The geometric link between Hardy and Clauser-Horne-Shimony-Holt

Johannes Seiler, 1 Thomas Strohm, 2 and Wolfgang P. Schleich 1, 3

1 Institut für Quantenphysik & Center for Integrated Quantum Science and Technology (IQST), Universität Ulm, D-89069 Ulm, Germany
2 Corporate Research, Robert Bosch GmbH, D-71272 Renningen, Germany
3 Hagler Institute for Advanced Study, Institute for Quantum Science and Engineering (IQSE), and Texas A&M AgrLife Research, Texas A&M University, College Station, Texas 77843-4242, USA

(Dated: May 16, 2022)

We show that the Hardy nonlocality condition is equivalent to the violation of the CHSH inequality with additional constraints. We adapt the geometrical optimization of the violation of the CHSH inequality to these additional constraints and show that the Hardy condition is equivalent to optimizing the length difference of two sides in a triangle. Furthermore, we discuss the effects of the different constraints.

I. INTRODUCTION

A typical Bell test scenario [1–10] consists of two qubits and two independent projective measurements on each qubit. The measurements are chosen in such a way that the sum of expectation values of the four possible correlated measurements violate the Clauser-Horne-Shimony-Holt (CHSH) inequality [11], which would hold classically.

A different approach to detect nonlocality was proposed by Hardy in 1993 [12]. In the Hardy scenario, the setup is similar, but four measurement outcomes are selected and three of their associated joint probabilities are fixed. Classically, the chosen probabilities lead to a vanishing outcome of the forth joint measurement. The nonlocality of quantum mechanics then manifests itself in a nonvanishing measurement outcome.

In this article we establish a connection between the Hardy scenario and the CHSH inequality: The nonvanishing outcome in the Hardy scenario is proportional to the violation in the CHSH inequality under three additional constraints on the measurements. In order to demonstrate this notion, we first start from the Hardy scenario and show that one of the probabilities is given by the expectation value of the CHSH inequality, which when optimized with respect to three constraints yields the Hardy scenario. In this way we gain a geometrical insight of the Hardy scenario, which further allows us to find and comprehend the optimal measurement settings to maximize the probability of finding a violation of the classical result.

When we start from the CHSH inequality and apply more and more constraints to the measurement settings we again reach the Hardy case. This approach brings out most clearly the crucial role of the constraints in obtaining the Hardy nonlocality condition.

FIG. 1. Schematic connection of the measurement results and contradiction to classical local theory. The two points $S = 1$ and $R = 1$ can be measured in single measurement on both subsystems with probability $w$. The conditional probabilities infer that one would also measure $Q = 1$ and $T = 1$ in a single run with the same probability. Instead this joint probability is zero and thus contradicts the nonlocality assumption.

A. Hardy scenario

The idea proposed by Hardy is to construct a set of measurements on a bipartite state, such that a specific measurement outcome is predicted by quantum mechanics, which cannot be explained by classical physics.

For this purpose, we consider a bipartite two qubit state. On the subsystem $A$ we either perform the measurement $Q$ or $R$ which each can yield the outcome +1 or −1. Analogously, on the subsystem $B$ we choose between the two measurements $S$ and $T$.

The measurements are defined such that the joint probability measuring $Q = 1$ and $R = 1$ is zero. Furthermore, if one measures $S = 1$, we always find $Q = 1$ and analogously, if we measure $R = 1$, one always finds $T = 1$. The question is then, what is the joint probability to obtain $R = 1$ and $S = 1$ in a single measurement.

In any classical, local theory, the measurements on each subsystem do not influence each other and we change the measurement on a subsystem without chang-
ing the measurement outcome on the other subsystem.

We now assume that we find an outcome $R = 1$ and $S = 1$ in a single measurement. Then changing the measurement on the subsystem $A$ from $R$ to $Q$, we would obtain $Q = 1$. In full analogy, if we change our measurement on the subsystem $B$ from $S$ to $T$, we would measure $T = 1$.

As a consequence, if we changed the measurements on both subsystems simultaneously, the outcome of our new measurements would be $Q = 1$ and $T = 1$, which is in contradiction to the requirement, that the joint probability of $Q = 1$ and $T = 1$ is set to zero. Thus, for any classical theory, a common measurement of $R = 1$ and $S = 1$ cannot occur. This contradiction is schematically depicted in Figure 1.

However, it can be shown that quantum theory allows to construct measurements such that they obey the all of three requirements mentioned above, and still have a nonvanishing probability $w$ of finding $R = 1$ and $S = 1$ in the same run. We denote this situation as Hardy nonlocality. For a vivid interpretation of this scenario we refer to the work of Kwiat and Hardy [13].

Therefore, in contrast to Bell inequalities, such as the CHSH inequality, in principle only a single measurement can suffice to violate the classical predictions. This behavior is similar to the famous Greenberger-Horne-Zeilinger (GHZ) scenario [14] for three qubits and its extensions to multipartite multidimensional systems [15]. However, in contrast to the GHZ model, the violation in the Hardy scenario only occurs probabilistically. Nevertheless, in any implementation, in order to guarantee the other three conditions, one theoretically needs to perform infinitely many measurements.

### B. Outline

Our article is organized in the following way. In Sec. III we formulate the Hardy scenario in a quantum mechanical formalism and show how it is related to the CHSH inequality plus additional constraints. We then develop a geometrical interpretation of the probability $w$ in Sec. III. We exploit this interpretation in Sec. IV to optimize the nonlocality probability $w$, and determine the optimal measurement settings in Sec. V. In Sec. VI we compare our findings for the Hardy scenario to the results for the CHSH scenario, and discuss how the different constraints in the Hardy scenario effect the results found. Finally, we provide a conclusion in VII.

In order to keep our article self-containing but focused on the main results, we provide three appendices. In Appendix A we calculate the relations between the different measurement vectors due to the additional constraints. We devote Appendix B to obtaining an analytical expression for the expected correlation value in terms of spherical coordinates. We finally provide a derivation of the violation probability in terms of a length appearing in our geometrical picture in Appendix C.

### II. CONNECTION BETWEEN THE HARDY SCENARIO AND THE CHSH INEQUALITY

In this section, we connect the Hardy scenario to the well-known Clauser-Horne-Shimony-Holt (CHSH) inequality. For this purpose, we first set up the quantum state and measurements. We then provide the quantum theoretical formulation of the CHSH inequality and the Hardy scenario. Thereafter, we show that these formulations are equivalent to one another with the exception that the Hardy scenario must obey three additional constraints on the measurements performed. We discuss the effects of these constraints on the measurement directions and the maximization of the nonlocality violation probability in the Hardy scenario. Finally, we numerically optimize this probability.

#### A. Quantum state and measurements

We first define the quantum mechanical state and measurements and translate the classical probabilities of the Hardy problem, presented in Section III, into quantum mechanical probabilities.

Throughout the remainder of this article, we consider a pure two-qubit state

$$|Ψ⟩ = \frac{1 + \sqrt{1 - C^2}}{2} |0⟩_A |0⟩_B - \sqrt{1 - \frac{1 - C^2}{2}} |1⟩_A |1⟩_B$$

in the Schmidt basis [16], where $|0⟩$ and $|1⟩$ are the eigenstates of the Pauli $σ_z$ operator with eigenvalue +1 and −1 in the respective subsystem. Furthermore, the concurrence $C$ [17, 18], with $0 ≤ C ≤ 1$ is a measure of the entanglement between the two subsystems of the two-qubit state. For $C = 0$ the state is separable, while $C = 1$ is a maximally entangled Bell state.

We note that the state $|Ψ⟩$ is an arbitrary state with given entanglement, that is any other state with the same entanglement can be transformed into Equation (1) by a change of basis.

Due to the symmetry of the state $|Ψ⟩$ in the Schmidt basis, the Bloch vectors $a_A = ⟨Ψ|σ_A|Ψ⟩$ and $a_B = ⟨Ψ|σ_B|Ψ⟩$ of the two subsystems $A$ and $B$, where $σ$ is the vector of the Pauli matrices $σ_x$, $σ_y$ and $σ_z$, have the same representation

$$a_A = a_B = \begin{pmatrix} 0 \\ 0 \\ \sqrt{1 - C^2} \end{pmatrix} = a$$

and we drop the indices in the following.

On each of the two subsystems we perform one of two possible measurements. On the subsystem $A$, we choose between the operators

$$\hat{Q} = q \cdot \hat{σ}_A \quad \text{or} \quad \hat{R} = r \cdot \hat{σ}_A,$$
where \( q \) and \( r \) are three-dimensional unit vectors that denote the measurement direction on the Bloch sphere of the subsystem \( A \).

In full analogy, on the subsystem \( B \) we perform either the measurement
\[
\hat{S} = s \cdot \hat{\sigma}_B \quad \text{or} \quad \hat{T} = t \cdot \hat{\sigma}_B, \tag{4}
\]
with the three-dimensional unit vectors \( s \) and \( t \).

### B. CHSH inequality

From these measurement settings, we derive the sum of expectation values
\[
S \equiv \langle \hat{Q} \otimes \hat{S} \rangle - \langle \hat{Q} \otimes \hat{T} \rangle + \langle \hat{R} \otimes \hat{S} \rangle + \langle \hat{R} \otimes \hat{T} \rangle \tag{5}
\]
of the individual measurement correlations. Inserting the state \( \langle \psi \rangle \), Equation (1), and the measurement operators, Equations (3) and (4), into Equation (5), we obtain the expectation value
\[
S = q^T K (s - t) + r^T K (s + t) \tag{6}
\]
in terms of the measurement vectors \( q \), \( r \), \( s \) and \( t \) and the correlation matrix
\[
K \equiv \langle \psi \vert \hat{\sigma}_A \otimes \hat{\sigma}_B \vert \psi \rangle, \tag{7}
\]
which is a property of the state alone. For the state \( \langle \psi \rangle \), Equation (1), the correlation matrix is explicitly given by
\[
K = \text{diag}(-C, C, 1), \tag{8}
\]
where the diagonal form of the matrix is a direct consequence of using the Schmidt decomposition. Note, that \( Ka = a \), since \( a \) points along the \( z \) axis.

For any classically correlated, local system, the CHSH inequality
\[
S \leq 2 \tag{9}
\]
holds. However, Equation (9) can be violated for entangled quantum mechanical states. Indeed, for a state with concurrence \( C \), the expectation value is bounded by
\[
S_{\text{CHSH}} \leq 2\sqrt{1 + C^2} \tag{10}
\]
for any CHSH like experiment. Only for separable states (\( C = 0 \)) a violation of the CHSH inequality is impossible. For maximally entangled states, that is \( C = 1 \), we find \( S \leq 2\sqrt{2} \), the well-known Tsirelson bound [19][20].

### C. Quantum probabilities in the Hardy setting

In contrast to the CHSH inequality, the idea of the Hardy scenario is to prepare the above measurements in such a way, that we guarantee them to fulfill the three probability distributions for the outcomes \( Q, R, S \) and \( T \) associated with a single measurement of the operator \( \hat{Q}, \hat{R}, \hat{S} \) and \( \hat{T} \), discussed in Section [2].

The first condition is that the joint probability
\[
\Pr_{Q,R,S}(1,1) = \frac{\langle \Psi | \frac{1}{2}(\hat{I} + \hat{Q}) \otimes \frac{1}{2}(\hat{I} + \hat{S}) | \Psi \rangle}{\langle \Psi | \frac{1}{2}(\hat{I} + \hat{S}) | \Psi \rangle} = 1, \tag{11}
\]
of simultaneous measurement of \( \hat{Q} \) and \( \hat{T} \) on the two subsystems of the state \( \langle \psi \rangle \) cannot yield the outcome 1 on both sides.

The second requirement is finding the outcome \( S = 1 \) when measuring \( \hat{S} \) on subsystem \( B \). This will definitely result in the outcome \( Q = 1 \) for a measurement of \( \hat{Q} \), that is we have the conditional probability
\[
\Pr_{Q|S}(1|1) = \frac{\langle \Psi | \frac{1}{2}(\hat{I} + \hat{Q}) \otimes \frac{1}{2}(\hat{I} + \hat{S}) | \Psi \rangle}{\langle \Psi | \frac{1}{2}(\hat{I} + \hat{S}) | \Psi \rangle} = 1, \tag{12}
\]
In full analogy to the second condition, the third is the conditional probability
\[
\Pr_{T|R}(1|1) = \frac{\langle \Psi | \frac{1}{2}(\hat{I} + \hat{R}) \otimes \frac{1}{2}(\hat{I} + \hat{T}) | \Psi \rangle}{\langle \Psi | \frac{1}{2}(\hat{I} + \hat{T}) | \Psi \rangle} = 1. \tag{13}
\]
which predicts that we find \( T = 1 \) on the subsystem \( B \), when we measured \( R = 1 \) on the subsystem \( A \).

We then ask the question, what is the probability
\[
w = \Pr_{R,S}(1,1) \equiv \langle \Psi | \frac{1}{2}(\hat{I} + \hat{R}) \otimes \frac{1}{2}(\hat{I} + \hat{S}) | \Psi \rangle \tag{14}
\]
of the joint measurement outcomes \( R = 1 \) and \( S = 1 \).

As we have discussed in the introduction, if our measurement settings fulfill Equations (11), (12) and (13), obtaining a single measurement outcome with \( R = 1 \) and \( S = 1 \) in the same measurement is a violation of any local hidden variable theory.

Inserting the definitions of our state \( \langle \psi \rangle \), Equation (1), and measurement operators, Equations (3) and (4), into the definition of the joint and conditional probabilities, Equations (11), (12), (13) and (14), we obtain a set of four equations:
\[
q^T K t = -(1 + a \cdot q + a \cdot t), \tag{15}
\]
\[
r^T K s = 1 - a \cdot q + a \cdot s, \tag{16}
\]
\[
r^T K t = 1 + a \cdot r - a \cdot t, \tag{17}
\]
\[
r^T K s = -(1 + a \cdot r + a \cdot s) + 4w, \tag{18}
\]
which only depend on the measurement vectors \( q, r, s \) and \( t \), the Bloch vector \( a \), and the correlation matrix \( K \).

### D. Relation to the CHSH inequality

We are now in the position to compare the Hardy scenario with the CHSH inequality. In order to do so, we
add Equations (16), (17) and (18) and subtract Equation (15), and obtain

\[ q^T K(s - t) + r^T K(s + t) = 4w - 2. \]  

(19)

We identify the left hand side of Equation (19) as the expectation value \( S \), Equation (6), at the heart of the CHSH inequality.

Hence, the probability

\[ w = \frac{S - 2}{4} \]  

(20)

of finding a measurement outcome that contradicts the locality condition is directly connected to the violation of the CHSH inequality, Equation (9). If and only if the measurement setting would violate this inequality, it is possible to obtain a measurement of \( Q = 1 \) and \( S = 1 \) in the same run. For this reason, we call \( w \) the violation probability.

In order to maximize \( w \), we simply have to maximize the expectation value \( S \), which corresponds to maximizing the violation of the CHSH inequality.

The maximization of \( S \) over the four independent measurements \( q, r, s \) and \( t \) for a pure state is known [21]. We even demonstrated a geometrical approach [22] which allows one to find all possible measurement that maximize \( S \) for a given concurrence \( C \).

However, in addition to Equation (20), the probabilities Equations (11), (12) and (13) between correlated measurement outcomes still must be fulfilled. As a consequence the measurement vectors \( q, r, s \) and \( t \) are no longer independent of each other.

Unfortunately, as it turns out none of the infinitely many optimal measurement settings for maximizing \( S \) for a given concurrence \( C \) without any further constraints fulfills the three additional constraints, Equations (24), (25) and (26), on the measurement vectors. In order to maximize \( w \), the expectation value thus has to be maximized again under these new constraints.

E. Constraints for measurement directions

We now take a closer look at the additional constraints, which arise from demanding the probabilities Equations (11), (12) and (13).

The first constraint, Equation (15), leads to

\[ (Kq + a) \cdot t = -(1 + a \cdot q), \]  

(21)

which we rewrite as

\[ |Kq + a| |t| \cos \alpha = -|1 + a \cdot q|, \]  

(22)

where \( \alpha \) is the angle between \( t \) and \( Kq + a \). The premise of \( t \) being a unit vector, together with the relation

\[ |Kq + a| = |1 + a \cdot q|, \]  

(23)

which is proven in Appendix A, then constitute that the measurement vector \( t \) is antiparallel to \( Kq + a \), leading to the connection

\[ t = -\frac{Kq + a}{|Kq + a|} \]  

(24)

between the vectors \( q \) and \( t \).

The second constraint, Equation (16), leads to the connection

\[ s = \frac{Kq - a}{|Kq - a|} \]  

(25)

between the vectors \( q \) and \( s \).

Analogously, we find

\[ r = \frac{Kt - a}{|Kt - a|}. \]  

(26)

between the vectors \( r \) and \( t \) from the third constraint, Equation (17).

As a consequence, instead of four independent vectors \( q, r, s \) and \( t \), the three vectors \( r, s \) and \( t \) are fully determined by the vector \( q \). We only have one independent vector \( q \) left.

F. Numerical maximization of the probability \( w \)

The expectation value \( S \), Equation (6), under the constraints Equations (24), (25) and (26), and hence the probability \( w \), Equation (20), is a function of only a single measurement vector.

In order to maximize the probability \( w \), we parameterize the measurement vector

\[ q = \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix} \]  

(27)

in spherical coordinates by two angles \( \theta \) and \( \phi \) and insert the resulting expression together with the expressions for the other three vectors \( r, s, t \), Equations (24), (25) and (26), and the correlation matrix \( K \), Equation (8), into the definition of the expectation value \( S \), Equation (30). The resulting expression for the expectation value, given in Appendix B, only depends on the angle \( \theta \) and the concurrence \( C \), but is rather cumbersome and difficult to maximize analytically.

Therefore, we first analyze this result numerically. Figure 2 shows the probability \( w \) as a function of the polar angle \( \theta \) and the concurrence \( C \).

As a result, we obtain that the violation probability \( w \) is vanishing for either separable states, that is \( C = 0 \), or maximally entangled states (\( C = 1 \)). While the vanishing probability for \( C = 0 \) is expected, since entanglement is a key ingredient in the violation of the CHSH inequality, the case of a maximally entangled state is more of a surprise, as the CHSH inequality is in general maximally
FIG. 2. The probability \( w \) as a function of the polar angle \( \theta \) of the vector \( q \) and the concurrence \( C \). For both separable \((C = 0)\) and maximally entangled states \((C = 1)\), as well as for \( q \) pointing along the positive \((\theta = 0)\) or negative \((\theta = \pi/2)\) \( z \) axis, the probability \( w \) vanishes. The black line marks the optimal \( w \) as a function of the concurrence \( C \). The grey line indicates the projection of the optimal line in the \( C - \theta \) plane.

violated for a maximally entangled state. Hence, the vanishing probability must occur due to the additional constraints on the measurement vectors.

Moreover, the probability \( w \) also vanishes if the polar angle of the measurement vector \( q \) is either \( \theta = 0 \) or \( \theta = \pi/2 \). For these cases the vector \( q \) points along the positive or negative \( z \) axis, respectively. This result is also not true for the general CHSH case, and thus must result from the additional constraints.

Furthermore, for any fixed concurrence \( 0 < C < 1 \) a unique angle \( \theta \) exists, such that the probability \( w \) is maximal. These values are indicated by the black line in Figure 2. For small concurrences the optimal angle is close to \( \pi \), that is the optimal vector \( q \) points mostly along the negative \( z \) axis. With increasing \( C \) the angle \( \theta \) decreases monotonically towards \( \pi/2 \), that is the vector \( q \) moves towards the \( x-y \) plane. For all concurrences the optimal vector \( q \) thus has a negative \( z \) component.

III. GEOMETRIC INTERPRETATION

In the previous section we have shown that the violation probability \( w \) in the Hardy scenario is related to the expectation value \( S \) of the CHSH inequality with three additional constraints on the measurement vectors. The optimization of the probability is thus a optimization of the CHSH setup under constraints.

Moreover, we have performed a numerical evaluation of the probability \( w \), in which we have seen that the probability is non-vanishing as long as the state is neither unentangled nor maximally entangled, that is \( 0 < C < 1 \), and the vector \( q \) does not point along neither the \( z \) axis nor in the \( x-y \) plane. Furthermore, there exists a unique maximum of the probability for any given concurrence \( C \), for which the vector \( q \) always has a negative \( z \) component.

In Ref. [22], we have demonstrated that the optimal expectation value \( S \) has a geometrical interpretation as half of the perimeter of a parallelogram enclosed by an ellipse whose semimajor and semiminor axes have lengths 1 and \( C \), respectively. This interpretation allows us to efficiently and analytically find all possible optimization strategies for the expectation value \( S \).

In this section we adapt our geometric picture to the Hardy scenario, and apply it to explain the results found in our numerical analysis in the previous section. In particular we first show that all the relevant measurement vectors lie in a common plane. This behavior explains the independence of the probability \( w \) on the azimuthal angle \( \phi \) of the vector \( q \). We then show that for all instances in which the probability \( w \) vanishes, the plane collapses to a single line. Moreover, we demonstrate how the relevant vectors are constructed geometrically. Finally, we exploit this construction to provide a geometrical interpretation of the probability \( w \) as the length difference in a triangle, and use this description to parameterize the probability.

A. Measurement vectors lie in a plane

In our numerical calculation of the probability \( w \), we have seen that it is independent of the azimuthal angle \( \phi \) of the vector \( q \). The probability is therefore symmetric under rotations around the \( z \) axis. In this section we show this symmetry by proofing that all relevant vectors lie in a common plane, which contains the \( z \) axis of our coordinate system.

In the expectation value \( S \), Equation (3), only combinations of a vector of the subsystem \( A \) and a vector of the subsystem \( B \) appear, while the correlation matrix \( K \) specifies how vectors from different subsystems are multiplied to each other.

Instead of using the correlation matrix as a connection between the two subsystems, we directly apply the correlation matrix to the vectors of the subsystem \( B \), defining the new vectors

\[
\mathbf{s}_K \equiv K \mathbf{s}
\]

and

\[
\mathbf{t}_K \equiv K \mathbf{t}.
\]

These new vectors are no longer of unit length. Instead, their length depends on their orientation, since their components in the \( x \) and \( y \) direction are contracted by the factor \( C \).

The new vectors \( \mathbf{s}_K \) and \( \mathbf{t}_K \) directly multiply to the vectors \( \mathbf{q} \) and \( \mathbf{r} \) of the subsystem \( A \) by the usual scalar product, that is we find

\[
S = \mathbf{q} \cdot (\mathbf{s}_K - \mathbf{t}_K) + \mathbf{r} \cdot (\mathbf{s}_K + \mathbf{t}_K)
\]

(30)
for the expectation value.

Our aim is to construct the vectors \( q, r, s_K \) and \( t_K \) geometrically. For this purpose we first show that these vectors lie in a common plane, defined by the vector

\[
K^2 q = \begin{pmatrix} C^2 q_x \\ C^2 q_y \\ q_z \end{pmatrix},
\]

which is in general not parallel to the vector \( q \), and the Bloch vector \( a \).

From the definitions of \( s_K \) and \( t_K \). Equations (28) and (29), and of \( s \) and \( t \), Equations (24) and (25), the new vectors \( s_K \) and \( t_K \) are given by the explicit expressions

\[
s_K = \frac{K^2 q - a}{|K q - a|}
\]

and

\[
t_K = -\frac{K^2 q + a}{|K q + a|}
\]

where we made use of the fact that the correlation matrix \( K \) does not change the Bloch vector \( a \), since the latter points along the \( z \) axis of our coordinate system.

In general the Bloch vector \( a \) is fixed, while the measurement vector \( q \) is our free parameter. For any arbitrary but fixed \( q \), it immediately follows form Equations (32) and (33), that \( s_K \) and \( t_K \) are in the plane \( P \) spanned by the vectors \( K^2 q \) and \( a \).

Furthermore, from Equations (26) and (29) it follows that the vector

\[
r = \frac{t_K - a}{|t_K - a|}
\]

is a linear combination of the two vectors \( t_K \) and \( a \). Therefore, the vector \( r \) also lies in the plane \( P \).

Moreover, since the measurement vector \( q \) can be rewritten as

\[
q = \frac{K^2 q - (q \cdot a) a}{C^2},
\]

which directly follows form evaluating the right hand side, \( q \) is also a linear combination of \( a \) and \( K^2 q \). The plane \( P \) thus contains the vector \( q \).

Therefore all the four vectors appearing in Equation (30) lie in \( P \) and we thus restrict our discussions and calculations for the remainder of our article to the plane \( P \).

We finally note that the Bloch vector \( a \) is parallel to the \( z \) axis of our coordinate system, and the plane \( P \) thus contains the \( z \) axis. Therefore, the second direction orthogonal on \( z \) is a vector in the \( x-y \) plane, which we denote by \( x' \). Hence, \( P \) is the \( x'-z \) plane. From this form of \( P \) it directly follows that the expectation value \( S \), Equation (20), is independent of the azimuthal angle \( \phi \) of the vector \( q \), which is consistent with our results from the previous section.

B. Degenerate plane leads to vanishing violation probability

The plane \( P \) is only uniquely defined as long as the vectors \( K^2 q \) and \( a \) are neither parallel nor either of them vanishes. We now examine the cases where this condition is not fulfilled.

The first case we consider is when \( a \) is vanishing. According to its definition, Equation (2), the Bloch vector only vanishes for \( C = 1 \), which furthermore implies \( K = \text{diag}(-1, 1, 1) \). Inserting these conditions into Equations (24), (25) and (26), we obtain \( q = s_K = -t_K = -r \), and thus all relevant vectors point along the same line. From our expression of the expectation value, Equation (30), it immediately follows under these conditions that \( S = 2 \), and therefore \( w = 0 \).

The second case we consider is that both \( K^2 q \) and \( a \) are parallel, which corresponds to the vector \( q \) pointing along the \( z \) axis of our coordinate system. It immediately follows from Equations (24), (25) and (26), that the three other vectors \( r, s_K \) and \( t_K \) also point along the \( z \)-axis of our coordinate system. By the same argument as for the case \( C = 1 \), it immediately follows, that \( w = 0 \).

Finally, we consider the case of \( K^2 q = 0 \), which is only achievable for \( C = 0 \). In this case the underlying state \( |\Psi\rangle \) is separable. Since a separable state never violates the CHSH inequality, the violation probability \( w \) is also vanishing.

Hence, the vectors \( q, r, s_K \) and \( t_K \) are either all located in a common well defined plane \( P \) or they point along a single line. In the latter case, the probability \( w \) always vanishes. Since we are only interested in obtaining a nonvanishing probability \( w \), we restrict ourselves in the following to the case of all four vectors \( q, r, s_K \) and \( t_K \) lying in a common plane \( P \).

C. Geometric construction of the measurement vectors

We are now in the position to construct the four vectors \( q, r, s_K \) and \( t_K \) in the plane \( P \) geometrically. This geometric interpretation is depicted in Figure 3.

We start with the vector \( q \), which is the only vector that can be chosen freely. Since \( q \) is a unit vector, it points from the origin of our new coordinate system to a unit circle.

The next vector we consider is the vector \( s_K \), which is given by Equation (32). When starting at the origin the vector \( s_K \) terminates in a point \( S \) on an ellipse with semi major axis of unit length along the \( z \) axis, and semi minor axis of length \( C \). Using the representation of the Bloch vector \( a \), Equation (2), the correlation matrix \( K \), Equation (8), and the fact that \( q \) is a unit vector, we show in Appendix A that the vector

\[
s_K = \xi q - a.
\]
defines a line spanned by the support vector $-\mathbf{a}$ and the directional vector $\mathbf{q}$. Here, we introduced the abbreviation

$$
\xi \equiv \frac{C^2}{1 - \sqrt{1 - C^2 q_z}} = |s_K + a| \tag{37}
$$

with $q_z$ being the component of $\mathbf{q}$ along the $z$ axis.

We therefore geometrically construct the vector $s_K$, and thus the point $S$, by going from the lower focal point $F_1$ in the direction of $\mathbf{q}$. The intersection with the ellipse constitutes the point $S$.

In an analogous way, we determine the point $T$ on the ellipse, which is defined by the vector $t_K$ pointing from the origin to the ellipse. By using the definition of $t_K$, Equation (24), as well as the Bloch vector $\mathbf{a}$, and the correlation matrix $K$, Equation (8), we demonstrate in Appendix A that the vector $t_K = -\tau \mathbf{q} - \mathbf{a}$.

$$
\tau \equiv \frac{C^2}{1 + \sqrt{1 - C^2 q_z}} = |t_K + a|, \tag{39}
$$

where we introduced the abbreviation.

is constructed by going from the lower focal point $F_1$ in the direction $-\mathbf{q}$ to the ellipse.

Our construction directly shows that, as a consequence of the additional constraints, Equations (24) and (25), the line $ST$, which also defines the vector $s_K - t_K$, intersects the $z$ axis in the focal point $F_1$ of the ellipse.

From the vectors $s_K$ and $t_K$ we construct the vectors $-s_K$ and $-t_K$, which point from the origin to the points $S$ and $T$, respectively, on the ellipse. Due to symmetry, the line $ST$, which also determines to the vector $s_K - t_K$, intersects the $z$ axis in the upper focal point $F_2$.

The only vector left is the vector $\mathbf{r}$. In Appendix A we derive that this vector is given by the linear combination

$$
\mathbf{r} = \mathbf{q} - \frac{2}{2 - \tau} (\mathbf{q} + \mathbf{a}) \tag{40}
$$

of the measurement vector $\mathbf{q}$ and the Bloch vector $\mathbf{a}$. This expression shows that the vector $\mathbf{r}$ can be constructed by going from the origin of our coordinate system to the end point of the vector $\mathbf{q}$ and then along the direction $-(\mathbf{q} + \mathbf{a})$, which is through the focal point $F_1$, until one again intersects the unit circle. This intersection is then the end point of the vector $\mathbf{r}$. As a consequence of this construction, it is obvious that the line between the two end points of the vectors $\mathbf{q}$ and $\mathbf{r}$ on the unit circle includes the focal point $F_1$.

This behaviour is similar to the construction of the points $S$ and $T$, where the line $ST$ also includes the same focal point $F_1$.

We furthermore notice, that the vector $\mathbf{r}$ is by its definition, Equation (24), parallel to the vector $t_K - \mathbf{a}$, and therefore the line $TF_1$ between the point $T$, defined by the vector $-t_K$, on the ellipse and the focal point $F_1$.

D. Violation probability is the length difference of two sides in a triangle

We now utilize the geometrical picture of the vectors $\mathbf{q}$, $\mathbf{r}$, $s_K$ and $t_K$ to find a geometrical interpretation of the violation probability $w$. We start by determining the expectation value $S$ in our geometrical picture. We then use this result to relate the probability $w$ to the length difference in a triangle.

The expectation value $S$, Equation (9), consist of the two terms $\mathbf{q} \cdot (s_K - t_K)$ and $\mathbf{r} \cdot (s_K + t_K)$.

The first scalar product

$$
\mathbf{q} \cdot (s_K - t_K) = \xi + \tau \tag{41}
$$

is the length of the line segment $\bar{ST}$, since $\mathbf{q}$ is parallel to $s_K - t_K$.

The measurement vector $\mathbf{r}$ is in general not parallel to $s_K + t_K$. Thus, the second scalar product

$$
\mathbf{r} \cdot (s_K + t_K) = |\bar{RT}|, \tag{42}
$$

is the distance between the points $\bar{T}$ and $R$, where the latter is defined by the parallel projection of $s_K + t_K$ on the vector $\mathbf{r}$.
The expectation value is the distance from the point $\bar{T}$ to the point $R$. We note, that $R$ can be in principle be further away from $\bar{T}$ than $F_1$. To include this effect, $\lambda$ is negative in this case.

When we insert this result into Equation (20), we find that the violation probability

$$w = \frac{\xi - \lambda}{4}$$  \hspace{1cm} (45)

is proportional to the difference between the two lengths $\xi$ and $\lambda$. Geometrically, the violation probability manifests itself as the difference between the hypothenuse and the leg of a right triangle. This interpretation is depicted in Figure 4(b).

In Appendix C, we further discuss the geometric interpretation of $w$ and derive the expression

$$w = \frac{C_2^2}{4} \left( 1 - \frac{\tau C_2^2}{(2 - \tau)(2\tau - C_2^2)} \right)$$  \hspace{1cm} (46)

which only depends on the concurrence $C$ and the distance $\tau$ between the point $\bar{T}$ and the focal point $F_u$.

IV. OPTIMIZATION OF THE PROBABILITY

We now use our geometric considerations to optimize the violation probability $w$ for a given entangled bipartite state $|\Psi\rangle$. We start by determining the optimal violation probability for an arbitrary but fixed concurrence $C$, and then discuss the special cases of small and large concurrences. Moreover, we identify the optimal concurrence which optimizes the violation probability $w$ over all concurrences, and discuss its relation to the golden ratio.

A. Optimal violation probability

We start by maximizing the probability $w$, Equation (46), for an arbitrary but fixed concurrence $C$ over all possible lengths $\tau$ by calculating the derivative

$$\frac{\partial w}{\partial \tau} = \frac{C_2^4}{2(2 - \tau)(2\tau - C_2^2)} \left( 1 + \frac{\tau}{2 - \tau} - \frac{2\tau}{2\tau - C_2^2} \right)$$  \hspace{1cm} (47)

and determining its roots. Straightforward algebra shows, that the roots are given by

$$\tau = \pm C,$$  \hspace{1cm} (48)

and since $\tau$ represents a length, only the positive solution

$$\tau_{\text{opt}} = C,$$  \hspace{1cm} (49)

allows us to optimize the probability $w$.

Inserting the optimal length $\tau_{\text{opt}}$, Equation (49), into the expression of $w$, Equation (46), we obtain

$$w_{\text{opt}}(C) = C_2^2 \frac{1 - C}{(2 - C)^2}$$  \hspace{1cm} (50)

as the optimal violation probability in the Hardy scenario, as a function of the concurrence $C$.  

FIG. 4. Geometric interpretation of the expectation value $\mathcal{S}$ and the probability $w$. (a) The expectation value $\mathcal{S}$ is given by the length of the two line segments $|ST|$ and $|RT|$ (solid red lines). Due to the focal point property $|F_u \bar{T}| + |\bar{T}F_1| = 2$, the expectation value $\mathcal{S} = 2 + \xi - |RF_1|$ can be expressed by $\xi$ and the line segment $\lambda$ (light green line). (b) The probability $w = (\xi - \lambda)/4$ is therefore four times the difference between the hypotenuse and the cathete of the rectangular triangle spanned by the points $S$, $R$ and $F_1$.  

From Figure 4(a) and the focal point property of the ellipse, we find the relation

$$2 - \tau = |RT| + \lambda,$$  \hspace{1cm} (43)

where $\lambda$ is the signed length of the line segment $|RF_1|$ between the point $R$ and the lower focal point $F_1$. Note, that $\lambda$ is defined positive if $R$ is between $\bar{T}$ and $F_1$, and negative otherwise.

Inserting the expressions for the two scalar products, Equations (41) and (42), together with the relation between the different lengths, Equation (43), into the expectation value $\mathcal{S}$, Equation (30), we find

$$\mathcal{S} = 2 + (\xi - \lambda).$$  \hspace{1cm} (44)
B. Special cases

We now consider two special cases of the optimal violation probability \( w_{\text{opt}} \). The first case is small concurrences, that is when the common bipartite state is nearly separable. The second case we consider is the case of \( C \lesssim 1 \), that is for almost maximally entangled states.

For small concurrences, we expand the optimal expectation value, given by Equation (50), into the Taylor series

\[
    w_{\text{opt}} \approx \frac{C^2}{4} + O(C^4)
\]

at \( C = 0 \). Thus, for small entanglement, the expectation value and thus the probability \( w \) grows quadratically with the concurrence. This behavior is similar to the optimal expectation value \( S_{\text{opt}} = 2\sqrt{1-C^2} \) for the CHSH scenario, which also grows quadratically in the concurrences for small \( C \). In fact, the violation probability \( w_{\text{opt}} \) of the Hardy scenario only differs in order \( O(C^3) \) from its corresponding value in the CHSH scenario. Therefore, for small concurrences \( C \), the additional constraints Equations (24), (25) and (26) have a negligible effect on the achievable expectation value \( S_{\text{opt}} \), and thus on the optimal violation probability \( w_{\text{opt}} \).

For \( C \lesssim 1 \), we expand Equation (50) around \( \epsilon = 1 - C \), and find

\[
    w_{\text{opt}} \approx \epsilon - 4\epsilon^2 + O(\epsilon^3),
\]

which decreases linearly with the concurrence when \( C \) goes to 1. For maximally entangled states, that is \( C = 1 \), the violation probability \( w \) vanishes. This behavior differs completely from the CHSH scenario, where the optimal expectation value \( S_{\text{opt}} \) obtains its maximal value for the maximally entangled state, that is for \( C = 1 \).

C. Optimal concurrence and the golden ratio

Since the probability \( w \) vanishes for both extreme cases \( C = 0 \) and \( C = 1 \) but is non vanishing for \( 0 < C < 1 \), an optimal concurrence \( C_{\text{opt}} \) must exist, which maximizes the optimal probability \( w_{\text{opt}}(C) \), Equation (50).

By differentiating \( w_{\text{opt}}(C) \), Equation (50), and determining the root of the resulting expression, the optimal concurrence reads

\[
    C_{\text{opt}} = 3 - \sqrt{5}.
\]

When we insert the optimal concurrence \( C_{\text{opt}} \) into the optimal length of \( \tau_{\text{opt}} \), Equation (49), and

\[
    \xi_{\text{opt}} = \frac{C^2}{2 - C^2},
\]

we find that their ratio

\[
    \frac{\tau_{\text{opt}}}{\xi_{\text{opt}}} = \frac{1 + \sqrt{5}}{2} = \Phi,
\]

is the golden ratio \( \Phi \). Due to the relation

\[
    \frac{\tau_{\text{opt}}}{\xi_{\text{opt}}} = \frac{2 - \tau_{\text{opt}}}{\tau_{\text{opt}}},
\]

which is valid for all concurrences \( C \), the ratio between the lengths \( \tau \) and \( 2 - \tau \) is also determined by the golden ratio.

From the definition of the golden ratio, the relation

\[
    \frac{\tau_{\text{opt}}}{\xi_{\text{opt}}} = \frac{\tau_{\text{opt}} + \xi_{\text{opt}}}{\tau_{\text{opt}}},
\]

follows directly. As a consequence, by comparing Equation (56) with Equation (57), we find that for the optimal concurrence, we have

\[
    \tau_{\text{opt}} + \xi_{\text{opt}} = 2 - \tau_{\text{opt}},
\]

that is the distance between the points \( \bar{S} \) and \( \bar{T} \), which are defined by the end points of the vectors \(-s_K\) and \(-t_K\) on the ellipse, is equal to the distance between the point \( \bar{T} \) and the lower focal point of the ellipse \( F_1 \).

V. OPTIMAL MEASUREMENT VECTORS

In the previous section, we have calculated the optimal violation probability for any given concurrence of the underlying state. Now, we determine the associated optimal measurement vectors. Furthermore, we compare these optimal measurements to a simpler measurement strategy and discuss the effect of using non optimal measurement vectors on the violation probability.

A. Optimal angles and resulting vectors

In order to determine the optimal vectors, we first parameterize the vector

\[
    q = \begin{pmatrix} \sin \theta \\ \cos \theta \end{pmatrix}
\]

by the polar angle \( \theta \) between the vector \( q \) and the \( z \) axis.

Inserting this representation of the vector \( q \) back into the definition of \( \tau \), Equation (39), and setting the result equal to the optimal length \( \tau_{\text{opt}} \), Equation (49), we determine the optimal angle

\[
    \theta_{\text{opt}} = \arccos \left( -\sqrt{\frac{1 - C}{1 + C}} \right)
\]

as a function of the concurrence \( C \). This function is depicted in Figure 5.

The optimal angle is therefore always in the interval \( \pi/2 \) to \( \pi \) and hence the vector \( q \) lives in the lower half of the Bloch sphere. For \( C \rightarrow 0 \) the optimal angle is close to \( \pi \), that is the vector \( q \) points mainly along the negative \( z \) axis. With increasing concurrence \( C \) the optimal angle
For small concurrences $C \rightarrow 0$ the optimal measurement vector $q$ points almost along the negative $z$ axis of the coordinate system, while for large concurrences, $C \rightarrow 1$, $q$ points along the $x'$ axis.

Indeed, when we insert Equation (60) into Equation (59), we obtain

$$q_{\text{opt}} = \frac{1}{\sqrt{1+C}} \left( \frac{\sqrt{2C}}{-\sqrt{1-C}} \right),$$

which when we also consider the azimuthal angle $\phi$ leads to the three dimensional measurement vector

$$q_{\text{opt}} = \frac{1}{\sqrt{1+C}} \left( \frac{\sqrt{2C} \cos \phi}{\sqrt{2C} \sin \phi} \right).$$

When we choose $\phi = 0$ that is the vector $q$ in the $x$-$z$ plane, inserting $q_{\text{opt}}$, Equation (62), into the definition of $t$, Equation (24), leads to

$$q_{\text{opt}} = t_{\text{opt}},$$

and the two optimal measurement vectors point along the same direction in their respective coordinate system.

By inserting this result into the expression for $s$, Equation (23), and comparing it with the expression for $r$, Equation (26), we furthermore find

$$r_{\text{opt}} = s_{\text{opt}} = \frac{1}{\sqrt{1-3C^2+C^4}} \left( \frac{-C \sqrt{2C}}{-(2+C) \sqrt{1-C}} \right).$$

As a consequence, the optimal measurement setting is to choose the measurements symmetrically on both subsystems $A$ and $B$ of the bipartite state.

We note that these measurement settings are assumptions explicitly made by Hardy in his original article [12]. Hence allowing $q$ and $t$ to be non equal measurement directions does not improve the probability $w$ to violate the nonlocality assumption.

In the previous sections we have determined the optimal measurements. We now ask the questions of how important the exact choice of these measurement settings is in obtaining a significant violation probability. For this purpose, we investigate the case of fixing the measurement vector $q$ to point along the $x'$ axis of our coordinate system. This situation has the advantage that we do not have to adjust this measurement operator to the underlying state, but only the other three measurements $r$, $s$ and $t$, where the latter two have the same $z$ component and differ in the $x'$ component only in the sign.

When $q$ points along the $x'$ axis, it follows from the definition of the length $\tau$, Equation (38), that $\tau = C^2$, which by inserting into the expression for the violation probability, Equation (46), leads to

$$w = \frac{C^2 - C^4}{4 - 2C^2}.$$  

In Figure 6 we depict this function and compare it to the optimal measurement strategy for the Hardy scenario. For small concurrences, that is in the limit $C \ll 1$, a Taylor expansion of Equation (65) gives

$$w \approx \frac{C^2}{4} + O(C^4),$$

and thus the expectation value grows quadratically in the concurrence. Compared to the Taylor expansion for small concurrences for the optimal violation probability $w_{\text{opt}}$, Equation (51), the approximation $w_{\tau}$ only differs up to fourth order in $C$, and thus for small concurrences the differences in the violation probability for the optimal and the case, where $q$ points along the $x'$ axis, are negligible.
This result seems quite surprising, since in the previous section, we have seen that the optimal angle for small concurrences is close to $\theta = \pi$. Thus, for small concurrences the choice of the measurement vector $q$ is not the relevant parameter for increasing the achievable violation probability $w$.

For $C \lesssim 1$, we perform a Taylor expansion of the violation probability, Equation (65), for the small parameter $\epsilon = 1 - C$

$$w \approx \epsilon - \frac{9}{2} \epsilon^2 + O(\epsilon^3), \quad (67)$$

at $\epsilon = 0$. The expectation value thus grows linearly with $\epsilon$, corresponding to a linear decay towards 2 for $C \rightarrow 1$. When we compare this expansion, Equation (67), to the Taylor series at $C = 1$ of the optimal expectation value, Equation (52), we find that they agree in first order and differ only slightly in the second order of $\epsilon$. Thus, for $C \lesssim 1$ the choice of $q$ pointing along the $x'$ axis provides a good approximation for the optimal measurement. This result is no surprise, since the optimal angle $\theta_{opt}$, Equation (60), converges to $\pi/2$, when $C$ goes to 1.

The difference $\Delta w = w_{opt} - w$ between the optimal violation probability and $q$ pointing along $x'$ is largest for $C \approx 0.75$ where the violation probability is decreased by $\Delta w \approx 0.44\%$ from the optimal probability.

VI. ROLE OF THE CONSTRAINTS

In the previous section, we optimized the violation probability $w$. We have demonstrated in Section III that this optimization corresponds to maximizing the expectation value $S$ familiar from the CHSH inequality under three additional constraints. In this section we study their effects. For this purpose, we consider the optimal expectation value $S_{opt}$, under all possible combinations of one or two of these constraints, and compare them to the CHSH case, without any constraints, and the Hardy scenario, discussed in the previous sections.

A. Single constraint

We start by discussing the case of a single constraint applied to the measurement vectors. We first show that all three constraints lead to the same optimal expectation value $S_{opt}$ and are therefore equivalent.

For this purpose, we note that relabeling the subsystems $A$ and $B$, as well as the respective measurements does not change the physical system. Thus, by exchanging the vectors $q$ and $t$ as well as the vectors $r$ and $s$, the constraint between the vectors $q$ and $s$, Equation (16), becomes the constraint between $t$ and $r$, Equation (17), and vice versa. If we only restrict the vector $s$ by Equation (16) while we choose the vectors $q$, $r$ and $t$ freely, this corresponds to the same situation as restricting $r$ by Equation (17), while choosing $q$, $s$ and $t$ independently.

Furthermore, by interchanging the vector $q$ with $-q$, the restriction between $q$ and $t$, Equation (15), is equivalent to the connection between the vectors $q$ and $s$, Equation (16). Thus, the case of $t$ being determined by $q$, while the other measurement vectors $q$, $s$ and $r$ are chosen freely, is equivalent to the case of only $s$ being restricted by $q$, while $q$, $t$ and $r$ are chosen independently. As a result, all three conditions lead to the same optimal expectation value $S_{opt}$.

Unfortunately, the optimization for only a single constraint is rather difficult to perform analytically, and we are therefore content with a numerical optimization. The optimal expectation value is shown by the red curve in Figure 7. In this picture, we compare it to the CHSH case (black curve) and the Hardy scenario (blue curve).

For small concurrences $C$, the optimal expectation value is similar to both the behavior of the CHSH inequality and the Hardy scenario. This result is no surprise, since we have already demonstrated in the previous section that the Hardy scenario and the CHSH inequality lead to almost the same behavior for small concurrences, and a single condition can neither lead to a result better than the CHSH case, nor worse than the Hardy scenario.

For medium concurrences, that is $0.5 \lesssim C \lesssim 0.8$, the single constraint case is still similar to the CHSH case, but starts to deviate clearly from the Hardy constraints. For large concurrences, there exists a maximum of the optimal expectation value $S_{opt} \approx 2.64$ at $C \approx 0.95$, while...
for larger concurrences the optimal expectation value decreases down to $S_{\text{opt}} = 2.5$ for $C = 1$.

Hence, a single constraint is sufficient to obtain the optimal expectation value $S_{\text{opt}}$ for a non maximally entangled state ($C < 1$). The associated entanglement of the underlying state is larger than for the Hardy scenario. In contrast to the Hardy scenario, the optimal expectation value does not decrease towards $S_{\text{opt}} = 2$, that is the classical boundary, for $C \to 1$. Thus, it is always possible to achieve a violation of the CHSH inequality with a single constraint for maximally entangled states.

### B. Two measurement vectors defined by a common measurement vector

We now consider the case of two restrictions. We start by lifting the restriction between the measurement vectors $r$ and $t$, Equation (17), compared to the Hardy scenario. Thus, the vectors $s$ and $t$ are still defined by the choice of $q$, while the vector $r$ is chosen independently.

In Section III we demonstrated that from the two constraints, Equations (15) and (16), on the vectors $s_K$ and $t_K$, it follows that the parallelogram defined by the points $S, T, S$ and $T$ intersects the $z$ axis in the focal points $F_u$ and $F_1$ of the ellipse.

Furthermore, we have seen that the measurement vector $q$ is parallel to the vector $s_K - t_K$. The other measurement vector $r$ on the subsystem $A$ is not constraint anymore, and we choose it parallel to the vector $s_K + t_K$. Then, maximizing the expectation value $S$ corresponds to maximizing the perimeter of a parallelogram enclosed by an ellipse which goes through the focal points of the ellipse.

From numerical optimization we find that the optimal parallelogram that fulfills these conditions is the rectangle whose sides are parallel to the $x'$ and $z$ axis. In this case, it is straightforward to see that the side parallel to the $z$ axis has length $2\sqrt{1-C^2}$, while the side parallel to the $x'$ axis is of length $2C^2$. The rectangle therefore has a perimeter of $4(C^2 + \sqrt{1-C^2})$.

Therefore, the optimal expectation value is given by

$$S_{\text{opt}} = 2(C^2 + \sqrt{1-C^2}), \quad (68)$$

and consists of an increasing and a decreasing term with the concurrence $C$.

In Figure 7 we depict the expectation value $S$ under these constraints (orange curve), alongside the expectation values $S$ for the CHSH case (black curve), that is without any constraints, and the Hardy scenario (blue curve).

The expectation value given by Equation (68) is not monotonic in $C$. Instead for both $C = 0$ and $C = 1$ the maximum $S_{\text{opt}} = 2$ cannot violate the CHSH inequality, while for all concurrences in between the maximum is strictly above 2, and a violation is possible.

The decrease of $S_{\text{opt}}$ for $C \to 1$ can be understood geometrically, which is depicted in Figure 8 and is a consequence of demanding the parallelogram to go through the focal points. Increasing $C$ increases the semi minor axis of the ellipse, and allows the side of the rectangle parallel to the semi minor axis to increase. At the same time, the focal points move in closer towards the center of the ellipse, decreasing the length of the sides parallel to the semi major axis. In the limit of approaching a maximally entangled state $\ket{\Psi}$, the focal points collapse onto the origin of the ellipse (which is distorted to a circle) and the rectangle collapse onto a line along the $x$ axis of our coordinate system.

Finally, we note that for the concurrence $C = \sqrt{3}/2$ the violation of the CHSH inequality is maximal, at which the expectation value, Equation (68), assumes the value $S_{\text{opt}} = 2.5$. For larger concurrences this value rapidly decreases towards the classical bound $S_{\text{opt}} = 2$.

We conclude the situation of having two measurement vectors depending on a single measurement vector by considering the case, where only constraints between the vectors $q, r$ and $t$, Equations (15) and (17) are present, while we lift the constraint between $q$ and $s$, Equation (16). When we interchange the vectors $q$ with $t$ and $r$ with $s$ in the conditions, Equations (15) and (17), they are identical to the conditions discussed above, Equations (15) and (16). Therefore, the two cases are equivalent, and we obtain the same result for the optimal expectation value $S_{\text{opt}}$.

### C. Two measurement vectors determined by two measurement vectors

In contrast to the two previous cases, the case of two constraints on $s$ and $r$, Equations (16) and (17), is different and cannot be mapped to the other cases of two constraints. We again optimize the expectation value $S_{\text{opt}}$ under these constraints numerically, and depict the result by the green curve in Figure 7.

As a result, the two constraints Equations (16) and (17) lead to a reduced optimal expectation value $S_{\text{opt}}$, compared to the case of applying the constraints, Equations (15) and (17) or Equations (15) and (16). For small concurrences the behavior is similar to the other cases, with various constraints. For large concurrences, the optimal expectation value again decreases towards the classical limit $S = 2$ for $C \to 1$. For $C \approx 0.84$ we obtain the maximal expectation value $S_{\text{opt}} \approx 2.43$. Thus, the optimal concurrence is larger than for the Hardy constraints, $C_{\text{opt}} = 3 - \sqrt{5}$, but lower than when we apply the other two constraints simultaneously.

We therefore conclude that two out of the three constraints for the Hardy scenario always lead to the decay of the optimal expectation value $S_{\text{opt}}$ to the classical regime for maximally entangled states. The main differences between the results for these different conditions is found in the regime $0.5 \leq C \lesssim 0.9$, where the achievable expectation value crucially depends on the underlying constraints.
In this article we have demonstrated that the Hardy scenario is equivalent to the CHSH inequality with three additional constraints.

We have studied the influence of each individual constraint by starting from the CHSH inequality and applying the constraints individually as well as in all different combinations. As a result we find that the constraints influence the optimal concurrence for a violation significantly. A single constraint still provides a violation for maximally entangled states, while for any two combinations of constraints maximally entangled states no longer violate the CHSH inequality.

Moreover, we developed a geometrical interpretation of the violation probability $w$ in the Hardy scenario, and applied this picture to optimize $w$ for all entangled pure states. We then determined the associated optimal measurement settings.

When we furthermore optimize the violation probability over all concurrences $C$ the value of the golden ratio appears, in complete agreement with Hardy’s original article. In our approach, the golden ratio appears twice inside our geometrical picture. In both cases between line segments determined by two points on the ellipse and the focal points of this ellipse. However, a fundamental geometric explanation of why the optimum is found for this ratio remains an open question for a future publication.

**Appendix A: Calculation of the vectors $s_K$ and $t_K$ in terms of $q$ and $a$**

In this appendix, we express the vectors $s_K$, $t_K$ and $r$ through the measurement vector $q$ and the Bloch vector $a$ of the state $|Ψ⟩$.

We first calculate the vector

$$s_K = \frac{K^2q - a}{|Kq − a|},$$

with help of the explicit expression for the correlation matrix $K$, Equation (8), and the decomposition of the vector

$$q = q_⊥ + q_∥$$

into a part that is orthogonal ($q_⊥$) and one that is parallel to the Bloch vector $a$, as

$$s_K = \frac{C^2q_⊥ + q_∥ - a}{|Kq − a|},$$

which we rearrange to

$$s_K = \frac{C^2q}{|Kq − a|} + \frac{(1 − C^2)q_∥ − a}{|Kq − a|}.$$  (A4)

The first term on the right hand side therefore points along the direction $q$, while the second term points along...
the Bloch vector $a$, since both $q_\parallel$ and $a$ point along the same direction.

From the definition of the Bloch vector, Equation (2), we find
\[(1 - C^2)q_\parallel - a = -\left(1 - \sqrt{1 - C^2}q_\parallel\right)a,\] 
(A5)
where $q_\parallel$ is the component of $q$ parallel to the Bloch vector.

As a last step we still have to evaluate the absolute value
\[|qK - a| = \sqrt{q^T K^2 q + |a|^2 - 2a^T K q}.\] 
(A6)

By inserting the definitions of the correlation matrix $K$, Equation (14), and the Bloch vector $a$, Equation (2), as well as using the decomposition of the vector $q$ into an orthogonal ($q_\perp$) and parallel ($q_\parallel$) component with respect to the Bloch vector, we obtain
\[|qK - a| = \sqrt{C^2 q_\perp^2 + q_\parallel^2 + 1 - C^2 - 2\sqrt{1 - C^2}q_\parallel}.\] 
(A7)

Since the measurement vector $q$ is a unit vector, it fulfills the relation $q_\perp^2 + q_\parallel^2 = 1$, leading to
\[|qK - a| = \sqrt{1 - 2\sqrt{1 - C^2}q_\parallel + (1 - C^2)q_\parallel^2},\] 
(A8)
which finally simplifies to
\[|qK - a| = 1 - \sqrt{1 - C^2}q_\parallel.\] 
(A9)

When we insert Equation (A9) together with Equation (A7) back into the expression for the vector $s_K$, Equation (A4), we obtain
\[s_K = \xi q - a,\] 
(A10)
where we defined the length
\[\xi \equiv \frac{C^2}{|qK - a|} = \frac{C^2}{1 - \sqrt{1 - C^2}q_\parallel}\] 
(A11)
of the component of $s_K$ along the vector $q$.

In complete analogy to the above decomposition of the vector $s_K$, the vector
\[t_K = \frac{qK + a}{|qK + a|},\] 
(A12)
is decomposed into
\[t_K = -\tau q - a,\] 
(A13)
with the length
\[\tau \equiv \frac{C^2}{|qK + a|} = \frac{C^2}{1 + \sqrt{1 - C^2}q_\parallel}.\] 
(A14)

We finally determine the vector $r$ in terms of the vectors $q$ and $a$. With help of its definition, Equation (20), and the explicit expression for the vector $t_K$, Equation (A13), we rewrite the vector
\[r = \frac{\tau q - 2a}{|t_K - a|},\] 
(A15)
in terms of the measurement vector $q$, the Bloch vector $a$ and the lengths $\tau$ and $|t_K - a|$.

From the focal point property of the ellipse, which constitutes that the distance from one focal point via any point on the ellipse to the other focal point is twice the length of the semimajor axis, and our geometrical picture, Figure 3, it immediately follows that
\[|t_K - a| = 2 - \tau,\] 
(A16)
and thus the vector $r$, Equation (A15), simplifies to
\[r = q - \frac{2}{2 - \tau} (q + a).\] 
(A17)

Appendix B: Parametrizing $S$ in terms of the polar angle $\theta$

The expectation value $S$ for the Hardy scenario depends only on a single measurement vector. In this appendix, we parametrize this vector in spherical coordinates and derive the resulting equation for $S$ in terms of these coordinates.

We start by writing the measurement vector
\[q = \begin{pmatrix} \cos \phi \sin \theta \\ \sin \phi \sin \theta \\ \cos \theta \end{pmatrix}\] 
(B1)
in the conventional spherical coordinates, where $\theta$ is the polar angle, and $\phi$ denotes the azimuthal angle.

From the connection between the vectors $s$ and $q$, Equation (25), we find
\[s = \frac{1}{1 - \sqrt{1 - C^2} \cos \theta} \begin{pmatrix} -C \cos \phi \sin \theta \\ C \sin \phi \sin \theta \\ \cos \theta - \sqrt{1 - C^2} \end{pmatrix}\] 
(B2)
in these coordinates, where we further made use of the definitions of the correlation matrix $K$, Equation (8), and the Bloch vector $a$, Equation (2), in terms of the concurrence $C$.

Analogously, we find the measurement vector
\[t = \frac{1}{1 - \sqrt{1 + C^2} \cos \theta} \begin{pmatrix} -C \cos \phi \sin \theta \\ C \sin \phi \sin \theta \\ \cos \theta + \sqrt{1 - C^2} \end{pmatrix}\] 
(B3)
from the connection between $t$ and $q$, Equation (24).

When we insert Equation (B3) into the relation between the vectors $r$ and $t$, Equation (26), we finally obtain the vector
\[ r = \frac{1}{\sqrt{C^2 \sin^2 \theta + ((2 - C^2) \cos \theta + 2\sqrt{1 + C^2})^2}} \left( \frac{C^2 \cos \phi \sin \theta}{C^2 \sin \phi \sin \theta} \right). \] (B4)

We now have determined all the measurement vectors in spherical coordinates. Inserting these expressions into the definition of the expectation value, Equation (6), and making use of the correlation matrix \( K \), Equation (8), we finally obtain

\[ S = \frac{2}{C^2 \cos^2 \theta + \sin^2 \theta} \left( C^2 + \frac{2\sqrt{1 - C^2} \sin^2 \theta ((2 - C^2)(2 - C^4) \cos \theta + 2\sqrt{1 - C^2})}{\sqrt{C^2 \sin^2 \theta + ((2 - C^2) \cos \theta + 2\sqrt{1 + C^2})^2}} \right), \] (B5)

which is only dependent on the polar angle \( \theta \), but not on the azimuthal angle \( \phi \).

### Appendix C: Derivation of \( w \) as a function of \( \tau \)

In this appendix, we derive the violation probability, which according to Equation (45), is given by

\[ w = \frac{\xi - \lambda}{4} \] (C1)

with the lengths \( \xi \) and \( \lambda \) as a function of the length \( \tau \) and the concurrence \( C \). From our geometrical picture, Figure 4, we can deduce that \( \ell \) is the leg of a right triangle with hypotenuse of length \( \xi \). When can therefore express the length of the leg

\[ \lambda = \xi \cos \gamma \] (C2)

in terms of the length \( \xi \) and the inner angle \( \gamma \) between this leg and the hypotenuse.

Inserting this result into Equation (45), the probability

\[ w = \frac{\xi}{4} \left(1 - \cos \gamma\right) \] (C3)

is completely determined by the length \( \xi \) of the hypotenuse and the inner angle \( \gamma \) of the triangle \( SRF’ \).

In order determine this angle \( \gamma \), we note from Figure 4, that the angle \( \gamma \) is also an inner angle of the triangle \( F_aT F’_l \). Since we know all the side lengths of this triangle,

\[ \cos \gamma = \frac{2C^2}{(2 - \tau)\tau} - 1. \] (C4)

When we insert this result back into our expression for the probability \( w \), Equation (C3), the probability reads

\[ w = \frac{\xi}{2} \left(1 - \frac{C^2}{(2 - \tau)\tau}\right) \] (C5)

as a function of the lengths \( \xi \) and \( \tau \).

Finally, from their definitions, Equations (37) and (38), we relate the two lengths \( \tau \) and \( \xi \) by

\[ \xi + \tau = \frac{2}{C^2} \xi \tau, \] (C6)

which allows us to rewrite the probability \( w \), Equation (C5), as

\[ w = \frac{C^2}{4} \left(1 - \frac{\tau C^2}{(2 - \tau)(2\tau - C^2)}\right), \] (C7)

which now only depends on the length \( \tau \).

We furthermore notice, that in order to obtain a more geometric interpretation of the violation probability \( w \), we rewrite Equation (C5) with help of Equation (C6) as

\[ w = \frac{C^2}{4} \left(1 - \frac{\xi}{2 - \tau}\right). \] (C8)

For a fixed concurrence \( C \), the violation probability \( w \) thus only depends on the ratio between the lengths \( \xi \) of the line segment \( SF’_l \) and \( 2 - \tau \) of the line segment \( TF’_l \). In order to maximize the violation probability for a fixed concurrence, we thus have to minimize the ratio between the lengths \( \xi \) and \( 2 - \tau \).

[1] J. S. Bell, On the Einstein Podolsky Rosen paradox, *Physics* **1**, 195 (1964)
[2] N. Brunner, D. Cavalcanti, S. Pironio, V. Scarani, and S. Wehner, Bell nonlocality, *Rev. Mod. Phys.* **86**, 419 (2014)
[3] S. J. Freedman and J. F. Clauser, Experimental test of local hidden-variable theories, *Phys. Rev. Lett.* **28**, 938 (1972)
[4] E. S. Fry and R. C. Thompson, Experimental test of local hidden-variable theories, *Phys. Rev. Lett.* **37**, 465 (1976)
[5] A. Aspect, P. Grangier, and G. Roger, Experimental realization of Einstein-Podolsky-Rosen-Bohm Gedankenexperiment: A new violation of Bell's inequalities, Phys. Rev. Lett. 49, 91 (1982).

[6] A. Aspect, J. Dalibard, and G. Roger, Experimental test of Bell's inequalities using time-varying analyzers, Phys. Rev. Lett. 49, 1804 (1982).

[7] G. Weihs, T. Jennewein, C. Simon, H. Weinfurter, and A. Zeilinger, Violation of Bell's inequality under strict Einstein locality conditions, Phys. Rev. Lett. 81, 5039 (1998).

[8] B. Hensen, H. Bernien, A. E. Dréau, A. Reiserer, N. Kalb, M. S. Blok, J. Ruitenberg, R. F. L. Vermeulen, R. N. Schouten, C. Abellán, et al., Loophole-free Bell inequality violation using electron spins separated by 1.3 kilometres, Nature 526, 682 (2015).

[9] M. Giustina, M. A. Versteegh, S. Wengerowsky, J. Handsteiner, A. Hochrainer, K. Phelan, F. Steinlechner, J. Kofler, J.-Å. Larsson, C. Abellán, et al., Significant-loophole-free test of Bell's theorem with entangled photons, Phys. Rev. Lett. 115, 250401 (2015).

[10] L. K. Shalm, E. Meyer-Scott, B. G. Christensen, P. Bierhorst, M. A. Wayne, M. J. Stevens, T. Gerrits, S. Glancy, D. R. Hamel, M. S. Allman, et al., Strong loophole-free test of local realism, Phys. Rev. Lett. 115, 250402 (2015).

[11] J. F. Clauser, M. A. Horne, A. Shimony, and R. A. Holt, Proposed experiment to test local hidden-variable theories, Phys. Rev. Lett. 23, 880 (1969).

[12] L. Hardy, Nonlocality for two particles without inequalities for almost all entangled states, Phys. Rev. Lett. 71, 1665 (1993).

[13] P. G. Kwiat and L. Hardy, The mystery of the quantum cakes, Am. J. Phys. 68, 33 (2000).

[14] D. M. Greenberger, M. A. Horne, A. Shimony, and A. Zeilinger, Bell’s theorem without inequalities, Am. J. Phys. 58, 1131 (1990).

[15] N. J. Cerf, S. Massar, and S. Pironio, Greenberger-Horne-Zeilinger paradoxes for many qudits, Phys. Rev. Lett. 89, 080402 (2002).

[16] M. H. János A. Bergou, Introduction to the Theory of Quantum Information Processing (SPRINGER NATURE, 2013).

[17] W. K. Wootters, Entanglement of formation of an arbitrary state of two qubits, Phys. Rev. Lett. 80, 2245 (1998).

[18] P. Rungta, V. Bužek, C. M. Caves, M. Hillery, and G. J. Milburn, Universal state inversion and concurrence in arbitrary dimensions, Phys. Rev. A 64, 042315 (2001).

[19] B. S. Cirel’son, Quantum generalizations of Bell’s inequality, Lett. Math. Phys. 4, 93 (1980).

[20] B. S. Tsirel’son, Quantum analogues of the Bell inequalities. The case of two spatially separated domains, J. Math. Sci. 36, 557 (1987).

[21] N. Gisin, Bell’s inequality holds for all non-product states, Phys. Lett. A 154, 201 (1991).

[22] J. Seiler, T. Strohm, and W. P. Schleich, Geometric interpretation of the Clauser-Horne-Shimony-Holt inequality of nonmaximally entangled states, Phys. Rev. A 104, 032218 (2021).