

This is a repository copy of *Proof of Heisenberg's error-disturbance relation*.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/75788/>

Version: Submitted Version

Article:

Busch, Paul orcid.org/0000-0002-2559-9721, Lahti, Pekka and Werner, Reinhard F (2013) Proof of Heisenberg's error-disturbance relation. *Physical Review Letters*. 160405. n/a. ISSN: 1079-7114

<https://doi.org/10.1103/PhysRevLett.111.160405>

Reuse

Items deposited in White Rose Research Online are protected by copyright, with all rights reserved unless indicated otherwise. They may be downloaded and/or printed for private study, or other acts as permitted by national copyright laws. The publisher or other rights holders may allow further reproduction and re-use of the full text version. This is indicated by the licence information on the White Rose Research Online record for the item.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

Proof of Heisenberg's error-disturbance relation

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*

(Dated: June 10, 2013)

While the slogan “no measurement without disturbance” has established itself under the name *Heisenberg effect* in the consciousness of the scientifically interested public, a precise statement of this fundamental feature of the quantum world has remained elusive, and serious attempts at rigorous formulations of it as a consequence of quantum theory have led to seemingly conflicting preliminary results. Here we show that despite recent claims to the contrary [Rozema *et al*, *Phys. Rev. Lett.* **109**, 100404 (2012)], Heisenberg-type inequalities can be proven that describe a trade-off between the precision of a position measurement and the necessary resulting disturbance of momentum (and vice versa). More generally, these inequalities are instances of an uncertainty relation for the imprecisions of *any* joint measurement of position and momentum.

PACS numbers: 03.65.Ta, 03.65.Db, 03.67.-a

In spite of their important role since the very beginning of quantum mechanics, uncertainty relations have recently become the subject of active scientific debates. On one hand, entropic versions of the information-disturbance trade-off [1] have become an important tool in security proofs [2] for continuous variable cryptography. On the other hand there were widely publicized [3] claims of a refutation [4–6] of the error-disturbance uncertainty relations heuristically claimed by Heisenberg [7]. A review of the literature on uncertainty relations is given in [8].

Heisenberg's 1927 paper [7] introducing the uncertainty relations is one of the key contributions to early quantum mechanics. It is part of virtually every quantum mechanics course, almost always in the version forwarded by Kennard [9] and Robertson [10]. What is often overlooked, however, is that this popular version is only one way of making the idea of uncertainty precise. The original paper begins with a famous discussion of the resolution of microscopes, in which the accuracy (resolution) of an approximate position measurement is related to the disturbance of the particle's momentum. This situation is no way covered by the standard relations, since in an experiment concerning the Kennard-Robertson inequality no particle meets with both a position and a momentum measurement. Heisenberg's semi-classical discussion has no immediate translation into the modern quantum formalism, particularly since the momentum disturbance *prima facie* involves the comparison of two non-commuting quantities, the momentum before and after the measurement. Such a translation does require some careful conceptual work, and one can get it wrong. This is shown by the example of Ozawa [4], who defines a relation and claims it to be a rigorous version of Heisenberg's ideas, only to show that it fails to hold in general. It is merely this failure which has recently been verified experimentally [5, 6].

In contrast, the purpose of this note is to describe and prove an inequality of the classic form

$$(\Delta Q)(\Delta P) \geq \frac{\hbar}{2}, \quad (1)$$

in which, however, the quantities ΔQ and ΔP are *not* given by the variances of the position and momentum distribution in the same state, as in the textbook inequality. Instead, following closely the suggestion of Heisenberg, they are explicitly defined figures of merit for a microscope-like measurement scenario: the accuracy ΔQ of a position measurement and the momentum disturbance ΔP incurred by it. Moreover, the inequality is sharp, and we will describe explicitly the cases of equality. We believe that the definitions and results are simple enough to use in a basic quantum mechanics course, although the full proof uses some tools beyond such a course.

The main progress over earlier work [11] is a simpler definition of the Δ quantities, using the idea of calibration [12]. This definition does not require the Monge transportation metric, which led in [11] to quantities akin to absolute deviations rather than root mean square deviations, and hence to a constant different from $\hbar/2$ in (1). A changed constant (even if optimal for the particular definitions of Δ) puts an undue burden on the memory of undergraduates. Using variances also for calibration solves this problem. The basic ideas of the proof in [11] can be taken over.

To keep matters simple, we stick to the classic situation of two canonically conjugate variables of a single quantum degree of freedom. For the sake of comparison, let us recall the scenario of the Kennard-Robertson inequality, which we call preparation uncertainty (see Fig. 1). The spreads $\Delta_\rho(A) = (\text{tr } \rho A^2 - (\text{tr } \rho A)^2)^{1/2}$ of position Q and momentum P are determined in separate experiments on the same source, given by a density operator ρ . The uncertainty relation $\Delta_\rho(Q)\Delta_\rho(P) \geq \hbar/2$ is a quantitative

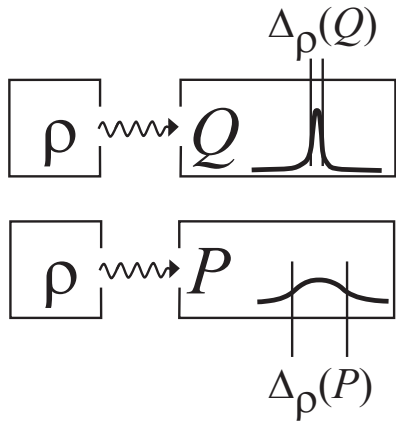


FIG. 1. Scenario of preparation uncertainty. Δ_ρ is the root of the variance of the distribution obtained for the indicated observable in the state ρ . In this pair of experiments no particle is subject to both a position and a momentum measurement.

version of the observation that there are no dispersion-free quantum states [13], as applied to a canonical pair of observables. It is *not* to be found in Heisenberg’s paper [7], except in a rough discussion of post-measurement states, which he assumes to be Gaussian with a spread related to the accuracy of a position measurement.

In contrast, Fig. 2 shows the scenario discussed by Heisenberg. The middle row shows an approximate position measurement Q' followed by a momentum measurement. How should we define the momentum disturbance and position error in this setup? The error of the approximate position measurement Q' clearly refers to the comparison with an ideal measurement Q as shown in the first row. For the momentum disturbance we can say the same: We have remarked that the momenta before and after the microscope interaction do not commute, so the difference makes no sense in the individual case. However, we can compare the *distributions* of the momenta measured after the position measurement (call this effective measurement P') with the *distribution* an ideal momentum measurement P would have given on the same input state. Come to think of it, this is precisely how we detect disturbance in other typical quantum settings. Consider, for example, the double slit experiment. It is well-known that illuminating the slits enough to detect the passage of a particle through one or the other hole makes the interference fringes go away. Clearly, the light used for observation disturbs the particles, and the evidence for this is once again the change of the *distribution* on the screen. Note that this way of looking at error and disturbance restores the symmetry between the position and momentum aspects of this scenario. The uncertainty relations we will prove therefore apply just as well to the position disturbance caused by an approximate momentum measurement and, more generally, to

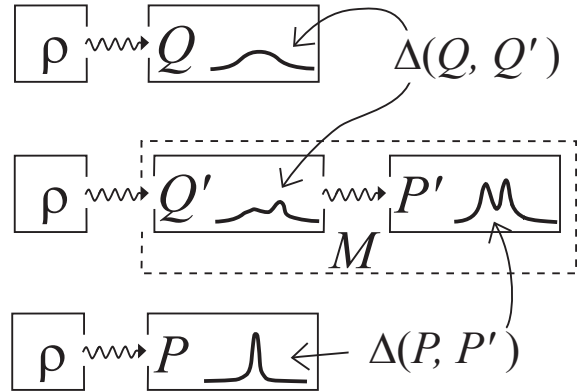


FIG. 2. Scenario of measurement uncertainty for successive measurements, as discussed by Heisenberg (middle row). An approximate position measurement Q' is followed by an ideal momentum measurement, effectively given a measurement P' on the initial state. The accuracy $\Delta(Q, Q')$ quantifies the difference between the output distributions of Q' and an ideal position measurement Q (first row). Similarly, the momentum disturbance $\Delta(P, P')$ quantifies the difference between the distributions obtained by P' and by an ideal momentum measurement P (last row). The definitions for these Δ quantities (see text) can be applied, more generally, to an arbitrary joint measurement M (dashed box). This can be any device producing, in every shot, a q value and a p value. Q' and P' are then defined as the marginals of M , obtained by ignoring the other output.

any measurement scheme M , which produces in every run a value p and a value q (see the dashed outline in Fig. 2). This generalization also covers any successive measurement scenario, in which one tries to *correct for* some of the momentum disturbance, perhaps using the detailed knowledge of how the position measuring device works. In principle, this could allow a reduction of uncertainties. However, the inequality holds without change, which gives a precise meaning and a proof to Heisenberg’s phrase “uncontrollable momentum disturbance”, which he himself uses without further justification.

Let us now discuss the definition of $\Delta(Q, Q')$ in more detail (the momentum case will be completely analogous). We think of this “microscope resolution” as a figure of merit for the device, a promise which might be advertised by the manufacturer, and which could be verified by a testing lab. $\Delta(Q, Q') = 0$ will mean that the “approximate” device Q' is completely equivalent to the ideal Q , i.e., for every input state ρ the output distributions will be the same. Similarly, a small value might indicate that the difference in the distributions will be small for every input state. This requires a definition for the distance of two general probability distributions, which we will give below (Section on “Uncertainty metrics”). However, we can also take a simpler approach, which avoids verifying a statement for *all* input states.

Instead the testing lab might concentrate on those states, which at least classically would seem to be the most demanding ones, namely states for which Q has a known and sharp value. We call this process “calibration”. Still, this requires testing of many states but no longer on very mixed states, or states which contain coherent superpositions of widely separated wave functions.

An advantage of the calibrated error is that we no longer need a quantitative evaluation of the distance between arbitrary probability distributions, but just between an arbitrary distribution and a known sharp value ξ . For this we naturally take the root mean square deviation from ξ

$$D(\rho, Q'; \xi) = \left\langle (q' - \xi)^2 \right\rangle_{\rho, Q'}^{1/2} \quad (2)$$

where the angle brackets denote the expectation of the indicated function of the output q' , in the distribution obtained on the preparation ρ with the device Q' . This statement allows for Q' to be a general POV (positive operator valued) measurement. For projection valued observables like Q we could simplify this to $D(\rho, Q; \xi)^2 = \text{tr}(\rho(Q - \xi\mathbb{I})^2)$. The latter quantity is to be small, say $\leq \varepsilon$, for the input states ρ used for calibration. Hence we set $\Delta_c(Q, Q')$ to be

$$\lim_{\varepsilon \rightarrow 0} \sup \left\{ D(\rho, Q'; \xi) \mid \rho, \xi; D(\rho, Q; \xi) \leq \varepsilon \right\}. \quad (3)$$

Here the set is non-empty since for any ξ and $\varepsilon > 0$ there is a ρ such that $\text{tr}(\rho Q) = \xi$ and $D(\rho, Q; \xi) < \varepsilon$; moreover, the limit exists, because with decreasing ε the supremum is over fewer and fewer states, so the function is non-increasing. In the case of a bad approximation, the supremum can be infinite, in which case we put $\Delta_c(Q, Q') = \infty$.

With this definition, and the corresponding one for P , we can state our main result. We just assume that the Q' and P' are the marginal observables of some joint measurement device M . As discussed above this covers also the case of a sequential measurement (Fig. 2). Then

$$\Delta_c(Q, Q') \Delta_c(P, P') \geq \frac{\hbar}{2}. \quad (4)$$

This inequality is sharp, and equality holds for an M for which the joint distribution of (q, p) -outputs is the so-called Husimi distribution [14] of the input state, which can be obtained by a Gaussian smearing of the Wigner function.

PROOF

The proof has two parts: The first is elementary and concerns the special case that M is a *covariant phase space observable*. These observables [14–17] can be described explicitly, including a very simple form of their

marginals Q' and P' , by which (4) can be reduced to the Kennard-Robertson preparation uncertainty. The second, more technical part of the proof reduces the general case to the covariant case by an averaging method, and is taken from [11]. We only sketch it.

By a covariant measurement we mean one which has a natural symmetry property for both position and momentum translations. That is, if we apply it to an input state shifted in position by δq and in momentum by δp , the output distribution will be the same as before, transformed by $(q, p) \mapsto (q + \delta q, p + \delta p)$. These symmetries are implemented by the Weyl operators (a.k.a. Glauber translations) $W(q, p) = \exp((iqP - ipQ)/\hbar)$. Then the whole observable can be reconstructed from its density at the origin, which must be [16, 17] a positive operator σ of trace 1, i.e., a density operator as for a quantum state. The probability for outcomes in a set $S \subseteq \mathbb{R}^2$ is then given by the positive operator

$$M(S) = \int_S \frac{dq dp}{2\pi\hbar} W(q, p)^* \sigma W(q, p). \quad (5)$$

A remarkable property of these joint measurements of position and momentum is that their marginals take a particularly simple form: The probability density of the outputs q' obtained on a state ρ is a convolution of the position distributions of ρ and σ . That is, we can model the output distribution by taking q distributed like the outputs of an ideal measurement Q on ρ , and adding a noise term q'' , which is independent of q and distributed according to the position distribution of σ . The same description applies to the marginal P' .

Therefore, for a covariant measurement we can immediately identify $\Delta_c(Q, Q')$ without further computation: The density σ is a fixed characteristic property of the measurement. Therefore, as the position distribution of ρ becomes sharply concentrated around some ξ , the outputs converge in distribution to $q' = \xi + q''$, so

$$\Delta_c(Q, Q') = D(\sigma, Q; 0), \quad (6)$$

which is the “size” (the root mean square deviation) of the “noise”. For example, if σ has sharp position distribution at some value a , this is equal to $|a|$, since the outputs will be off by a shift a (i.e., $q' \approx q + a$). Hence one will choose σ with zero mean. The uncertainty product then becomes $\Delta_c(Q, Q') \Delta_c(P, P') = \Delta_\sigma(Q) \Delta_\sigma(P)$, which is $\geq \hbar/2$ by the preparation uncertainty relation applied to σ . This proves (4) for the case of covariant measurements, and at the same time provides examples of minimum uncertainty measurements: all we have to do is to choose σ as a centered minimum uncertainty state, i.e., as $\sigma = |\Psi\rangle\langle\Psi|$ with Ψ a real valued centered Gaussian wave function. The phase space distribution associated to an input state ρ by this measurement M is then the Husimi distribution [14].

The more technical part of the proof of (4) is to show that for any measurement M there is a covariant one,

say \overline{M} with at most the same Δ 's. Basically, \overline{M} is obtained from M by averaging, the technical problem being that the parameter range of (q, p) over which one has to “average” is infinite (see [11]). Let us introduce $\mathcal{M}_\varepsilon(\Delta Q, \Delta P)$ as the set of measurements M such that, for $A = Q, P$, $D(\rho, A'; \xi) \leq \Delta A$ whenever $D(\rho, A; \xi) \leq \varepsilon$ for given ΔA and ε . This is a convex set, and compact in a suitable weak topology. We can write the covariance condition as a fixed point equation for some transformations on the set of all observables, namely a unitary transformation by a Weyl operator combined with a shift in the argument. These transformations commute, and leave $\mathcal{M}_\varepsilon(\Delta Q, \Delta P)$ invariant. Therefore, by the Markov-Kakutani fixed point theorem this set, if non-empty, must also contain a covariant element, which by construction has at most the same uncertainties. This concludes our sketch of the proof of (4).

UNCERTAINTY METRICS

The calibration criterion only involves highly concentrated states so that, in principle, on general input states the optimal joint measurement might produce output distributions quite different from the ideal ones. One can easily give examples of a projection valued observable A and an “approximation” A' for which the calibrated distance is a rather optimistic estimate. That is if we denote by $\Delta(Q, Q')$ a figure of merit based on comparison of *all* states we might have $\Delta(Q, Q') \gg \Delta_c(Q, Q')$. Note first that in the covariant case this cannot happen: The statement that Q' can be simulated by adding fixed independent noise to Q is valid for arbitrary input states, and any reasonable definition of $\Delta(Q, Q')$ should give the size of the noise. However, in the general case we would need a definition which is independent of that special form. The idea is to define a metric D on probability distributions which extends (2) in the sense that $D(\rho, Q'; \xi)$ becomes the metric distance between the output distribution of Q' and a point measure at ξ . Then we set

$$\Delta(Q, Q') = \sup_{\rho} D(\rho, Q; \rho, Q'), \quad (7)$$

where the expression on the right is the metric distance of the two output distributions. Since Δ_c takes the supremum over the smaller set of highly concentrated states, we have $\Delta(Q, Q') \geq \Delta_c(Q, Q')$. The metric D on probability distributions is basically fixed by our requirements as what is technically known as the Wasserstein-2 distance, which is a variant of the the Monge-Kantorovich transport or “earth mover’s” distance (see [18] for a study of such metrics). The problem addressed by Monge was the cost of transforming a hill (earth distribution μ) into some fortifications (earth distribution η), when the workers had to be paid by the bucket and the distance covered. A transport plan, also known as a *coupling* between the

measures μ and η would be a measure γ on $\mathbb{R} \times \mathbb{R}$ describing how much earth was to be moved from x to y . This entails that the marginals of γ must be μ and η . The cost in the Monge problem is then $\int \gamma(dx dy) |x - y|$, which is then minimized by choosing an optimal γ . In the Wasserstein-2 distance the cost function is chosen to be quadratic in the distance and an overall root is taken to bring the units back to a length:

$$D(\mu; \eta) = \inf_{\gamma} \left(\int \gamma(dx dy) |x - y|^2 \right)^{1/2}, \quad (8)$$

where the infimum is over all couplings γ . Consider now the case that η arises from μ by adding independent noise with distribution ν , which amounts to the convolution $\eta = \mu * \nu$. This immediately suggests a transport plan, namely shifting each individual element of the μ distribution by the amount suggested by the noise (formally: $\gamma(dx dy) = \mu(dx) \nu(dy - x)$). This may not be optimal, but gives the estimate $D(\mu; \mu * \nu) \leq D(\nu; 0)$, the size of the noise, where once again the second argument stands for the point measure at zero. This says that the largest distance is attained for a point measure μ , and therefore

$$\Delta(Q, Q') = \Delta_c(Q, Q') \quad (9)$$

whenever Q' is the marginal of a covariant measurement. To summarize this section: if we define the deviation between Q and Q' by a worst case figure of merit over *all* states, the uncertainty relation once again holds. Moreover, the two notions coincide on all covariant measurements, and in particular for the cases of equality.

CONCLUSION AND OUTLOOK

With the inequality (4) we have provided a general, quantitative quantum version of Heisenberg’s original semiclassical uncertainty discussion. This is a remarkable vindication of Heisenberg’s intuitions, far beyond the usual view, which takes the quantitative content of the paper to be summarized entirely by the Kennard-Robertson inequality, and sees the discussion of the microscope as no more than a heuristic order of magnitude argument.

Our conceptual framework applies to any pair of observables, which are not jointly measurable. However, evaluating the respective uncertainty bounds, which will typically not be expressed in terms of the product of uncertainties, is another matter. Explorations of some of these generalizations are under way.

We have already expressed our dissatisfaction with the proposed formalization of Heisenberg’s ideas by [4]. A more detailed analysis is in preparation [19].

ACKNOWLEDGEMENTS

Part of this work (P.L.) is supported by the Academy of Finland, project no 138135. R.F.W. acknowledges support from the European network SIQS. This work has benefited from discussions at a conference organised as part of COST Action MP1006, *Fundamental Problems in Quantum Physics*.

* paul.busch@york.ac.uk

† pekka.lahti@utu.fi

‡ reinhard.werner@itp.uni-hannover.de

- [1] M. Tomamichel and R. Renner, Phys. Rev. Lett. **106**, 110506 (2011).
- [2] F. Furrer, T. Franz, M. Berta, A. Leverrier, V. B. Scholz, M. Tomamichel, and R. F. Werner, Phys. Rev. Lett. **109**, 100502 (2012).
- [3] A. Furuta, Scient. Am. **March 8** (2012), see also <http://phys.org/news/2012-09-scientists-renowned-uncertainty-principle.html>.
- [4] M. Ozawa, Ann. Phys. **311**, 350 (2004).
- [5] J. Erhart, S. Sponar, G. Sulyok, G. Badurek, M. Ozawa, and Y. Hasegawa, Nature Phys. **8**, 185 (2012).
- [6] L. A. Rozema, A. Darabi, D. H. Mahler, A. Hayat, Y. Soudagar, and A. M. Steinberg, Phys. Rev. Lett. **109**, 100404 (2012).
- [7] W. Heisenberg, Zeitschr. Phys. **43**, 172 (1927).
- [8] P. Busch, T. Heinonen, and P. Lahti, Phys. Rep. **452**, 155 (2007).
- [9] E. H. Kennard, Zeitschrift für Physik **44**, 326 (1927).
- [10] H. P. Robertson, Phys. Rev. **34**: 163–64 (1929).
- [11] R. F. Werner, Quant. Inform. Comput. **4**, 546 (2004), quant-ph/0405184.
- [12] P. Busch and D. B. Pearson, J. Math. Phys. **48**, 082103 (2007).
- [13] J. von Neumann, *Mathematische Grundlagen der Quantenmechanik* (Springer, Berlin, 1932).
- [14] K. Husimi, Proc. Physico-Mathem. Soc. Japan **22**, 264 (1940).
- [15] E. Davies, *Quantum theory of open systems* (Academic Press, 1976).
- [16] A. Holevo, *Probabilistic and Statistical Aspects of Quantum Theory* (North Holland, Amsterdam, 1982).
- [17] R. F. Werner, J. Math. Phys. **25**, 1404 (1984).
- [18] C. Villani, *Optimal transport: old and new* (Springer, 2009).
- [19] P. Busch, P. Lahti, and R. Werner, in preparation (2013).