SCISPACE
formerly Typeset

Open access • Journal Article • DOI:10.1103/PHYSREVA.89.012129

Heisenberg uncertainty for qubit measurements — Source link

Paul Busch, Pekka Lahti, Reinhard F. Werner

Institutions: University of York, University of Turku, Leibniz University of Hanover

Published on: 31 Jan 2014 - Physical Review A (American Physical Society)

Topics: Uncertainty principle

Related papers:

- Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik
- The Uncertainty Principle
- Proof of Heisenberg's Error-Disturbance Relation
- Zur Quantenmechanik einfacher Bewegungstypen
- Universally valid reformulation of the Heisenberg uncertainty principle on noise and disturbance in measurement

Share this paper:

View more about this paper here: https://typeset.io/papers/heisenberg-uncertainty-for-qubit-measurements-2a5frbaflx

PHYSICAL REVIEW A **89**, 012129 (2014)

Heisenberg uncertainty for qubit measurements

Paul Busch,[1,*] Pekka Lahti,[2,†] and Reinhard F. Werner[3,‡]
[1]*Department of Mathematics, University of York, York, United Kingdom*
[2]*Turku Centre for Quantum Physics, Department of Physics and Astronomy, University of Turku, FI-20014 Turku, Finland*
[3]*Institut für Theoretische Physik, Leibniz Universität, Hannover, Germany*

(Received 6 December 2013; published 31 January 2014)

Reports on experiments recently performed in Vienna [Erhard *et al*., Nature Phys. **8**, 185 (2012)] and Toronto [Rozema *et al*., Phys. Rev. Lett. **109**, 100404 (2012)] include claims of a violation of Heisenberg's error-disturbance relation. In contrast, we have presented and proven a Heisenberg-type relation for joint measurements of position and momentum [Phys. Rev. Lett. **111**, 160405 (2013)]. To resolve the apparent conflict, we formulate here a general trade-off relation for errors in qubit measurements, using the same concepts as we did in the position-momentum case. We show that the combined errors in an approximate joint measurement of a pair of ±1-valued observables A,B are tightly bounded from below by a quantity that measures the degree of incompatibility of A and B. The claim of a violation of Heisenberg is shown to fail because it is based on unsuitable measures of error and disturbance. Finally we show how the experiments mentioned may directly be used to test our error inequality.

DOI: 10.1103/PhysRevA.89.012129 PACS number(s): 03.65.Ta, 03.67.−a

I. INTRODUCTION

Heisenberg's error-disturbance relation [1] for measurements of incompatible quantities has recently become a popular subject of attack and proposed "correction" [2–9]. Thus it appears as if one of the fundamental tenets of quantum mechanics is called into question; if the challenge proved tenable, it would have far-reaching ramifications for the status of the Heisenberg limit in precision measurements studied in the booming field of quantum-enhanced metrology [10,11]. In contrast, in Ref. [12] we presented a Heisenberg-type error-disturbance relation for position and momentum. This result appears to contradict claims of an experimental violation of Heisenberg's relation made in Refs. [2–5]. A direct comparison is made difficult by the fact that the experiments were performed on qubits rather than continuous variable systems. Therefore, we will describe here the qubit variant of Ref. [12].

The apparent conflict is then resolved by analyzing the meaning of the quantity, $\varepsilon_{\mathrm{NO}}$, proposed by Ozawa (e.g., Ref. [13]), and adopted by the authors of Refs. [2–5] and others. This quantity is defined suggestively as the square root of the expectation of a squared *noise operator*. However, we will see that it does *not* meet its intended purpose of representing state-specific experimental errors but something else. Therefore $\varepsilon_{\mathrm{NO}}$ provides *no basis for claims of a theoretical or experimental violation of Heisenberg-type error-disturbance relations*. Actually the experiments confirm Ozawa's inequality and demonstrate a violation of the (incorrect) inequality $\varepsilon_{\mathrm{NO}}(A,\rho)\varepsilon_{\mathrm{NO}}(B,\rho) \geqslant |\langle [A,B]\rangle_\rho|$, which is attributed wrongly to Heisenberg (who never gave a quantum mechanical definition of measurement errors or proposed a precise inequality of this generality).

In contrast, our approach represents measurement error as an overall figure of merit of the measuring device, giving a worst-case estimate of the inaccuracy applicable to all possible input states. Our error measure Δ, introduced in Ref. [12], is an *operationally significant* quantum version of the classic root-mean-square error, obtained by an adaptation of the so-called Wasserstein distance (of order 2) between probability distributions [14]. It can be applied seamlessly to the qubit case, yielding our main result, a Heisenberg-type error uncertainty relation (Sec. II): *any joint measurement of two-outcome observables C,D has combined approximation errors that are constrained by a measure of the degree of incompatibility of the target observables A,B to be approximated.* Symbolically:

$$\Delta(C,A)^2 + \Delta(D,B)^2 \;\geqslant\; (\text{incompatibility of } A,B).$$

The additive form of this trade-off relations offers itself given that an error product cannot have a nonzero bound. This raises the question of whether the traditional uncertainty relation for the spreads of two observables in a quantum state can be supplemented with an additive version. We answer this in the positive (Sec. III), with an inequality for the sum of the variances of A,B in state ρ,

$$\Delta(A,\rho)^2 + \Delta(B,\rho)^2 \;\geqslant\; (\text{noncommutativity of } A,B),$$

where the bound is state independent and is nontrivial also for eigenstates of A or B.

The proofs of these inequalities are based on simple geometric considerations, which makes it possible to teach them in a basic quantum mechanics course.

Ironically, Ozawa's measure $\varepsilon_{\mathrm{NO}}$ is actually state independent in the class of qubit measurements under consideration here and thus overestimates badly the state-dependent errors (Sec. V). In fact, *rather than helping to prove Heisenberg wrong, the quantity $\varepsilon_{\mathrm{NO}}$ itself satisfies a Heisenberg-type trade-off inequality*, with the same bound as for our quadratic error inequality:

$$\varepsilon_{\mathrm{NO}}(A,\rho) + \varepsilon_{\mathrm{NO}}(B,\rho) \;\geqslant\; \tfrac{1}{2}(\text{incompatibility of } A,B).$$

(For a more general, detailed critique of the noise-operator based approach of attempting to quantify measurement errors,

---

[*] paul.busch@york.ac.uk
[†] pekka.lahti@utu.fi
[‡] reinhard.werner@itp.uni-hannover.de

©2014 American Physical Society

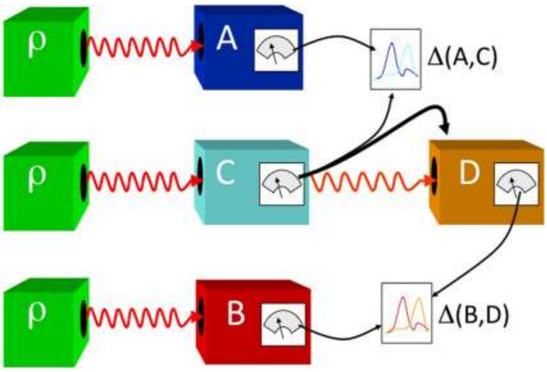

FIG. 1. (Color online) Sequence of compatible measurements C and D. The statistics are compared with control measurements A and B, respectively, defining the approximation errors.

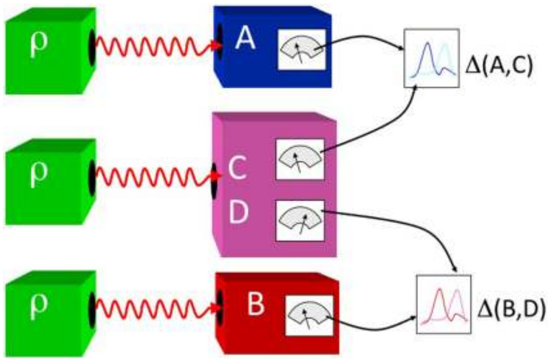

FIG. 2. (Color online) Scheme of a joint measurement of compatible observables C,D, each being used as an approximation of A,B, respectively.

we refer the reader to our forthcoming investigation [15]; there we also amplify on the fact that it is historically incorrect to associate Heisenberg with the above wrong inequality.)

We finally (Sec. VI) proceed to demonstrate the possibility of using the setups of the Vienna and Toronto experiments to test our qubit measurement error relation. The Toronto experiment allows the realization of the tight error bound. Importantly, the experiments and their analyses reported so far are in fact *incomplete*: they investigate and confirm a mathematical relation, Ozawa's inequality, between two quantum mechanical expectation values, $\varepsilon_{\mathrm{NO}}(A,\rho)$ and $\varepsilon_{\mathrm{NO}}(B,\rho)$, and this result is accompanied with the statement that hence the incorrect error-disturbance relationship attributed to Heisenberg has been tested and violated. There is no independent evaluation of the claim that these quantities do represent approximation errors; this assertion is adopted on faith from Ozawa. What is required for a test of measurement error trade-off relations is an error analysis in which the actual measurement statistics are compared to those of (more) precise reference measurements.

II. A TRADE-OFF RELATION FOR QUANTUM RMS ERRORS

We will consider a pair of sharp qubit observables A,B. Our aim is to characterize positive operator-valued measurements (observables) C, D which are *compatible*, that is, they can be performed simultaneously, and which will be considered as approximations to A and B, respectively. (We recall that two observables are compatible or jointly measurable if there is another, joint, observable of which they are marginals.)

The problem of measurement disturbance and simultaneous approximation is illustrated in Figs. 1 and 2. A measurement C as an approximation to A makes itself felt by changing the state of the system, so that a subsequent measurement of an observable D will be an approximation of B only with limited accuracy if A and B are not compatible (Fig. 1). Such a scheme is a special case of a device in which the boxes C and D are merged, giving a truly "joint" measurement (Fig. 2). Thus, a measure of disturbance is conceptually an instance of an approximation error.

We use Bloch sphere notation to write the spectral projections of A as $A_\pm = E^{\mathbf{a}}_\pm = \frac{1}{2}(\mathbf{1} \pm \mathbf{a}\cdot\boldsymbol{\sigma})$, and similarly for B, where $\mathbf{a},\mathbf{b}$ are unit vectors. To be specific, we scale these measurements such that their outcomes are ± 1, so that, for example, the observable C is given as a map $\pm 1 \mapsto C_\pm$, with the positive operators $C_+ = \frac{1}{2}(c_0\mathbf{1} + \mathbf{c}\cdot\boldsymbol{\sigma})$, $C_- = \mathbf{1} - C_+$. (Positivity is equivalent to $\|\mathbf{c}\| \leqslant \min\{c_0, 2-c_0\} \leqslant 1$.)

The first task is to specify an error measure to quantify the quality of such approximations. We follow the choice made in Ref. [12]. For a pair of observables $E:\pm 1 \mapsto E_\pm$ with $E_+ = \frac{1}{2}(e_0\mathbf{1} + \mathbf{e}\cdot\boldsymbol{\sigma})$ and $F:\pm \mapsto F_\pm$, $F_+ = \frac{1}{2}(f_0\mathbf{1} + \mathbf{f}\cdot\boldsymbol{\sigma})$, a state-dependent distance that scales with the values and quantifies the difference between the probability distributions E_ρ and F_ρ (where ρ denotes the state) is given by the Wasserstein distance:

$$\Delta(E_\rho,F_\rho) = \left[\inf_\gamma \iint (x-x')^2\, d\gamma(x,x')\right]^{\frac{1}{2}},$$

where the infimum is taken over all *couplings* γ of E_ρ, F_ρ (i.e., all joint distributions with marginals E_ρ, F_ρ). By maximizing the state-dependent error over all states ρ one has the worst-case error estimate, $\Delta\big(E,F\big) = \sup_\rho \Delta\big(E_\rho,F_\rho\big)$.

In the present case of ± 1-valued qubit observables it is straightforward to write all possible couplings and determine this infimum (Appendix A). Writing the states as $\rho = \frac{1}{2}(\mathbf{1} + \mathbf{r}\cdot\boldsymbol{\sigma})$, one obtains

$$\Delta(E_\rho,F_\rho)^2 = 2|e_0 - f_0 + \mathbf{r}\cdot(\mathbf{e}-\mathbf{f})|.$$

By maximizing this over all states ρ one has the worst-case error estimate

$$\Delta(E,F)^2 = 2|e_0 - f_0| + 2\|\mathbf{e}-\mathbf{f}\|.$$

We pause to emphasize that one could use the state-dependent error measure for the study of precision measurements in which one is interested in an error-disturbance trade-off in a specific state. However, for an assessment of the quality of a joint measurement device, it is also important to note that one can always arrange for situations where the state-dependent errors are both zero: for example, one can take C and D both sharp and identical, with $\mathbf{c} = \mathbf{d}$ in the plane spanned by $\mathbf{a},\mathbf{b}$. Then for a state ρ with $\mathbf{r}$ perpendicular to that plane one has $\Delta(A_\rho,C_\rho) = \Delta(B_\rho,D_\rho) = 0$.

We could also have chosen different distance measures for the comparison of two observables. For comparison we note the distance induced by the total variation norm (also known as 1-norm), which was used in Ref. [16] for a similar purpose and turns out to be given as $D\big(E,F\big)=\frac{1}{4}\Delta\big(E,F\big)^2$ in this special qubit situation.

The Heisenberg-type joint measurement error trade-off relation that we present here gives a tight lower bound for the sum of the squared approximation errors: for any pair of observables C and D that are jointly measurable, their errors of approximation relative to A and B are tightly bounded as follows:

$$\begin{aligned}\Delta(C,A)^2+\Delta(D,B)^2 &\geqslant \sqrt{2}[\|\mathbf{a}-\mathbf{b}\|+\|\mathbf{a}+\mathbf{b}\|-2]\\ &=\frac{1}{\sqrt{2}}[\Delta(A,B)^2+\Delta(A,B^{(-)})^2-4].\end{aligned} \tag{1}$$

Here $B^{(-)}$ is the observable obtained from B by swapping the outcomes ±1. We will see that the lower bound represents the incompatibility of A and B.

The proof procedure (which we sketch in Appendix B) follows the same steps as that of the position-momentum case [12]. First, one reduces the inequality to the special case where the estimating observables are *covariant* under value translations (swaps). For such estimators we can then directly give a simple geometric proof of the tight bound. We find this bound by minimizing the left-hand side of (1) under the constraint of compatibility of the covariant approximators $\overline{C}_\pm=\frac{1}{2}(\mathbf{1}+\mathbf{c}\cdot\boldsymbol{\sigma})$, $\overline{D}_\pm=\frac{1}{2}(\mathbf{1}+\mathbf{d}\cdot\boldsymbol{\sigma})$, a criterion of which is given by the following inequality [17] (see also Appendix C):

$$\|\mathbf{c}-\mathbf{d}\|+\|\mathbf{c}+\mathbf{d}\|\leqslant 2. \tag{2}$$

For an incompatible pair of observables A,B, it is thus natural to define their *degree of incompatibility* by the (positive) number $\|\mathbf{a}-\mathbf{b}\|+\|\mathbf{a}+\mathbf{b}\|-2$.

According to (2), compatibility does not require commutativity. It is only when at least one of the observables is sharp (projection valued) that compatibility implies commutativity. The lower bound in (1) reaches its minimal value zero exactly when the sharp observables A,B are compatible. (For a pair of compatible unsharp observables this number can be negative.) The bound reaches its maximum $2(2-\sqrt{2})$ when $\mathbf{a},\mathbf{b}$ are orthogonal unit vectors; then equality in (1) is achieved for the covariant observables $\overline{C},\overline{D}$ with $\mathbf{c}=\mathbf{a}/\sqrt{2}$ and $\mathbf{d}=\mathbf{b}/\sqrt{2}$. In all other cases, the optimal approximations are obtained for vectors $\mathbf{c},\mathbf{d}$ that are not collinear with $\mathbf{a},\mathbf{b}$.

We will show in Sec. IV that the covariant estimators $\overline{C},\overline{D}$ are smearings of their sharp counterparts, and analysis of the smearing operation shows that the error trade-off can be reduced to a form of *preparation* uncertainty relation. This observation, which was also made in the position-momentum case, corroborates an intuition held by the pioneers of quantum mechanics: *the possibilities of measurement cannot exceed the possibilities of preparation.*

We thus see how inequality (1) limits the combined approximation accuracies, and the tight bound is given by the incompatibility degree of the target observables A,B being approximated. The minimum is taken under the constraint of joint measurability of the approximating observables. The incompatibility degree is determined by the average of the squared distances between A,B and $A,B^{(-)}$, respectively, reflecting the fact that compatibility is independent of the choice of scaling of the outcomes.

We can also give a somewhat different error trade-off relation that is closer in form and spirit to the position-momentum inequality of Ref. [12]. Note that one has

$$\begin{aligned}\Delta(C,A)^2&=2[|c_0-1|+\|\mathbf{c}-\mathbf{a}\|]\\ &\geqslant 2(1-\|\mathbf{c}\|)\geqslant 1-\|\mathbf{c}\|^2\equiv U(\overline{C})^2,\end{aligned}$$

and similarly $\Delta(D,B)^2\geqslant U(\overline{D})^2$, where $U(\overline{C})$ is a measure of the degree of *unsharpness* of the covariant observable $\overline{C}$, that is, its deviation from being projection valued. A simple calculation [16] shows that the compatibility condition (2) is equivalent to

$$U(\overline{C})^2U(\overline{D})^2\geqslant\|\mathbf{c}\times\mathbf{d}\|^2=4\|[\overline{C}_+,\overline{D}_+]\|^2. \tag{3}$$

This inequality says that two noncommuting (covariant) observables are compatible if and only if they are sufficiently unsharp. Sharpness of one of them forces commutativity. Thus we also obtain a bound for the error product:

$$\Delta(\overline{C},A)^2\,\Delta(\overline{D},B)^2\geqslant 4\|[\overline{C}_+,\overline{D}_+]\|^2. \tag{4}$$

Here we see how the noncommutativity of the *compatible* approximators limits the accuracies. However, one may choose to approximate A,B using commuting observables C,D. In this case the approximation will not be optimal but the bound for the error product vanishes. This highlights the relative strength of the bound (1) for the sum of squared errors.

III. PRODUCTS OR SUMS OF UNCERTAINTIES?

It has become accepted wisdom that uncertainty relations have the form of a lower bound for an uncertainty *product*. In contrast, (1) gives a lower bound to a *sum* of uncertainties. From the discussion above it is evident that there is no nontrivial lower bound for the product of errors. To help appreciate this less conventional perspective, we note here an *additive* version of a preparation uncertainty relation. The standard deviation of a sharp ±1-valued qubit observable A in a state ρ is $\Delta(A,\rho)=[1-(\mathbf{r}\cdot\mathbf{a})^2]^{1/2}$. Then

$$\Delta(A,\rho)+\Delta(B,\rho)\ \geqslant\ \|\mathbf{a}\times\mathbf{b}\|=2\|[A_+,B_+]\|. \tag{5}$$

The left-hand side is equal to $\|\mathbf{r}\times\mathbf{a}\|+\|\mathbf{r}\times\mathbf{b}\|$, so one can see that the bound is attained for $\mathbf{r}=\pm\mathbf{a}$ or $\mathbf{r}=\pm\mathbf{b}$.

We can also minimize the sum of the variances:

$$\begin{aligned}\Delta(A,\rho)^2+\Delta(B,\rho)^2&\geqslant 1-|\mathbf{a}\cdot\mathbf{b}|=1-\sqrt{1-\|\mathbf{a}\times\mathbf{b}\|^2}\\ &=1-\sqrt{1-4\|[A_+,B_+]\|^2}.\end{aligned} \tag{6}$$

Again, the bound is tight, but this time it is attained at $\mathbf{r}=(\mathbf{a}\pm\mathbf{b})/\|\mathbf{a}\pm\mathbf{b}\|$ for $\mathbf{a}\cdot\mathbf{b}\geqslant 0$ and $\leqslant 0$, respectively.

Inequalities (5) and (6) are stricter than the state dependent bound for the product of the standard deviations: here we obtain a nontrivial lower bound also when ρ is an eigenstate of A or B. The lower bounds in both (5) and (6) vanish exactly when A and B commute.

It is interesting to compare this situation with the case of position Q and momentum P. Let x_0 be an arbitrary positive constant of the dimension of length. It is an easy exercise to show that for any value of x_0 the inequality

$$\frac{4\hbar^2}{x_0^2}\Delta(Q,\rho)^2 + x_0^2\Delta(P,\rho)^2 \geqslant 2\hbar^2 \tag{7}$$

is a consequence of

$$\Delta(Q,\rho)\Delta(P,\rho) \geqslant \hbar/2. \tag{8}$$

Conversely, using the reciprocal behavior of position and momentum under scale transformations, it can be shown that if inequality (7) is assumed to hold for only one value of x_0 and all states ρ, then it holds for all values of x_0 and entails the standard uncertainty relation (8). Inequality (7) can also be proven directly [i.e., without making use of (8)] by observing that finding the minimum of the left-hand side is equivalent to finding the minimum energy eigenstate of the harmonic oscillator Hamiltonian.

IV. ERROR BOUNDED BY UNCERTAINTY

There is a general connection between the limitations of preparations and the limitations of measurement: the possibilities of measurement should not exceed the possibilities of preparation, and hence a limitation of the latter should entail a limitation of the former. We can see this principle at work in the present case of qubit measurements, in much the same way as it played a role in the case of position and momentum [12].

We consider the case $\mathbf{a} \perp \mathbf{b}$. If the approximator C is a smearing of A, so that $C_+ = \frac{1}{2}(\mathbf{1} + \lambda\mathbf{a}\cdot\boldsymbol{\sigma}) = \mu_+ A_+ + \mu_- A_-$ for a probability distribution μ with $\mu_+ + \mu_- = 1$, then we find $\Delta(A,C)^2 = 2(1-\lambda) = 4\mu_- \geqslant \Delta(\mu)^2$, since $\lambda = \mu_+ - \mu_-$. Similarly we get $D_+ = \frac{1}{2}(\mathbf{1} + \lambda\mathbf{b}\cdot\boldsymbol{\sigma}) = \nu_+ B_+ + \nu_- B_-$, and $\Delta(B,D)^2 = 2(1-\lambda) = 4\nu_- \geqslant \Delta(\nu)^2$. Now we observe that we can identify the distributions μ and ν with distributions of A and B with one and the same quantum state $\rho_\mathbf{s}$, with $\mathbf{s} = \lambda(\mathbf{a}+\mathbf{b})$, $\lambda = \mathbf{s}\cdot\mathbf{a} = \mathbf{s}\cdot\mathbf{b}$:

$$\mu_- = \tfrac{1}{2}(1 - \mathbf{s}\cdot\mathbf{a}) = \tfrac{1}{2}(1-\mathbf{s}\cdot\mathbf{b}) = \nu_-,$$

so that $\Delta(\mu)^2 = \Delta(A,\rho_\mathbf{s})^2 \leqslant \Delta(A,C)^2$ and $\Delta(\nu)^2 = \Delta(B,\rho_\mathbf{s})^2 \leqslant \Delta(B,D)^2$. Taking $\lambda = 1/\sqrt{2}$, the largest value allowed by the compatibility of C,D, and using (6) we get

$$\begin{aligned}\Delta(A,C)^2 + \Delta(B,D)^2 &= 4\mu_- + 4\nu_- = 2(2-\sqrt{2})\\ &\geqslant \Delta(A,\rho_\mathbf{s})^2 + \Delta(B,\rho_\mathbf{s})^2 \geqslant 1.\end{aligned}$$

Thus, if one did not know already that $2(2-\sqrt{2})$ is the optimal bound for the combined squared errors, the uncertainty relation for the state $\rho_\mathbf{s}$ would guarantee a bound. Moreover, the tight bound, given above in the form $4(\mu_- + \nu_-)$, is itself a characteristic of the state operator $\rho_\mathbf{s}$.

The role of the operator $\rho_\mathbf{s}$ becomes more transparent by constructing a joint observable for the approximators C,D. A general expression is given in Appendix C; it is easy to see that if C,D (with $\mathbf{c} \perp \mathbf{d}$ and $c = \|\mathbf{c}\| = \|\mathbf{d}\| = d$) are compatible (that is, $c = d \leqslant 1/\sqrt{2}$), then the following is a joint observable:

$$\begin{aligned}G_{+\pm} &= \tfrac{1}{4}[\mathbf{1} + (\mathbf{c}\pm\mathbf{d})\cdot\boldsymbol{\sigma}],\\ G_{-\pm} &= \tfrac{1}{4}[\mathbf{1} - (\mathbf{c}\mp\mathbf{d})\cdot\boldsymbol{\sigma}].\end{aligned}$$

We specify Cartesian coordinates with orthogonal unit vectors $\mathbf{e}_1,\mathbf{e}_2,\mathbf{e}_3$, such that $\mathbf{c} = c\,\mathbf{e}_1$, $\mathbf{d} = d\,\mathbf{e}_3$. Then $(k,\ell) \mapsto G_{k\ell}$ is covariant under the unitary group acting on operators, with elements

$$\begin{aligned}\mathcal{U}_{++} &= \mathbf{1}(\cdot)\mathbf{1}, & \mathcal{U}_{--} &= \sigma_2(\cdot)\sigma_2,\\ \mathcal{U}_{+-} &= \sigma_2(\cdot)\sigma_2, & \mathcal{U}_{-+} &= \sigma_3(\cdot)\sigma_3,\end{aligned}$$

where $\sigma_1,\sigma_2,\sigma_3$ are the Pauli operators associated with coordinate axes x,y,z. This group can be cast as a representation of a discrete Heisenberg-Weyl group, and it is straightforward to verify that the joint observable can be given in the form

$$G_{k\ell} = \tfrac{1}{2}\mathcal{U}_{k\ell}(\rho_\mathbf{s}), \quad \rho_\mathbf{s} = \tfrac{1}{2}[\mathbf{1} + c(\mathbf{e}_1 + \mathbf{e}_3)\cdot\boldsymbol{\sigma}].$$

This explains why the approximation errors in such a covariant measurement are determined by the uncertainties inherent in the state operator $\rho_\mathbf{s}$. Further discussion of error trade-off relations for discrete Heisenberg-Weyl covariant observables and their mutually unbiased marginals can be found in Ref. [18].

V. INTERPRETATION OF THE NOISE-OPERATOR-BASED MEASURES

The Vienna and Toronto experiments make use of covariant observables as approximators, and it turns out that the "disturbed" observables are covariant as well. The disturbance measure η_{NO} used there is in fact a variant of $\varepsilon_{\mathrm{NO}}$, so that we can use unified notation. For a covariant approximator $\overline{C}$ of A one obtains (we are using the notation $A[x^n] = \int x^n\, dA(x)$ for the nth moment operator of an observable A):

$$\begin{aligned}\varepsilon_{\mathrm{NO}}(A,\rho)^2 &= \mathrm{tr}[\rho(\overline{C}[x^2] - \overline{C}[x]^2)] + \mathrm{tr}[\rho(\overline{C}[x] - A[x])^2]\\ &= 1 - \|\mathbf{c}\|^2 \;+\; \|\mathbf{c}-\mathbf{a}\|^2 = U(\overline{C})^2 + \tfrac{1}{4}\Delta(\overline{C},A)^4.\end{aligned}$$

Here we see that $\varepsilon_{\mathrm{NO}}$ is a mix of an error contribution and the intrinsic unsharpness of the estimator observable, which is already accounted for in the Δ term; it is not hard to see that $\varepsilon_{\mathrm{NO}}(A,\rho) \leqslant \Delta(\overline{C},A)$. For approximators that are smearings of the target observable, for which $\mathbf{c} = \gamma\mathbf{a}$, one has in fact $\varepsilon_{\mathrm{NO}}(A,\rho) = \Delta(\overline{C},A)$. This situation arises in the Toronto experiment (see below).

What is most striking is that in this particular case of covariant qubit observables, $\varepsilon_{\mathrm{NO}}$ has lost what the advocates of this measure consider to be one of its virtues: its state dependence. Thus $\varepsilon_{\mathrm{NO}}$ is a bad overestimate of the state-dependent error; in particular, it cannot capture the peculiar situation arising in both the Vienna and Toronto experiments where the input and output distributions are identical, so that the state-dependent error vanishes.

The inequalities (1) and (3) immediately yield similar trade-off relations for the $\varepsilon_{\mathrm{NO}}$ quantities. In fact, using $\varepsilon_{\mathrm{NO}}(A,\rho) \geqslant \frac{1}{2}\Delta(\overline{C},A)^2$ (and similarly for $\varepsilon_{\mathrm{NO}}(B,\rho)$), then (1) gives

$$\varepsilon_{\mathrm{NO}}(A,\rho) + \varepsilon_{\mathrm{NO}}(B,\rho) \geqslant \frac{1}{\sqrt{2}}[\|\mathbf{a}-\mathbf{b}\| + \|\mathbf{a}+\mathbf{b}\| - 2];$$

and using $\varepsilon_{\mathrm{NO}}(A,\rho) \geqslant U(\overline{C})$, then (3) entails

$$\varepsilon_{\mathrm{NO}}(A,\rho)^2\varepsilon_{\mathrm{NO}}(B,\rho)^2 \geqslant \|\mathbf{c}\times\mathbf{d}\|^2 = 4\|[\overline{C}_+,\overline{D}_+]\|^2.$$

Thus, not surprisingly, the quantities $\varepsilon_{\rm NO}$, which comprise a mix of contributions from error and unsharpness, are seen to be subject to Heisenberg-type trade-off constraints.

As argued in Refs. [19,20] and elaborated further in Ref. [15], $\varepsilon_{\rm NO}$ is a problematic generalization of Gauss's root-mean-square error into the quantum context. This quantity does not, in general, provide an operationally significant estimate of measurement errors in a single state. One can see this already from the general defining expression for $\varepsilon_{\rm NO}$ given above: the operator $\overline{C}[x]$ does not, in general, commute with $A[x]$, so that the difference $\overline{C}[x] - A[x]$ is in fact incompatible with both. Therefore it is not evident that a comparison of the statistics of the observables A and C can be obtained from studying their difference operator. This apparent deficiency has been addressed with the observation that $\varepsilon_{\rm NO}^2$ can be expressed as a combination of expectation values of first or second moments of the approximator observable C in *three different states* instead of just one. Accordingly, in the Vienna experiment the quantity $\varepsilon_{\rm NO}$ is measured using the so-called three-state method. The fact that three distinct states are required makes evident the impossibility of interpreting this quantity as the error relevant to a *single* state. This is illustrated in the present qubit case by the above expression for $\varepsilon_{\rm NO}$, which shows it to be state-independent and in fact related to our maximized error.

In higher dimensional Hilbert spaces it is not hard to construct examples of measurements where $\varepsilon_{\rm NO}$ vanishes although the input and output distributions to be compared are not identical. There are also examples where these distributions do coincide but the quantity $\varepsilon_{\rm NO}$ can be made arbitrarily large. Similar observations apply to the use of this quantity as a measure of disturbance, showing that these quantities are unreliable as indicators of error or disturbance [15].

VI. PROPOSED EXPERIMENTAL TESTS

We consider first an experiment of the kind performed in Vienna, where a projective (or von Neumann-Lüders) measurement of a sharp observable C (with $\|\mathbf{c}\| = 1$) is considered as an approximate measurement of A. Such a measurement causes the state change $\rho \to C_+\rho C_+ + C_-\rho C_-$ or, equivalently, distorts an observable B into D as

$$\begin{aligned} B_\pm \to D_\pm &= C_+ B_\pm C_+ + C_- B_\pm C_- \\ &= B_{\rho_{\mathbf{c}}}(\pm 1) C_+ + B_{\rho_{\mathbf{c}}}(\mp 1) C_- \\ &= \tfrac{1}{2}(\mathbf{1} \pm \mathbf{d}\cdot\boldsymbol{\sigma}), \quad \mathbf{d} = (\mathbf{c}\cdot\mathbf{b})\mathbf{c}. \end{aligned}$$

[Here $\rho_{\mathbf{c}}$ denotes a pure state with unit Bloch vector $\mathbf{c}$, and $B_{\rho_{\mathbf{c}}}(\pm 1) = \mathrm{tr}[\rho_{\mathbf{c}} B_\pm] = \frac{1}{2}(1 \pm \mathbf{c}\cdot\mathbf{b})$.] This scheme defines a joint observable M (necessarily of the product form since C is sharp), with positive operators $M_{k,\ell} = C_k D_\ell, \quad k,\ell = \pm$, which can be considered as an approximate joint measurement of A and B, with the characteristic errors $\Delta(C,A)$ and $\Delta(D,B)$.

The (squared) state-dependent error and disturbance are given by ($\rho = \frac{1}{2}(\mathbf{1} + \mathbf{r}\cdot\boldsymbol{\sigma})$)

$$\Delta(C_\rho, A_\rho)^2 = 2|\mathbf{r}\cdot(\mathbf{c}-\mathbf{a})|, \ \Delta(D_\rho, B_\rho)^2 = 2|\mathbf{r}\cdot(\mathbf{d}-\mathbf{b})|\,.$$

We observe that if $\mathbf{r}\cdot\mathbf{a} = \mathbf{r}\cdot\mathbf{c}$, then $C_\rho = A_\rho$, so that the state-dependent $\Delta(C_\rho, A_\rho) = 0$ in this case. Since $\Delta(D_\rho, B_\rho) \leqslant 2$, the state dependent uncertainty product $\Delta(C_\rho, A_\rho)\Delta(D_\rho, B_\rho) = 0$ for all such states.

The maximized error and disturbance are

$$\begin{aligned} \Delta(C,A)^2 &= 2\|\mathbf{c}-\mathbf{a}\| = 2\sqrt{2}\sqrt{1-\mathbf{c}\cdot\mathbf{a}}\,, \\ \Delta(D,B)^2 &= 2\|\mathbf{d}-\mathbf{b}\| = 2\|\mathbf{b}\times\mathbf{c}\|. \end{aligned}$$

These are nonzero if $\mathbf{c} \neq \mathbf{a}$ and $\mathbf{b} \neq \mathbf{c}$, respectively.

It is straightforward to show that the following uncertainty relation holds for this experiment:

$$\begin{aligned} \Delta(C,A)^2 + \Delta(D,B)^2 &= 2\|\mathbf{c}-\mathbf{a}\| + 2\|\mathbf{b}\times\mathbf{c}\| \\ &\geqslant 2\|\mathbf{a}\times\mathbf{b}\| = 4\|[A_+, B_+]\|. \quad (9) \end{aligned}$$

The minimum is achieved for $\mathbf{c} = \mathbf{a}$.

This kind of sharp measurement as an approximate joint measurement is not an optimal joint approximation: for example, in the case of orthogonal $\mathbf{a},\mathbf{c}$, the lower bound is $2 > 2(2-\sqrt{2})$.

For comparison we give the squared quantities $\varepsilon_{\rm NO}$:

$$\begin{aligned} \varepsilon_{\rm NO}(A,\rho)^2 &= \mathrm{tr}[\rho(C[x]-A[x])^2] = \|\mathbf{c}-\mathbf{a}\|^2 \\ \varepsilon_{\rm NO}(B,\rho)^2 &= \mathrm{tr}[\rho(D[x^2]-D[x]^2)] + \mathrm{tr}[\rho(D[x]-B[x])^2] \\ &= 1-(\mathbf{b}\cdot\mathbf{c})^2 + \|\mathbf{b}-\mathbf{c}(\mathbf{b}\cdot\mathbf{c})\|^2 = 2\|\mathbf{b}\times\mathbf{c}\|^2. \end{aligned}$$

These are state-independent, as expected.

With the choices $\mathbf{a} = (1,0,0), \mathbf{b} = (0,1,0), \mathbf{c} = (\cos\alpha, \sin\alpha, 0), \mathbf{r} = (0,0,1)$ the above scenario is just the experiment studied and realized by the Vienna group [2]. Then, in particular, $\mathbf{r}\cdot\mathbf{a} = \mathbf{r}\cdot\mathbf{b} = \mathbf{r}\cdot\mathbf{c} = 0$, so that both state-dependent errors become zero: $\Delta(C_\rho, A_\rho) = 0$ and $\Delta(D_\rho, B_\rho) = 0$. By contrast, $\varepsilon_{\rm NO}(A,\rho) = \varepsilon_{\rm NO}(\sigma_x,\rho) = 2\sin\frac{\alpha}{2}$, $\varepsilon_{\rm NO}(B,\rho) = \varepsilon_{\rm NO}(\sigma_y,\rho) = \sqrt{2}\cos\alpha$; these are bad overestimates of the state-dependent error and disturbance for most values of α. Curiously, the experimenters do not report a comparison of the values obtained for the quantities $\varepsilon_{\rm NO}$ with an actual estimation of the error in measuring observable C as an approximation of A; this would be of particular interest as the target and estimator observables do not commute; yet in the given state, the two observables are indistinguishable, while $\varepsilon_{\rm NO}$ does not recognize this. This discrepancy should show up in an error analysis.

Instead of using a projective measurement of a "misaligned" sharp observable C as an approximator to A one may construct an explicit measurement scheme $\mathcal{M}$ as an approximate A measurement. Such a strategy was followed in the Toronto experiment [3], which we reconstruct next. We take the parameters as used in that experiment. Thus, we fix $\mathbf{a} = (0,0,1) = \mathbf{k}$ and consider a measurement scheme $\mathcal{M} = (\mathbb{C}^2, \sigma_z, U, |\phi\rangle\langle\phi|)$, where σ_z is the pointer, the coupling U is the CNOT gate (in the canonical basis of $\mathbb{C}^2 \otimes \mathbb{C}^2$), $\phi = \alpha|0\rangle + \beta|1\rangle$, $\alpha,\beta \in \mathbb{R}$, $\alpha^2 + \beta^2 = 1$ (again in the canonical basis). The measured observable C is then an unsharp version of the observable $A = E^{\mathbf{k}}$, the spectral measure of σ_z,

$$C_\pm = \tfrac{1}{2}[\mathbf{1} \pm (2\alpha^2 - 1)\sigma_z].$$

The distortion exerted by $\mathcal{M}$ on the observable $B = E^{\mathbf{i}}$ ($\mathbf{b} = (1,0,0) = \mathbf{i}$) then results in an observable D, where

$$D_\pm = \tfrac{1}{2}(\mathbf{1} \pm 2\alpha\beta\sigma_x).$$

These observables can also be written in terms of Bloch vector parametrization for ϕ using $\mathbf{s} =$

$(\sin\theta\cos\varphi, \sin\theta\sin\varphi, \cos\theta)$:

$$C_\pm = \tfrac{1}{2}(\mathbf{1} \pm \cos\theta\sigma_z) = \tfrac{1}{2}[\mathbf{1} \pm (\mathbf{s}\cdot\mathbf{k})\,\mathbf{k}\cdot\boldsymbol{\sigma}],$$
$$D_\pm = \tfrac{1}{2}(\mathbf{1} \pm \sin\theta\cos\phi\sigma_x) = \tfrac{1}{2}[\mathbf{1} \pm (\mathbf{s}\cdot\mathbf{i})\,\mathbf{i}\cdot\boldsymbol{\sigma}].$$

The sequential joint observable $M_{k,\ell} = \mathfrak{I}(k)^*(B_\ell)$ [here $\mathfrak{I}(k)$ denotes the conditional output channel associated with the outcome k of $\mathcal{M}$, and $\mathfrak{I}(k)^*$ its dual channel] thus realizes an approximate joint measurements of $A = E^{\mathbf{k}}$ and $B = E^{\mathbf{i}}$. This gives the following expressions for the state-dependent and maximized errors:

$$\Delta(C_\rho, A_\rho)^2 = |\mathbf{r}\cdot\mathbf{k}|\,\|\mathbf{s}-\mathbf{k}\|^2 \leqslant \|\mathbf{s}-\mathbf{k}\|^2 = \Delta(C,A)^2,$$
$$\Delta(D_\rho, B_\rho)^2 = |\mathbf{r}\cdot\mathbf{i}|\,\|\mathbf{s}-\mathbf{i}\|^2 \leqslant \|\mathbf{s}-\mathbf{i}\|^2 = \Delta(D,B)^2.$$

If the initial state of the system is $\rho = \rho_{\mathbf{j}}$, with $\mathbf{r} = \mathbf{j}$, then, again, both state-dependent errors vanish.

By contrast, the $\varepsilon_{\rm NO}$ quantities are again state-independent and coincide, in fact, with the Δ errors:

$$\varepsilon_{\rm NO}(\sigma_z,\rho)^2 = \big\langle(1-\mathbf{s}\cdot\mathbf{k})^2\sigma_x^2\big\rangle_\rho + 1 - (\mathbf{s}\cdot\mathbf{k})^2 = \|\mathbf{s}-\mathbf{k}\|^2,$$
$$\varepsilon_{\rm NO}(\sigma_x,\rho)^2 = \big\langle(1-\mathbf{s}\cdot\mathbf{i})^2\sigma_z^2\big\rangle_\rho + 1 - (\mathbf{s}\cdot\mathbf{i})^2 = \|\mathbf{s}-\mathbf{i}\|^2,$$

again badly overestimating the state-dependent errors.

The uncertainty relation for the maximized errors becomes here

$$\Delta(C,A)^2 + \Delta(D,B)^2 = \|\mathbf{s}-\mathbf{k}\|^2 + \|\mathbf{s}-\mathbf{i}\|^2 \geqslant 2(2-\sqrt{2}). \tag{10}$$

This is the optimal lower bound of (1); it is reached with $\varphi = 0$ and $\theta = \pi/4$, hence $\mathbf{s} = (\mathbf{i}+\mathbf{k})/\sqrt{2}$.

In the actual experiment [3] the numbers $\varepsilon_{\rm NO}$ are determined using the weak measurement strategy suggested by Ref. [21], thus confirming rather indirectly the quantum predictions for the expectations of second moments of the relevant difference observables. Again, no error analysis is reported in Ref. [3] to check whether the $\varepsilon_{\rm NO}$ numbers in question reflect the actual measurement errors.

In any case, the data that have been obtained in these experiments or could be obtained in variations of them can easily be used to test the error trade-off inequality (1) since the Δ errors are here found to be directly related to the corresponding $\varepsilon_{\rm NO}$ numbers.

VII. CONCLUSION

With the inequality (1) we have provided a general error trade-off relation for joint measurements of qubit observables in the spirit of Heisenberg's ideas of 1927. The additive form of this inequality can be matched with an additive form of preparation uncertainty relation, with a *state-independent* lower bound that only vanishes when the observables commute. We have also exhibited the true operational meaning of the quantities, $\varepsilon_{\rm NO}$, in the qubit context, which were taken to represent error and disturbance in these experiments. Our analysis shows that Ozawa's inequality does not admit an interpretation as a trade-off between error and disturbance for individual states. Rather than leading to a violation of a Heisenberg bound, the $\varepsilon_{\rm NO}$ quantities were found themselves to obey Heisenberg-type trade-off relations. Finally we have identified possible tests of our error relation that could be performed using the Vienna and Toronto experiments. We emphasize that such tests are not complete by simply measuring the $\varepsilon_{\rm NO}$ or Δ quantities: a genuine test of error-error or error-disturbance trade-off relations must compare these data with an error analysis carried out for the joint measurements of C and D as approximations of A and B, respectively, as indicated in Figs. 1 and 2.

ACKNOWLEDGMENTS

This work is partly supported (P.B., P.L.) by the Academy of Finland, Project No. 138135, and EU COST Action MP1006. R.F.W. acknowledges support from the European network SIQS.

APPENDIX A: CALCULATION OF WASSERSTEIN DISTANCE

We consider a slightly more general problem, that of minimizing the quantity

$$\Delta^\gamma(E_\rho,F_\rho)^2 = \iint (x-y)^2\, d\gamma(x,y)$$

when E has values ±1 and F has values a_+,a_-, where we assume $a_+ > a_-$. Our use of this will be to consider F as an approximation to E. The Wasserstein distance should in fact vanish when the probabilities of E and F coincide for their corresponding values, $\pm1 \leftrightarrow a_\pm$.

A general coupling is given by four positive numbers:

$$(1,a_+) \mapsto \gamma_{++} \equiv \gamma,$$
$$(1,a_-) \mapsto \gamma_{+-} = E_\rho(+1) - \gamma,$$
$$(-1,a_+) \mapsto \gamma_{-+} = F_\rho(a_+) - \gamma,$$
$$(-1,a_-) \mapsto \gamma_{--} = 1 - E_\rho(+1) - F_\rho(a_+) + \gamma.$$

It is then straightforward to obtain

$$\Delta^\gamma(E_\rho,F_\rho)^2 = (1+a_-)^2 - 4\gamma(a_+-a_-) - 4E_\rho(+1)a_- + F_\rho(a_+)[(1+a_+)^2 - (1+a_-)^2].$$

In order to minimize this quantity, γ must be chosen as large as allowed by the positivity constraints (given that $a_+ - a_- > 0$), hence $\gamma = \min\{E_\rho(+1), F_\rho(a_+)\}$. Now it is easy to see that the minimum, $\Delta(E_\rho,F_\rho)$, can only vanish for $E_\rho(+1) = F_\rho(a_+)$ if $a_+ = 1$ and $a_- = -1$. In this case one obtains

$$\Delta(E_\rho,F_\rho)^2 = 4|E_\rho(+1) - F_\rho(+1)| = 2\big|e_0 - f_0 + \mathbf{r}\cdot(\mathbf{e}-\mathbf{f})\big|.$$

APPENDIX B: PROOF SKETCH FOR THE ERROR TRADE-OFF INEQUALITY (1)

This inequality is a direct translation, here for the Δ measure, of an equivalent form proven in Ref. [16] for the D measure, using the proportionality of Δ^2 with D. We sketch the steps of its derivation. One first makes use of the reduction of (1) to the case where $c_0 = 1 = d_0$. If $C = \frac{1}{2}[c_0\mathbf{1} + \mathbf{c}\cdot\boldsymbol{\sigma}]$ and $D_+ = \frac{1}{2}[d_0\mathbf{1} + \mathbf{d}\cdot\boldsymbol{\sigma}]$ are jointly measurable, then so are C', D' with $C'_+ = \frac{1}{2}[(2-c_0)\mathbf{1} + \mathbf{c}\cdot\boldsymbol{\sigma}]$ and $D'_+ = \frac{1}{2}[(2-d_0)\mathbf{1} + \mathbf{d}\cdot\boldsymbol{\sigma}]$. It follows that the convex combinations of these

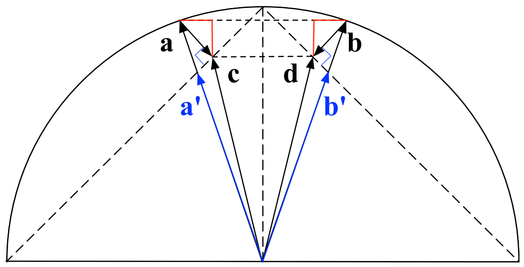

FIG. 3. (Color online) Optimal compatible approximations of sharp observables A,B by covariant unsharp observables C,D. The compatibility of the optimal pair C,D has $\mathbf{c},\mathbf{d}$ located on the dotted lines and the vectors $\mathbf{c}-\mathbf{a}$ and $\mathbf{d}-\mathbf{b}$ orthogonal to these dotted lines. Vectors $\mathbf{a}'$, $\mathbf{b}'$ represent the best compatible approximators among the "smeared" versions of A,B.

observables are also jointly measurable [16], in particular $\overline{C}$ and $\overline{D}$ with $\overline{C}_+ = \frac{1}{2}(C_+ + C'_+) = \frac{1}{2}(\mathbf{1}+\mathbf{c}\cdot\boldsymbol{\sigma})$ and $\overline{D}_+ = \frac{1}{2}(D_+ + D'_+) = \frac{1}{2}(\mathbf{1}+\mathbf{d}\cdot\boldsymbol{\sigma})$. In addition we have that the errors do not increase:

$$\Delta(C,A)^2 \geqslant 2\|\mathbf{c}-\mathbf{a}\| = \Delta(\overline{C},A)^2,$$

$$\Delta(D,B)^2 \geqslant 2\|\mathbf{d}-\mathbf{b}\| = \Delta(\overline{D},B)^2.$$

This process of averaging can be understood as the transition to observables that are covariant under the shift group $\pm1 \mapsto \mp1$ acting on the set $\{-1,+1\}$ [16]. This group acts on $\overline{C}$ and $\overline{D}$ via the unitary operator $U = \mathbf{u}\cdot\boldsymbol{\sigma}$, with $\mathbf{u}$ a unit vector perpendicular to $\mathbf{c}$ and $\mathbf{d}$, so that the covariance $U\overline{C}_\pm U^* = \overline{C}_\mp$ and $U\overline{D}_\pm U^* = \overline{D}_\mp$ holds. We may therefore refer to the observables $\overline{C},\overline{D}$ as *covariant*. The compatibility of these covariant observables is equivalent to inequality (2).

A similar convexity argument shows that if $\mathbf{c},\mathbf{d}$ are not already in the plane spanned by $\mathbf{a},\mathbf{b}$, then their projections into that plane define new observables which are again compatible and no worse approximations to A,B than C,D. Hence we can assume that $\mathbf{c},\mathbf{d}$ are in the plane spanned by $\mathbf{a},\mathbf{b}$.

A simple geometric consideration shows that the minimum of the left-hand side of (1) must be attained for approximators C,D whose vectors $\mathbf{c},\mathbf{d}$ have equal length and are located symmetrically relative to $\mathbf{a},\mathbf{b}$, as shown in Fig. 3. (In fact, using once more the preservation of compatibility under convex mixings of observables, it is straightforward to see that any asymmetric constellation of vectors $\mathbf{c},\mathbf{d}$ can be transformed into a symmetric one for which the errors are not greater.) Analysis of the right-angled triangle with vertices given by the end points of $\mathbf{a}$, $\mathbf{c}$ and the intersection between the vertical line through $\mathbf{c}$ and the horizontal line connecting $\mathbf{a}$ and $\mathbf{b}$ (and similarly on the side of $\mathbf{b}$) immediately gives the relations $\frac{1}{2}[\|\mathbf{a}-\mathbf{b}\| - \|\mathbf{c}-\mathbf{d}\|] = \frac{1}{2}[\|\mathbf{a}+\mathbf{b}\| - \|\mathbf{c}+\mathbf{d}\|] = \|\mathbf{c}-\mathbf{a}\|/\sqrt{2} = \|\mathbf{d}-\mathbf{b}\|/\sqrt{2}$; hence the lower bound in (1) follows via the compatibility constraint $\|\mathbf{c}-\mathbf{d}\| + \|\mathbf{c}+\mathbf{d}\| = 2$.

APPENDIX C: COMPATIBILITY CRITERION AND JOINT OBSERVABLE

Compatible observables C,D with $C_\pm = \frac{1}{2}(\mathbf{1}\pm\mathbf{c}\cdot\boldsymbol{\sigma})$, $D_\pm = \frac{1}{2}(\mathbf{1}\pm\mathbf{d}\cdot\boldsymbol{\sigma})$ arise as marginals of the operator measure $G: k,\ell \mapsto G_{k\ell}$, $k,\ell = \pm1$, where

$$G_{+,\pm} = \tfrac{1}{4}(1\pm\mathbf{c}\cdot\mathbf{d})\mathbf{1} + \tfrac{1}{4}(\mathbf{c}\pm\mathbf{d})\cdot\boldsymbol{\sigma},$$

$$G_{-,\pm} = \tfrac{1}{4}(1\mp\mathbf{c}\cdot\mathbf{d})\mathbf{1} - \tfrac{1}{4}(\mathbf{c}\mp\mathbf{d})\cdot\boldsymbol{\sigma}.$$

Note the marginality relation $C_\pm = G_{\pm,+} + G_{\pm,-}$ and $D_\pm = G_{+,\pm} + G_{-,\pm}$. For G to be an observable, the operators $G_{k\ell}$ must be positive, that is, $1\pm\mathbf{c}\cdot\mathbf{d} \geqslant \|\mathbf{c}\pm\mathbf{d}\|$. This implies immediately, and is in fact equivalent to, (2). [Equivalence follows easily via (3).] The proof of the necessity of (2) for the compatibility of C,D is slightly more involved [17].

[1] W. Heisenberg, Zeitschr. Phys. **43**, 172 (1927).
[2] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, Nature Phys. **8**, 185 (2012).
[3] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, Phys. Rev. Lett. **109**, 100404 (2012).
[4] S.-Y. Baek, F. Kaneda, M. Ozawa, and K. Edamatsu, Sci. Rep. **3**, 2221 (2013).
[5] G. Sulyok, S. Sponar, J. Erhart, G. Badurek, M. Ozawa, and Y. Hasegawa, Phys. Rev. A **88**, 022110 (2013).
[6] C. Branciard, Proc. Natl. Acad. Sci. USA **110**, 6742 (2013).
[7] M. M. Weston, M. J. W. Hall, M. S. Palsson, H. M. Wiseman, and G. J. Pryde, Phys. Rev. Lett. **110**, 220402 (2013).
[8] A. Di Lorenzo, Phys. Rev. Lett. **110**, 010404 (2013).
[9] K. Fujikawa, Phys. Rev. A **85**, 062117 (2012).
[10] R. Demkowicz-Dobrzanski, J. Kolodynski, and M. Guta, Nature Comm. **3**, 1063 (2012).
[11] J. Borregaard and A. S. Sørensen, Phys. Rev. Lett. **111**, 090801 (2013).
[12] P. Busch, P. Lahti, and R. F. Werner, Phys. Rev. Lett. **111**, 160405 (2013).
[13] M. Ozawa, Ann. Phys. **311**, 350 (2004).
[14] C. Villani, *Optimal Transport: Old and New* (Springer, Berlin, 2009).
[15] P. Busch, P. Lahti, and R. Werner, arXiv:1312.4393.
[16] P. Busch and T. Heinosaari, Quantum Inf. Comput. **8**, 0797 (2008).
[17] P. Busch, Phys. Rev. D **33**, 2253 (1986).
[18] P. Busch, P. Lahti, and R. Werner, arXiv:1312.4392.
[19] P. Busch, T. Heinonen, and P. Lahti, Phys. Lett. A **320**, 261 (2004).
[20] R. F. Werner, Quant. Inform. Comput. **4**, 546 (2004).
[21] A. Lund and H. Wiseman, New J. Phys. **12**, 093011 (2010).