rohrlich. Quantum nonlocality as an axiom. Foundations of Physics, 24(3):379–385, 1994.
- [29] Nicolas Brunner, Daniel Cavalcanti, Stefano Pironio, Valerio Scarani, and Stephanie Wehner. Bell nonlocality. *Reviews of modern physics*, 86(2):419–478, 2014.
- [30] Mateus Araújo, Flavien Hirsch, and Marco Túlio Quintino. Bell nonlocality with a single shot. *Quantum*, 4:353, 2020.
- [31] Michael A. Nielsen and Isaac L. Chuang. Quantum Computation and Quantum Information: 10th Anniversary Edition. Cambridge University Press, 2010.
- [32] George Casella and Roger L Berger. Statistical inference, volume 2. Duxbury Pacific Grove, CA, 2002.
- [33] V. Scarani. Bell Nonlocality. Oxford Graduate Texts. Oxford University Press, 2019.
- [34] John Watrous. Lecture 6: Nonlocal games and Tsirelson's theorem, 2020. Advanced Topics in Quantum Information Theory, CS 798/QIC 890, University of Waterloo.
- [35] Scott Aaronson. Lecture 14: Nonlocal games, 2018. Introduction to Quantum Information Science, Fall 2018, University of Texas at Austin.
- [36] B. S. Cirel'son. Quantum generalizations of bell's inequality. Letters in Mathematical Physics, 4(2):93–100, 1980.
- [37] Jonathan Barrett, Daniel Collins, Lucien Hardy, Adrian Kent, and Sandu Popescu. Quantum nonlocality, bell inequalities, and the memory loophole. *Phys. Rev. A*, 66:042111, Oct 2002.
- [38] István Márton, Erika Bene, and Tamás Vértesi. Bounding the detection efficiency threshold in bell tests using multiple copies of the maximally entangled two-qubit state carried by a single pair of particles. *Physical Review A*, 107(2):022205, 2023.

- [39] The MathWorks Inc. Matlab version: 9.13.0 (r2022b), 2022.
- [40] Guido Van Rossum and Fred L Drake Jr. *Python reference manual*. Centrum voor Wiskunde en Informatica Amsterdam, 1995.
- [41] J.R. Johansson, P.D. Nation, and Franco Nori. Qutip: An open-source python framework for the dynamics of open quantum systems. *Computer Physics Communications*, 183(8):1760–1772, August 2012.
- [42] Renato Renner. Security of quantum key distribution. *International Journal of Quantum Information*, 6(01):1–127, 2008.
- [43] Miguel Herrero-Collantes and Juan Carlos Garcia-Escartin. Quantum random number generators. *Reviews of Modern Physics*, 89(1):015004, 2017.
- [44] Charles H Bennett, Herbert J Bernstein, Sandu Popescu, and Benjamin Schumacher. Concentrating partial entanglement by local operations. *Physical Review A*, 53(4):2046, 1996.
- [45] Charles H Bennett, Gilles Brassard, Sandu Popescu, Benjamin Schumacher, John A Smolin, and William K Wootters. Purification of noisy entanglement and faithful teleportation via noisy channels. *Physical review letters*, 76(5):722, 1996.
- [46] Ryszard Horodecki, Paweł Horodecki, Michał Horodecki, and Karol Horodecki. Quantum entanglement. Reviews of modern physics, 81(2):865–942, 2009.
- [47] Timo Holz, Hermann Kampermann, and Dagmar Bruß. Device-independent secret-key-rate analysis for quantum repeaters. *Physical Review A*, 97(1):012337, 2018.
- [48] David P Nadlinger, Peter Drmota, Bethan C Nichol, Gabriel Araneda, Dougal Main, Raghavendra Srinivas, David M Lucas, Christopher J Ballance, Kirill Ivanov, EY-Z Tan, et al. Experimental quantum key distribution certified by bell's theorem. *Nature*, 607(7920):682–686, 2022.
- [49] Wei Zhang, Tim van Leent, Kai Redeker, Robert Garthoff, René Schwonnek, Florian Fertig, Sebastian Eppelt, Wenjamin Rosenfeld, Valerio Scarani, Charles C-W Lim, et al. A device-independent quantum key distribution system for distant users. *Nature*, 607(7920):687–691, 2022.
- [50] Erik Woodhead, Jędrzej Kaniewski, Boris Bourdoncle, Alexia Salavrakos, Joseph Bowles, Antonio Acín, and Remigiusz Augusiak. Maximal randomness from partially entangled states. *Physical Review Research*, 2(4):042028, 2020.
- [51] Eric Chitambar and Gilad Gour. Quantum resource theories. Reviews of modern physics, 91(2):025001, 2019.
- [52] Bob Coecke, Tobias Fritz, and Robert W Spekkens. A mathematical theory of resources. *Information and Computation*, 250:59–86, 2016.
- [53] Elie Wolfe, David Schmid, Ana Belén Sainz, Ravi Kunjwal, and Robert W Spekkens. Quantifying bell: The resource theory of nonclassicality of common-cause boxes. *Quantum*, 4:280, 2020.
- [54] Beata Zjawin, David Schmid, Matty J Hoban, and Ana Belén Sainz. Quantifying epr: the resource theory of nonclassicality of common-cause assemblages. Quantum, 7:926, 2023.

XIII. (APPENDIX) 2-CHSH-OPT Score Derivation

$$\begin{split} \operatorname{Tr}(\mathbf{P}_{\mathbf{p}}\mathbf{\bar{B}}_{\mathbf{q}'}^{T}) &= \operatorname{Tr}\left[\frac{1}{16}\left(\sum_{a,b=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))|e_{a}\rangle\langle e_{b}|\right)\left(\sum_{i,j=1}^{16}\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{j}\rangle\langle e_{i}|\right)\right] \\ &= \frac{1}{16}\operatorname{Tr}\left[\sum_{a,b=1}^{16}\sum_{i,j=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{a}\rangle\langle e_{b}|e_{j}\rangle\langle e_{i}|\right] \\ &= \frac{1}{16}\operatorname{Tr}\left[\sum_{a,b=1}^{16}\sum_{i,j=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{a}\rangle\langle e_{b}|e_{j}\rangle\langle e_{i}|\right] \\ &= \frac{1}{16}\operatorname{Tr}\left[\sum_{a,b=1}^{16}\sum_{i,j=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{a}\rangle\langle e_{i}|\right] \\ &= \frac{1}{16}\operatorname{Tr}\left[\sum_{a,b=1}^{16}\sum_{i,j=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{a}\rangle\langle e_{i}|\right] \\ &= \frac{1}{16}\sum_{a,b=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{a}\rangle\langle e_{i}|\right) \\ &= \frac{1}{16}\operatorname{Tr}\left(\sum_{a,b=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}|e_{a}\rangle\langle e_{i}|\right) \\ &= \frac{1}{16}\operatorname{Tr}\left(\sum_{a,b=1}^{16}\operatorname{Tr}(\rho_{\text{tr}}(A_{a}\otimes B_{b}))\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}\right) \\ &= \frac{1}{16}\operatorname{Tr}\left(\sum_{a,b=1}^{16}\left(A_{a}\otimes B_{b}\right)\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}\right) \\ &= \frac{1}{16}\operatorname{Tr}\left(\rho_{\text{tr}}(A_{a}\otimes B_{b})\mathbf{\bar{B}}_{\mathbf{q}'}^{ij}\right) \\ &= \frac{$$

XIV. (APPENDIX) Convincingness Computations for Optimized 2-CHSH Games

Table 3: Gapped Expressions for Non-Local Games

G	ω_{η_G} raw expressions
CHSH	$0.35355\eta^2 + 0.5$
2-CHSH	$0.12500\eta^4 + 0.35355\eta^2 + 0.25000$
2-CHSH-OPT $\eta' = 1$	$0.63748\eta^4 + 0.74686\eta^2(1-\eta^2) + 0.21875(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.95$	$0.61447\eta^4 + 0.97677\eta^2(1 - \eta^2) + 0.36133(1 - \eta^2)^2$
2-CHSH-OPT $\eta' = 0.90$	$0.61447\eta^4 + 0.97677\eta^2(1-\eta^2) + 0.36133(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.89$	$0.58682\eta^4 + 0.93057\eta^2(1-\eta^2) + 0.34375(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.88$	$0.58682\eta^4 + 0.93057\eta^2(1-\eta^2) + 0.34375(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.87$	$0.58682\eta^4 + 0.93057\eta^2(1-\eta^2) + 0.34375(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.86$	$0.58682\eta^4 + 0.93057\eta^2(1-\eta^2) + 0.34375(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.85$	$0.58682\eta^4 + 0.93057\eta^2(1-\eta^2) + 0.34375(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.84$	$0.53125\eta^4 + 1.06250\eta^2(1-\eta^2) + 0.53125(1-\eta^2)^2$
2-CHSH-OPT $\eta' = 0.83$	$0.53125\eta^4 + 1.06250\eta^2(1-\eta^2) + 0.53125(1-\eta^2)^2$
MSG	$0.27778\eta^4 + 0.222222\eta^2 + 0.5$

Note: 4-qubit games are in the form $k_{\text{ideal}}\eta^4 + k_{\text{part-mixed}}\eta^2(1-\eta^2) + k_{\text{mixed}}(1-\eta^2)^2$.

XV. (APPENDIX) 2-CHSH-OPT Local Bound Computation

The local bound of a 2-CHSH-OPT game is computed by exhaustively evaluating the score of all possible deterministic local strategies. The local strategies are represented as 16-bit strings corresponding to the input-output behaviour of both parties. For each strategy, we compute its score using the game's $\bar{\mathbf{B}}_{\eta'}$ matrix, and finally identify the maximum achievable score across all bit-strings. This approach is complete since any local strategy can be written as a probabilistic mixture of deterministic strategies, and the maximum score over deterministic strategies gives the local bound by convexity.

Algorithm 2: Compute Maximum Classical Score for a 2-CHSH-OPT Game

Input: Normalized Bell matrix $\mathbf{B}_{\eta'}$ encoded in the form B with elements B[x,y][a,b] where:

- $x \in [0, 1, 2, 3]$ represents input pairs (x_1, x_2) as (0, 0), (0, 1), (1, 0), (1, 1) for Alice
- $y \in \{0, 1, 2, 3\}$ represents input pairs $\{y_1, y_2\}$ as $\{0, 0\}, \{0, 1\}, \{1, 0\}, \{1, 1\}$ for Bob
- $a \in \{0, 1, 2, 3\}$ represents output pairs $\{a_1, a_2\}$ as $\{0, 0\}, \{0, 1\}, \{1, 0\}, \{1, 1\}$ from Alice
- $b \in [0, 1, 2, 3]$ represents output pairs (b_1, b_2) as (0, 0), (0, 1), (1, 0), (1, 1) from Bob

Each strategy s is a 16-bit string where:

- Bits [0-3]: Alice's first output (a_1) for inputs $x \in [0,1,2,3]$
- Bits [4-7]: Alice's second output (a_2) for inputs $x \in [0,1,2,3]$
- Bits [8-11]: Bob's first output (b_1) for inputs $y \in [0,1,2,3]$
- Bits [12-15]: Bob's second output (b_2) for inputs $y \in [0,1,2,3]$

```
Output: Maximum normalized score
```

return max score

```
Generate all possible deterministic strategies S = \{0, 1\}^{16};
\max \ \text{score} \leftarrow -\infty;
foreach strategy s \in S do
    joint prob \leftarrow 0;
    for x \leftarrow 0 to 3 do
         for y \leftarrow 0 to 3 do
              a_1 \leftarrow s[x];
               a_2 \leftarrow s[x+4];
              b_1 \leftarrow s[y+8];
               b_2 \leftarrow s[y+12];
               a \leftarrow 2a_1 + a_2;
               b \leftarrow 2b_1 + b_2;
              joint prob \leftarrow joint prob + B[x, y, a, b];
         end
    end
    \max \text{ score} \leftarrow \max(\max \text{ score}, \text{joint prob});
end
```

XVI. (APPENDIX) Proof of Noise-Tolerance Relation to Convincingness

Lemma 3. The noise-tolerance η_G^* for a non-local game, G, corresponds to a completely unconvincing state ($\mathfrak{C}_G = 1$), under asymptotic resource conditions $(N_{res} \to \infty)$.

Proof. By definition, the noise-tolerance, η_G^* , for a non-local game, G, is the visibility at which $\omega_{c_G} = \omega_{\eta_G}$. Rearranging, we see that:

$$\omega_{c_G} = \omega_{\eta_G} \iff 0 = \omega_{\eta_G} - \omega_{c_G} = \Delta_G(\eta_G^*) \implies \mathfrak{C}_G(\eta_G^*) = 1,$$

where the last implication is true for $N_{res} \to \infty$ by Eqn. 19. In other words, we can make two statements from different perspectives: (1) the value of G played with an input state with visibility η_G^* equals the local bound ω_{c_G} , so by the definition of non-local games, there is no non-locality in the input state and, (2) since $\mathfrak{C}_G = 1$ under $N_{res} \to \infty$, there is no likelihood of non-locality under these conditions. That is, η_G^* is the value at which the game is completely unconvincing for asymptotic resources. As desired.

XVII. (APPENDIX) Impossibility of DI-LOSR Conversion from the Magic-Square to Tsirelson Box

This proof was contributed by Dr. Elie Wolfe, and written in this thesis with his permission.

Theorem XVII.7. Under device-independent local operations and shared randomness (DI-LOSR), the perfect bipartite Magic-Square correlation ("Magic-Square box") cannot be converted into the unique quantum correlation that attains the Tsirelson bound for the CHSH game ("Tsirelson box").

Proof. A box (also called a correlation or behaviour) is a conditional probability distribution P(ab|xy) with binary inputs $x, y \in \{0, 1\}$ and outputs $a, b \in \{0, 1\}$. A DI-LOSR wiring is a box-to-box map that uses only local pre-processing of the inputs and post-processing of the outputs, together with shared randomness. Operationally it is a convex combination $\Lambda = \sum_r p_r \Lambda^{(r)}$ of deterministic local wirings $\{\Lambda^{(r)}\}_r$, each selected with probability p_r that depends solely on the shared random seed r. A deterministic DI local operation is any single extremal element $\Lambda^{(r)}$; equivalently, it is a DI-LOSR wiring with no further random choice once the seed is fixed (the special case of one seed). Importantly, applying such a deterministic LOSR wiring to a quantum-realizable box always yields another quantum-realizable box. The quantum realization of the second box will be different from the realization of the first box, but it will still be a quantum realization—possibly requiring different measurements or a different underlying state.

We note also that the Magic-Square box, \mathcal{M} , wins its game with probability 1, whereas the Tsirelson box, \mathcal{T} , achieves the quantum-maximal CHSH value $2\sqrt{2}$.

Assume for contradiction that a DI-LOSR wiring Λ satisfies $\Lambda(\mathcal{M}) = \mathcal{T}$. Several basic facts rule this out:

- (i) Uniqueness of the Tsirelson box. Up to local relabellings of inputs and outputs, the Tsirelson box \mathcal{T} is the unique non-local quantum-realizable box that attains the maximal CHSH value $2\sqrt{2}$.
- (ii) Rationality of M. The Magic Square box, M, has all probabilities given by rational numbers.
- (iii) Reduction to a deterministic operation. Write the wiring as a convex combination $\Lambda = \sum_r p_r \Lambda^{(r)}$ of deterministic device-independent local operations $\Lambda^{(r)}$ with weights $p_r \geq 0$ and $\sum_r p_r = 1$. Because the CHSH score is a linear function of the box, we know that:

(CHSH score of
$$\Lambda(\mathcal{M})$$
) = $\sum_{r} p_r (\text{CHSH score of } \Lambda^{(r)}(\mathcal{M})).$

The left-hand side is $2\sqrt{2}$, the maximum quantum CHSH value. Since every term in the sum is bounded above by this maximum, each coefficient with $p_r > 0$ must itself achieve the value $2\sqrt{2}$. By the uniqueness statement in (i), this forces every such deterministic wiring to output the Tsirelson box \mathcal{T} . In particular, there is at least one deterministic wiring, call it Λ_{det} , with $\Lambda_{\text{det}}(\mathcal{M}) = \mathcal{T}$.

- (iv) Deterministic wirings preserve rationality. A deterministic DI operation always maps rational probabilities to rational probabilities. Hence, by (ii), $\Lambda_{\text{det}}(\mathcal{M})$ must have only rational entries.
- (v) Contradiction. The Tsirelson box \mathcal{T} contains irrational probabilities (i.e., terms involving $1/\sqrt{2}$). Step (iv), thus, contradicts $\Lambda_{\text{det}}(\mathcal{M}) = \mathcal{T}$.

Therefore no DI-LOSR wiring can convert the Magic-Square box into the Tsirelson box.